UNIVERSIDAD DE PANAMÁ
VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
CENTRO REGIONAL UNIVERSITARIO DE PANAMÁ OESTE
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y TECNOLOGÍA
PROGRAMA DE MAESTRÍA EN MATEMÁTICA EDUCATIVA

Algunas Aplicaciones del Álgebra Lineal a la Ingeniería Biomédica

.

Por

Edwin Cedeño

8-823-682

Presentado como uno de los requisitos para obtener el grado de Maestría en Matemática Educativa

Profesor Asesor

Dr. José Solanilla

PANAMÁ, REPÚBLICA DE PANAMÁ

2024

DEDICATORIA

A mi familia, a los estudiantes de la Facultad de Ciencias Naturales, Exactas y Tecnología, a los estudiantes de ingeniería Biomédica.

AGRADECIMIENTO

Ante todo, agradezco a Dios, ser supremo, infinito y eterno por darme la sabiduría para culminar este trabajo. Gracias por ser dador de bendiciones y bondades, por enviar a su hijo Jesús, nuestro Señor, para mostrarnos los caminos de salvación. Por Él todas las cosas subsisten, sin Él nada somos y nada podemos hacer.

A mi familia, esposa María Ubarte, mis padres Edwin y Velkys Cedeño que de una u otra forma me colaboraron a lo largo de la carrera universitaria.

A mis amigos, Adela Arosemena, Nelson Mendoza, Yarabis Marín, Nelson Sanjur, Norma Sosa y Ángela Flores por orientarme a optar por el grado académico de magíster.

A mi asesor, doctor José Solanilla, pilar de esta investigación académica, por su guía y orientación durante mi estancia como maestrando; de igual manera, a su esposa la doctora María Corrales e hijos.

A la memoria de la profesora Ariscela Díaz que en paz descanse por, coordinar este programa educativo.

Finalmente, agradezco a la profesora Edilsa Alveo, Elisa Villar y Leticia Gómez por todo el apoyo que me brindaron en la realización de este trabajo.

RESUMEN

Con el objetivo de dar relevancia a la teoría del “álgebra lineal” y sus aplicaciones, en la ingeniería biomédica, el siguiente trabajo documental expone una recopilación de elementos de esta teoría presentes en algunos estudios contemporáneos sobre equipos biomédicos. La importancia de esta indagación es fundamentar que la teoría en mención es necesaria para el soporte y funcionamiento de los equipos biomédicos que se utilizan, estudian e investigan en la actualidad. Es por ello, que se muestran elementos importantes del “álgebra lineal” utilizando el código R y lenguaje de programación estadístico; el cual, es capaz de realizar múltiples capacidades algorítmicas. Esto es posible por los elementos de la teoría de vectores y matrices que la teoría del “álgebra lineal” facilita a programas importantes como: el perceptrón y aprendizaje de máquina vectorial, la función clasificadora, matrices, los espacios vectoriales duales, transformaciones lineales y las funciones clasificadoras. Todos estos programas son necesarios para la planificación y aprendizaje de las máquinas investigadas; las cuales se utilizan en la actualidad por los especialistas de la salud, con la finalidad de mejorar sus servicios. Se resaltan investigaciones de equipos en los que el “álgebra lineal” aporta diagnósticos de tejidos cancerígenos. En la cirugía, el uso de la robótica para controlar la velocidad de la sangre con el uso de nanorobots. El robot Origami que utiliza geometría vectorial para el movimiento y configuración utilizando el pivote vectorial. Lo anterior, evidencia que son muchas las aplicaciones del “álgebra lineal” en la biomédica, las cuales se pueden retomar para futuras investigaciones.

Palabras Claves: espacio vectorial, hiperespacios, matrices, función clasificadora, código R.

ABSTRACT

With the aim of giving relevance to the theory of "linear algebra" and its applications in biomedical engineering, the following documentary work presents a compilation of elements of this theory present in some contemporary studies on biomedical equipment. The importance of this research is to establish that the theory mentioned is necessary for the support and operation of the biomedical equipment that is currently used, studied and investigated. Therefore, important elements of "linear algebra" are shown using the R code and statistical programming language, which is capable of performing multiple algorithmic capabilities. This is made possible by the elements of vector and matrix theory that the theory of "linear algebra" facilitates important programs such as: the perceptron and vector machine learning, the classifier function, matrices, dual vector spaces, linear transformations and classifier functions. All these programs are necessary for the planning and learning of the machines investigated, which are currently used by health specialists in order to improve their services. Research on equipment in which "linear algebra" provides diagnoses of cancerous tissues is highlighted. In surgery, the use of robotics to control blood velocity with the use of nanorobots. The Origami robot that uses vector geometry for movement and configuration using the vector pivot. The above shows that there are many applications of "linear algebra" in biomedicine, which can be taken up for future research.

Keywords: vector space, hyperspaces, matrices, classifier function, R code.

ÍNDICE

ÍNDICE DE FIGURAS

PÁGINA

INTRODUCCIÓN

En el siglo XXI se cuenta con importantes tecnologías que han mejorado el estilo de vida de las personas en varias áreas, tales como: ingeniería, arquitectura, economía, área científica y también en el cuidado de la salud. Dentro de este panorama, encontramos las matemáticas; como ciencia que siempre han estado presentes en todo nuestro estilo de vida; aunque, no podamos advertirlas a simple vista, ella está equilibrando el estilo de vida actual.

En las siguientes páginas, se presenta un trabajo investigativo y documental, dividido en cuatro capítulos, en donde se fundamenta teórica y científicamente que la matemática del “álgebra lineal” se ha implementado en el soporte, avance y ejecución de los equipos utilizados en la ingeniería biomédica contemporánea. Esto, con el objetivo de modernizar el funcionamiento del instrumental; y, por ende, mejorar la salud de las personas.

En el primer capítulo, se presentan aspectos primordiales de la investigación: objetivos generales y específicos, justificación, alcance, aportaciones, limitaciones y proyecciones del “álgebra lineal” aplicada a la ingeniería biomédica. Además, se presentan algunos datos interesantes de estudios e investigaciones contemporáneas que resaltan la presencia del “álgebra lineal” en la ingeniería biomédica; los cuales, motivaron la exploración en este tema innovador.

En el segundo capítulo, se definen los conceptos empleados en los sistemas algebraicos computacionales y algorítmicos: código R. utilizados en la programación de los

equipos biomédicos. Estos aspectos teóricos son esenciales para la comprensión de las posteriores secciones de esta investigación.

En el capítulo tercero, se abordan temas como: hiperplanos separados, perceptrón o neurona artificial, convergencia, variables duales, algoritmo de aprendizaje, el análisis del PCA para el código R, entre otros elementos del “álgebra lineal” que se han convertido en herramientas de soporte en la programación de software y hardware en la ingeniería biomédica.

En el cuarto capítulo**,** se exponen investigaciones sobre los avances tecnológicos que se están implementando en la ingeniería biomédica y que aplican la teoría del “álgebra lineal”. Con ello, se busca demostrar que las teorías explicadas en los capítulos dos y tres sobre el “álgebra lineal”, son un soporte fundamental para el funcionamiento adecuado del equipo que se utiliza para el diagnóstico de células cancerígenas, a través de la función clasificadora. Además, se presenta la aplicación del “álgebra lineal” utilizada para regular los equipos usados en las cirugías, como la velocidad de la sangre. También se muestran los contenidos del “álgebra lineal” útiles para el funcionamiento de la robótica, los cuales, son importantes para el especialista quirúrgico que utiliza estos innovadores equipos.

Para concluir con esta investigación se resumen los datos más relevantes encontrados, después de analizar el amplio panorama de las teorías del “álgebra lineal” que se utilizan en la ingeniería biomédica.

CAPÍTULO 1. Aspectos Preliminares

Las técnicas innovadoras computacionales usadas en la biomédica es el resultado de los primeros matemáticos creadores de las computadoras como: Wilhelm Schickard, Babbage, Ada Lovelace, Turing Machine y otros. Estos estudiosos aplicaron los elementos y algoritmos matemáticos para crear computadoras de cálculos, que luego evolucionaron programaciones de “Inteligencia Artificial, el Machine Learning y el Deep Learning, la Ciencia de Datos, la Robótica y el tema innovador de los nanorobots. Sus descubrimientos, han evolucionado con resultados exactos en la informática, y son muy necesarios para la innovación de la tecnología, donde se encuentra inmersa la ingeniería biomédica.

Lo anterior demuestra, que las matemáticas han estado presentes en el desarrollo tecnológico desde la primera computadora, programadores, aprendizaje de máquina software, robótica, nanorobots que se aplican en la actualidad y favorecen a la “ingeniería biomédica”. (Babini, 2002, p. 63; Vargas 2012, p.43-63; Mora 2022, p. 294-306) Esta ingeniería, se apoya en la inteligencia artificial, un área innovadora con fundamento en las herramientas de la ciencia de datos; a las cuales, el “álgebra lineal” contribuye para su avance y desarrollo.

Los antes expuesto, ha sido la fuente de motivación para establecer los objetivos de este trabajo investigativo; puesto que ha impulsado el interés por conocer sobre qué contenidos del “álgebra lineal” se han implementado para el soporte de equipos programados computacionalmente, en la ingeniería biomédica.

Tomando en cuenta la información descrita, se ha elaborado este primer capítulo con aspectos primordiales, como lo son: objetivos generales y específicos, justificación, alcance, aportaciones, limitaciones y proyecciones del tema investigado: “álgebra lineal aplicada a la ingeniería biomédica”.

1.1. Objetivo General

Presentar un compendio de los modelos, métodos y algoritmos sobre los elementos del “álgebra lineal”, que han aportado al avance de la tecnología de las máquinas contemporáneas; y cómo estos trabajan, con exactitud, en la ingeniería biomédica proporcionando resultados satisfactorios en la mejoría de la salud del hombre.

1.1.1. Objetivos Específicos

- Exponer los elementos del “álgebra lineal” importantes en el desarrollo de las máquinas que han ayudado a la innovación de los equipos de la ingeniería biomédica.
- Plasmar cómo los espacios vectoriales son importantes para la comprensión de las máquinas computacionales y robóticas, como soportes usados en la ingeniería biomédica.
- Compilar modelos del “álgebra lineal” importantes para el desarrollo de la inteligencia artificial necesaria para la programación y funcionamiento de las máquinas, usadas por los especialistas de la salud para diagnósticos, cirugías y rehabilitación de los pacientes.
- Reunir conceptos, elementos y algoritmos fundamentales del “álgebra lineal” que han contribuido a la robótica; y son importantes para la tecnología de la biomédica.

1.2. Justificación

1.2.1. Importancia

Esta investigación documental centra su importancia en conocer los elementos del “álgebra lineal” que han contribuido a los avances tecnológicos de la ingeniería biomédica; con la finalidad, de entender su contribución en el mejoramiento de la salud.

1.2.2. Aporte

Este trabajo tiene la finalidad de ser un compendio de información sobre los elementos del “álgebra lineal”, que se tienen en el siglo XXI, y se han utilizado en la innovación tecnológica, proporcionando resultados significativos en el mejoramiento de la vida de las personas.

1.2.3. Alcance

Los avances en la informática e ingeniería han surgido con múltiples proyecciones y creaciones de tecnología de punta; los cuales, han solucionado varios problemas de la vida cotidiana. El acceso a esta información es valioso debido a su amplio campo de estudio.

Es decir, la recopilación de modelos matemáticos, en especial del “álgebra lineal”, que han contribuido a la programación de las nuevas creaciones tecnológicas y han logrado resultados satisfactorios es un valioso compendio; el cual permitirá nuevos análisis de las contribuciones del “álgebra lineal” a las ciencias biomédicas.

1.2.4. Limitaciones

Existe pocos documentos con información de las aplicaciones del “álgebra lineal” en la ingeniería biomédica. Por ello, en primera instancia, se realiza un estudio examinando los avances que el “álgebra lineal” ha aportado a la computadora, a las programaciones de inteligencia artificial, a la robótica y a otros; para luego, relacionar estos aportes con los que se han utilizado en la ingeniería biomédica.

1.2.5. Proyecciones

El uso innovador de equipos en la tecnología biomédica ha evolucionado por la presencia de contenidos del “álgebra lineal”. Esta investigación se proyecta en conocer los

conceptos de la teoría del “álgebra lineal”, que han favorecido el avance y desarrollo de esta tecnología, logrando resultados exactos y favorables a la salud de la humanidad.

Actualmente, la teoría más utilizada de las matemáticas es la aplicada en el “álgebra lineal”. Hay una gran contribución de esta teoría a los avances tecnológicos. De allí, que temas tradicionales como el cálculo de valores y vectores propios han encontrado aplicaciones en problemas tan diversos como la visión por computador o la clasificación de sitios en la red.

Nuevos algoritmos surgen como respuesta a las necesidades de aplicaciones específicas y esto hace que sea necesario volver a la teoría para comprender sus fundamentos. (Ossa, 2016, p. XVII)

Las áreas del desarrollo científico tecnológico en las que la matemática se ha mostrado especialmente útil es el campo de las ciencias de la computación, denominado “inteligencia artificial” y conocido como “aprendizaje de máquina”. Esto se trata de programar computadores para optimizar un criterio de desempeñó usando datos como ejemplos o experiencias pasadas. Se han desarrollado sofisticadas técnicas para construir máquinas de aprendizaje lineal que han resultado exitosas en muchas aplicaciones.

Entre los diversos problemas del aprendizaje de máquina uno de los más importantes es el problema de la clasificación que, con base en ese conocimiento, clasifique adecuadamente cualquier nuevo dato que esté sin clasificar. (Ossa, 2016, p. XVIII)

En estos últimos años, es muy fácil tener acceso a grandes bases de datos que consisten en colecciones de estos, adecuadamente clasificados (exámenes médicos, clientes bancarios, industria, arquitecturas, otros.) y el principal problema que se plantea es construir una función clasificadora que, con base en ese conocimiento, clasifique adecuadamente cualquier nuevo dato que esté sin clasificar.

El problema no es simple; y, por lo tanto, una primera forma de atacarlo es considerando el conjunto de posibles “funciones clasificadoras” como un conjunto de “funciones lineales”, pues la teoría de las “funciones lineales” es una de las teorías matemáticas más completas. (Ossa, 2016, p. XVIII)

Es por ello, que para empezar lo mejor es considerar el conjunto $\mathcal{M}$ formado por los objetos a clasificar como elementos de un “espacio vectorial”. Esto lleva directamente a conceptos como medida de similitud y matrices de similitud. Una medida de similitud es una función:

$$\kappa : \mathcal{M}\ x\ \mathcal{M} \rightarrow \mathbb{R}$$

con algunas propiedades especiales. Un primer ejemplo de una medida de similitud sencilla es el producto punto (o interno) canónico $\boldsymbol{u} \cdot \boldsymbol{v}$ en $\mathbb{R}^n$. Es interesante observar cómo nos guía rápidamente a la construcción de algoritmos. Usando la teoría del “álgebra lineal”, se van desarrollando teorías y técnicas para tratar el problema de la clasificación. Su modesto propósito es lograr que alguien interesado en estos temas, por ejemplo, un estudiante de ingeniería, que no tenga una formación sólida en matemáticas, pueda comprender de forma acelerada cómo se aplica el “álgebra lineal” a los problemas básicos del aprendizaje de máquina. Y de este modo, que pueda avanzar sin muchas dificultades en la comprensión de esos temas. (Ossa, 2016, p. XVIII)

El “álgebra lineal” es importante en el aprendizaje de máquina, porque los modelos del aprendizaje automático se pueden expresar en forma de matriz, donde los conjuntos de datos a menudo se representan como una matriz. (Wayne Richards, 2022, p.1646-1647) Los equipos biomédicos necesitan programación de software y hardware para que los resultados sean exactos; por lo tanto, el “álgebra lineal” aplica el preprocesamiento de datos, la transformación de datos y la evaluación de modelos necesarios para la Biomédica.

(Wayne Richards, 2022, p.1648) Estos equipos diseñados a base de computadora, inteligencia artificial, y requieren de programación para que puedan funcionar, de aquí, la importancia de temas del álgebra lineal como: vectores, matrices, transpuesta de una matriz, inversa de una matriz, determinante de una matriz, rastro de una matriz, producto punto, valores propios y vectores propios (Wayne Richards, 2022,p. 1645). Demostraciones que se contemplan en el capítulo cuarto, con la finalidad de presentar que en los equipos actuales se utiliza el contenido del “álgebra lineal”.

En los últimos años, las redes neuronales artificiales y el deep learning, son dos de las herramientas más poderosas del aprendizaje de máquina. Ellas tienen por objetivo desarrollar sistemas que aprenden automáticamente, reconocen patrones, predicen comportamientos y generalizan información a partir de conjuntos de datos.

Estas dos herramientas se han convertido en un potencial campo de investigación con aplicaciones a las ingenierías, y la ingeniería biomédica no es la excepción. Al usar las redes neuronales y el deep learning (aprendizaje profundo) en la ingeniería biomédica en las ramas de: la ómica, la imagenología, las interfaces cerebro-máquina y hombre-máquina, y la gestión y administración de la salud pública; ramas extendidas desde el estudio de procesos a nivel molecular, hasta procesos que involucran grandes poblaciones. (Ramos, 2020, p.1)

Estas investigaciones planteadas con fines de mejorar la salud social se deben a las grandes aportaciones de la teoría del “álgebra lineal” como herramienta para la programación, aprendizaje y entrenamiento de las máquinas. Su uso a través de las computadoras permite examinar y diagnosticar inmediatamente los resultados de los estudios realizados a los pacientes, con el fin de intervenir a tiempo y controlar la enfermedad, reduciendo así el número de mortandad. Es aquí, donde el “álgebra lineal” es considerada un puente de soporte para la tecnología biomédica con la finalidad de que los

resultados de los exámenes aplicados a pacientes sean más precisos y poder ayudar a su mejoría. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010,p. 34-35)

Obsérvese el caso de los especialistas oncólogos, que utilizan tecnología con: máquinas de vectores de soporte, redes neuronales y bosques aleatorios con el propósito de actualizar el entrenamiento, aprendizaje de la máquina, la clasificación; de manera que, los resultados de los exámenes aplicados a pacientes con tejidos lesionados puedan darse con exactitud y más rápido. Como es el caso de El método de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa.

Este método tiene el propósito de distinguir lesiones benignas y malignas en la piel humana a partir de espectros de reflexión difusa, se ha requerido del análisis de diferentes algoritmos de clasificación usando el software de aprendizaje automático y reconocimiento de patrones software WEKA[1]. Para realizar las tareas de clasificación en las técnicas mencionadas se ha utilizado,

$$WEKA\ (Waikato\ Environment\ for\ Knowledge\ Analysis),$$

que es una aplicación ampliamente usada para la experimentación e investigación en áreas como el reconocimiento de patrones y la minería de datos, ya que incluye una gran cantidad de algoritmos de preprocesado, clasificación y clustering. Además, dada la alta dimensionalidad de la señal espectral, fue empleada una técnica de selección de atributos para determinar las variables que aporten la mayor cantidad de información.

Se probó la clasificación de la señal usando los algoritmos: de máquinas de vectores de soporte, redes neuronales y bosques aleatorios el desempeño fue evaluado usando el promedio de la $k - fold\ cross - validation$ tomando en cuenta los porcentajes de

[1] Son máquinas de vectores de soporte, redes neuronales y bosques aleatorios.

instancias clasificadas correctamente, $el\ índice\ kappa$, el área bajo la curva ROC, la sensibilidad, y la especifidad. Finalmente, se demuestra que el algoritmo de "redes neuronales" con los parámetros $momentum\ y\ learning\ rate\ en\ 0{,}6\ y\ 0{,}3$ respectivamente, es el que mejor se adapta al problema de reconocimiento de patrones ya que clasifica correctamente al 89,89% de los casos. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.36-39)

Las técnicas espectroscópicas se han utilizado médicamente con el propósito de solucionar lesiones de diferentes índoles, con el principio fundamental para diagnosticar enfermedades con técnicas espectroscópicas, construyendo algoritmos robustos que permitan extraer las características más importantes de la señal espectral y correlacionarlas con su respectiva patología. Entre las técnicas se puede destacar: algoritmo WEKA para, detectar lesiones cancerígenas en el cuello uterino; también algoritmos como $el\ ensemble\ AdaBoostM1 - Reptree$ para diferenciar entre imágenes (obtenidas mediante un colposcopio) de cuello uterino sano y cuello uterino con lesiones precancerosas, logrando un porcentaje de acierto del 89,4737%; usando una red neuronal formada por siete neuronas; en la capa oculta fue posible obtener un 83,3% de sensibilidad y 88,9% de especificidad en la clasificación de la señal espectral.

Alternativamente, para detectar lesiones en la piel, también se ha empleado el análisis de imágenes obtenidas mediante epiluminescencia, en conjunto con $algoritmos\ como\ el\ clasificador\ k - nearest\ neighbors\ (KNN)$ se evaluaron atributos en la imagen como: tamaño de la lesión, media del borde de la lesión. Probada por $Wallace\ et\ al$ en detección de lesiones en la piel mediante espectrometría de reflexión difusa y $algoritmos\ de\ clasificación$. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p. 34)

En el trabajo citado en el párrafo anterior, se ha evaluado el desempeño de diferentes técnicas computacionales para clasificar espectros de reflexión difusa entre muestras de tejido sano y maligno, usando el $software\ WEKA$ de investigación en aprendizaje automático y minería de datos, las técnicas implementadas.

Después de aplicar estas técnicas de clasificación, se ha demostrado que el algoritmo correspondiente a las redes neuronales es el que tiene mejor desempeño, alcanzando 89,89% de aciertos en la clasificación de los casos estudiados. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.35)

Además, no solo en la rama de la oncología, sino que también en la mayoría de las áreas médicas, hoy contamos con tecnología innovadora que ayuda a mejorar con rapidez a los pacientes; permitiéndoles a los especialistas de la salud, diagnosticar inmediatamente e intervenir en la rama de las cirugías que utilizan importantes aparatos para controlar la velocidad de la sangre, aparatos que utilizan las herramientas del "álgebra lineal" en la teoría de matrices, producto punto y vectores.

En la cirugía, se utiliza la nanorobótica, una de las innovaciones más recientes, en la que el álgebra lineal es puente para el desarrollo y funcionamiento de estos equipos; los cuales son muy importantes para los especialistas de la salud al permitir que los pacientes se recuperen prontamente. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010; Do, 2017, p. 40)

Esta cirugía de nueva generación implica que se podrá perforar un orificio muy pequeño, para hacer biopsias en el caso de cáncer de próstata o en diferentes tipos de cáncer, con estos robots. En la actualidad, este tipo de procedimiento ya lo realizan robots, el más conocido de ellos es el robot Da Vinci. Básicamente, el cirujano permanece detrás de la consola y dispone de un robot que está preajustado y que hace las incisiones.

Inclusive, utilizando una plataforma de origami, si es necesario, se puede añadir un nuevo instrumento simplemente reconfigurando la forma. La teoría que aporta el “álgebra lineal” a la reconfiguración de este robot es gracias a los teoremas de vector normal. (Belke y Paik, 2017, p. 2153-2164)

El “álgebra lineal” es una herramienta esencial en la ciencia de datos y el aprendizaje automático es importante y necesario para el funcionamiento, reconfiguración y programación del software y los hardware utilizados en los equipos biomédicos. Por lo tanto, los principiantes interesados en la ciencia de datos deben familiarizarse con los conceptos esenciales del “álgebra lineal” (Wayne, 2022, p. 1646-1648). Las teorías más usadas, actualmente, en la matemática aplicada provienen del “álgebra lineal”. Hay una gran contribución de esa teoría y los avances tecnológicos. Temas tradicionales como el cálculo de valores y vectores propios han encontrado aplicaciones en problemas tan diversos como la visión por computador o la clasificación de sitios en la red.

Cómo la matemática del “álgebra lineal ayuda a la programación de software de esta tecnología para, que su funcionalidad sea con exactitud a través de la comprensión de algoritmos, por ende, es de importancia demostrar e implementar estos algoritmos para que se puedan probar con diferentes bases de datos. (Ossa, 2016, p. xvii)

Lo anterior, sustenta el propósito de esta investigación que es demostrar cómo a través del uso de la teoría del “álgebra lineal” se desarrollan nuevas teorías y técnicas, que aportan a la programación de las herramientas, que la ingeniería biomédica, utiliza al momento de realizar: un examen médico, cirugías, fisioterapias, biopsias, otras.

CAPÍTULO 2. Fundamentación del Álgebra Lineal y Empleo de Sistemas Algebraicos Computacionales y Algoritmos: Código R

En este capítulo se presentan las definiciones y contenidos algebraicos que se requieren para la ejecución de la tecnología de la biomédica. El mismo, se divide en dos secciones: la primera aborda las definiciones de elementos fundamentales como: vectores, espacios vectoriales, espacios duales, hiperplanos, matrices, combinaciones lineales, transformaciones linéales, geometría de vector normal (ortogonal), entre otros; y la segunda parte, define el Código R desde el lenguaje de probabilidad estadística con las aplicaciones del contenido de los fundamentos del "álgebra lineal".

2.1. Aspectos Esenciales de la Teoría del Álgebra Lineal

El álgebra lineal como la definen, en forma simple, los autores, es aquella rama de la matemática que trata de las propiedades comunes de los sistemas algebraicos, que constan de un conjunto más una noción razonable de "combinación lineal" de los elementos del conjunto. (Hoffman y Kunze, 1973, p. 28)

2.1.1. Vectores

Dados un número natural n (fijo) y un cuerpo F (en particular F= $\mathbb{R}$ O F= $\mathbb{C}$), a cuyos elementos llamaremos escalares, se llaman "vectores de n componentes" del cuerpo F a los elementos que se especifican en (I) entre las que hay establecidas las dos operaciones que se detallan en (II):

I. Se llaman vectores de n componentes del cuerpo F a las sucesiones o sistemas ordenados de n escalares de F, es decir, a los elementos de F^n. Estos vectores son, pues, los objetos:

$$\bar{u} = (u_1, u_2, \ldots, u_n).$$

$$\text{Donde } \begin{cases} u_1, u_2, \ldots, u_n \; son \; escalares \; del \; cuerpo \; F \\ u_i \; se \; llama \; componente \; \; i - \acute{e}sima \; de \; \bar{u} \end{cases}$$

II. Se llama suma $\bar{u} + \bar{v}$, de dos vectores $\bar{u}\ y\ \bar{v}$, y se llama producto $\lambda\bar{u}$, donde λ es un escalar, a los siguientes vectores:

$$\left.\begin{matrix}\bar{u} = (u_1, u_2, \ldots, u_n) \\ \bar{v} = (v_1, v_2, \ldots, v_n)\end{matrix}\right\} \rightarrow \bar{u} + \bar{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n),$$ se entiende aquí que $u + v$ es también producto.

$$\left.\begin{matrix}\bar{u} = (u_1, u_2, \ldots, u_n) \\ \lambda \text{ es un escalar}\end{matrix}\right\} \rightarrow \lambda\,\bar{u} = (\lambda\,u_1, \lambda\,u_2, \ldots, \lambda\,u_n)$$

(Burgos, 1997, p.23)

2.1.2. Propiedades entre Vectores

Para los vectores de n componentes y en relación con la suma y el producto por escalar, se verifica:

Si $\bar{u}, \bar{v}\ y\ \bar{w}$ son vectores de n componentes (n escalares de un cuerpo F; en particular, reales o complejos) y si $\lambda\ y\ \mu$ son dos escalares, entonces:

1. $(\bar{u} + \bar{v}) + \bar{w} = \bar{u} + (\bar{v} + \bar{w})$ propiedad asociativa.
2. $\bar{u} + \bar{v} = \bar{v} + \bar{u}$ propiedad conmutativa.
3. $\bar{u} + \bar{0} = \bar{u}, donde\ \bar{0} = (0, 0, \ldots, 0)$ se llama $vector\ nulo$
4. Para cada $\bar{u}$, existe un vector $-\bar{u}$ tal que $\bar{u} + (-\bar{u}) = \bar{0}$; Si $\bar{u} = (u_1, u_2, \ldots, u_n)$, entonces $-\bar{u} = (-u_1, -u_2, \ldots, -u_n)$. Al vector $-\bar{u}$ se le llama opuesto de $\bar{u}$.
5. $\lambda\,(\bar{u} + \bar{v}) = \lambda\,\bar{u} + \lambda\,\bar{v}$
6. $(\lambda + \mu)\bar{u} = \lambda\,\bar{u} + \mu\bar{u}$
7. $\lambda(\mu\bar{u}) = (\lambda\,\mu)\,\bar{u}$
8. $1\bar{u} = \bar{u}$

Además se verifican las siguientes propiedades:

a) $0\,\bar{u} = \bar{0}\ y \lambda\bar{0} = \bar{0}$

b) $\lambda\bar{u} = \bar{0} \Rightarrow [\lambda = 0 \text{ o } \bar{u} = \bar{0}]$

c) $(-\lambda)\bar{u} = \lambda(-\bar{u}) = -(\lambda\,\bar{u})$ (Burgos, 1997, p.25).

2.2. Matrices

Sea dado un sistema de m ecuaciones lineales con n incógnitas, al que llamaremos $[F]$:

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{array}\right\}$$

Se llama matriz de los coeficientes, matriz de los términos independientes y matriz del sistema (o ampliada) a las siguientes tablas rectangulares $A, B\ y\ AB$ formadas por escalares:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}; B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}; AB = \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right] \begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_m \end{array}$$

La matriz A es una tabla rectangular de $m \cdot n$ escalares dispuestos en m filas y n columnas. La fila $i - ésima$ de A es la sucesión de escalares $a_{i1}, a_{i2} \cdots a_{in}$; la columna $j - ésima$ de A es la sucesión de escalares $a_{1j}, a_{2j} \cdots a_{mj}$; (que en A aparecen dispuestos verticalmente). Se dice que el elemento a_{ij}, que está situado en la fila i y en la columna j, es el elemento que ocupa el lugar espacio $(i, j) o\ ij$.

De manera abreviada y, por el momento, también simbólicamente, el sistema F se expresará poniendo:

$AX = B, donde\ X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_3 \end{bmatrix}$ se llama columna o vector columna de la incógnita.

(Burgos, 1997, p.25)

Ejemplo 1

Si n es un entero positivo, $\mathbb{R}^n$ (léase “erre ene”) denota la colección de todas las listas ordenadas de n números reales, generalmente escritas como matrices columna de orden $n\ x\ 1$ del tipo

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

El vector cuyas entradas son todas cero se llama **vector cero** y se denota con **0**. (El número de entradas en **0** será evidente a partir del contexto). La igualdad de vectores en $\mathbb{R}^n$ y las operaciones de multiplicación escalar y suma vectorial en $\mathbb{R}^n$ se definen entrada por entrada como en $\mathbb{R}^2$. Esas operaciones sobre vectores tienen las siguientes propiedades, las cuales pueden verificarse directamente a partir de las propiedades correspondientes de los números reales. (Lay, 2007, p. 27)

2.3. Concepto de Espacios Vectoriales

Sea V un conjunto dado, a cuyos elementos llamaremos $vectores$. Se considera también un cuerpo F (en particular $F = \mathbb{R}\ y\ F = \mathbb{C}$), a cuyos elementos llamaremos $escalares$. Se dice que V es un espacio vectorial sobre F si se dispone de las siguientes operaciones de suma y producto por escalar y se satisfacen las siguientes propiedades:

1. Una operación (+) interna en V (suma de vectores) tal que se cumplen las siguientes propiedades de la suma:

- $(\bar{u}+\bar{v})+\bar{w}=\bar{u}+(\bar{v}+\bar{w}) \quad \forall \quad \bar{u},\bar{v},\bar{w} \in V$ (asociativa).
- $\bar{u}+\bar{v}=\bar{v}+\bar{u} \qquad \forall \quad \bar{u},\bar{v} \in V$ (conmutatividad).
- $\exists \bar{o} \in V\ (vector\ nulo)$ tal que $\bar{u}+\bar{o}=\bar{u}$ para cualquier $\bar{u} \in V$.
- Para cada $\bar{u} \in V$ existe $-\bar{u} \in V$ tal que $\bar{u}+(-\bar{u})=\bar{o}$ (cada vector $\bar{u} \in V$ tiene un vector opuesto $-\bar{u} \in V$).

2. Una operación externa producto por escalar, que a cada pareja $\lambda \in F$, $\bar{u} \in V$ asociar un vector $\lambda\bar{u}$, tal que se cumplan los siguientes axiomas del producto por escalar:
   - $\lambda(\bar{u}+\bar{v})=\lambda\,\bar{u}+\lambda\,\bar{v} \qquad \forall \quad \bar{u},\bar{v} \in V \quad \forall \quad \lambda \in F$.
   - $(\lambda+\mu)\bar{u}=\lambda\,\bar{u}+\mu\bar{u} \qquad \forall \quad \bar{u} \in V \quad \forall \quad \lambda,\mu \in F$.
   - $\lambda\,(\mu\bar{u})=(\lambda\,\mu)\bar{u} \qquad \forall \quad \bar{u} \in V \quad \forall \quad \lambda,\mu \in F$.
   - $1\bar{u}=\bar{u}$ (1= unidad de F) $\forall \quad \bar{u} \in V$. (Burgos, 1997, p.116)

Ejemplo 2. El espacio de $n-tuples$, F^n

Sea $\boldsymbol{F}$ cualquier cuerpo y sea $\boldsymbol{V}$ el conjunto de todos los $\boldsymbol{n-tuples}$ $\boldsymbol{\alpha}=(x_1,x_2,\ldots,\ x_n)$ de escalares x_i de F. Si $\beta=(y_1,y_2,\ldots,\ y_n)$ con $y_i\ de\ F$, la suma de $\alpha\ y\ \beta$ se define por:

$$\alpha+\beta=(x_1+y_1,x_2+y_2,\ldots,x_n+y_n).$$

El producto de un escalar c y el vector α se define por

$$c\alpha=(cx_1\,,c\,x_2,\ldots,cx_n)$$

Que esta adición vectorial y multiplicación por escalar cumplen las condiciones usando las propiedades semejantes de la adición y multiplicación de los elementos del cuerpo escalar F. (Hoffman y Kunze, 1973, p. 29)

Ejemplo 3. El espacio de matrices $m\ x\ n$, todas las $F^{m\ x\ n}$

Sea $\boldsymbol{F}$ cualquier cuerpo y sean $m\ y\ n$ enteros positivos. Sea $\boldsymbol{F^{m\ x\ n}}$ el conjunto de todos los matrices $m\ x\ n$ sobre el cuerpo F. La suma de dos vectores $A\ y\ B$ en $\boldsymbol{F}^{\boldsymbol{m\ x\ n}}$ Se define por

$$(A + B)_{i\ j} = A_{i\ j} + B_{i\ j}$$

El producto de un escalar c y de la matriz A se define por

$$(cA)_{i\ j} = cA_{i\ j}$$

Obsérvese que $F^{1\ x\ n} = F^{\ n}$. (Hoffman y Kunze, 1973, p.29)

Ejemplo 4. El espacio de funciones de un conjunto en un cuerpo

Sea F cualquier cuerpo y sea S cualquier conjunto no vacío. Sea V el conjunto de todas las funciones de S en F. La suma de dos vectores $f\ y\ g$ de V es el vector $f + g$; es decir, la función de S en F definida por

$$(f + g)\ (s) = f(s) + g(s).$$

El producto del escalar c y la función f es la función cf definida por

$$(cf)(s) = cf(s).$$

Ejemplo 5. El espacio de las funciones polinomios sobre el cuerpo F

Sea F un cuerpo y sea V el conjunto de todas las funciones f de F en F definidas en la forma

$$f(x) = c_0 + c_1x + \cdots + c_nx^n$$

Donde $c_0, c_1, \ldots, c_n$ son escalares fijos de F (independiente de x). Una función de este tipo se llama función polinomio sobre F . Sean la adición y la multiplicación de escalares. Se debe observar que si $f\ y\ g$ son funciones polinomios y c está en F, entonces $f +$ $g\ y\ cf$ son también funciones polinomios. (Hoffman y Kunze, 1973, p. 30)

2.3.1. Subespacios Vectorial

Se llama subespacio de un espacio vectorial V a aquellos subconjuntos de V que son, a su vez, espacios vectoriales respecto a las mismas operaciones de V. Al concretar y desarrollar lo aquí dicho, obtenemos:

Sea V un espacio vectorial sobre F (cuerpo) y sea U un subconjunto de V.

Se dice que U es un subespacio vectorial de V si las operaciones de V son, también, operaciones para U y, con ellas, U es un espacio vectorial sobre F. Caracterización, se verifica que, U es un subespacio vectorial de V si, siendo $U \neq \emptyset$, se cumplen las dos condiciones a) y b):

a) $\bar{u}, \bar{v} \in U \Rightarrow \bar{u} + \bar{v} \in U$

b) $\lambda \in F, \bar{u} \in U \Rightarrow \lambda\bar{u} \in U$.

Estas condiciones a) y b) se pueden sustituir, ambas, por la condición c):

c) $[\bar{u}, \bar{v} \in U \; y \; \lambda, \; \mu \in F] \Rightarrow \lambda\bar{u} + \mu\bar{v} \in U$. (se dice que el vector $\lambda\bar{u} + \mu\bar{v}$, para cualesquiera que sean los escalares $\lambda \; y \; \mu$ es, una combinación lineal de los vectores $\bar{u} \; y \; \bar{v}$. (Burgos, 1997, p.118)

Ejemplo 6

(a) Si V es cualquier espacio vectorial, V es un subespacio de V; el subconjunto que consta solo del vector nulo es un subespacio vectorial de V, llamado subespacio nulo de V^*.

(b) En F", el conjunto de los $n - tuples\ (x_1, \cdots, x_n)$ con $x_1 = 0$ es un subespacio, pero el conjunto de los $n - tuples$ con $x_1 = 1 + x_2$ no es un subespacio $(n \geq 2)$.
(c) El espacio de las funciones polinomios sobre el cuerpo F es un subespacio del espacio de todas las funciones de F en F.

d) Una matriz (cuadrada) $n\ x\ n$, sobre el cuerpo F es simétrica si $A_{ij} = A_{ji}$ para todo $i\ y\ j$. Las matrices simétricas forman un subespacio del espacio de las matrices $n\ x\ n$ sobre F. (Hoffman y Kunze, 1973, p. 35)

2.3.2. Geometría de Vectores

Ciertas partes del "álgebra lineal" están íntimamente relacionadas con la geometría. La misma palabra espacio sugiere algo geométrico, como lo hace el vocablo vector. Considérese el espacio vectorial $\mathbb{R}^3$. En geometría analítica se identifican las ternas (x_1, x_2, x_3) de números reales con los puntos del espacio euclidiano tridimensional. Un vector se suele definir como un segmento dirigido PQ, del punto P del espacio a otro punto Q.

Ello equivale a una formulación cuidadosa de la idea de flecha de P a Q. Tal como son usados los vectores, se ha tenido en mente que queden definidos por su longitud y dirección. Así, se deben identificar dos segmentos dirigidos si tienen la misma longitud y dirección.

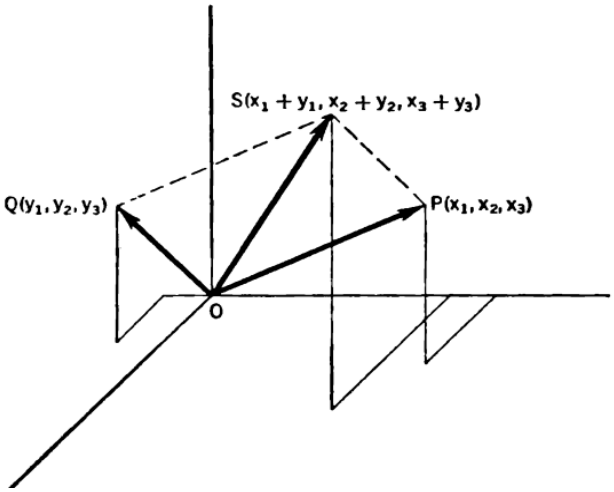
El segmento dirigido PQ, del punto $P = (x_1, x_2, x_3)$ al punto $Q = (y_1, y_2, y_3)$ tiene la misma longitud y dirección que el segmento dirigido del origen $0 = (0, 0, 0)$ al punto $(y_1 - x_1,\ y_2 - x_2,\ y_3 - x_3)$. Además, este es el único segmento que partiendo del origen tiene la misma longitud y dirección que PQ. Con lo que, si se conviene en operar solo con vectores aplicados al origen, hay exactamente un **vector asociado** con cada longitud y dirección dadas. (Hoffman y Kunze, 1973, p. 32)

El vector OP, del origen a $P = (x_1, x_2, x_3)$ está completamente determinado por P, y es por ello posible identificar este vector con el punto P. En la definición del espacio vectorial $\mathbb{R}^3$, los vectores se definen simplemente como las ternas (x_1, x_2, x_3) .

Dados los puntos $P = (x_1, x_2, x_3)$ y $Q = (y_1, y_2, y_3)$, la definición de suma de los vectores OP y OQ puede hacerse geométricamente. Si los vectores no son paralelos, entonces los segmentos OP y OQ determinan un plano y estos segmentos son dos de los lados de un paralelogramo en aquel plano. A continuación, se observa la Figura 1,

Figura 1. ***Diagonal de paralelogramos*** $\mathbb{R}^3$

Diagonal de Paralelogramos $\mathbb{R}^3$



Tomada de: "Álgebra lineal", (p. 33) por, Hoffman, K., Kunze, R., & Finsterbusch, H. E. 1973. Prentice-Hall Internacional.

La diagonal de este paralelogramo, que va de 0 al punto S, define la suma de OP y OQ como el vector OS. Las coordenadas del punto S son $(y_1 + x_1, y_2 + x_2, y_3 + x_3)$, por la definición de suma del espacio vectorial. La multiplicación escalar tiene una interpretación geométrica simple. Si c es un número real, entonces el producto de c por el vector OP es el vector que parte del origen de longitud $|c|$ veces a longitud de OP y dirección que coincide con la de OP si $c > 0$, y que es opuesta a la dirección de OP si $c < 0$. Esta multiplicación escalar da justamente el vector OT donde $T = (c\,x_1, cx_2, cx_3)$, y es, por tanto, compatible con la definición algebraica dada en $\mathbb{R}^3$. (Hoffman y Kunze, 1973, p. 33)

2.4. Combinación Lineal

Un vector β de V se dice "combinación lineal" de los $vectores\ \alpha_1, . \ . \ ., \alpha_n$ de F tales que

$$\beta = c_1\alpha_1 + \cdots + c_n\alpha_n = \sum_{i=1}^{n} c_i\alpha_i.$$

Las propiedades asociativas de la adición vectorial y las propiedades distributivas de la multiplicación escalar se aplican en la combinación lineal. (Hoffman y Kunze, 1973, p. 31)

2.5. Vector Linealmente Dependiente e Independiente

Sea V un espacio vectorial sobre F. Un subconjunto S de V se dice linealmente dependiente (o simplemente, dependiente) si existen vectores distintos $\alpha_1, \alpha_2, \cdots \alpha_n$ de S y escalares $c_1, c_2, \cdots c_n$ de F, no todos nulos, tales que

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = 0$$

Un conjunto que no es linealmente dependiente se dice linealmente independiente. (Hoffman y Kunze, 1973, p. 40)

Si el conjunto S tiene solo un número finito de vectores $\alpha_1, \alpha_2, \cdots \alpha_n$,, se dice, a veces, que los $\alpha_1, \alpha_2, \cdots \alpha_n$son dependientes (o independientes), en vez de decir que S es dependiente (o independiente).

Las siguientes son fáciles consecuencias de la definición:

a) Todo conjunto que contiene un conjunto linealmente dependiente es linealmente dependiente.

b) Todo subconjunto de un conjunto linealmente independiente es linealmente independiente.

c) Todo conjunto que contiene el vector 0 es linealmente dependiente; en efecto,

$1 \cdot 0 = 0.$

d) Un conjunto S de vectores es linealmente independiente si, y solo si, todo subconjunto finito de S es linealmente independiente; es decir, si, y solo si, para vectores diferentes $\alpha_1, \alpha_2, \cdots \alpha_n$ de S, arbitrariamente elegidos $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$ implica que todo $c_i = 0$. (Hoffman y Kunze, 1973, p. 41)

Ejemplo 7

Sea F un subcuerpo de los números complejos. En F^3los vectores

$$\alpha_1 = (3, 0, -3)$$

$$\alpha_2 = (-1, 1, 2)$$

$$\alpha_3 = (4, 2, -2)$$

Son linealmente dependientes, pues

$$2\alpha_1 + 2\alpha_2 - \alpha_3 + 0 \cdot \alpha_4 = 0.$$

Los vectores

$$\epsilon_1 = (1, 0, 0)$$

$$\epsilon_2 = (0, 1, 0)$$

$$\epsilon_3 = (0, 0, 1)$$

Son linealmente independiente. (Hoffman y Kunze, 1973, p. 41)

2.6. Espacio Vectorial de Dimensión Finita

Un espacio vectorial V, sobre un cuerpo F, se dice que es de tipo finito sí esta generado por un número finito de vectores, es decir, si en V existe algún sistema de vectores $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ tal que V = $\mathcal{V}$ (S).

Si V es de tipo finito, se dice que un sistema de vectores $B = (\beta_1, \beta_2, \ldots, \beta_n)$ es una base de V si se verifica cualquiera de las dos condiciones siguientes, que son equivalentes entre sí como:

a) B es un sistema generador de V que, además, es sistema linealmente independiente.
b) Todo vector de V se puede expresar de una sola manera como combinación lineal de los vectores de B. (Burgos, 1997, p.127)

2.7. Nociones de Dimensión

2.7.1. Existencia de Bases

Cualquier sistema generado de un espacio vectorial de tipo finito, $V \neq 0$, incluye a una base de V. En consecuencia, todo espacio vectorial $V \neq 0$ de tipo finito tiene alguna base. (Burgos, 1997, p.129)

Teorema 1. Dimensión

Todas las bases de un espacio vectorial $V \neq 0$, de tipo finito, tiene igual número de vectores. A este número se le llama $dimensión$ del espacio V y se le representa poniendo $\dim V$. Se conviene en que el espacio $V = 0$ tiene dimensión 0. (Burgos, 1997, p.129)

2.7.2. Propiedades de las Bases y la Dimensión

Si $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ es un sistema de vectores de un espacio vectorial V de dimensión finita, entonces se verifica que:

a) Si $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ es un sistema generador de V, entonces $n \geq \dim V$.

b) Si $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ es un sistema independiente, entonces $n \leq \dim V$.

c) Si $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ es generador de V y $\dim V = n$, entonces S es base de V.

d) Si $S = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ es independiente y $\dim V = n$, entonces S es base de V. (Burgos, 1997, p.130)

2.7.3. Dimensión de un Subespacio

Si U es un subespacio de un espacio vectorial V y si V tiene dimensión finita, entonces U también tiene dimensión finita y se verifica que $\dim U \leq \dim V$, donde el signo $=$ sólo es válido en el caso de se $U = V$. (Burgos, 1997, p.132)

Ejemplo 8

Si F es un cuerpo, la dimensión de F^n es n, porque la base canónica de F^n contiene n vectores. El espacio de las matrices $F^{m\,x\,n}$ tiene dimensión mn. Ello es claro por analogía con el caso de F^n, ya que las mn matrices que tienen un 1 en el lugar i, j, con ceros en los demás, forman una base de $F^{m\,x\,n}$.

Si A es una matriz $m\ x\ n$, entonces el espacio solución de A tiene dimensión $n - r$, donde r es el número de filas no nulas de una matriz escalón reducida por filas, que es equivalente por filas a A. (Hoffman y Kunze, 1973, p. 45)

Teorema 2. Dimensión

Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea W un subespacio de V. Entonces

$$\dim W + \dim W^C = \dim V$$

Corolario 1

Si W es un subespacio de dimensión k de un espacio vectorial V de dimensión n, entonces W es la intersección de $(n - k)$ hiperespacios en V. (Hoffman y Kunze, 1973, p.100-101)

En un espacio vectorial de dimensión n, un subespacio de dimensión $n - 1$ se llama un hiperespacio. Tales espacios se llaman también a veces hiperplanos o subespacios de codimensión. El espacio nulo de una función lineal es todo hiperespacio.

2.8. Nociones de Diagonización

Sea T un operador lineal sobre un espacio vectorial de dimensión finita V. Se dice que T es diagonalizable si existe una base de V tal que cada vector suyo sea vector propio de T. (Corrales, Martín y Solanilla, 2016, pág. 81-82)

La razón del nombre es clara; en efecto, si existe una base ordenada $\mathfrak{B} = \alpha_1\ ,\ \alpha_2, \ldots, \alpha_n\ de\ V$ en la que cada α_i es un vector propio de T, entonces la matriz de T en la base ordenada $\mathfrak{B}$ es diagonal. Si $T\alpha_i = c_i\alpha_i, entonces,$

$$[T]_{\mathfrak{B}} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}$$

Por cierto, no se requiere que los escalares $c_1, \ldots, c_n$ sean distintos, incluso pueden ser todos iguales (cuando Tes un múltiplo escalar del operador identidad).
Se podría también definir T como diagonalizable cuando los vectores propios de T generan V. Esto es solo en apariencia distinto a la definición dada. Ya que se puede formar una base tomándola de cualquier conjunto generador de vectores.

(Hoffman y Kunze, 1973, p. 184)

2.9. Nociones de Transformaciones Lineales

Sean $V\ y\ W$ dos espacios vectoriales sobre el cuerpo F. Una transformación lineal de V en W es una función T de V en W tal que

$$T(c\alpha + \beta) = c(T\alpha) + T\beta$$

Para todos los vectores $\alpha\ y\ \beta\ de\ V$ y todos los escalares c de F.

Ejemplo 9

Sea F un cuerpo y sea V el espacio vectorial de las funciones polinomios f de F en F, dado por

$$f(x) = c_0 + c_1 x + \ldots + c_k x^k.$$

Sea

$$(Df)(x) = c_1 + 2c_2 x + \ldots + k c_k x^{k-1}.$$

Entonces D es una transformación lineal de V en V: la transformación derivada.

(Hoffman y Kunze, 1973, p. 67)

Teorema 3

Sea V un espacio vectorial de dimensión finita sobre el cuerpo F, sea $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ una base ordenada V. Sean W un espacio vectorial sobre el mismo cuerpo F y $\beta_1, \beta_2, \ldots, \beta_n$ vectores cualesquiera de W. Entonces existe una transformación lineal de T de V en W tal que

$$T\alpha_j = \beta_j, \qquad j = 1, \ldots, \qquad n$$

(Hoffman y Kunze, 1973, p. 69)

Ejemplo 10

Sea T una transformación lineal del espacio de los $n - tuples\ F^n$ en el espacio de $n - tuples\ F^n$. El Teorema 3 dice que T está unívocamente determinado por la sucesión de vectores $\beta_1, \beta_2, \ldots, \beta_n$, donde $\beta_i = T\epsilon_i, \qquad i = 1, \ldots, m.$

Dicho brevemente, T está unívocamente determinado por las imágenes de los. vectores de la base canónìca. Esta determinación es

$$\alpha = (x_1, \ldots, x_m)$$

$$T\alpha = x_1\beta_1 + \ldots + x_m\beta_m.$$

Si B es la matriz $m\ x\ n$ que tiene por filas los vectores β_1 , $\beta_2, \ldots, \beta_n$, esto quiere decir que

$$T\alpha = \alpha B.$$

O sea que si $\beta_i = (\beta_{i1}, \ldots, \beta_{in})$, entonces

$$T\ (x_1\ ,\ \ldots\ , x_n\ ) = \ [x_1\ ,\ \ldots\ , x_m\ ] \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}$$

Esta es una descripción muy explícita de **la transformación lineal.** (Hoffman y Kunze, 1973, p. 70)

2.9.1. Representación de Transformaciones por Matrices

Sea V un espacio vectorial de dimensión n sobre el cuerpo F, y sea W un espacio vectorial de dimensión m sobre F. Sea $\mathfrak{B} = \{\alpha_1\ ,\ \ldots\ , \alpha_n\}$ una base ordenada de W . Si T es cualquier transformación lineal de V en W, entonces T está determinada por su efecto sobre los vectores α_j. Cada uno de los n vectores $T\alpha_j$ se expresa de manera única como combinación lineal

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\ \beta_i$$

De los β_i, los escalares $A_{1j}, \ \ldots\ , A_{mi}$ son las coordenadas de $T\alpha_j$ en la base ordenada $\mathfrak{B}'$. Por consiguiente, la transformación T está determinada por los mn escalares de A_{ij} en la expresión. La matriz $m\ x\ n, A,$ definida por

$$A\ (\ i, j) = A_{ij}$$

se llama matriz de T respecto al par de **bases ordenadas** $\mathfrak{B}\ y\ \mathfrak{B}'$. La tarea inmediata es comprender claramente cómo la matriz A determina la transformación lineal T.

Si $\alpha = \ x_1\alpha_1 + \ \ldots + x_n\ \alpha_n$ es un vector de V, entonces

$$T\alpha = T\left(\sum_{j=1}^{n} x_j \alpha_j\right)$$

$$= \sum_{j=1}^{n} x_j (T\alpha_j)$$

$$= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\, \beta_i$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}\, x_j\right) \beta_i$$

(Hoffman y Kunze, 1973, p. 86)

2.9.2. Transpuesta de una Transformación Lineal

Sean V y W espacios vectoriales sobre el cuerpo F. Para toda transformación lineal T de V en W. Existe una única transformación lineal T^t de W^* en V^* tal que

$$(T^t g)(x) = g\,(T\alpha)$$

Para todo g de W^* y todo α de V. (Hoffman y Kunze, 1973, p. 111)

Ejemplo 11

Sean $\boldsymbol{W = F^{nxn}}$, el espacio de todas las matrices $n\ x\ n$ sobre F. Entonces w es isomorfo a F^{n^2} de un modo natural, se sigue, pues, Ejemplo 1, que la igualdad

$$\boldsymbol{A \cdot B = \sum_{j,k} A_{jk} \overline{B}_{jk}}$$

Define un producto interno sobre W. Además, si se introduce la matriz transpuesta conjugada $\boldsymbol{B^*}$, donde $\boldsymbol{B^*_{jk} = \overline{B}_{jk}}$. Se puede expresar este producto interno sobre $\boldsymbol{F^{nxn}}$, en términos de la función traza:

$$\boldsymbol{(A \cdot B) = t\,r\,(AB^*) = tr\,(B^*A)}$$

En efecto,

$$\boldsymbol{t\,r}\,(\boldsymbol{AB}^{*}) = \sum_{\boldsymbol{j}} (\boldsymbol{AB}^{*})_{\boldsymbol{jj}}$$

$$= \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} \boldsymbol{A_{jk}}\, \boldsymbol{B^{*}_{jk}}$$

$$= \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} \boldsymbol{A_{jk}}\, \overline{\boldsymbol{B}}_{\boldsymbol{jk}}$$

Ejemplo 12

$\boldsymbol{F^{nx1}}$ el espacio de las matrices (columnas) $n\ x\ 1$ sobre F, y sea Q una matriz inversible $n\ x\ n$, sobre F. Para X, Y en F^{nx1} se hace

$$\boldsymbol{X}\ \cdot \boldsymbol{Y} = \ \boldsymbol{Y}^{*}\ \boldsymbol{Q}^{*}\boldsymbol{QX}$$

Se identifica la matriz $1\ X\ 1$ del segundo miembro con su solo elemento. Cuando Q es la matriz unidad, este producto interno es esencialmente el mismo del ejemplo 1, se llama el producto interno canónico sobre $\boldsymbol{F^{nx1}}$. (Hoffman y Kunze, 1973, p. 269)

Ejemplo 13

Por el siguiente método se pueden construir nuevos productos internos a partir de uno dado. Sean $V\ y\ W$ espacios vectoriales sobre F y supóngase que $(\ |\ )$ es un producto interno sobre W. Si T es una "transformación lineal" no singular de V en W, entonces la igualdad

$$p_{\mathrm{T}}(\alpha\ , \beta) = (\mathrm{T}\alpha|\mathrm{T}\beta)$$

Define un producto interno p_{T} sobre V. El producto interno del ejemplo 3 es un caso especial de esta situación. También lo son los siguientes:

(a) Sea V un espacio vectorial de dimensión finita y sea

$$\mathfrak{B} = \{\alpha_1 , \quad \alpha_2, \quad . . \quad ., \alpha_n\}$$

Una base ordenada de V. Sean $\epsilon_1, \ . . ., \ \epsilon_n$ los vectores de las bases canónicas en F^n, y sea T la transformación lineal de V en F^n tal que $T\alpha_j = \epsilon_j, \ J = 1, \ . . ., n$. En otras palabras, sea T el **isomorfismo natural** de V sobre el producto interno canónico sobre F^n, entonces

$$p_{\mathrm{T}}\left(\sum_j x_j\alpha_j \, , \sum_k y_k\alpha_k\right) = \sum_{j=1}^{n} x_i\bar{y}_j$$

Así para cada base para V existe un producto interno sobre V con la propiedad de que $(\alpha_j|\alpha_k)$; en efecto, es fácil ver que existe exactamente uno solo de tales productos internos. (Hoffman y Kunze, 1973, p. 270)

2.10. Nociones de Función Lineal

Si V es un espacio vectorial sobre el cuerpo F, una transformación lineal f de V en el cuerpo de escalares F se llama también un funcional lineal sobre V. Si se comienza desde el principio, esto quiere decir que f es una función de V en F, tal que

$$f(c\alpha + \beta) = cf(\alpha) + f(\beta)$$

para todos los vectores $\alpha \ y \ \beta$ de V y todos los escalares c de F. El concepto de funcional lineal es importante para el estudio de los espacios vectoriales de dimensión finita, pues ayuda a organizar y clarificar el estudio de los subespacios, las ecuaciones lineales y las coordenadas. (Hoffman y Kunze, 1973, p. 96)

Ejemplo 14

Sea F un cuerpo y sean $a_1, \ldots, a_n$ escalares pertenecientes a F. Defínase una función f en F^n por

$$f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n.$$

Entonces f es una función lineal sobre F^n. Es el funcional lineal representado por la matriz $[a_1, \ldots, a_n]$ respecto a la base ordenada canónica de F^n y la base de F:

$$a_j = f(\epsilon_j), \quad j = 1, \ldots, n.$$

Toda función lineal sobre F^n es de esta forma para ciertos escalares $a_1, \ldots, a_n$. Ello se sigue de la definición de funcional lineal, ya que definimos

$a_j = f(\epsilon_j)$ y empleando la linealidad.

$$f(x_1, \ldots, x_n) = f\left(\sum_j x_j \epsilon_j\right)$$

$$= \sum_j x_j f(\epsilon_j)$$

$$= \sum_j a_j \epsilon_j$$

(Hoffman y Kunze, 1973, p. 97)

2.10.1. El Espacio de Funciones de un Conjunto en un Cuerpo

Sea F cualquier cuerpo y sea S cualquier conjunto no vacío. Sea V el conjunto de todas las funciones de S en F. La suma de dos vectores f y g de V es el vector $f + g$; es decir, la función de S en F definida por

$$(f + g)(s) = f(s) + g(s).$$

El producto del escalar c y la función f es la función cf definida por

$$(cf)(s) = cf(s).$$

2.10.2. El Espacio de las Funciones Polinomios sobre el Cuerpo F

Sea F un cuerpo y sea V el conjunto de todas las funciones f de F en F definidas en la forma

$$f(x) = c_0 + c_1x + \cdots + c_nx^n$$

Donde $c_0, c_1, \ldots, c_n$ son escalares fijos de F (independiente de x). Una función de este tipo se llama función polinomio sobre F . Sean la adición y la multiplicación de escalares. Se debe observar que si $f\ y\ g$ son funciones polinomios y c está en F, entonces $f + g\ y\ cf$ son también funciones polinomios. (Hoffman y Kunze, 1973, p. 31)

2.11. Nociones de Espacios Duales

Si V es un espacio vectorial, el conjunto de todos los funcionales lineales sobre V forman, naturalmente, un espacio vectorial; es el espacio $L(V, F)$. Se designa este espacio por V^* y se le llama espacio dual del V:

$$V^* = \mathrm{L(V, F)}.$$

Teorema 4

Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea

$\mathfrak{B} = \{x_1, x_2, x_3\}$ una base de V. Entonces existe una única base dual

$\mathfrak{B}^* = \{f_1, \ldots, f_n\}$ de V^* tal que $f_i(\alpha_j) = \delta_{ij}$. Para cada función lineal f sobre V se tiene

$$f = \sum_{i=1}^{n} f(\alpha_j) f_i$$

Y para cada vector α de V se tiene

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i.$$

(Hoffman y Kunze, 1973, p. 98)

2.11.1. Espacio Dual

Sea $[a, b]$ un intervalo cerrado del eje real y sea $C([a, b])$ el espacio de las funciones reales continuas sobre $[a, b]$. Entonces

$$L\,(g) = \int_a^b g\,(t)\;dt\text{f}$$

Define un funcional lineal L en $C([a, b])$.

Si V es de dimensión finita se puede obtener una descripción muy explícita del espacio dual V^*. Por la definición de dimensión se conoce acerca del espacio vectorial V^*:

$$\dim\, V^* = \dim V.$$

Sea $\mathfrak{B} = \{\alpha_1\,,\ \alpha_2, \ldots, \alpha_n\}$ una base de V. Existe para cada i una función lineal único f_i en V tal que

$$f_i(\alpha_j) = \delta_{ij}$$

De esta forma se obtiene de $\mathfrak{B}$ un conjunto de n funcionales lineales distintos, $f_1, \ldots, f_n$ sobre V.

Estos funcionales son también **linealmente independientes**, pues supóngase que

$$f = \sum_{i=1}^{n} c_i f_i.$$

Entonces

$$f(\alpha_j) = \sum_{i=1}^{n} c_i f_i\,(\alpha_i)$$

$$= \sum_{i=1}^{n} c_i \delta_{ij}.$$

$$= c_j.$$

En particular, si f es el funcional cero $f(\alpha_j) = 0$ para cada j y por tanto los escalares c_j son todos ceros.

Entonces los $f_1, \ldots, f_n$ son n funcionales linealmente independientes, y como se sabe que V^* tiene dimensión n, deben ser tales que

$\mathfrak{B}^* = \{f_1, \ldots, f_n\}$ es una base de V^*.

Esta base se llama **base dual** de $\mathfrak{B}$. (Hoffman y Kunze, 1973, p. 98)

2.12. Nociones de Producto Punto o Escalar

Las magnitudes en los conceptos geométricos de longitud, distancia y ángulo en $\mathbb{R}^3$ término de vectores pueden ser descritas usando la idea de producto punto entre dos vectores sean,

$$\boldsymbol{\alpha} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad y \quad \boldsymbol{\beta} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

el producto punto $\boldsymbol{v}$ se define mediante:

$$(\boldsymbol{\alpha}|\boldsymbol{\beta}) = x_1y_1 + x_2y_2 + x_3y_3 \cdots + x_ny_n$$

Geométricamente, este producto escalar es el producto de la longitud de $\boldsymbol{\alpha}$, la longitud de $\boldsymbol{\beta}$ y el coseno del ángulo entre $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$. Es, por tanto, posible definir los conceptos geométricos de longitud y ángulo en $\mathbb{R}^3$ por la definición algebraica de un producto escalar.

En pocas palabras, $\boldsymbol{\alpha}|\boldsymbol{\beta}$ es la suma de los productos de los componentes correspondientes de $\boldsymbol{\alpha}$ y de $\boldsymbol{\beta}$, entonces para que dicho producto esté definido, ambos vectores deben tener el mismo número de componentes. El resultado de realizar esta operación es un escalar, por esto $\boldsymbol{\alpha}|\boldsymbol{\beta}$ también es conocido como producto escalar canónico

de $\boldsymbol{\alpha}\ y\ \boldsymbol{\beta}$. El producto punto de vectores en es un caso especial de producto interno que se llama espacio vectorial euclídeo usual $\mathbb{R}^n$.

2.12.1. Propiedades del Producto Punto

Sea $\boldsymbol{F}$ el cuerpo de los números reales, o de los complejos, $\boldsymbol{V}$ un espacio vectorial sobre $\boldsymbol{F}$. Un producto interno sobre $\boldsymbol{V}$ es una función que asigna a cada par ordenado de vectores, $\boldsymbol{\alpha, \beta}$ de $\boldsymbol{V}$ un escalar $(\boldsymbol{\alpha}|\boldsymbol{\beta})$ de $\boldsymbol{F}$ de tal modo que para cualesquiera $\boldsymbol{\alpha, \beta}$, $\boldsymbol{\gamma}$ de $\boldsymbol{V}$ y todos los escalares $\boldsymbol{c}$.

1) $(\boldsymbol{\alpha}+\boldsymbol{\beta}|\boldsymbol{\gamma}) = (\boldsymbol{\alpha}|\boldsymbol{\gamma}) + (\boldsymbol{\beta}|\boldsymbol{\gamma})$;
2) $(\boldsymbol{c\alpha}|\boldsymbol{\beta}) = \boldsymbol{c}(\boldsymbol{\alpha}|\boldsymbol{\beta})$;
3) $(\boldsymbol{\beta}|\boldsymbol{\alpha}) = \boldsymbol{c}\left(\overline{\boldsymbol{\alpha}|\boldsymbol{\beta}}\right)$; $\quad conjugación\ completa$
4) $(\boldsymbol{\alpha}|\boldsymbol{\alpha}) > \mathbf{0}$, $si\ \boldsymbol{\alpha} \neq \mathbf{0}$.

Debe observarse que las condiciones (1), (2) y (3) implican que

5) $(\boldsymbol{\alpha}|\boldsymbol{c\beta}+\boldsymbol{\gamma}) = \bar{\boldsymbol{c}}(\boldsymbol{\alpha}|\boldsymbol{\beta}) + (\boldsymbol{\alpha}|\boldsymbol{\gamma})$. (Hoffman y Kunze, 1973, p. 269)

Ejemplo 15

Sea F^n existe un producto interno que se llama “producto interno canónico”. Está definido sobre $\boldsymbol{\alpha} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ y $\boldsymbol{\beta} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ por $(\boldsymbol{\alpha}|\boldsymbol{\beta}) = \sum_j \boldsymbol{x_j}\overline{\boldsymbol{y}}_{\boldsymbol{j}}$.

Sí $F = R$ puede escribirse $(\boldsymbol{\alpha}|\boldsymbol{\beta}) = \sum_j \boldsymbol{x_j y_j}$.

2.12.2. Expresión del Producto Escalar

En el espacio vectorial $V = \mathcal{M}_n$ de las matrices reales de tamaño $n\ x\ n$, se puede definir un producto escalar de dos matrices $A = [a_{ij}]$ $B = [b_{ij}]$ mediante

$$\langle A, B\rangle = traza\left(A\ B^{\prime}\right)$$

Vemos que para esta forma de definir $\langle A, B\rangle$ se verifican los tres axiomas del producto escalar:

- $\langle A, B\rangle = \sum_{ij} a_{ij} b_{ij} = \langle B, A\rangle$
- $\langle \lambda A + \lambda' A', B\rangle = \sum_{ij} \left(\lambda a_{ij} + \lambda' a'_{ij}\right) b_{ij} = \lambda_i \sum_{ij} a_{ij} b_{ij} + \lambda' \sum a'_{ij} =$

  $\lambda\langle A, B\rangle + \lambda'\langle A', B\rangle$
- $\langle A, A\rangle = \sum_{ij} a_{ij} b_{ij} = \sum_{ij} a^2{}_{ij} > 0$ salvo si $A = [a_{ij}] = 0$

Sea V un espacio vectorial euclídeo. Si la $\dim V = n,\ si\ (\bar{e}_1, \bar{e}_2, \cdots \bar{e}_n)$ es una base de V y si se conocen los n^2 productos escalares $\bar{e}_i \cdot \bar{e}_j$ (para $i, j = 1, 2, \cdots, n$), entonces el producto escalar $\bar{x} \cdot \bar{y}$, de dos vectores cualesquiera $\bar{x}, \bar{y}, \epsilon V$, igual a:

$$\bar{x} \cdot \bar{y} = \sum_{i,j=1}^{n} g_i\, x_i y_j = X' \text{ siendo } \begin{cases} g_i = \bar{e}_i \cdot \bar{e}_j\ (i, j = 1, 2, \dots,\ n \\ G = [g_{ij}]\ matriz\ nxn \end{cases}$$

Donde $(x_1, \dots, x_n)\ e\ (y_1, \dots, y_n)$ son las coordenadas de $\bar{x}\ e\ \bar{y}\ \ y\ X\ e\ Y$ son sus columnas de coordenadas. La matriz $G = [g_{ij}]$, que se llama $matriz\ métrica$ (o de Gram) del producto escalar en la base dada, es una matriz simétrica positiva.

Al realizar un cambio de base, llamado P a la matriz del correspondiente cambio de coordenadas $\left(X = PX'\right)$, la matriz métrica en la nueva base es la

$$G' = P'GP \quad (G\ y\ G' \text{ congruentes}) \text{ (Burgos, 1997, p.256)}$$

Sea V espacio vectorial, y cuando se desee especificar el cuerpo, se dirá que V es un espacio **vectorial sobre el cuerpo** $\boldsymbol{F}$. El nombre $vector$ se da a los elementos del conjunto V más bien por conveniencia.

Teorema 6

Sea V un espacio de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Entonces T es triangulable si, y solo si, el polinomio minimal de T es producto de polinomios lineales sobre F.

2.12.3. Signo

Signo:

Se define el signo de una permutación por

$$sgn\ \sigma = \begin{cases} 1, & si\ \sigma\ \ es\ par \\ -1, & si\ \sigma\ es\ impar \end{cases}$$

representando aquí el símbolo 1 el entero 1. Para las funciones determinísticas en particular, si D es una función determinística $D\ (\epsilon_{\sigma 1},\ \ldots, \epsilon_{\sigma n}) = (-1)^m$. (Hoffman y Kunze, 1973, p.151)

Teorema 7

Sean V un espacio vectorial de dimensión n sobre el cuerpo F, y W un espacio vectorial de dimensión m sobre F. Sean $\mathfrak{B}$ una base ordenada de V y $\mathfrak{B}'$ una base ordenada de W. Para cada transformación lineal T de V en W, está una matriz $m\ x\ n$, A, cuyos elementos pertenecen a F, tal que

$$[T\alpha]_{\mathfrak{B}'} = A[\alpha_{\mathfrak{B}}]$$

Para todo vector α en V. Además, $T \rightarrow A$ es una correspondencia biyectiva entre el conjunto de todas las transformaciones lineales de V en W y el conjunto de todas las matrices $m\ x\ n$ sobre el cuerpo F. (Hoffman y Kunze, 1973, p. 87)

2.12.4. Espacio de Producto Interno

Un espacio producto interno es un espacio real o complejo junto con: un producto interno definido sobre ese espacio.

Un espacio producto interno real de dimensión finita se llama a menudo espacio euclidiano. Un espacio con producto interno complejo se llama a veces espacio unitario. (Hoffman y Kunze, 1973, p. 275)

2.12.4.1. Propiedades

Si V es un espacio producto interno, entonces para vectores, β cualesquiera de V y cualquier escalar c.

i. $\|c\,\alpha\| = \|c\|\|\alpha\|$

ii. $\|\alpha\| > 0\ para\ \alpha \neq 0$

iii. $|(\alpha|\beta)| \leq \|\alpha\|\,\|\beta\|$

iv. $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

Demostración

Las afirmaciones (i) y (ii) se desprenden casi inmediatamente de las varias definiciones vistas. La desigualdad en (iii) es claramente válida para $\alpha = 0$. Si $\alpha \neq 0$, hágase $\gamma = \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\,\alpha$.

Entonces, $(\gamma|\alpha) = 0$ y

$$0 \leq \|\gamma\|^2 = (\beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\,\alpha \;\Big|\; \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\,\alpha)$$

$$= (\beta|\beta) - \frac{(\beta|\alpha)\,(\alpha|\beta)}{\|\alpha\|^2}$$

$$= \|\alpha\|^2 - \frac{|(\alpha|\beta)|^2}{\|\alpha\|^2}$$

Luego, $|(\alpha|\beta)|^2 \leq \|\alpha\|^2\,\|\beta\|^2$. Ahora usando (c) se tiene que

$$\|\alpha + \beta\|^2 = \|\alpha\|^2 + (\alpha|\beta) + (\beta|\alpha) + \|\beta\|^2$$

$$= \|\alpha\|^2 + 2R(\alpha|\beta) + \|\beta\|^2$$

$$\leq \|\alpha\|^2 + 2\|\alpha\|\,\|\beta\| + \|\beta\|^2$$

$$= (\|\alpha\| + \|\beta\|)^2$$

Con lo que $\|\alpha + \beta\| \leq \|\alpha\|\,\|\beta\|$.

La desigualdad en (iii) se llama **desigualdad de Cauchy-Schwarz** y tiene una n muestra que si (por ejemplo), es cero, entonces

$|(\alpha|\beta)| \leq \|\alpha\|\,\|\beta\|$ a menos que

$$\beta = \frac{(\beta|\alpha)}{\|\alpha\|^2}\,\alpha$$

Así, la igualdad se tiene en (iii) si, y solo si, α y β son linealmente dependientes. (Hoffman y Kunze, 1973, p. 275).

Ejemplo 16

El vector (x, y) de R^2 es ortogonal a $(-y, x)$ con respecto al producto interno canónico; en efecto, $[(x, y)|(-y, x)] = -xy + yx = 0$.

2.12.5. Nociones de la Forma Cuadrática

Un producto interno sobre un espacio vectorial, real o complejo está determinado por la llamada función de la forma cuadrática determinada por el producto interno. Para definirla, primero se representa la raíz cuadrada positiva de $(\alpha|\alpha)$ por $\|\alpha\|$; llamada “norma” de α respecto al producto interno, en un espacio vectorial real o complejo. Por ende, es apropiado pensar en la norma de α como la longitud o magnitud de α.

La **forma cuadrática** determinada por el producto interno es la función que asigna a cada vector α el escalar $\|\alpha\|^2$. Se sigue de las propiedades del producto interno que

$$\|\alpha \pm \beta\|^2 = \|\alpha\|^2 \pm 2\,Re\,(\alpha|\beta) + \|\beta\|^2$$

Para todos los vectores $\alpha\ y\ \beta$. Así

$$(\alpha|\beta) = \frac{1}{4}\|\alpha + \beta\|^2 - \frac{1}{4}\|\alpha - \beta\|^2.$$

En el caso de los complejos se expresa

$$(\alpha|\beta) = \frac{1}{4}\|\alpha + \beta\|^2 - \frac{1}{4}\|\alpha - \beta\|^2 + \frac{i}{4}\|\alpha + i\beta\|^2 - \frac{i}{4}\|\alpha - i\beta\|^2.$$

Las igualdades son llamadas las identidades de polarización puede también ser escrita de la siguiente manera: $(\alpha|\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n \ \|\alpha + i^n\beta\|^2$.

(Hoffman y Kunze, 1973, p. 271)

Las propiedades obtenidas anteriormente son válidas para cualquier producto interno sobre un espacio real o complejo V, independientemente de su dimensión. Se vuelve al caso en que V es de dimensión finita.

Como es de pensarlo, un producto interno en un espacio de dimensión finita puede ser descrito siempre en términos de una base ordenada por medio de una matriz.

Supóngase que V es de dimensión finita, de modo que

$$\mathfrak{B} = \{\alpha_1\ ,\quad \alpha_2,\quad .\ .\quad .,\alpha_n\}$$

es una base ordenada de V y que se tiene dado un producto interno particular sobre V; se verá que el producto interno está completamente determinado por los valores

$$G_{ik} = (\alpha_k|\ \alpha_j)$$

que tienen los pares de vectores en $\mathfrak{B}$. Si $\alpha = \sum_k x_k\alpha_k\ \ y\ \beta = \sum_j y_j\alpha_j$, entonces

$$(\alpha|\beta) = \left(\sum_k x_n\ \alpha_k|\ \beta\right)$$

$$= \sum_k x_k \, (\alpha_k | \, \beta)$$

$$= \sum_k x_k \sum_j \bar{y}_j (\alpha_k | \, \alpha_j)$$

$$= \sum_{j,k} \bar{y}_j G_{j\,k} x_k$$

$$= \; Y^* G X.$$

donde X, Y son las matrices de coordenadas de $\alpha \; y \; \beta$ y en la base ordenada $\mathfrak{B}$, y G es la matriz con elemento $G_{ik} = \; (\alpha_k | \, \alpha_j)$.

Se llama a G **la matriz del producto interno en la base ordenada** $\mathfrak{B}$. Entonces $G = \; G^*$; sin embargo, G es una clase de matriz hermítica más bien especial.

(Hoffman y Kunze, 1973, p. 272)

En efecto, G debe satisfacer, además, la condición

$$X^* G X \; > 0, \qquad X \neq 0.$$

En particular, G debe ser inversible. En forma más explícita dice que para los escalares $x_1, \; \ldots \, , x_n$, no todos nulos,

$$= \sum_{j,k} \bar{x}_j G_{j\,k} x_k \; > 0$$

con esto se observa inmediatamente que cada elemento de la diagonal de G debe ser positivo.

El proceso anterior es reversible; esto es, si G es cualquier matriz n x n sobre F que satisface $X^* G X \; > 0, \qquad X \neq 0$ y satisface la condición de que $G = \; G^*$, entonces G es la

matriz n x n en la base ordenada $\mathfrak{B}$ de un producto interno sobre V. Este producto interno este dado por

$$(\alpha|\beta) = Y^*GX$$

donde $X\ e\ Y$ son las matrices de coordenadas de α y β en la base ordenada $\mathfrak{B}$.

(Hoffman y Kunze, 1973, p. 272)

2.13. Vector Ortogonal

Sean α y β vectores de un espacio producto interno V. Entonces α es **ortogonal** a β si $(\alpha|\beta) = 0$; como esto implica que β es ortogonal a α, a menudo solo se dirá que α y β son ortogonales. Si S es un conjunto de vectores de V, se dice que S es un conjunto ortogonal siempre que todos los pares de vectores distintos de S sean ortogonales.

Un conjunto ortonormal es un conjunto ortogonal S con la propiedad además de que $\|\alpha\| = 1$ para todo α de S. El vector cero es ortogonal a todo vector en V y es el único vector con esa propiedad. Es apropiado pensar un conjunto ortonormal como conjunto de vectores mutuamente perpendiculares que tienen longitud 1.

(Hoffman y Kunze, 1973, p. 66)

Teorema 8

Un conjunto ortogonal de vectores no nulos es linealmente independiente.

(Hoffman y Kunze, 1973, p. 277)

Corolario 2

Si un vector β es combinación lineal de una sucesión ortogonal de vectores no nulos α_1, $\alpha_2, \ldots, \alpha_n$ entonces β es igual a la combinación lineal particular

$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{\|\alpha\|^2}\ \alpha_k.$ (Hoffman y Kunze, 1973, p. 277)

Teorema 9

Sea V un espacio con producto interno y sean $\beta_1, \ \dots\ , \beta_n$vectores independientes cualesquiera de V. Entonces se pueden construir vectores ortogonales $\alpha_1, \ \dots\ , \alpha_n$, en V tales que para cada $k\ =\ 1, 2, \dots, n$ el conjunto

$$\{\alpha_1, \quad \dots\ , \alpha_k\}$$

sea una base del subespacio generado por $\beta_1, \ \dots\ , \beta_k$. (Hoffman y Kunze, 1973, p. 278).

Corolario 3

Todo espacio con producto interno de dimensión finita tiene una base ortonormal.

En esta sección se define el código R lenguaje de programación estadística cada uno de sus algoritmos, propiedades y características importantes para la programación y evolución de la innovadora tecnología biomédica.

2.14. Código R en el Lenguaje de Programación Estadístico

2.14.1 ¿Qué es R?

Es un lenguaje de programación interpretado, de distribución libre, bajo Licencia GNU, y se mantiene en un ambiente para el cómputo estadístico y gráfico. Este software corre en distintas plataformas Linux, Windows, MacOS, e incluso en PlayStation 3. (Sepúlveda & Farfán, 2014, p.7)

El término ambiente pretende caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación gradual de herramientas muy específicas y poco flexibles, como suele ser con otro software de análisis de datos. (Sepúlveda & Farfán, 2014, p.7)

2.14.2. Los Datos y sus Tipos

Todas las cosas que manipula R se llaman objetos. En general, éstos se construyen a partir de objetos más simples, que se clasifican en cinco clases a las que se denomina atómicas y que son las siguientes:

- character (cadenas de caracteres)
- numeric (números reales)
- integer (números enteros)
- complex (números complejos)
- logical (lógicos o booleanos, que sólo toman los valores True o False).

En el lenguaje, sin embargo, cada una de estas clases de datos no se encuentran ni se manejan de manera aislada, sino encapsulados dentro de la clase de objeto más básica del lenguaje: **el vector**. Un vector puede contener cero o más objetos, pero todos de la misma clase. En contraste, la clase denominada list, permite componer objetos también como una secuencia de otros objetos, pero, a diferencia del **vector**, cada uno de sus componentes puede ser de una clase distinta. (Sepúlveda & Farfán, 2014, p.13)

2.14.3. Los Datos Numéricos

El lenguaje agrupa estos datos en tres categorías, a saber: numeric, integer y complex, pero cuando se introduce algo que puede interpretar como un número, su inclinación es tratarlo como un dato de tipo numeric, es decir, un número de tipo real, a no ser que explícitamente se indique otra cosa como:

```
x <- 2                    Se asigna el valor 2 a x
print(x)                  Se imprime el valor de x
## [1] 2
```

$class(x)$ *Muestra cuál es la clase de x*

[1] "numeric"

$x < -\frac{6}{2}$ *Se asigna el valor de la operación dividir* $\frac{6}{2}$ *a x print(x)*

[1] 3

$class(x)$

[1] "numeric"

(Sepúlveda & Farfán, 2014, p.14)

Las dos asignaciones que se hacen, mediante el operador de asignación, $< -$, a la $variable\ x$, es de los enteros $2\ y\ 3$ respectivamente. Sin embargo, al preguntar, mediante la $función\ class(\ \ )$, cuál es la $clase\ de\ x$, la respuesta es $numeric$, esto es, un número real. Para asignar explícitamente un entero, integer, a una variable, se agrega la letra L al final del número, como sigue:

$x < -23L;\ print(x)$

[1] 23

$class(x)$

[1] "integer"

Aquí la variable x tendrá como valor el entero 23.

Como una nota adicional del lenguaje, nótese que se han escrito dos expresiones de R en un mismo renglón. En este caso, las expresiones se separan mediante ';'.

Para lograr que una expresión, como la operación de división $\frac{6}{2}$, arroje como resultado un entero, se tiene que hacer una $conversión$; ello se logra mediante la función $as.integer$, como sigue:

$x <- as.integer\left(\frac{6}{2}\right);\ print(x)$

[1] 3

$class(x)$

[1] "integer"

(Sepúlveda & Farfán, 2014, p.14)

Los números complejos, $complex$ en el lenguaje, tienen una sintaxis muy particular; misma que se tiene que emplear para indicar explícitamente que un número introducido corresponde a ese tipo:

$x <- 21 + 2i$

$y <- -2i + 21$ *El mismo valor que x*

$z <- -(-1 + 0i)$ *Corresponde a* -1

$tt <- sqrt(z)$ *raíz cuadrada de* -1

$print(x);\ print(y);\ print(z);\ print(tt)$

[1] $21 + 2i$

[1] $21 + 2i$

[1] $-1 + 0i$

[1] $0 + 1i$

$class(tt)$

[1] "complex"

En los ejemplos anteriores a la variable tt se le asigna el resultado de una función, $sqrt(\ )$, que es la raíz cuadrada de el número -1. Nótese que ésta es la forma correcta de

calcular esa raíz, por ejemplo, $sqrt(-1)$, hubiera arrojado como resultado un error. (Sepúlveda & Farfán, 2014, p.15)

También, existe un valor numérico especial, Inf, que representa el infinito y que puede resultar en algunas expresiones, como:

$$x < -\frac{1}{0}$$ *División por cero*

```
x
```

```
## [1] Inf
```

Tambien dividir un número por Inf da cero:

$$y < -\frac{1}{Inf}$$

```
y
```

```
## [1] 0
```

Finalmente, algunas operaciones pueden resultar en algo que no es un número, esto se representa por el valor NaN. Como sigue:

$$x < -\frac{0}{0}$$

```
x
```

```
## [1] NaN
```

(Sepúlveda & Farfán, 2014, p.15)

2.14.4. Vectores

Las clases atómicas de datos no se manejan de manera individual. En efecto, en todos los ejemplos anteriores, el lenguaje ha creado implícitamente vectores de longitud 1, y son esos los que se han asignado a las variables, por consiguiente:

```
x <- 2          # Se asigna el valor 2 a x
print(x)        # Se imprime el valor de x
```

```
## [1] 2
```

Aquí, la impresión del valor de x tiene una forma muy particular: "[1]2". El '[1]' que precede al valor, indica que se trata del primer elemento y único, en este caso, del vector que se muestra.

Hay diversas maneras de crear vectores de otras longitudes, que, como se ha dicho antes, son secuencias de objetos de la misma clase atómica. (Sepúlveda & Farfán, 2014, p.16)

2.14.4.1. Creación de Vectores a partir de Archivos de Texto - la Función $Scan(\ \ )$

La primera manera de crear vectores es a partir de los elementos individuales que compondrán el vector. Para esto se utiliza la función $c(\ \ )$ como se muestra a continuación.

```
c(4, 2, -8)                 Creación de un vector sin asignarlo a una variable

## [1] 4 2 -8

## ------------------
## Distintas formas de asignar un vector a una variable
u <- c(4,2,-8)                                  Usando el operador <-
c(4, 2, -8) -> v                                Usando el operador ->
                                                Usando la función assign:
assign(w, c(4, 2, -8))
p = c(4,2,-8)                                   Usando el operador =
print(u); print(v); print(w); print(p)

## [1] 4 2 -8
## [1] 4 2 -8
## [1] 4 2 -8
## [1] 4 2 -8
```

(Sepúlveda & Farfán, 2014, p.16)

La función $c(\ )$ sirve para concatenar varios elementos del mismo tipo. En todos los ejemplos mostrados, la impresión del vector se hace en un renglón que comienza con el símbolo '[1]', indicando con ello que el primer elemento del renglón corresponde al primer elemento del vector.

2.14.4.2. Creación de Vectores a partir de Secuencias y otros Patrones

Un vector, inicializado en ceros, o $FALSE$, y de longitud determinada, se puede crear con la función $vector(\ )$. Es esta misma función la que permite crear vectores sin elementos, a continuación:

$$v\ <-\ vector(\text{integer}, 0)$$

$$v \qquad\qquad Un\ vector\ de\ enteros\ \text{sin}\ elementos$$

$$\#\#\ integer(0)$$

$$w\ <-\ vector(\text{numeric}, 3)$$

$$w \qquad\qquad Un\ vector\ de\ tres\ ceros$$

[1] 0 0 0

$$u\ <-\ vector(\text{logical}, 5)$$

$$u \qquad\qquad Un\ vector\ de\ 5\ FALSE$$

$$\#\#\ [1]\ FALSE\ FALSE\ FALSE\ FALSE\ FALSE$$ (Sepúlveda & Farfán, 2014, p.18)

El operador ':' permite generar un vector entero a partir de una secuencia creciente o decreciente de enteros, cuyos extremos se indican, tal como se muestra en seguida:

$$1{:}\,3$$

[1] 1 2 3

$v <- 40:13$

$print(v);\ class(v)$ *El vector y su clase*

[1] 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23

[19] 22 21 20 19 18 17 16 15 14 13

[1] "integer"

Nótese que el desplegado o impresión del vector v, se ha tenido que hacer en dos renglones. Cada uno de esos renglones comienza, indicando entre corchetes [ ], el índice del primer elemento en el renglón. El alcance del operador ': ' no se limita, sin embargo, sólo a números enteros, como:

$v <- pi:6$

$print(v);\ class(v)$ *El vector y su clase*

[1] 3.142 4.142 5.142

[1] "numeric"

pi simboliza el valor de la constante matemática π $'3.1416$, y la secuencia de números reales que se produce es con incrementos de 1 a ese valor hasta mientras que no se rebase el límite superior, 6, en este caso.

Por otra parte, este operador es un caso particular de la función $seq(\ \ )$ que permite generar una mayor variedad de secuencias numéricas, a continuación:

$v <- seq(from = 5, to = 15, by = 2)$

$print(v)$ *secuencia desde 5 hasta 15 de 2 en 2*

[1] 5 7 9 11 13 15 (Sepúlveda & Farfán, 2014, p.19)

2.14.4.3 Acceso a los Elementos Individuales de un Vector

Dentro de un vector, sus elementos se pueden identificar mediante un índice entero, que en el caso de este lenguaje empieza con el 1 así,

```
v <- c(8, 7, -3, 2, 182)

v[5]                                   El quinto elemento

## [1] 182

print(v[1]); print(v[3])

## [1] 8 #
# [1] -3

v[4] + v[2]                    La suma del cuarto y segundo elementos de v

## [1] 9
```

Primeramente, se ha accedido al quinto elemento, mediante $v[5]$; si el intérprete de R se está usando interactivamente, el valor de ese elemento, 182 , se imprime implícitamente. Luego se manda imprimir explícitamente, los elementos 1 y 3 del vector.

Finalmente, se suman los elementos 4 y 2 del vector; el resultado de esa operación se imprime implícitamente, es decir, de manera automática, si la operación se solicita al intérprete en su modo interactivo. El acceso a los elementos individuales de un vector no solamente es para *consulta o lectura*, sino también para su *modificación o escritura*, Supóngase que se tiene el registro de cantidades de ciertas frutas en un vector, como:

```
frutas <- c(15,  100,  2,  30)
frutas
## [1] 15  100  2  30
```

(Sepúlveda & Farfán, 2014, p.20)

Supóngase ahora que se quiere asociar esos valores con el nombre de la fruta correspondiente:

```
names(frutas) <- c(naranja, pera, manzana, durazno)
```

Si ahora se manda desplegar el vector:

$Frutas$

$\#\# naranja \;\; pera \;\; manzana \;\; durazno$

$\#\#\ 15 \qquad 100 \qquad 2 \qquad 30$ (Sepúlveda & Farfán, 2014, p.21)

2.14.4.4. Operaciones Sencillas con Vectores

Las operaciones aritméticas más comunes están definidas para vectores: la suma, la resta, la división y la exponenciación, todas ellas se definen elemento a elemento entre dos vectores como:

$v <- 2 + 3$ *Resulta en un vector de longitud 1*

v

$\#\#\ [1]\ 5$

$v <- c(2,3) - c(5,1)$ *Resulta en un vector de longitud 2*

v

$\#\#\ [1]\ -3\ 2$

$v <- c(2,3,4) * c(2,1,3)$ *Resulta en un vector de longitud 3*

v

$\#\#\ [1]\ 4\ 3\ 12$

$v <- c(2,3,4)^{3:1}$ *Eleva a potencias 3, 2, 1*

v

$\#\#\ [1]\ 8\ 9\ 4$

En todos los casos, la operación indicada se aplica elemento a elemento entre los dos vectores operandos. (Sepúlveda & Farfán, 2014, p.22)

2.14.5 Matriz

El punto de vista del lenguaje, una matriz es un vector con un atributo adicional: dim. Para el caso de las matrices, este atributo es un vector entero de dos elementos, a saber: el número de renglones y el número de columnas que componen a la matriz.

Las matrices también se pueden crear de manera flexible por medio de la función primitiva $matrix(\quad)$, que permite alterar la secuencia por $default$ de armado de la matriz; esto es, ahora, si se quiere, se puede armar la matriz por renglones en vez de columnas:

```
(m <- matrix( 11:30, nrow = 5, ncol = 4, byrow = TRUE) )

##      [,1]  [,2] [,3]  [,4]
## [1,]   11   12   13   14 .
## [2,]   15   16   17   18 .
## [3,]   19   20   21   22 .
## [4,]   23   24   25   26 .
## [5,]   27   28   29   30 .
```

Adicionalmente, a los renglones y las columnas de una matriz se les pueden asignar nombres, que pueden ser después consultados o usados como índices:

```
rownames(m) <- c("uno","dos","tres","cuatro","cinco")
colnames(m) <- c("UNO","DOS","TRES","CUATRO")
m
##        UNO  DOS  TRES  CUATRO .
## uno     11   12    13     14 .
## dos     15   16    17     18 .
## tres    19   20    21     22.
## cuatro  23    24   25     26.
## cinco   27   28   29     30.
```

(Sepúlveda & Farfán, 2014, p.28)

2.14.5.1. Data $Frames$.

Un data $frame$ es una lista, cuyos componentes pueden ser vectores, matrices o factores, con la única salvedad de que las longitudes, o número de renglones, en el caso

de matrices, deben coincidir en todos los componentes. La apariencia de un data $frame$ es la de una tabla y una forma de crearlos es mediante la función data.$frame(\ \ )$. Como:

$$\left(m\ <-\ cbind\left(ord\ =\ 1{:}3, edad\ =\ c(30L, 26L, 9L)\right)\right)$$

$$\#\#\quad ord\quad edad$$

$$\#\#\ [1,]\quad 1\quad 30$$

$$\#\#\ [2,]\quad 2\quad 26$$

$$\#\#\ [3,]\quad 3\quad 9$$

$$\left(v\ <-\ c(1.80, 1.72, 1.05)\right)$$

$$\#\#\ [1]\ 1.80\ 1.72\ 1.05$$

$$ff\ <-\ data.frame(familia\ =\ c(\text{Padre}, \text{Madre}, \text{Hijo}), m, estatura\ =\ v)$$

$$ff$$

##	familia	ord	edad	estatura
## 1	Padre	1	30	1.80
## 2	Madre	2	26	1.72
## 3	Hijo	3	9	1.05

Una gran ventaja de los $data\ frames$, es que R tiene diversas funciones para leer y guardar las tablas que representan, en archivos de texto, y otros formatos. (Sepúlveda & Farfán, 2014, p.41)

2.14.5.2. Matrices y $Data\ Frames$

Las matrices y los $data\ frames$, son estructuras bidimensionales; es decir, tienen renglones y columnas, y por consiguiente, su comportamiento bajo el operador [ ] es similar. El acceso a los elementos individuales a estas dos estructuras consta de dos índices, separados por una coma, en el operador, así: $x[i, j]$; donde, x, es la estructura en cuestión,

i, representa el número o identificador de renglón y j, el número o identificador de columna. Para las explicaciones y ejemplos que siguen se usarán las matrices y $data\ frames$ que se generan a continuación:

```
(m <- matrix( 11:30, nrow = 5, ncol = 4, byrow = TRUE) )

##      [,1]  [,2] [,3]  [,4]
## [1,]  11   12   13   14 .
## [2,]  15   16   17   18 .
## [3,]  19   20   21   22 .
## [4,]  23   24   25   26 .
## [5,]  27   28   29   30 .
# Se convierte en la matriz a una data frame
df.mt <- as.data.frame (mt)

# Se le asignan nombres a renglones y columnas de df.mt
rownames(m) <- c("uno","dos","tres","cuatro","cinco")
colnames(m) <- c(UNO, DOS, TRES, CUATRO)
m

##        UNO  DOS  TRES  CUATRO .
## uno     11   12    13     14 .
## dos     15   16    17     18 .
## tres    19   20    21     22.
## cuatro  23   24    25     26.
```

(Sepúlveda & Farfán, 2014, p.53)

2.14.6. Funciones de Clasificación, Transformación y Agregación de Datos

La riqueza de este lenguaje, sin embargo, está en el manejo de cada una de las distintas estructuras de información, implementadas a través de las clases de datos estructuradas, como $vectores, factores, data\ frames$, etc., como un todo a través de funciones que las contemplan de esa manera.

Las funciones de transformación, clasificación y agregación de datos, se puede comprender con el cálculo del módulo o magnitud de un vector de números reales y el cálculo del promedio, o media, de un conjunto de datos provistos como un vector numérico.

2.14.6.1. Módulo o Magnitud

Dado un vector n-dimensional:

$$\mathbf{V} = \langle v_1, v_2, \ldots, v_n \rangle$$

su módulo o magnitud está dado por:

$$|\mathbf{v}| = \sqrt{\sum_{i=1}^{n} {v_i}^2}$$

La estructura del operador $\square^{\square}$, sirve para elevar a una potencia dada y la función $sqrt(\ \ )$, extrae la raíz cuadrada, las operaciones aritméticas se pueden distribuir a lo largo de los elementos de un vector, y esta es una característica del lenguaje en la que se puede sacar provecho; recurriendo adicionalmente a la función de agregación $sum(\ \ )$, que toma como argumento un vector y suma uno a uno sus elementos y entrega esa suma como resultado. (Sepúlveda & Farfán, 2014, p.70)

Entonces en R, la función $modulo(\ \ )$, se puede programar así:

```
modulo0 <- function(v){sqrt(sum(v^2))}
```

Puede quedar en una línea, así:

```
modulo0 <- function(v)sqrt(sum(v^2))
```

y la utilizamos igual:

```
modulo0(vv)
```

```
## [1] 5.385
```

Esta forma de programación se apega mucho más a la fórmula matemática.

2.14.6.2. La Media

Para el mismo vector, **v**, definido con anterioridad, la siguiente fórmula sirve para obtener la media:

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

En un lenguaje de programación, esto se haría así:

```
# Primeramente construyamos un vector con elementos numéricos
                                    arbitrarios, con un generador de números aleatorios
                                    (distribución uniforme), entre 10.5 y 40.8, generamos 32
                                    números, así:
nums <- runif(32, 10.5, 40.8)
        suma <- 0                   Iniciamos la suma como cero for (elt in nums)
{ suma <- suma + elt # se agrega cada elemento }
                                                    Se calcula e imprime el promedio
(promedio <- suma/length(nums))

## [1] 28.66
```

(Sepúlveda & Farfán, 2014, p.71)

2.14.6.3. La Función $accumulate = TRUE$

La función entrega lo que la operación da cada vez que se añade un elemento del objeto a la operación validando los datos vectoriales de tal manera que estos sean

equivalentes. Esta misma idea de operar con los elementos de los vectores, puede ser llevada más allá, a operar con los elementos de datos más estructurados como pueden ser las listas y los $data\ frames$. Para ello, el lenguaje cuenta con un conjunto de funciones de transformación, clasificación y agregación de datos, como se muestra a continuación:

$$mif\ <-\ function(a,b)\left\{return\left(a\ *\frac{b}{a\ +\ b}\right)\right\}$$

$$\#$$

$$miOp\ <-\ function(v)Reduce(mif,v)$$

$$\#\ Note{:}\ mif\ va\ \text{sin}\ comillas\ \#\ Apliquemos{:}\ miOp(vv)$$

$$\#\#\ [1]\ 12$$

$$Otra\ versión,\ con\ resultados\ parciales\ y\ en\ la\ que\ incluso\ la\ función\ se\ da\ 'en\ linea'{:}$$

$$miOpV2\ <-\ function(v)Reduce\left(function(a,b)a\ *\frac{b}{a\ +\ b},v,accumulate\ =\ TRUE\right)miOpV2(vv)$$

$$\#\#\ [1]\ -2\ -6\ 12$$

El lenguaje cuenta con un conjunto de funciones de transformación, clasificación y agregación de datos. (Sepúlveda & Farfán, 2014, p.73)

2.14.6.4. Operaciones Marginales en Matrices y la Función $apply(\ \ )$

Como objetos numéricos, las matrices resultan útiles para hacer diversidad de cálculos. Una de sus principales características es que tanto sus renglones como sus columnas pueden ser tratados como elementos individuales. De esta forma, hay operaciones que se efectúan para todas sus columnas o para todos sus renglones; a estas se les denominará operaciones marginales.

El lenguaje tiene algunas de estas operaciones implementadas directamente, entre ellas están las funciones:

$rowSums(\ \ ), colSums(\ \ ), rowMeans(\ \ ) y\ colMeans(\ \ ).$

A continuación, se calcularán las sumas de las columnas de una matriz y los promedios de sus renglones.

```
# Se hace una matriz arbitraria de 3 renglones y 5 columnas,
                                    con rbind, que la construye por renglones:
(mm <- rbind(5:9,      runif(5, 10, 20),      c(2, -2, 1, 6, -8)))
##          [,1]    [,2]     [,3]     [,4]    [,5]
## [1,]    5.00   6.00     7.00     8.00    9.00
## [2,]   18.76  17.41   10.05    18.43   17.36
## [3,]    2.00  -2.00     1.00     6.00   -8.00
colSums(mm)

## [1]  25.76   21.41  18.05  32.43  18.36

rowMeans(mm)

## [1]    7.0   16.4   -0.2
```

Estas operaciones, que se tienen implementadas directamente, evitan mucha programación que se tendría que hacer con ciclos y operaciones individuales con los elementos de la matriz. El lenguaje provee además la posibilidad de construir el mismo tipo de operación marginal para el caso general de cualquier función, mediante la función $apply(\ \ )$. Por ejemplo, la función $sd(\ \ )$, calcula la desviación estándar para un conjunto de números dados como un vector numérico. (Sepúlveda & Farfán, 2014, p.75)

2.14.6.5. Estructura Formal de una Función

R trata las funciones prácticamente igual que cualquier otra variable. Así, ellas se pueden manipular de manera semejante a como se hace con otros objetos de R: se pueden pasar como argumentos de otras funciones, se pueden regresar como el valor final de una función, se pueden definir en el interior de otra función, etc. Para crear o definir una función, se emplea la directiva "$function$", en general, asociándola con un símbolo mediante una operación de asignación, como se muestra en la Figura.2. Esta directiva, tiene dos partes, la definición de los argumentos formales de la función, y el cuerpo de la función. El cuerpo de la función está constituido por una o más expresiones válidas del lenguaje.

Figura 2. ***Definición de la función***

Tomada de: "*El arte de programar en R: un lenguaje para la estadística".* Santana Sepúlveda, S., & Mateos Farfán, E. 2014. P.88.

2.14.6.6. Revisión de los Argumentos de una Función

Una vez definida una función, hay dos funciones de R que permiten revisar su lista de argumentos formales, a saber: $args(\ \ )y\ formals(\ \ )$. En el siguiente código se ve el comportamiento de cada una de ellas.

$args(MiFunc.v2)$

$\#\#\ function\ (x, yyy, z\ =\ 5, t)$

```
## NULL
```

```
(ar <- formals(MiFunc.v2))
```

puede o no llevar comillas

```
## $x
##
##
## $yyy
##
##
## $z
## [1] 5
##
## $t
```

Si se quiere revisar el argumento z, se hace así:

```
ar$z
## [1] 5
```

La función *args*( ), entrega lo que sería el encabezado de la función, desprovisto de su cuerpo; prácticamente, permite ver la lista de argumentos de la función tal como fue definida. Esta función le resulta útil a un usuario del lenguaje para revisar la lista de argumentos de cualquier función. (Sepúlveda & Farfán, 2014, p.91)

La función *formals*( ), digiere un poco más la información ya que entrega una lista con cada uno de los argumentos formales y los valores asignados por omisión. Esta función le resulta más útil a un programador que desea manipular esta lista. En efecto, esta lista es una clase de objeto de R conocido como "*pairlist*", cada uno de cuyos elementos tiene, por un lado, como nombre el nombre de uno de los argumentos y como valor, el valor correspondiente por omisión, sin evaluar aún. Estas funciones resultan útiles para revisar los argumentos de funciones de biblioteca, que pueden ser muchos, y sus valores por omisión.

Si se quiere revisar esto para la función $lm(\ \ )$, que se usa para ajustes lineales, se puede hacer con:

```
args(lm).
## function (formula, data, subset, weights, na.action,
## method = "qr", model = TRUE, x = FALSE, y = FALSE,
## qr = TRUE, singular.ok = TRUE, contrasts = NULL,
## offset, ...)
## NULL
```

(Sepúlveda & Farfán, 2014, p.92)

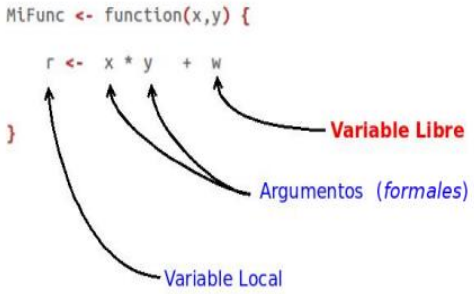
R mantiene por separado los espacios de nombres de las funciones y de todos los otros objetos que no son funciones. De este modo, se pudiera definir un símbolo "t", asociado a un vector, por ejemplo, y aún así tener acceso a la función $t(\ \ )$, veamos:

```
t <- c(1,2,3)          definimos vector
t(mx)                  usamos la funcion t
##      [,1] [,2]
## [1,]    4    5
## [2,]    6    7
t                      pero al consultar t, tenemos ...
## [1] 1 2 3
```

2.14.6.7. Reglas de Alcance

Se refieren a la forma como el lenguaje resuelve la asociación entre una variable o símbolo libre y su correspondiente valor. En una función, una variable libre es aquella que no se ha definido en el código de esta, y cuyo valor, sin embargo, se solicita, al incluir su nombre o símbolo correspondiente, como la figura 3.

Figura 3. ***Tipos de símbolos en el interior de una función***

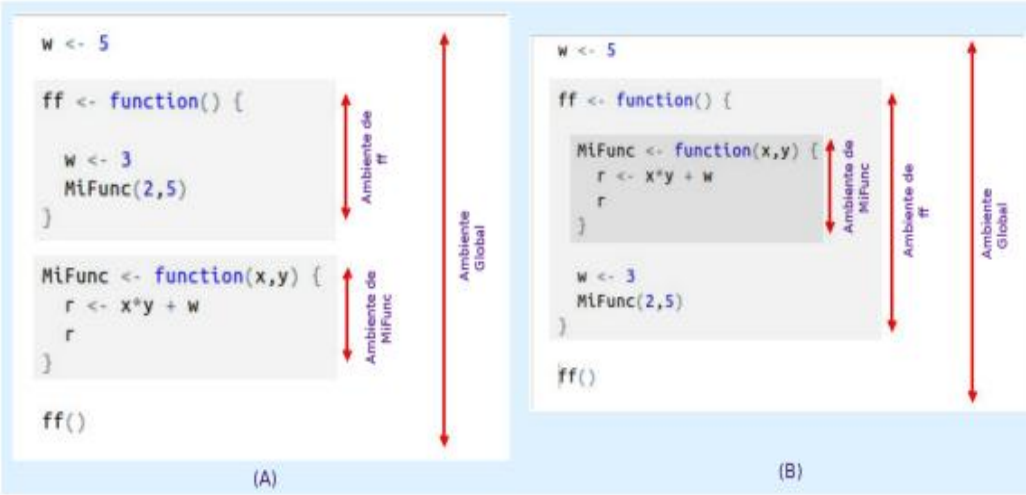


Tomada de: "*El arte de programar en R: un lenguaje para la estadística".* Santana Sepúlveda, S., & Mateos Farfán, E. 2014. p.98.

Para asociar las variables libres con un valor, R utiliza lo que se conoce como $alcance\ léxico$. En este tipo de alcance se establece que *los valores de las variables se buscan en el ambiente en el cuál la función se definió*. Como se puede intuir de lo dicho antes, un ambiente es una colección de pares, donde, por ejemplo, un símbolo podría ser $MiSimbolo$, y su valor 5, o el símbolo t y su valor el $vector\langle 1,\ 2,\ 3\rangle$. Ahora bien, en R los ambientes se organizan en una jerarquía; esto es, cada ambiente tiene un ambiente padre y a su vez puede tener cero o más hijos. El único ambiente que no tiene padre es el ambiente vacío, que se encuentra en lo más alto de la jerarquía.

Así pues, el lugar dónde R buscará el valor de una variable libre, dependerá del lugar dónde se encuentre escrito el código de la función en el que se hace referencia a ella. Por ejemplo, el caso de la función que se ha mostrado en la figura. 3, y veamos la diferencia en el valor que tomará la variable "w", al insertar el código de la función en distintas partes.

Figura 4. ***Jerarquía en los ambientes de las funciones***



Tomada de: "*El arte de programar en R: un lenguaje para la estadística".* Santana Sepúlveda, S., & Mateos Farfán, E. 2014. P.99.

La Figura. 4 muestra la inserción de la misma función, $MiFunc(\ \ )$, en dos distintas partes del código global. La jerarquía de los ambientes es diferente en cada caso: en la situación marcada como (A), el ambiente global tiene dos hijos el correspondiente a la función $\mathrm{ff}(\ \ )$ y el correspondiente a $MiFunc(\ \ )$, mientras que en la situación (B), el ambiente global tiene un solo hijo, el ambiente correspondiente a la función $\mathrm{ff}(\ \ )$, que a su vez tiene un hijo, el correspondiente a la función $MiFunc(\ \ )$. Una nota adicional aquí es que una función puede ser definida en el interior de otra función y eso es precisamente lo que ilustra la situación (B) de la figura. Para resolver el valor de "w", o sea de una variable libre, en cualquier caso, R busca primero en el ambiente de la misma función, si no encuentra el símbolo ahí, procede a buscar en el ambiente padre, y así lo sigue haciendo con todos los predecesores (Sepúlveda & Farfán, 2014, p.92) hasta encontrar una asociación. Si en este proceso, llega al ambiente vacío sin encontrar una asociación posible, el lenguaje emitirá un mensaje de error. En el código que sigue se muestra el comportamiento de R las situaciones mostradas en la Figura 4, y se añade una situación más para ejemplificar lo que se ha explicado aquí.

```
w <- 5                                  Primera w

ff <- function( ){w <- 3 # Segunda w MiFunc(2,5)}

MiFunc <- function(x,y){r <- x*y + w r }

ff( )

## [1] 15
```

En este caso, a pesar de que en el interior de la función ff( ) se ha definido un valor para w de 3, el lenguaje ha resuelto, de acuerdo con la regla explicada previamente, que el valor de w en $MiFunc(\ )$ es el del ambiente global, esto es, 5, y por ello el resultado que se despliega es 15. (Sepúlveda & Farfán, 2014, p.99)

2.14.6.8. Gráfico de la Función

2.14.6.8.1. La función más Básica de Graficación: $plot(\ )$

Se elaborará también una función particular, como se hizo para el método anterior, que pueda ser graficada mediante la función $curve(\ )$, de la siguiente manera

```
# Los parámetros encontrados se encuentran en los componentes
                                        de la variable 'ff', como sigue:
                                        (1)shape: ff$estimate[[1]]
                                        (2)scale: ff$estimate[[2]]
                                Entonces la función particular para estos parámetros es:

dgammaXX <- function(x){ dgamma(x, shape = ff$estimate[[1]], scale
                                = ff$estimate[[2]])}
```

(Sepúlveda & Farfán, 2014, p.122)

Ahora se procede a hacer la gráfica comparativa con las dos curvas como sigue:

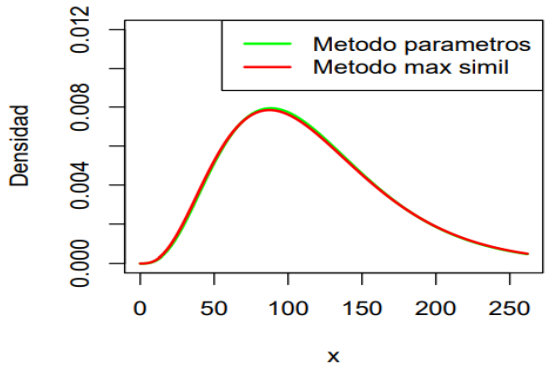
```
                                        La primera curva; agregaremos límites al
                                        eje X para su graficado

curve(dgammaX, col = "green", lwd = 2,
```

```
    ylab = "Densidad",
                                    ylim = c(0,0.012),          Límites de Y
    xlim = c(0, max (pp$Precip))) # Límites de X
                                    La otra curva se agrega a la anterior,
                                    por eso el argumento add = T
curve(dgammaXX, col = red, lwd = 2, add = T)
                                    Se agregará un leyenda para hacer más
                                    comprensible el gráfico:
```

(Sepúlveda & Farfán, 2014, p.122)

Figura 5. ***Comparación de dos métodos de ajuste de la curva de densidad de probabilidades***



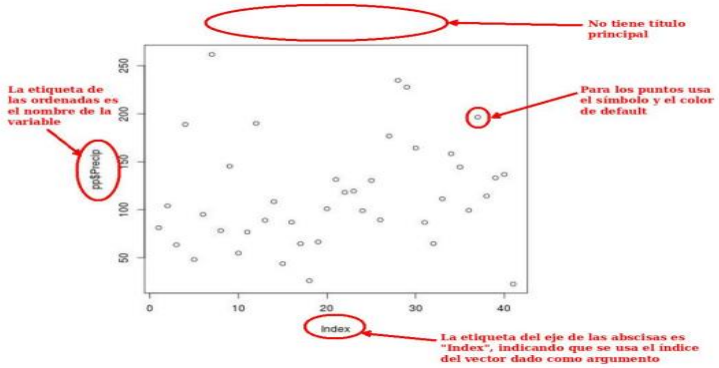
Tomada de: "*El arte de programar en R: un lenguaje para la estadística".* Santana Sepúlveda, S., & Mateos Farfán, E. 2014. p.123.

```
legend("topright",
       c(Metodo parametros, Metodo max simil),
       lwd = 2, col = c("green","red"))
```

La función más simple para graficar es `plot( )`. Antes de explicar sus posibilidades, veremos qué es lo que hace en el caso más sencillo. Del ejemplo anterior

```
plot(pp$Precip)                    Produce un gráfico de dispersión
```

Figura 6. ***Un sencillo gráfico de dispersión***



Tomada de: "*El arte de programar en R: un lenguaje para la estadística".* Santana Sepúlveda, S., & Mateos Farfán, E. 2014. p.126.

El resultado de la operación anterior se puede ver en la Fig. 6., con algunas anotaciones acerca de los valores que ha tomado el lenguaje por omisión. En la función $plot(\ \ )$ toma como argumentos dos vectores de la misma dimensión, uno para los valores de las abscisas, x, y otro para los valores de las ordenadas, y. Sin embargo, cuando se omite uno de ellos, el lenguaje entiende que las abscisas serían simplemente los índices de los elementos en el vector provisto, y que las ordenadas serán, por tanto, cada uno de los elementos del vector. De igual manera, el lenguaje supone que lo que se quiere graficar son los puntos como tal y con el color y tipo de símbolo seleccionados por omisión. (Sepúlveda & Farfán, 2014, p.126)

La función $plot(\ \ )$ tiene una gran variedad de argumentos que permiten cambiar ese comportamiento por omisión y que pueden ser consultados en línea mediante la instrucción "$?plot$".

Las definiciones, conceptos, ejemplos y teoremas, expuestos en las páginas anteriores, se emplean en la actualidad para garantizar la programación y ejecución de los equipos que la ingeniería biomédica utiliza para detectar cualquier problema de salud en los pacientes. Además, tener clara estas definiciones será fundamental para la comprensión

de los temas a tratar en el siguiente capítulo referente al marco metodológico. Puesto que allí, se muestra la utilidad del Código R (Lenguaje de programación estadístico), y los elementos del “álgebra lineal” que se utilizan para su programación.

La teoría del “álgebra lineal” es esencial para poder explicar, desarrollar los avances mencionados, por ende, son fundamentales los temas conceptuales del: producto punto y longitud, vectores unitarios, espacio y sub espacio de vectores, la desigualdad triangular, distancia entre vectores, ángulos, ortogonalidad de vectores, álgebra de vectores, ecuaciones de lineales, determinantes, matrices simétricas, entre otros importantes para el desarrollo, uso y avance de la ingeniería biomédica. (Lay, 2012)

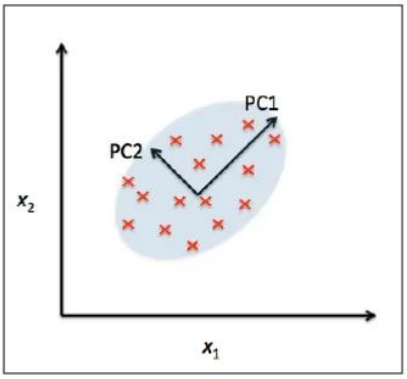
CAPÍTULO 3. Implementación del Álgebra Lineal en la Función Clasificadora

En esta sección, se presentan, explican y demuestran los algoritmos, teoremas y funciones del “álgebra lineal” que son utilizados como herramientas de soporte en la ingeniería biomédica. En la actualidad, se estudia, demuestra y desarrolla los avances innovadores de los equipos usados por los especialistas de la salud en el campo de la ingeniería biomédica; en donde se aplica el “álgebra lineal” a la programación de estos software y hardware, para que den resultados precisos de los pacientes y se puedan intervenir inmediatamente. A continuación, se aborda temas como: Hiperplanos Separados, Perceptrón o Neurona Artificial, Convergencia, Variable Duales, Algoritmo de Aprendizaje, el análisis del PCA, entre otros relacionados al código R.

3.1. Fundamentos Teóricos del Análisis de Componentes Principales (PCA) con implementación de Código R (lenguaje de programación estadístico)

El análisis de componentes principales (ACP)) es un método estadístico que se utiliza para la extracción de características. Principal Component Analysis (PCA) se utiliza para datos de alta dimensión y altamente correlacionados. Es un método estadístico cuya utilidad radica en la reducción de la dimensionalidad de la base de datos (BDD) con la que estamos trabajando. Simplificando la base de datos, ya sea para elegir un menor número de predictores para pronosticar una variable objetivo, o para comprender una BDD de una forma más simple. La idea básica de PCA es transformar el espacio original de características en el espacio de componentes principales. (Obi Tayo, 2022)

Figura 7. ***El algoritmo PCA se transforma del espacio de características antiguo al nuevo para eliminar la correlación de características***



Tomada de: *"Matemáticas del Análisis de Componentes Principales con Implementación de Código R", (p. 1) por Tayo Benjamín, 2020, Towardas AI.*

Una transformación PCA logra lo siguiente:

a) Reducir la cantidad de características que se utilizarán en el modelo final centrándose solo en los componentes que representan la mayor parte de la variación en el conjunto de datos.

b) Elimina la correlación entre características. (Obi Tayo, 2022)

3.2. Base Matemática de PC

Al hacer la suposición de que tenemos una matriz de características altamente correlacionada con 4 características y n observaciones como se muestra.

Tabla 1. *Matriz de características con* 4 *variables y* n *observaciones.*

$X_1^{(1)}$	$X_2^{(1)}$	$X_3^{(1)}$	$X_n^{(1)}$
$X_1^{(2)}$	$X_2^{(2)}$	$X_3^{(2)}$	$X_n^{(2)}$
$X_1^{(3)}$	$X_2^{(3)}$	$X_3^{(3)}$	$X_n^{(3)}$
.	.	.	.
.	.	.	.
.	.	.	.
$X_1^{(n)}$	$X_2^{(n)}$	$X_3^{(n)}$	$X_3^{(n)}$

Nota: Para visualizar las correlaciones entre las características, podemos generar un gráfico de dispersión, y para cuantificar el grado de correlación entre las características, podemos calcular la **matriz de covarianza** mediante la ecuación: (Obi Tayo, 2022).

$$\sigma_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_J{}^{(i)} - \mu_j}{\sigma_j}\right)\left(\frac{X_k{}^{(i)} - \mu_k}{\sigma_k}\right)$$

Dónde μ_i y ρ_j son la media y la desviación estándar de la característica X_j, respectivamente.

En forma de matriz, la matriz de covarianza se puede expresar como una matriz simétrica $4\ x\ 4$:

$$\sum\begin{bmatrix} \sigma_1{}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2{}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3{}^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4{}^2 \end{bmatrix}$$

Esta matriz se puede diagonalizar realizando una transformación unitaria (transformación PCA) para obtener lo siguiente:

$$\sum\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$$

Dado que la traza de una matriz permanece invariable bajo una transformación unitaria, observamos que la suma de los valores propios de la matriz diagonal es igual a la varianza total contenida en las características $\mathrm{X}_{1,}\mathrm{X}_2, \mathrm{X}_{3,}\mathrm{X}_4$. (Obi Tayo, 2022). Por lo tanto, podemos definir las siguientes cantidades:

$$Relación\ de\ varianza\ explicada = \frac{\lambda_j}{\sum_{j=1}^{4}\lambda_j}$$

$$varianza\ acumulada = \frac{\sum_{j=1}^{p} \lambda_j}{\sum_{j}^{4} \lambda_j}$$

Observe que cuando $p = 4$*, la varianza acumulada se vuelve igual a* 1 *como se esperaba.* (Obi Tayo, 2022)

R Implementada al PC

Tabla 2

Ilustración de cómo funciona el PCA utilizando el conjunto de datos del Iris[2]*.*

1	*Código R para realizar el análisis de componentes principales (PCA) con el fin de eliminar las correlaciones de características.*
2	
3	*# autor: Benjamín O. Tayo*
4	
5	*# fecha: 25-10-2018*
6	
7	*# importar bibliotecas necesarias*
8	*biblioteca (GGally)*
9	*biblioteca (tidyverse)*
10	*biblioteca (readr)*
11	
12	*importar conjunto de datos de la base de datos uci.edu*
13	
14	*ruta <- " https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data "*
15	*df <- leer.csv (ruta)*
16	*nombres (de) <- c ("longitud_sepalo ", " anchura_sepalo ", "longitud pétalo ", "anchura_pétalo ", " clase ")*
17	*df % > %cabeza()*

[2] Método de identificación biométrica con los rasgos característicos de la pupila del ojo. Identifica la persona en el ámbito digital.

18	
19	
20	*# calcular matriz de covarianza*
21	
22	*df % > %seleccionar(- clase)% > %cor()% > %redondear(3)*
23	
24	*# visualizar diagramas de densidad, diagramas de dispersión y coeficientes de correlación entre las características originales*
25	
26	*ggpairs(df , columnas = c(" longitud_sepalo ","nchura_sepalo" , "longitud_pétalo " , " anchura_pétalo "))*
27	
28	*# realizar PCA*
29	
30	*pca <- prcomp(df [, 1 : 4], escala. = TRUE)*
31	
32	*# extraer componentes importantes*
33	*resumen (pca)*
34	
35	*# calcular la matriz de covarianza en el espacio PCA*
36	
37	*cor(pca $ x [, 1 : 4])% > %redondo(3)*
38	
39	*# convertir la matriz PCA en un marco de datos*
40	
41	*pcdf <- como.datos.marco(pca $ x [, 1 : 4])*
42	*Nombres (pcdf) <- c(" PC1 " , " PC2 " , " PC3 " , " PC4 ")*
43	

44	*# visualizar diagramas de densidad, diagramas de dispersión y coeficientes de correlación entre características transformadas*
45	*ggpairs(pcdf , columnas = c(" PC1 " , " PC2 " , " PC3 " , " PC4 "))*

Nota: (Obi Tayo, 2022)

Se observa la matriz de covarianza:

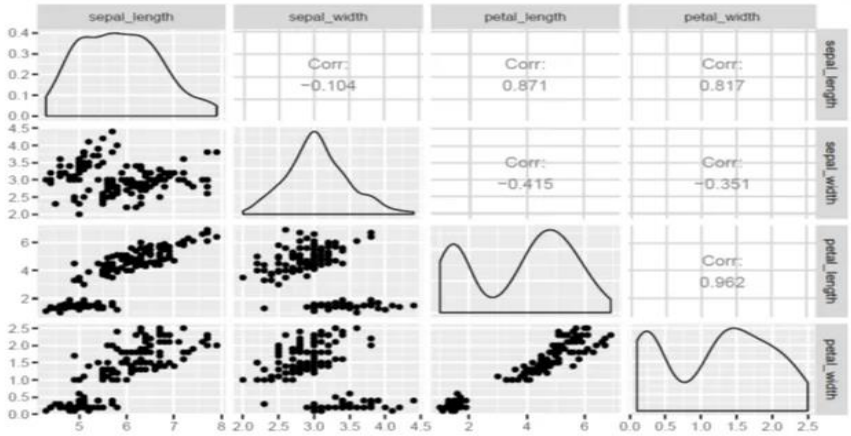
Tabla 3.

Matriz de correlación para el conjunto de datos del iris.

	Longitud del sépalo	*Anchura del sépalo*	*Longitud del pétalo*	*Ancho del pétalo*
Longitud del sépalo	*1.000*	*-0.104*	*0.871*	*0.871*
Anchura del sépalo	*-0.104*	*1.000*	*-0,145*	*-0.351*
Longitud del pétalo	*0.871*	*-0.415*	*1.000*	*0.962*
Ancho del pétalo	*0.817*	*-0.351*	*0.962*	*1.000*

Nota: muestra fuertes correlaciones entre las características originales en el conjunto de datos del iris (Obi Tayo,2022)

Figura 8. ***Gráfico de pares para el conjunto de datos de iris en el espacio de características original***



Nota: es un diagrama de pares que muestra diagramas de dispersión, diagramas de densidad y coeficientes de correlación entre las características originales. Observe las fuertes correlaciones entre las características originales. Tomado de: Tomada de: "Matemáticas del Análisis de Componentes Principales con Implementación de Código R", (p. 3) por Tayo Benjamín, 2020, Towardas AI.

Examinemos ahora la matriz de covarianza transformada:

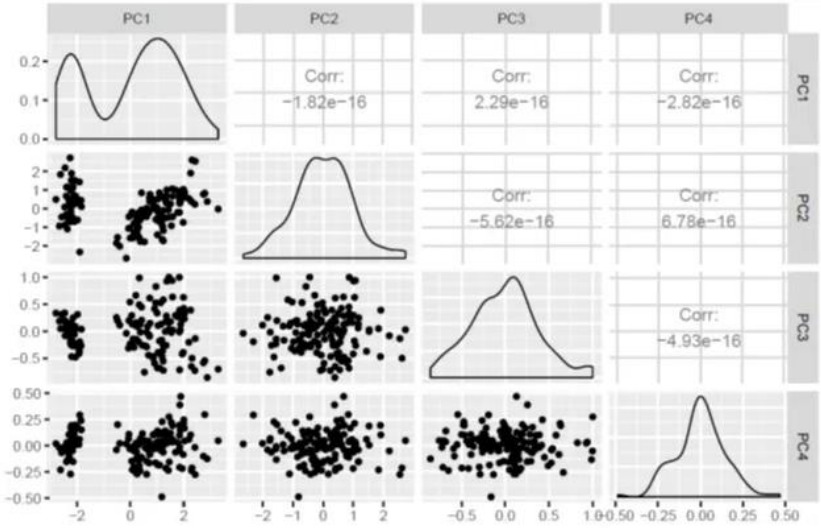
Tabla 4.

Matriz de Convergencia del Espacio PCA.

	$PC1$	$PC2$	$PC3$	$PC3$
$PCA1$	1	0	0	0
$PCA2$	0	1	0	0
$PCA3$	0	0	1	0
$PCA4$	0	0	0	1

Nota: muestra correlaciones cero entre las características transformadas. (Obi Tayo, 2022).

Figura 9. ***Gráfico de pares para el conjunto de datos de iris en el espacio PCA***



Nota: muestra el diagrama de pares en el espacio PCA. Vemos que se ha eliminado la entre características. Tomada de: "Matemáticas del Análisis de Componentes Principales con Implementación de Código R", (p. 3) por Tayo Benjamín, 2022, Towardas AI.

Tabla 5.

Resumen de indicadores útiles del cálculo de PCA

Importancia de los componentes				
	$PC1$	$PC2$	$PC3$	$PC4$
$Desviación\ Estándar$	1.704	0.962	0.385	0.144
$Proporción\ de\ varianza$	0.726	0.232	0.037	0.005
$Proporción\ acumulada$	0.726	0.958	0.995	1.000

Nota: Con base en este resumen, se observa que los tres primeros componentes principales (p = 3) contribuyen al 99,5 por ciento de la varianza. Esto significa que, en el modelo final, el cuarto componente principal $PC4$ podría eliminarse ya que su contribución a la varianza es insignificante. (Obi Tayo, 2022)

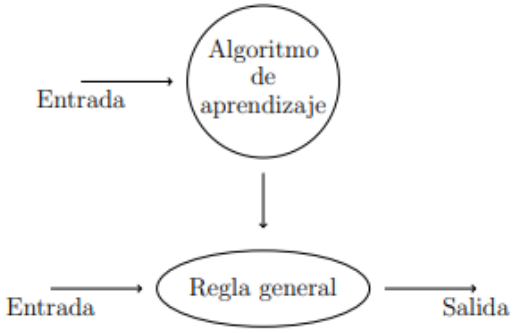
Los fundamentos matemáticos de PCA demuestran cómo se puede implementar el algoritmo de PCA en R utilizando el conjunto de datos de iris con fines ilustrativos (Obi Tayo, 2022).

En el capítulo segundo se presentaron elementos fundamentales del “álgebra lineal” para el soporte del código R; de ellos, el principal es el contenido “vectorial” que permiten la alineación en las funciones para el funcionamiento del código R al momento de insertar y programar.

3.3. Elementos del Álgebra Lineal en la Función Clasificadora

El camino inicia con algo llamado aprendizaje de máquina para llegar a la realización del concepto Inteligencia Artificial. Importante para la innovación de la tecnología que en la actualidad se utiliza. Donde el ser humano proporciona los datos de entrada o entrenamiento; es decir, tiene la facultad de introducir datos para la programación de esta tecnología y entrenarlos a que realicen tareas específicas. (Ossa, 2016, p. xv)

Figura 10. ***Abstracción del proceso de aprendizaje de máquina***



Nota: Se puede comparar mejor la dificultad de esta tarea con el problema de los niños aprendiendo a ver y hablar a partir del flujo continuo de imágenes y sonidos que aparecen en la vida cotidiana." Tomado de: "elementos del álgebra lineal en el aprendizaje de máquina". (p. XV), por Jorge Ossa, 2016.

También, el ser humano diseña el algoritmo de aprendizaje o datos de entrenamiento; y, éste último genera una *regla general* que explica algún aspecto sobre los datos de entrada. Uno de los algoritmos de aprendizaje más importantes, tanto en los inicios de la era computacional como hoy día es el perceptrón llamado también *Máquinas de Vectores de Soportes*; el cual, es un elemento de procesamiento que puede "aprender" una tarea relativamente simple y, combinado con otras unidades similares, puede aprender problemas arbitrariamente complejos (Ossa, 2016, p. XV), esto posible gracias a las aportaciones de algunos matemáticos como Alan Turing quien logró en su tesis descubrir, definir y plantear el aprendizaje de máquina. (Mora, 2022, p.295)

El problema del aprendizaje se puede establecer de la siguiente manera: dada una muestra de tamaño limitado, encontrar una descripción concisa de los datos. Una propuesta de la clasificación como función para la ejecución del funcionamiento del aprendizaje de máquinas según (Ossa, 2016, p. xx) es:

$$Aprendizaje\ de\ máquina \Longrightarrow \begin{cases} \triangleright\ Supervisado \\ \rightarrow Clasificación \\ \rightarrow Preferencia \\ \rightarrow Función \\ \triangleright\ No\ supervisado \\ \triangleright\ Reforzado \end{cases}$$

El aprendizaje supervisado se caracteriza por los datos que recibe, estos se componen de duplas (objeto-etiqueta) y el algoritmo de aprendizaje deberá generar una función que etiquete correctamente tanto las muestras de entrenamiento como las muestras que no tienen etiqueta. Cuando se trata de preferencias, la idea es ordenar las muestras de acuerdo con un criterio específico y la clase de aprendizaje supervisado devuelve una función que realiza la aproximación de los datos en forma de regresión. (Ossa, 2016, p.xx)

En el aprendizaje no supervisado los datos de entrenamiento no tienen etiqueta y el objetivo primario es encontrar la estructura subyacente a los objetos similares. (Ossa, 2016, p.20)

El aprendizaje reforzado tiene como entrada un conjunto de triplas (estado-acción-recompensa) y debe aprender a identificar cuál acción elegir en un estado actual con el objeto de maximizar la recompensa a posteriori. (Ossa, 2016, p.xx)

El problema del aprendizaje de máquina está relacionado con la estadística en el campo de la inferencia, esto porque partiendo de observaciones particulares se quiere llegar a descripciones generales. La única diferencia entre el aprendizaje de máquina y el enfoque estadístico es que el último considera la descripción de los datos en términos de medidas de probabilidad en lugar de una función determinística.

El aprendizaje de máquina para estimar no requiere un modelo probabilístico de los datos, en su lugar asume que el único interés está en la predicción posterior de nuevas instancias, esto a diferencia del enfoque estadístico, el cual pretende clasificar el total de

los datos. Los enfoques mencionados son complementarios y su punto de encuentro es la teoría de la probabilidad. (Ossa, 2016, p. xx).

Por ende, en este capítulo se presentan elementos del "álgebra lineal" sobre los elementos de hiperplanos que contiene teorías del perceptrón o llamado también máquinas de vectores de soporte donde se involucran la acepción de Rosembaltt y el teorema de convergencia de la regla de aprendizaje y la descripción del algoritmo que permite entrenar una función.

3.4. Hiperplanos Separados

Sea un conjunto de muestras de entrenamiento $\{(x_1, y_1), (x_2, y_2), \ldots., (x_n y_n)\}$ con $x_i \in \mathbb{R}^p$. Cada una de dichas muestras puede pertenecer a una de dos clases etiquetadas con $1\ y\ -1$, por ende, la variable $y_i\ \in\ \{-1,1\}$. Se define un hiperplano por:

$$H(\beta, \beta_0) = \{x: f(x) = x^T\beta + \beta_0 = 0\} \qquad (1)$$

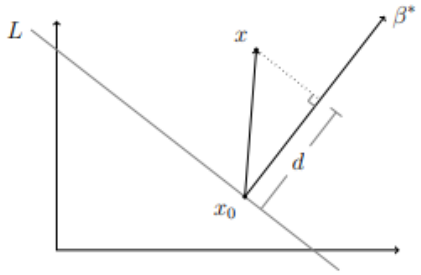
Donde β es un vector normal al hiperplano y β_0 es un escalar. Una regla de clasificación indicada por $f(x)$ es:

$$G(x) = signo\ [x^T\beta + \beta_0].$$ (Ossa, 2016, p.1)

Esta clasificación tiene como propósito determinar un hiperplano separado entre muestras de dos tipos diferentes con la restricción que el hiperplano es la frontera que separa correctamente las muestras. Se puede reescribir el enunciado del problema de la siguiente manera: Dado un conjunto de muestras de entrenamiento de dos clases, el objetivo es construir el hiperplano descrito por la ecuación (1) tan bien como sea posible (separabilidad lineal).

Para construir $f(x)$ se usarán algunos elementos de "álgebra lineal", la línea L representa un hiperplano o conjunto afín definido por la ecuación $f(x) = x^T\beta + \beta_0 = 0$; en esta ilustración se trabaja en $\mathbb{R}^2$ por tanto L es una línea. (Ossa, 2016, p.1).

Figura 11. ***Elementos vectoriales del hiperplano***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.2), por Jorge Ossa, 2016

Algunas propiedades de L son:

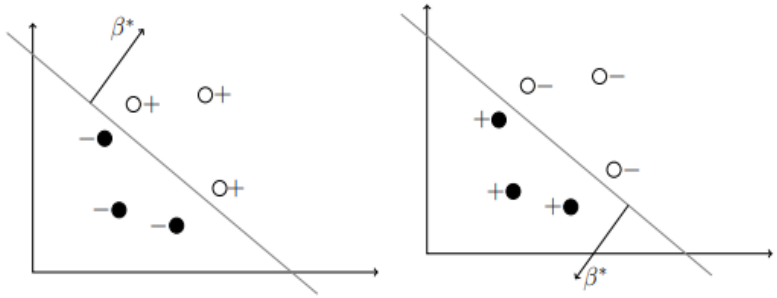
1. Para cualquier par de puntos $x_1, x_2 \in L,\ \beta^T\ (x_1 - x_2) = 0, por\ tanto\ \beta^* = \frac{\beta}{\|\beta\|}$ es el vector unitario normal al hiperplano L.
2. Para cualquier punto $x_0 \in L, \beta^T x_0 = \beta_0$. (Ossa, 2016, p.2)
3. La distancia con signo desde algún punto $x\ a\ L$ está dada por:

$$\beta^{*T}(x - x_0) = \frac{1}{\|\beta\|}\ (\beta^T x + \beta_0)$$

$$= \frac{1}{\|f'(x)\|} f(x)$$ (Ossa, 2016, p.2)

De la propiedad 3, se tiene que $f(x)$ es proporcional a la distancia con signo desde x al hiperplano definido por $f(x) = 0$

Figura 12. ***Hiperplanos separados con diferente vector directo***

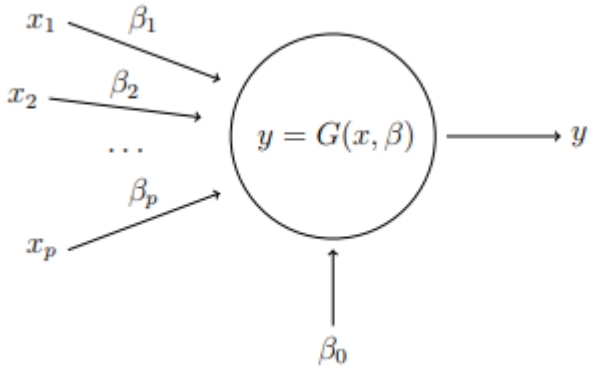


Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.2), por Jorge Ossa, 2016.

3.5. El Perceptrón: Generalidades

Un Perceptrón es un elemento que procesa datos con base en unas entradas, una función y una salida, también suele ser llamado **neurona artificial**. Se muestra el siguiente esquema.

Figura 13. ***Esquema general de un perceptrón***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.2), por Jorge Ossa, 2016.

Existen varios mecanismos para calcular la salida de un perceptrón. El valor de salida del elemento de procesamiento de la figura 13 se calcula en términos de un vector de entrada $x = (1, x_1, x_2, \ldots, x_p)$ y los pesos dados por un vector intrínseco $\beta =$

$\left(\beta_0, \beta_1, \beta_2, \ldots, \beta_p\right)$. Matemáticamente, la salida del perceptrón es una función de sus entradas y sus pesos,

$$y = G(x, \beta). \quad (2)$$

Una función particular de salida es una combinación lineal (producto punto) entre el vector de entrada x y el vector de pesos β. Para un elemento como el de la figura 13 la salida se calcula mediante la ecuación, (Ossa, 2016, p.3)

$$y = G\left(\sum_{i=0}^{p} x_i \beta_i\right) = G(x \cdot \beta), \quad (3)$$

Donde G es alguna función lineal o no lineal. El "producto punto" tiene una cualidad muy atractiva. Usando la relación

$$\text{Cos}(x, \beta) = \frac{x \cdot \beta}{\|x\| \|\beta\|} \quad (4)$$

se puede ver que entre mayor es el producto punto (asumiendo longitudes fijas de $x\ y\ \beta$) más similares son los dos vectores, pues es menor el ángulo que forman; debido a esto el producto punto se puede ver como una medida de similaridad. (Ossa, 2016, p.4)

Como ya se ha visto, un perceptrón simple tiene un número p de entradas x_i y una salida y. Cada una de las entradas tiene un peso asociado β_i. Hay también un parámetro adicional β_0 llamado $bias$ el cual siempre tiene un peso asignado de 1.

El perceptrón, o neurona artificial, es un dispositivo anticipativo, es decir, las conexiones están dirigidas desde la entrada hacia la salida del perceptrón. (Ossa, 2016, p.4)

3.5.1. Regla de Activación del Perceptrón

La regla de activación tiene dos pasos, a saber:

1. De acuerdo con el vector de entrada $x = (x_1, x_2, \ldots, x_p)$ se calcula la $activación$ del perceptrón,

$$a = \sum_{i=0}^{p} x_i \beta_i \quad (5)$$

donde p es la dimensión del vector x y está implícito el peso 1 para el $bias$; si este último no se incluye, la suma inicia en $i = 1$. (Ossa, 2016, p.5)

2. La salida y se establece como una función de a, $f(a)$, que depende del valor calculado en el primer paso (activación). Existen varias posibilidades para $f(a)$ o $función\ de\ activación$, entre ellas están:

    a) Funciones de activación determinística

        i. Lineal.

        ii. Sigmoide (función logística).

        iii. Sigmoide (tanh).

        iv. Función de umbral (Threshold function).

    b) Funciones de activación estocásticas: y es seleccionada estocásticamente de ± 1

        i Baño caliente.

        ii La regla Metrópolis.

La función de activación determinística lineal, $y(a) = a$ (Ossa, 2016, p.5)

3.5.2. El Perceptrón de Rosenblatt

3.5.2.1. Algoritmo de Aprendizaje

El algoritmo de aprendizaje del Perceptrón intenta encontrar un hiperplano separador al minimizar la distancia de los puntos mal clasificados a la frontera de decisión. Si un dato x_i está mal clasificado, entonces

$$(\beta^T x_i + \beta_0) y_i < 0$$

ya sea que la etiqueta correcta para x_i sea 1 o −1. Ahora, teniendo en cuenta las distancias con signo, el objetivo será minimizar:

$$D(\beta, \beta_0) = -\sum_{i \in M} y_i(\beta^T x_i + \beta_0) \text{ (Ossa, 2016, p.6)} \quad (6)$$

donde M indexa el conjunto de puntos mal clasificados. La cantidad $D(\beta, \beta_0)$ es no-negativa y proporcional a la distancia de los puntos mal clasificados a la frontera de decisión definida por $\beta^T x_i + \beta_0 = 0$. En la figura 14 se visualiza como es la función D, para un conjunto arbitrario de muestras de entrenamiento, en términos de β_0 y el ángulo de inclinación (del hiperplano). Al optimizar dicha función se debe encontrar un par β, β_0 tal que la distancia, de los puntos mal clasificados a la frontera de decisión, determinada por β, β_0, sea mínima. El gradiente (asumiendo que M está fijo) está dado por: (Ossa, 2016, p.6)

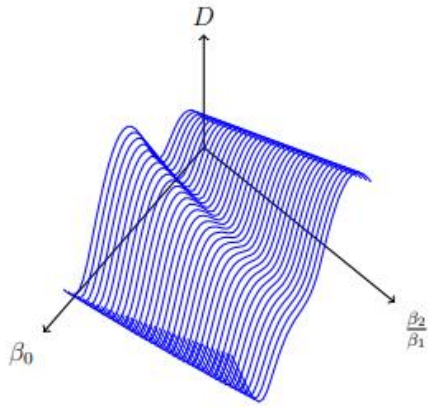
$$\frac{\partial D}{\partial \beta} = -\sum_{i \in M} y_i x_i$$

$$\frac{\partial D}{\partial \beta_0} = -\sum_{i \in M} y_i$$

El algoritmo de Rosenblatt usa descenso de gradiente estocástico para minimizar por partes este criterio lineal. Esto significa que, en lugar de calcular la suma de las contribuciones al gradiente de cada muestra seguida por un paso en la dirección del gradiente negativo, se da un paso después de visitar cada muestra. Por tanto, las muestras mal clasificadas se visitan en alguna secuencia y los parámetros β se actualizan así:

$$\begin{pmatrix} \beta \\ \beta_0 \end{pmatrix} \leftarrow \begin{pmatrix} \beta \\ \beta_0 \end{pmatrix} + \rho \begin{pmatrix} y_i x_i \\ y_i \end{pmatrix}, \quad \text{(Ossa, 2016, p.6)}$$

Figura 14. ***Función*** $\boldsymbol{D(\beta, \beta_0)}$



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.7), por Jorge Ossa, 2016.

donde ρ es la razón de aprendizaje la cual, en este caso, se puede tomar como 1 sin pérdida de generalidad.

3.5.3. Nociones Separabilidad Lineal

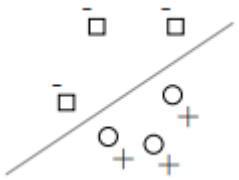
Supóngase la existencia de N puntos $x_i \in \mathbb{R}^p$ en cualquier posición con etiquetas de clase $y_i \in \{-1, 1\}$ nótese un hiperplano por

$$f(x) = x^T \beta_i + \beta_0 = 0$$

ó en una notación más compacta $\beta^T x^* = 0$, donde $x^* = (x, 1)$ y $\beta = (\beta_1, \beta_0)$.

Sea $z_i = \frac{x_i{}^*}{\|x_i{}^*\|}$. Si existe un β_{sep} tal que $\forall i \left(y_i \beta^T{}_{sep} \mathcal{Z}_i \geq 1\right)$, el conjunto de datos se dice $Linealmente\ Separable$. (Ossa, 2016, p.6)

Figura 15. ***Hiperplano separador***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.7), por Jorge Ossa, 2016.

Supóngase que β_{sep} existe, por tanto, todas las muestras en forma de cuadrado (ver figura 15) cuya etiqueta es −1 se notan.

$$(\mathrm{x_i}\, \mathrm{y_i}) \text{ tal que } \mathrm{y_i} = -1$$

al realizar el producto punto $\beta^{\mathrm{T}}{}_{\text{sep}}\, \mathrm{x_i}^* \leq 0$ porque produce la distancia con signo desde la posición de la muestra hasta el hiperplano determinado por β_{sep}.

Al multiplicar la etiqueta y_i por la distancia con signo da como resultado un número positivo. De igual manera, pasa con las muestras O cuyas etiquetas son $+1$, al multiplicar su etiqueta y_i por la distancia asignada se produce un número positivo.

Se puede concluir entonces que las muestras son separables linealmente cuando existe un hiperplano separador que cumple con $y_i\, \beta^T{}_{sep} x_i{}^* \geq 0, \forall_i$. (Ossa, 2016, p.8)

3.5.3.1. Teorema de Convergencia de la Regla de Aprendizaje del Perceptrón

Teorema 10: Si existe un hiperplano indicado por β_{sep} que cumple con la condición de separabilidad lineal entonces, para cualquier vector $\beta^0{}_{sep}$ inicial, la regla de aprendizaje del Perceptron convergerá a un vector β_{sep} (no necesariamente único) que proporciona la

etiqueta correcta para todos los puntos (muestras de entrenamiento) y esta convergencia ocurrirá en un número finito de pasos.

Demostración

Sean los siguientes conjuntos:

$$L^{+} = \left\{x: {\beta^{T}}_{sep^{x}} > 0\right\}$$

$$L^{-} = \left\{x: {\beta^{T}}_{sep^{x}} < 0\right\}$$

Las muestras se dividen en dos clases, la positiva L^{+} cuyos elementos tienen distancia con signo positivo hasta el hiperplano separador, y la negativa L^{-} cuyos elementos cumplen la condición contraria.

Sea $-L^{-} = \{-x : x \in L^{-}\}$. Se define un nuevo conjunto de muestra.

$$M = L^{+} \cup -L^{-}$$ (Ossa, 2016, p.8)

que abarca la totalidad de las muestras de entrenamiento y que cumple con

$${\beta^{T}}_{sep} x_i > 0 \; si \; x_i \in M$$

luego la respuesta de la función $f(x) = {\beta^{T}}_{sep^{x}}$ será positiva mientras que x esté en el conjunto M. (Ossa, 2016, p.9)

Si la respuesta de la función es incorrecta para una muestra de entrenamiento x, el hiperplano determinado por β_{sep}

se deberá actualizar de acuerdo con la siguiente fórmula:

$${\beta^{j+1}}_{sep} = {\beta^{j}}_{sep} + \rho x.$$

La esencia de la demostración es mostrar que la secuencia de vectores de entrenamiento, que obligan a una actualización de β_{sep} , es finita.

Denótese el hiperplano inicial como ${\beta^0}_{sep}$ y ${\beta^i}_{sep}, i \in \mathbb{Z}^+$ como las modificaciones subsiguientes. Si x_0 es la primera muestra de entrenamiento que produjo una respuesta incorrecta de la función, entonces

$${\beta^i}_{sep} = {\beta^0}_{sep} + x_0, \text{ cumpliendo con } \left({\beta^0}_{sep}\right)^T x_0 \leq 0. \text{ (Ossa, 2016, p.9)}$$

Cuando se examina la muestra x_i pueden ocurrir dos posibilidades:

$$\left({\beta^i}_{sep}\right)^T x_i > 0$$

En cuyo caso ${\beta^i}_{sep}$ no será modificado, ó

$$\left({\beta^i}_{sep}\right)^T x_i < 0$$

Caso cuando se deberá modificar ${\beta^i}_{sep}$.

En ambos casos se genera ${\beta^{i+1}}_{sep}$ el cual puede ser el mismo ${\beta^i}_{sep}$ o ${\beta^i}_{sep}$ modificado.

El siguiente β_{sep} después de la inicial, lo proporciona la fórmula de recurrencia, ${\beta^{i+1}}_{sep} = {\beta^i}_{sep} + x_i$, del algoritmo, ${\beta^2}_{sep} = {\beta^1}_{sep} + x_i$ cumpliendo con $\left({\beta^1}_{sep}\right)^T x_1 \leq 0$. (Ossa, 2016, p.9)

Llevando la fórmula de recurrencia hasta un paso arbitrario k, ${\beta^{k-1}}_{sep}$ si y solo si la muestra de entrenamiento x_{k-1} produce una respuesta incorrecta al clasificarla ${\beta^{k-1}}_{sep}$, produciendo así ${\beta^k}_{sep}$. Combinando los cambios sucesivos desde ${\beta^0}_{sep}$ se obtiene

$$\beta^{k}{}_{sep} = \beta^{0}{}_{sep} + x_0 + x_1 + \cdots x_{k-1}.$$

Ahora se muestra que k no puede ser arbitrariamente grande. (Ossa, 2016, p.10)

Sean $\beta^{*}{}_{sep}$ un hiperplano tal que $\left(\beta^{*}{}_{sep}\right)^{T} x > 0, \forall x: x \in M$.

Sea $m = \min\left\{\left(\beta^{*}{}_{sep}\right)^{T} x\right\}$, donde el mínimo se toma teniendo en cuenta todas las muestras de entrenamiento en M, m es la distancia con signo más corta desde la muestra hasta $\beta^{*}{}_{sep}$ Este mínimo existe siempre y cuando solo haya un número finito de muestras de entrenamiento. Ahora

$$\beta^{k}{}_{sep} \cdot \beta^{0}{}_{sep} = \left[\beta^{0}{}_{sep} + x_0 + x_1 + \cdots x_{k-1}\right] \cdot \beta^{*}{}_{sep} \tag{7}$$

$$\beta^{k}{}_{sep} \cdot \beta^{*}{}_{sep} \geq \beta^{0}{}_{sep} \cdot \beta^{*}{}_{sep} + km \tag{8}$$

Esto porque $\left(\beta^{*}{}_{sep}\right)^{T} x_i > m, \qquad x_i \in M.$

Ahora se tiene $\beta^{k}{}_{sep} \cdot \beta^{*}{}_{sep}$ y por la desigualdad de Cauchy-Schwartz:

$$\frac{\left(\beta^{k}{}_{sep} \cdot \beta^{*}{}_{sep}\right)^{2}}{\left\|\beta^{*}{}_{sep}\right\|^{2}} \leq \left\|\beta^{k}{}_{sep}\right\|^{2}. \tag{9}$$

Esto muestra que el cuadrado de la norma de $\beta^{k}{}_{sep}$ crece más rápido que k^2, donde k representa la cantidad de veces que cambio $\beta^{0}{}_{sep}$. (Ossa, 2016, p.10)

Ahora se debe acotar $\left\|\beta^{*}{}_{sep}\right\|^{2}$ para mostrar que no crece indefinidamente. Sea

$$\beta^{k}{}_{sep} = \beta^{k-1}{}_{sep} + x_{k-1} \tag{10}$$

y esta modificación está ocurriendo porque

$$\left(\beta^{k-1}{}_{sep}\right)^T x_{k-1} \leq 0 \tag{11}$$

es decir, quedó mal clasificada la muestra de entrenamiento. De (10) se tiene,

$$\left\|\beta^{k}{}_{sep}\right\|^2 = \left\|\beta^{k-1}{}_{sep}\right\|^2 + 2\left(\beta^{k-1}{}_{sep}\right)^T x_{k-1} + \|x_{k-1}\|^2 \quad \text{(Ossa, 2016, p.9),}$$

y por (11) se tiene que,

$$\left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{k-1}{}_{sep}\right\|^2 + \|x_{k-1}\|^2 \tag{12}$$

Ahora, sea $N = \max_{x \in M}\{\|x\|^2\}$, y expandiendo la fórmula recursiva:

$$\left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{k-1}{}_{sep}\right\|^2 + \|x_{k-1}\|^2 \tag{13}$$

$$\left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{k-2}{}_{sep}\right\|^2 + \|x_{k-2}\|^2 + \|x_{k-1}\|^2 \tag{14}$$

$$\vdots$$

$$\left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{o}{}_{sep}\right\|^2 + \|x_0\|^2 + \cdots + \|x_{k-1}\|^2 \tag{15}$$

$$\left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{o}{}_{sep}\right\|^2 + k\,N \tag{17}$$

Así, el cuadrado de la norma crece menos rápido que una función lineal en k.

Combinando (9) y (17) se tiene (Ossa, 2016, p.11)

$$\frac{\left(\beta^{k}{}_{sep}\cdot\beta^{*}{}_{sep}+km\right)^2}{\left\|\beta^{*}{}_{sep}\right\|^2} \leq \left\|\beta^{k}{}_{sep}\right\|^2 \leq \left\|\beta^{o}{}_{sep}\right\|^2 + k\,N. \tag{18}$$

De aquí se puede concluir que $\left\|\beta^{k}{}_{sep}\right\|^2$ está acotado, por ende, lo está la cantidad de veces que $\beta^{o}{}_{sep}$ cambia para llegar a ser $\beta^{*}{}_{sep}$.

Ahora, en aras de simplificar, asúmase (sin pérdida de generalidad) que $\beta^{o}{}_{sep} = 0$.

Entonces el número máximo de veces que $\beta^{o}{}_{sep}$ puede cambiar está dado por:

$$\frac{(km)^2}{\left\|\beta^{*}{}_{sep}\right\|^2} \leq kN \tag{19}$$

ó despejando k:

$$k \leq \frac{N\left\|\beta^{*}{}_{sep}\right\|^2}{m^2} \tag{20}$$

Asumiendo que $\beta^{*}{}_{sep}$ existe y su norma es 1, el número máximo de actualizaciones realizadas es $\frac{N}{m^2}$, sin embargo, se debe tener en cuenta que se pueden requerir muchos cálculos más porque un número muy pequeño de vectores pueden generar errores durante cualquier iteración del entrenamiento. Como β_{sep} es desconocido, y por ende m también se ignora, el número de actualizaciones no puede preestablecerse a partir de la desigualdad (20). (Ossa, 2016, p.11)

Debe tenerse en cuenta que no hay requisito que limite el número de muestras de entrenamiento. De otro lado, cuando la norma de estos vectores es muy pequeña, causa que m sea pequeño lo cual redunda en un número grande de actualizaciones.

Los valores reales de los vectores tampoco importan, la regla de aprendizaje requiere que β_{sep} sea actualizado por el vector de entrada x_i (ó un múltiplo) cada vez que la respuesta de la función sea incorrecta. (Ossa, 2016, p.12)

Algunas variaciones del tamaño del paso de aprendizaje son:

1. Asignar a la razón de aprendizaje ρ alguna constante no negativa.
2. Si se trabaja con el valor de $\rho = \frac{1}{\|x\|}$ la actualización implicara un vector unitario.

3. Usar $\rho = \frac{x \cdot \beta}{\|x\|^2}$ ocasionará que el hiperplano cambie lo justo para que la muestra x quede bien clasificada.

Minsky[3] establece el valor de ${\beta^o}_{sep}$ a un vector de entrenamiento arbitrario. Otros autores indican, usualmente, valores aleatorios. (Ossa, 2016, p.12)

Es importante resaltar que los elementos de hiperplanos, el teorema de convergencia y el perceptrón de Rosenblatt son de gran contribución al algoritmo del perceptrón o las máquinas de vectores de soportes necesarias para la aplicación en las áreas de la Biomédica. Se necesitan para la programación de software que tiene que ver con la clasificación del entrenamiento de la máquina para que realicen la función. El "álgebra lineal" es fundamental porque estos algoritmos y propiedades permiten hoy en día realizar pruebas cancerígenas que ayudan a detectar células malignas como las benignas en corto tiempo. (Guillén, E. O., Garcia, G. I., Vázquez-Montiel, S., Atencio, J. A. D., Ramos, J. C., & Delgado, F. G, 2010, pág. 34-35)

Existen ciertos problemas con este algoritmo $a = \sum_{i=0}^{p} x_i \beta_i$:

1. Cuando los datos son separables, hay muchas soluciones y aquella que se encuentra depende de los valores iniciales.
2. El número finito de pasos puede ser muy grande. Entre más pequeño sea el margen de separación, mayor será el tiempo necesario para encontrarlo.
3. Cuando los datos no son separables, el algoritmo no convergerá y hará muchos ciclos. Cada ciclo puede ser muy largo y difícil de detectar.

[3] Marvin Lee Minsky, es considerado uno de los padres de la inteligencia artificial. Fue cofundador del laboratorio de inteligencia artificial del Instituto de Tecnología de Massachusetts. Contribuyó al desarrollo de la descripción gráfica simbólica, geometría computacional, representación computacional, semántica computacional, percepción mecánica, aprendizaje simbólico y conexionista. En 1951 creó SNARC, el primer simulador de redes neuronales.

El tercer problema puede eliminarse, en muchas ocasiones, buscando un hiperplano en un espacio distinto al original, mucho más grande, el cual se puede obtener llevando los datos a un espacio de características de dimensión mucho mayor.

Esta idea es la base de **la teoría de las máquinas de vectores de soporte (SVM).** Para solucionar el primer problema, se deben agregar restricciones adicionales al hiperplano separador. (Ossa, 2016, p.12)

Para hablar del algoritmo del perceptrón o los algoritmos de máquinas de soportes, redes neuronales, bosques aleatorios y otros aspectos importantes se deben analizar las teorías del "álgebra lineal" que fundamentan otras propiedades significativas para el desarrollo de la tecnología biomédica como lo son:

3.5.3.2. El Perceptrón en Variables Duales

De acuerdo con algoritmo del perceptrón, al inicio se tiene un vector $\boldsymbol{\beta^0}_{\boldsymbol{sep}} = \mathbf{0}$ y modificándolo sucesivamente mediante adición, solamente cuando ocurre un error de clasificación $\boldsymbol{\beta^*}_{\boldsymbol{sep}}$ con la muestra actual, se llega al hiperplano separador. Sin embargo, mirando desde otro punto de vista cada solución, después de la primera iteración $\boldsymbol{\beta^i}_{\boldsymbol{sep}}$, tiene la forma:

$$\boldsymbol{\beta^i}_{\boldsymbol{sep}} = \alpha_1 \boldsymbol{x_1} + \alpha_1 \boldsymbol{x_1} + \cdots + \alpha_N \boldsymbol{x}_N. \qquad (21)$$

Es decir, el vector solución es una *"combinación lineal"* de las muestras de entrenamiento que han influido en su formación, vía suma, mediante el criterio de clasificación.

Ahora, en lugar de buscar los n pesos para establecer el hiperplano en el espacio de muestras $\mathbb{R}^n$

$$\boldsymbol{\beta^*}_{\boldsymbol{sep}} = (\beta_1, \beta_2, \cdots, \beta_n)$$

se buscarán los M pesos correspondientes para configurar la combinación lineal adecuada para encontrar $\boldsymbol{\beta^*}_{\boldsymbol{sep}}$. Expresando (21) de la siguiente manera: (Ossa, 2016, p.13)

$$\boldsymbol{\beta^*}_{\boldsymbol{sep}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \alpha_i \, \boldsymbol{x_i} \tag{22}$$

y a esto se le llamará el " $espacio\ dual\ de\ variables$".

Ahora, para determinar si una muestra está bien clasificada se utiliza (22).

$$y_j \left\langle \boldsymbol{x}_j, \boldsymbol{\beta^k}_{\boldsymbol{sep}} \right\rangle = y_j \left\langle \boldsymbol{x}_j \, , \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \alpha_i \, \boldsymbol{x_i} \right\rangle = y_j \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \alpha_i \langle \boldsymbol{x_j}, \boldsymbol{x_i} \rangle \tag{23}$$

Esta operación requiere conocer la función de producto punto entre las muestras de entrenamiento. Con base en (23) se puede inferir la existencia de una matriz de la siguiente forma:

$$k = \begin{pmatrix} \langle \boldsymbol{x_1}, \boldsymbol{x_1} \rangle & \cdots & \langle \boldsymbol{x_1}, \boldsymbol{x}_N \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{x_N}, \boldsymbol{x_1} \rangle & \cdots & \langle \boldsymbol{x}_N, \boldsymbol{x}_N \rangle \end{pmatrix} \tag{24}$$

la matriz (24) contiene todos los productos internos de todas las muestras de entrenamiento entre sí. (Ossa, 2016, p.13)

De manera simbólica, cada posición K_{ij} de la matriz está dada por

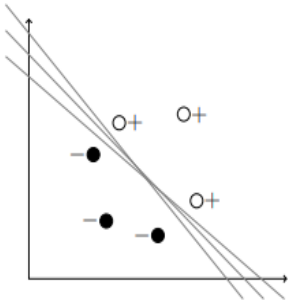
$$K_{ij} = \langle \boldsymbol{x}_i \boldsymbol{x}_j \rangle$$

y esta matriz es un kernel, noción en la cual se basa el macroalgoritmo del perceptrón en variables duales. (Ossa, 2016, p.14)

3.5.4. Nociones de Hiperplanos Separadores Óptimos

1. El hiperplano separador encontrado, cuando los datos son separables, hay muchas soluciones y aquella que se encuentra depende de los valores iniciales.
2. El número finito de pasos puede ser muy grande. Entre más pequeño sea el margen de separación, mayor será el tiempo necesario para encontrarlo.
3. Cuando los datos no son separables, el algoritmo no convergerá y hará muchos ciclos. Cada ciclo puede ser muy largo y difícil de detectar.

Figura 16. ***Varios hiperplanos separados***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.14), por Jorge Ossa, 2016.

Esto ocurre porque el único criterio que se tuvo en cuenta para encontrarlo fue la correcta clasificación de las muestras mas no la mejor clasificación; esto sin contar con que el resultado final depende de la asignación inicial, arbitraria, que se le haga a β. Cuando se habla de $hiperplano\ separador\ óptimo$ se está queriendo encontrar un hiperplano que, además de separar correctamente las muestras de entrenamiento, sea único y, también, que maximice el margen entre las dos clases de muestras; todo esto para obtener un mejor rendimiento en la clasificación sobre todos los datos de prueba. (Ossa, 2016, p.14)

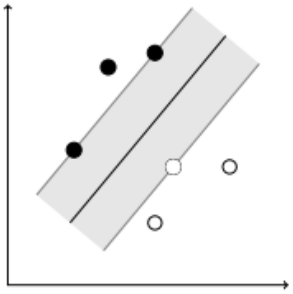
De acuerdo con Vapnik, el hiperplano separador óptimo separa las dos clases y maximiza la distancia al punto o puntos más cercanos de cada clase. La minimización de la función (6) se debe generalizar. Considérese el problema de optimización

$$\max_{\beta,\beta_0,\|\beta\|=1} M \tag{25}$$

sujeto a, $y_i(x_i{}^T\beta + \beta_0) \geq M,\ i = 1,\cdots;N.$

¿Cuál es el hiperplano, dado por $\beta\ y\ \beta_0$, siendo β un "vector unitario" y β_0 un escalar, tal que la distancia con signo de cualquier muestra de entrenamiento a dicho plano sea como mínimo M?

Figura 17. ***Distancia mínima M de las muestras al hiperplano***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.15), por Jorge Ossa, 2016.

Las condiciones del problema aseguran que todos los puntos están, al menos, a una distancia M de la frontera de decisión definida por el hiperplano y se busca el M más grande. (Ossa, 2016, p.15)

Partiendo de la restricción $\|\beta\| = 1$,

$$y_i(x_i{}^T\beta + \beta_0) \geq M, \quad M > 0 \tag{26}$$

$$y_i\left(x_i{}^T\frac{\beta}{M} + \frac{\beta_0}{M}\right) \geq 1. \tag{27}$$

Sea $\tilde{\beta} = \beta\, M$, además se tiene que $\|\beta\| = 1$, por las condiciones del problema, entonces

$$\left\|\tilde{\beta}\right\| = \frac{1}{M}. \tag{28}$$

$\tilde{\beta}$ es el vector β escalado, entonces buscar cualquiera de los dos es equivalente. Ahora, entre más grande sea M, más pequeña será la norma de $\tilde{\beta}$. (Ossa, 2016, p.16)

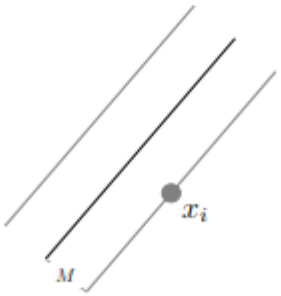
Teniendo esto en cuenta, (25) es equivalente a

$$\min_{\beta,\beta_0} \frac{1}{2}\|\beta\|^2 \tag{29}$$

$$\text{sujeto a, } y_i({x_i}^T\beta + \beta_0) \geq M,\ i = 1, \cdots; N.$$

Ya que, el requisito de que un vector β en $\mathbb{R}^n$ sea unitario se puede establecer en varias formas equivalentes: $\|\beta\| = 1, \beta^T\beta = 1$ ó $\|\beta\|^2 = 1\ por\ (7)$. Basándose en la propiedad (3) de L (la distancia con signo desde alguna muestra x_i al hiperplano) se puede decir que la restricción define un margen, de $\frac{1}{\|\beta\|}$ de ancho, alrededor del hiperplano. (Ossa, 2016, p.16)

Figura 18. ***Margen del hiperplano***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.16), por Jorge Ossa, 2016.

Por lo anterior, se deben encontrar β y β_0 que maximicen dicho ancho. Este es un problema de optimización convexa (criterio cuadrático con restricciones de desigualdad

lineales) llamado problema primal. Tiene la desventaja de ser computacionalmente costoso. (Ossa, 2016, p.16)

La función de Lagrange (primal) que será minimizada con respecto a β y β_0, es

$$L_P = \frac{1}{2}\|\beta\|^2 - \sum_{i=1}^{N} \alpha_i [\, y_i(x_i{}^T\beta + \beta_0) - 1] \tag{30}$$

Derivando parcialmente e igualando dichas derivadas a cero, se obtiene:

$$\beta = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{31}$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{32}$$

y al substituir estas en (30),

De la primera derivada parcial $\beta = \sum_{i=i}^{N} \alpha_i y_i x_i$:

$$\begin{aligned} \tfrac{1}{2}\|\beta\|^2 &= \tfrac{1}{2}\beta^T\beta \\ &= \frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i\right)^T \sum_{i=1}^{N} \alpha_k y_k x_k \\ &= \frac{1}{2}\left(\sum_{i=1}^{N} \alpha_i y_i x_i{}^T\right)\sum_{i=1}^{N} \alpha_k y_k x_k \\ &= \frac{1}{2}\sum_{i=1}^{N} \alpha_k y_k x_k (x_i{}^T x_k). \end{aligned} \tag{33}$$

Usando la segunda derivada parcial $0 = \sum_{i=i}^{N} \alpha_i y_i$:

$$-\sum_{i=1}^{N} \alpha_i [\, y_i(x_i{}^T\beta + \beta_0) - 1]$$

$$-\sum_{i=1}^{N}\alpha_i y_i(x_i{}^T\beta+\beta_0)-\alpha_i \qquad (34)$$

$$-\sum_{i=1}^{N}\alpha_i y_i x_i{}^T\beta-\sum_{i=1}^{N}\alpha_i y_i\beta_0-\sum_{i=1}^{N}\alpha_i$$

$$-\sum_{i=1}^{N}\alpha_i y_i x_i{}^T\sum_{k=1}^{N}\alpha_k y_k x_k-\beta_0\sum_{i=1}^{N}\alpha_i y_i-\sum_{i=1}^{N}\alpha_i$$

$$-\sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i y_i\alpha_k y_k(x_i{}^T x_k)-\sum_{i=1}^{N}\alpha_i$$

$$\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\sum_{k=1}^{N}\alpha_i y_i\alpha_k y_k(x_i{}^T x_k).$$

después de los cálculos (33) y (34) se obtiene la función dual

$$L_P=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}\alpha_i\, y_i\alpha_k y_k\, x_i{}^T x_k \qquad sujeto\ a \qquad \alpha_i\geq 0. \qquad (35)$$

Es importante observar que en la función dual solamente aparecen los productos punto entre todos los patrones x_i de “aprendizaje”. (Ossa, 2016, p.17)

La solución que se obtiene maximizando L_P en el ortante positivo, (Ossa, 2016, p.18) (en geometría, un ortante o hiperoctante es el análogo, en un espacio Euclidiano n-dimensional, de un cuadrante en el plano o de un octante en tres dimensiones), un problema de optimización convexa (en esencia la función convexa significa que tiene un único mínimo global), más sencillo el cual puede ser solucionado con $software\ estándar$. Además, la solución debe satisfacer las condiciones de $Karush-Kuhn-Tucker$, las cuales incluyen (31), (32) y (35) y

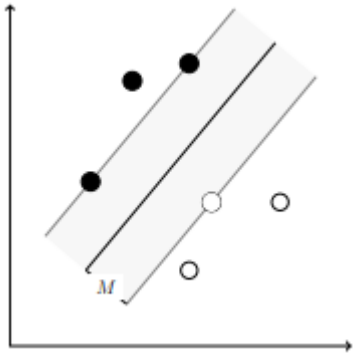
$$\alpha_i[y_i(x_i{}^T\beta+\beta_0)-1]=0 \quad \forall_i \qquad (36)$$

En (36) se puede ver que:

1. Si $\alpha_i > 0, entonces\ y_i(x_i{}^T\beta + \beta_0) = 1\ o, en\ otras\ palabras, x_i$ está sobre el borde del margen;
2. Si $y_i(x_i{}^T\beta + \beta_0) > 1, x_i$ no está sobre dicho borde y $\alpha_i = 0$.

En (31) se puede ver que el vector solución β está definido en términos de una "combinación lineal" de los $vectores\ de\ soporte\ x_i$ – es decir, los puntos que están sobre el borde del margen para los cuales cada $\alpha_i > 0$. (Ossa, 2016, p.18)

Figura 19. ***Hiperplano separador óptimo con el margen máximo de separación entre las dos clases***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.19), por Jorge Ossa, 2016.

Se muestra un "hiperplano separador óptimo" con tres puntos de soporte.

Siguiendo esta idea, β_0 se calcula al resolver (36) para cualquiera de los puntos de soporte.

El "hiperplano separador óptimo" produce una función $\hat{f}(x) = x^T\hat{\beta} + \hat{\beta}_0$ para clasificar nuevas muestras; entonces se construye la función

$$\hat{G}(x) = signo\ \hat{f}(x) \qquad (37)$$

la cual devuelve la etiqueta calculada para la muestra x.

Debe tenerse en cuenta que, al construir el hiperplano con su margen maximizado, no quedan muestras de entrenamiento al interior de dicho margen; sin embargo, cuando se

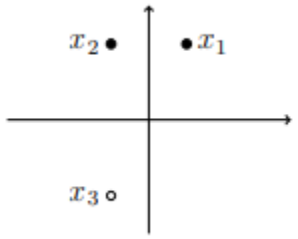
están clasificando muestras nuevas, es posible que caigan en el interior, es por esto por lo que surge la idea de "a mayor margen sobre las muestras de entrenamiento, mejor separación sobre las muestras nuevas". La descripción de la solución en términos de los vectores de soporte parece sugerir que el hiperplano óptimo se enfoca más sobre las muestras que importan y soporta mejor una mala especificación del modelo. A continuación, se muestra un ejemplo que pretende ilustrar el enfoque de la solución del problema dual. (Ossa, 2016, p.19)

Ejemplo 6.

Solución del problema dual.

El siguiente ejemplo juguete es la construcción de un hiperplano separador (3). Para el conjunto de muestras de entrenamiento de la figura 20, encontrar el hiperplano separador.

Figura 20. ***Muestras de entrenamiento***



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.20), por Jorge Ossa, 2016.

$$X = \{[(1,2), -1], [(-1,2), -1] \ [(-1, -2), 1] \} \quad (38)$$

Las ecuaciones necesarias para resolver el ejercicio son las (31, 32, 33 y 36). Se inicia con la *"función dual"* la cual se calcula por partes. (Ossa, 2016, p.20)

La primera se expresa así:

$$\sum_{i=1}^{3} \alpha_i = \alpha_1 + \alpha_2 + \alpha_3. \quad (39)$$

La segunda parte implica algunos cálculos adicionales:

$$\sum_{i=1}^{3}\sum_{K=1}^{3} \alpha_i\, y_i \alpha_k y_k\, {x_i}^T x_k =$$

$$\alpha_1 y_1 \alpha_1 y_1 {x_1}^T x_1 + \alpha_1 y_1 \alpha_2 y_2 {x_1}^T x_2 + \alpha_1 y_1 \alpha_3 y_3 {x_1}^T x_3 +$$

$$\alpha_2 y_2 \alpha_1 y_1 {x_2}^T x_1 + \alpha_2 y_2 \alpha_2 y_2 {x_2}^T x_2 + \alpha_2 y_2 \alpha_3 y_3 {x_2}^T x_3 +$$

$$\alpha_3 y_3 \alpha_1 y_1 {x_3}^T x_1 + \alpha_3 y_3 \alpha_2 y_3 {x_3}^T x_2 + \alpha_3 y_3 \alpha_3 y_3 {x_3}^T x_3.$$ (Ossa, 2016, p.20)

Después de hacer los reemplazos, con base en el conjunto de muestras (38), y simplificar se obtiene:

$$\sum_{i=1}^{3}\sum_{K=1}^{3} \alpha_i\, y_i \alpha_k y_k\, {x_i}^T x_k$$

$$= 5{\alpha^2}_1 + 5{\alpha^2}_2 + 5{\alpha^2}_3 + 6\alpha_1\alpha_2 + 10\alpha_1\alpha_3 + 6\alpha_2\alpha_3 \quad (40)$$

Al reunir las expresiones (39) y (40) en la expresión completa se obtiene:

$$L_D = \alpha_1 + \alpha_2 + \alpha_3$$

$$= -\frac{1}{2}[5{\alpha^2}_1 + 5{\alpha^2}_2 + 5{\alpha^2}_3 + 6\alpha_1\alpha_2 + 10\alpha_1\alpha_3 + 6\alpha_2\alpha_3]. \quad (41)$$

Esta es la función que se debe minimizar y normalmente se ingresa a un software especializado que realiza el proceso de optimización. Trabajando de forma manual se utilizarán las condiciones de Karush-Kuhn-Tucker para así reducir el número de variables. (Ossa, 2016, p.21) Con base en (32) se tiene:

$$0 = \sum_{i=1}^{3} \alpha_i y_I$$

$$0 = \alpha_1(-1) + \alpha_2(-1) + \alpha_3(1)$$

$$0 = -\alpha_1 - \alpha_2 + \alpha_3$$

$$\alpha_3 = \alpha_1 + \alpha_2 \qquad (42)$$

Al reemplazar (42) en (41) y simplificar queda una función en dos variables mucho más sencilla de optimizar:

$$L_D = 2\alpha_1 + 2\alpha_2 - \frac{1}{2}[20{\alpha^2}_1 + 32\alpha_1\alpha_2 + 16{\alpha^2}_2]. \qquad (43)$$

(Ossa, 2016, p.21)

La función (43) se puede optimizar con derivadas parciales (11):

$$\frac{\partial L_D}{\partial \alpha_1} = 2 - 20\alpha_1 - 16\alpha_2 = 0, \qquad (44)$$

$$\frac{\partial L_D}{\partial \alpha_2} = 2 - 16\alpha_1 - 16\alpha_2 = 0$$

Cuando se resuelve el sistema (44), $\alpha_1 = 0$ y $\alpha_2 = \frac{1}{8}$; y teniendo en mente (42), $\alpha_3 = \frac{1}{8}$. . Como se puede ver, *"los vectores de soporte"* son la segunda y tercera muestras $(x_2\ y\ x_3)$. Ahora, con (31) se puede establecer β como combinación lineal de los vectores de soporte:

$$\beta = \sum_{i=1}^{3} \alpha_i y_i x_i$$

$$\beta = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \alpha_3 y_3 x_3$$

$$\beta = 0(-1)\binom{1}{2} + \frac{1}{8}(-1)\binom{-1}{2} + \frac{1}{8}(1)\binom{-1}{-2}$$

$$\beta = \begin{pmatrix} \frac{1}{8} \\ -\frac{1}{4} \end{pmatrix} + \begin{pmatrix} -\frac{1}{8} \\ -\frac{1}{4} \end{pmatrix}$$

$$\beta = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}. \qquad (45)$$

Ahora se utiliza (28) y se tiene

$$M = \frac{1}{\|\beta\|}$$

$$M = 2 \qquad (46)$$

lo cual indica que el margen de separación de las muestras es de 2 unidades. (Ossa, 2016, p.22)

Lo último que falta por calcular del hiperplano es β_0. Este cálculo se realiza utilizando la restricción (36) reemplazando alguna muestra arbitraria x y α sea diferente de cero $(un\ vector\ de\ soporte)$, en este caso x_2:

$$\alpha_2\left[y_2\left(x^T{}_2\beta + \beta_0\right) - 1\right] = 0$$

$$\frac{1}{8}\left\{(-1)\left[(-1 \quad 2)\begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix} + \beta_0 - 1\right]\right\} = 0$$

$$1 - \beta_0 - 1 = 0$$

$$\beta_0 = 0$$

Entonces la función indicadora del hiperplano separador es

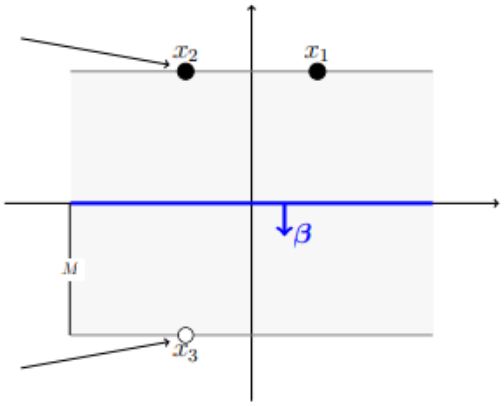
$$f(x) = x^T\beta+\beta_0 = x^T\begin{pmatrix}0\\-\frac{1}{2}\end{pmatrix} + 0 = x^T\begin{pmatrix}0\\-\frac{1}{2}\end{pmatrix} \quad (47)$$

y la regla de clasificación inducida por $f(x)$ es:

$$G(x) = signo\left[x^T\begin{pmatrix}0\\-\frac{1}{2}\end{pmatrix}\right]. \quad (48)$$

(Ossa, 2016, p.22)

Figura 21. ***Hiperplano calculado que separa las muestras de entrenamiento***



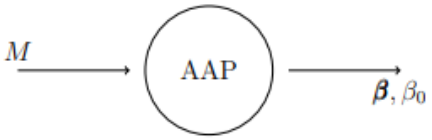
Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.23), por Jorge Ossa, 2016.

El hiperplano generado por $f(x)$ se muestra el márgen es de dos unidades y los vectores de soporte están señalados por las flechas.

3.5.5. Implementación del algoritmo de Aprendizaje del Perceptrón

3.5.5.1. Diagrama de Contexto

Diagrama de contexto del algoritmo de aprendizaje del Perceptrón



Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.25), por Jorge Ossa, 2016.

Donde M representa el conjunto de las muestras de entrenamiento y se especifica en duplas (x_i, y_i).

$$M = \{(x_1, y_1), (x_2, y_2), \cdots; (n, y_N\}$$

La primera parte de la dupla, x_i, representa un vector en el espacio de características que a su vez representa la muestra específica susceptible de clasificación. La segunda parte, y_i , representa la etiqueta preestablecida para la muestra en cuestión. Los elementos β y β_0 son los componentes de la función clasificadora $f(x)$ que especificará la frontera de decisión una vez finalizado $el\ algoritmo\ de\ aprendizaje$, esto solamente si las muestras en M representan un conjunto linealmente separable. (Ossa, 2016, p.26)

$$f(x) = \beta^T \cdot x + \beta_0$$

3.6. Especificación de la Función

3.6.1. {$\boldsymbol{Precondición}$:

$$M = \{(x_1, y_1), (x_2, y_2), \cdots; (n, y_N\}$$

$$\beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\beta_0 = 0$$

$$\rho = 1 \qquad (0 < \rho \leq 1) \qquad \}$$

3.6.2. {$\boldsymbol{Invariante}$:

Calcular la respuesta para la muestra: $y_{ic} = \beta^T \cdot x_i + \beta_0$

Si la respuesta a es diferente a la etiqueta $y_{ic} \neq y_i$

Entonces actualizar β , β_0 $\qquad \boldsymbol{\beta} \leftarrow \boldsymbol{\beta} + \rho y_i \boldsymbol{x_i}$,

$$\beta_0 \leftarrow \beta_0 + \rho y_i$$

$$1 \leq i \leq N \qquad \}$$

3.6.3. $\{\boldsymbol{Postcondición}$:

$$y_{ic} > 0 \ \forall i \ : 1 \leq i \ \leq N$$

El hiperplano separa correctamente todas las muestras} (Ossa, 2016, p.26)

3.6.4. Macroalgoritmo

Paso 0 Inicializar el hiperplano.

(Por simplicidad, establecer el hiperplano en cero)

Establecer tasa de aprendizaje $\rho : (0 < \rho \leq 1)$

(Por simplicidad, ρ puede ser establecido en 1)

Paso 1 Mientras la condición de parada sea falsa, hacer los pasos 2-6.

Paso 2 Para cada muestra (x_i, y_i) realizar los pasos 3-5

Paso 3 Establecer la muestra a procesar x_i

Paso 4 Calcular la respuesta del hiperplano actual a la muestra en proceso

$$y_{ic} = \beta_0 + \sum_j x_j \beta_j$$

Paso 5 Actualizar el hiperplano si ocurrió un error para la muestra actual

(Si$y_{ic} \cdot y_i, < 0$, ACTUALIZAR HIPERPLANO) (Ossa, 2016, p.27)

$$\rightarrow \quad \left(\boldsymbol{\beta} = \boldsymbol{\beta} + \ \rho y_i \boldsymbol{x}\right)$$

$$\rightarrow \left(\beta_0 = \beta_0 + \rho y_i\right)$$

Paso 6 Verificar condición de parada

$(si\ los\ pesos\ no\ cambiaron\ en\ el\ \boldsymbol{paso\ 2}, FIN)$

$(de\ lo\ contrario, CONTINUAR)$

El algoritmo solamente actualiza el hiperplano cuando la muestra actual no queda bien clasificada. La expresión y_{ic} se refiere a la etiqueta calculada, de acuerdo con el hiperplano actual, para la muestra x_i. (Ossa, 2016, p.27)

3.7. Implementación en Lenguaje C

Figura 22. ***Muestra en software de la función perceptrón***

```
void perceptron(double M[N][2], double y[N], int cantMuestras,
                double Beta[2], double Beta0) {
Paso 0 Inicializar el hiperplano
     Beta[0] = 0;
     Beta[1] = 0;
     Beta0 = 0;
     rho = 1;
     contadorIteraciones = 0;
Paso 1 Ciclo principal
     do
     {
          existeError = false;
     Paso 2 Recorrido sobre las muestras
          for(i=0; i<cantMuestras;i++)
          {
          Paso 3 Muestra a procesar
              /* La muestra en la posición i será procesada */
          Paso 4 Calcular respuesta hiperplano actual
              Yic = Beta0 + (M[i][0] * Beta[0] + M[i][1] * Beta[1]);
          Paso 5 Actualizar hiperplano
              if (Yic * y[i] < 0)
              {
                   Beta[0] = Beta[0] + rho * y[i] * M[i][0];
                   Beta[1] = Beta[1] + rho * y[i] * M[i][1];
                   Beta0 = Beta0 + rho * y[i];
                   existeError = true; }
          }
          contadorIteraciones++;
     Paso 6 Fin ciclo principal
     }
     while(existeError && contadorIteraciones < MAX_ITER);

     } /* Fin de la función perceptron */
```

Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.28), por Jorge Ossa, 2016.

Se presentó una "función" que recibe como parámetros de entrada un arreglo con muestras en $\mathbb{R}^2$, otro arreglo paralelo al primero con etiquetas y una variable que describe la cantidad de muestras. Como parámetro de salida se encuentra el arreglo β y el bias β_0, $Beta\ y\ Beta0$ respectivamente. Se introduce la variable $contadorIteraciones$ con el objeto de llevar la cuenta de las iteraciones del ciclo $do \cdots while$. Esto para evitar que el algoritmo se quede en un loop infinito en caso de las muestras no ser linealmente separables. Para el caso de la implementación se fija la

constante de máximo de iteraciones $(MAX\ ITER)$ en 5000, sin embargo, se puede configurar dicho valor.

La constante N es arbitraria, fijando la capacidad de los arreglos para contener las muestras de entrenamiento en 100; Esto porque se trata de una implementación didáctica, de querer trabajar con un conjunto de muestras significativa mente grande se deben buscar alternativas de memoria dinámica. (Ossa, 2016, p.29)

Figura 23. ***Microalgoritmo (en variables duales)***

Paso 0 Inicializar el vector de pesos $\boldsymbol{\alpha}$.
($Establecer\ el\ vector\ de\ pesos\ en\ cero$)
Establecer tasa de aprendizaje $\rho : (0 < \rho \leq 1)$
($Por\ simplicidad,\ \rho\ puede\ ser\ establecido\ en\ 1$)

Paso 1 Mientras la condición de parada sea *falsa*, hacer los pasos 2-6.

Paso 2 Para cada muestra $(\boldsymbol{x}_i, y_i)$ realizar los pasos 3-5

Paso 3 Establecer la muestra a procesar $\boldsymbol{x}_i$

Paso 4 Calcular la respuesta del hiperplano actual a la muestra en proceso $y_{ic} = \beta_0 + \sum_{j=1}^{N} \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$

Paso 5 Actualizar el i-ésimo componente del vector de pesos si ocurrió un error para la muestra actual
($Si\ y_{ic}.y_i < 0$, ACTUALIZAR VECTOR DE PESOS)
$\rightarrow \quad (\alpha_i = \alpha_i + \rho\, y_i)$
$\rightarrow \quad (\beta_0 = \beta_0 + \rho\, y_i)$

Paso 6 Verificar condición de parada
($Si\ los\ pesos\ no\ cambiaron\ en\ el$ **Paso 2**, FIN)
($de\ lo\ contrario$, CONTINUAR)

Tomada de: "elementos del álgebra lineal en el aprendizaje de máquina". (p.23), por Jorge Ossa, 2016.

Al demostrar el uso e importancia de los elementos como: vectores normales, matrices, espacios duales, hiperespacios, espacios de funciones lineales, combinaciones lineales, transformaciones lineales; se ha constatado que su estudio es fundamental para el aprendizaje de máquina vectorial (perceptrón), el algoritmo de máquina, la convergencia y los hiperplanos separados, que tienen la finalidad de entrenar para el aprendizaje de máquina.

CAPÍTULO 4. Aplicaciones a través de Estudios e Investigaciones que el Álgebra Lineal ha Aportado e Importado en la Ingeniería Biomédica.

En esta sección se muestran algunas investigaciones de la ingeniería biomédica, en las cuales el "álgebra lineal" se hace presente como herramienta de soporte para el funcionamiento de los equipos.

El "álgebra lineal" a través de su teoría de la función clasificadora y la función del perceptrón ha ayudado en la programación del software que controlan la maquinaria utilizada en la detección de: tejidos cancerígenos, el autismo, estimar la velocidad de la sangre en cirugías, entre otras aportaciones. Estás técnicas innovadoras le han permitido al especialista quirúrgico mejorar su praxis en pro de la salud del paciente.

4.1. Investigación que emplea técnicas ópticas no invasivas que permiten diagnosticar lesiones y extraer parámetros ópticos en tejidos biológicos

El método de clasificación para identificar lesiones en la piel humana, a partir de espectros de reflexión difusa, tiene el propósito de distinguir las lesiones benignas y malignas. Para ello, se ha requerido del análisis de diferentes algoritmos de clasificación, usando el software de aprendizaje automático y reconocimiento de patrones WEKA[4]. Además, dada la alta dimensionalidad de la señal espectral, fue empleada una técnica de selección de atributos para determinar las variables que aporten la mayor cantidad de información. Se probó la clasificación de la señal usando los algoritmos:

$$de\ máquinas\ de\ vectores\ de\ soporte, redes\ neuronales\ y\ bosques\ aleatorios,$$[5]

el desempeño fue evaluado usando el promedio de la $k - fold\ cross - validation$ tomando en cuenta los porcentajes de instancias clasificadas correctamente, $el\ índice\ kappa$, el área bajo la curva ROC, la sensibilidad, y la especifidad.

[4] Software analiza datos a través del aprendizaje automático de máquina. (Corso, 2009, 2)
[5] Función clasificadora capítulo 3, pág. 85, párr. 3.

La espectroscopia, de reflexión difusa, aunada a los soportes de las técnicas de reconocimiento de patrones y la respectiva correlación con la regla de oro (la biopsia) en el diagnóstico de enfermedades de la piel, es una técnica prometedora y no invasiva, de respuesta rápida que puede ser empleada en jornadas masivas de detección de cáncer de piel por médicos no expertos en el área dermatológica. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p. 35).

Sin embargo, una vez obtenida la señal espectral, no existe un criterio único para la categorización de los espectros; por ello para llevar a cabo la tarea de clasificación es necesaria la utilización de técnicas de reconocimiento de patrones, que aprendan las posibles diferencias entre patrones de tejido sano y lesionado, y logren con un alto porcentaje de éxito, diagnosticar a cuál grupo pertenece un espectro. Es importante señalar que el "álgebra lineal" proporciona herramientas importantes para que el software pueda funcionar y realizar la tarea de clasificación, esto se atribuye a la teoría de hiperplanos separados de vectores mencionado en el capítulo tercero, en su página 104.

Para llevar a cabo tareas de reconocimiento de patrones es necesario contar con un conjunto de datos certificados, en este caso los datos han sido valorados por médicos especialistas en oncología y dermatología lo que proporciona herramientas a *la función de la máquina de vector de soporte para que el software*[6] para que pueda clasificar datos aprendidos, con la técnica de una red neuronal. Esto ha permitido clasificar tejidos en la piel y obtener un mejor diagnóstico al tener más información de una lesión maligna o benigna. Este procedimiento ayuda a los médicos especialistas a recopilar un conjunto de datos de reconocimiento de patrones, los cuales asimilen las posibles diferencias entre los tejidos sanos y lesionados.

[6] El álgebra lineal aporta la teoría de vectores y matrices para el uso del código R y algoritmo PCA, capítulo 3, pág. 90, párr. 1.

De este modo, los expertos obtienen resultados inmediatos y pueden intervenir a tiempo las lesiones en los tejidos de la piel. Con el propósito de disminuir el mayor porcentaje de muerte relacionada con el cáncer de piel. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p. 35)

Alternativamente, para detectar lesiones en la piel también se ha empleado el análisis de imágenes obtenidas mediante epiluminescencia, en conjunto con $algoritmos\ como\ el\ clasificador\ k-nearest\ neighbors\ (KNN)$ se evaluaron atributos en la imagen como: tamaño de la lesión y, media del borde de la lesión.

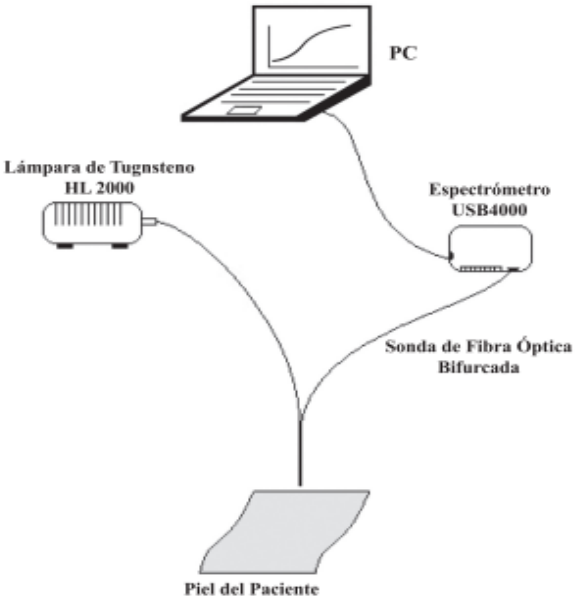
Probada por $Wallace\ et\ al$ en detección de lesiones en la piel mediante espectrometría de reflexión difusa y $algoritmos\ de\ clasificación$. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.38-39)

En este trabajo se ha evaluado el desempeño de diferentes técnicas computacionales para clasificar espectros de reflexión difusa entre muestras de tejido sano y maligno, usando el $software\ WEKA$ de investigación en aprendizaje automático y minería de datos, las técnicas implementadas

$son\ máquinas\ de\ vectores\ de\ soporte, redes\ neuronales\ y\ bosques\ aleatorios.$

Se muestra a continuación en la figura 24 como la función perceptrón aporta técnicas para el soporte del $software\ WEKA.$

Figura 24. ***Esquema del montaje experimental para obtener los espectros de reflexión difusa***



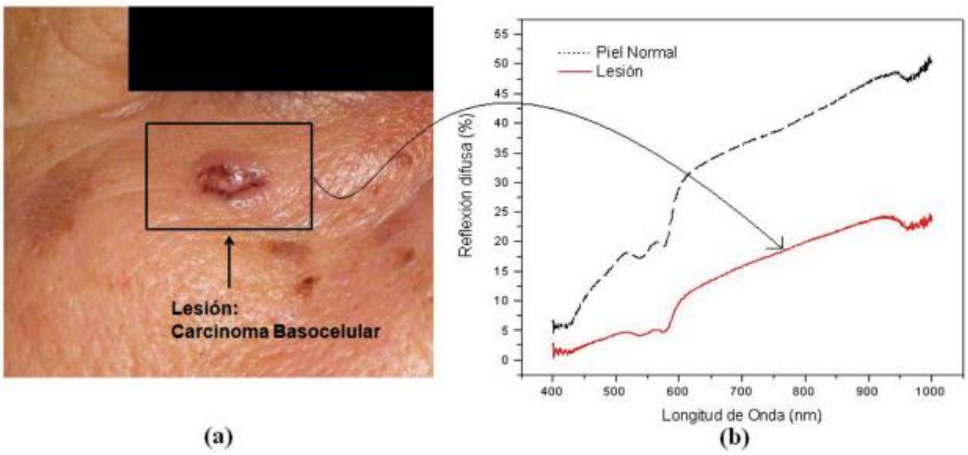
Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.36), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

Se presenta el esquema del arreglo experimental empleado para capturar los espectros, el cual consiste de un espectrómetro USB4000 fabricado por la empresa Ocean Optics que está equipado con un detector CCD Toshiba de 3648 elementos, una fuente de luz HL2000 optimizada para el VIS-NIR (360 nm-2000 nm), una sonda de fibra óptica bifurcada (R600\7\VIS\125F) de la misma firma comercial, un patrón de reflexión (Teflón) y un computador con el software SpectraSuite (Ocean Optics) que calcula automáticamente el porcentaje de luz reflejada mediante la siguiente expresión "matemática"

$$\%R(\lambda) = \frac{S(\lambda)-D(\lambda)}{R_{mr}-D(\lambda)} x100\% \qquad (a)$$

Donde $S(\lambda)$ es la señal del medio analizado $(Piel)$ para cada longitud de onda $(\lambda), D(\lambda)$ es la señal de oscuridad y $R_{mr}(\lambda)$ es la muestra de referencia. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.36)

Figura 25. ***(a) Fotografía de la lesión: carcinoma basocelular, (b) Espectros de reflexión difusa para piel normal y lesión***



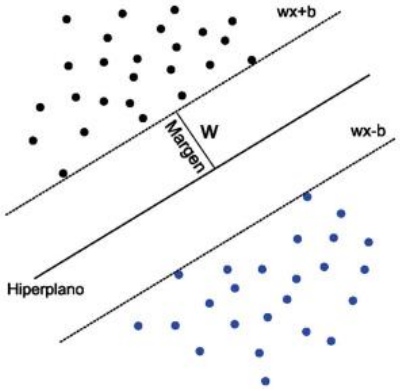
Nota: es mostrada una fotografía con una lesión de tipo carcinoma Basocelular y dos curvas espectrales una correspondiente a la lesión y otra a un tejido sano o piel normal. Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.36), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

Las etapas de clasificación consisten en normalizar los datos para evitar que la función de decisión sea influenciada por variables de magnitudes considerablemente mayores que otras; y posteriormente, aplicar una técnica de selección de atributos para analizar la influencia de los atributos de los patrones de entrenamiento en el proceso de su propia categorización y así poder determinar un subconjunto óptimo de atributos para evitar redundancias de información. La clasificación de los espectros fue realizada mediante las máquinas de vectores de soporte, árboles aleatorios y redes neuronales artificiales. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.37)

Las máquinas de soporte vectoriales (MSV) creadas por Vapnik se popularizaron debido a: sus características, rendimiento, a que permiten enfrentar problemas de clasificación en dominios complejos y pueden ser usadas para extraer información relevante a partir de conjuntos de datos y construir algoritmos de clasificación eficientes y rápidos para datos masivos. Son un modelo de clasificación cuyo funcionamiento se basa en la

búsqueda de un margen de separación máximo entre un hiperplano y los patrones de las diferentes clases que comprenden el conjunto de entrenamiento, el modelo de optimización busca establecer la frontera de decisión, mediante los patrones que más resaltan la distribución de clases, estos son los llamados vectores de soporte. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.37)

Figura 26. Hiperplano de separación



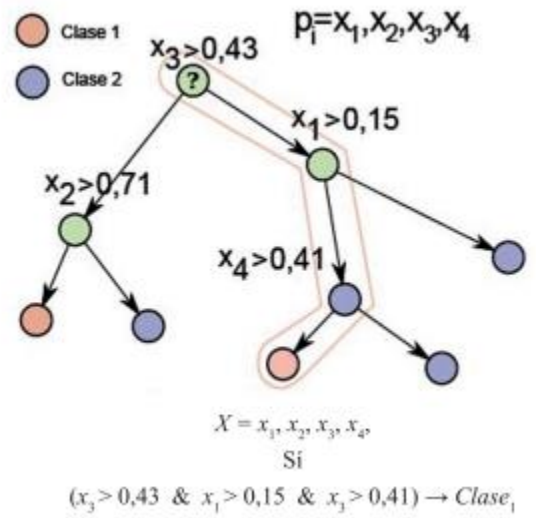
Nota: muestra el hiperplano de separación que es calculado maximizando la distancia de los patrones más cercanos. Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.37), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

La teoría de los Hiperplanos Separados del "álgebra lineal" aporta a las máquinas de soporte vectoriales para que las mismas puedan clasificar e identificar lesiones en la piel programando software para el análisis de datos. [7].

Los árboles de decisión (AD) son árboles dirigidos en los que en cada nodo se realiza una consulta a una de las características de un patrón con el objetivo de asignarle una categoría. Partiendo desde el nodo raíz hasta algún nodo hoja se van considerando las posibilidades de que el patrón pertenezca a una u otra clase, la decisión final depende del nodo hoja en el que se termine, ya que cada uno de estos nodos tiene asociada una categoría. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.33)

[7] Los Hiperplanos Separados definidos en el capítulo 3, pág. 88, párr. 2
Código R capítulo 2, pág. 53, párr.1 y algoritmo PCA capítulo 3, pág. 78, párr. 2

Figura 27. ***Ejemplo árbol de clasificación***



Nota: el uso de un árbol de clasificación como discriminador se puede interpretar mediante la conjunción lógica de las decisiones tomadas en cada nodo. Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.37), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

Entre los métodos empleados con esta técnica se tiene el Random Forest (bosques aleatorios) que es un algoritmo compuesto por numerosos árboles de clasificación, en él se definen una cantidad de árboles a desarrollar y una cantidad de atributos m tal que sea menor a la cantidad total de atributos.

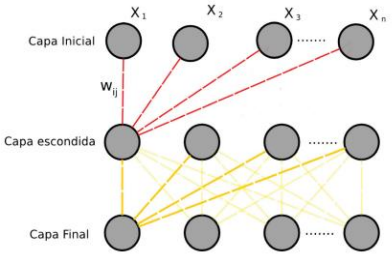
Entre los árboles se reparten *k* patrones con reemplazo y se desarrollan los árboles, el resto de los patrones son usados para la prueba.

Al desarrollar cada nodo se eligen *m* atributos y se determina el mejor atributo para desarrollar el nodo. Para el entrenamiento los patrones son repartidos aleatoriamente con repetición entre cada árbol. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.37)

Las redes neuronales (RN) consisten en un sistema de procesamiento de información planteado inicialmente por inspiración biológica, está compuesta por unidades de procesamiento llamadas neuronas que están separadas en capas donde se recibe,

procesa y transmite la información. Las unidades de procesamiento reciben como entrada los elementos del patrón con el que es alimentada la red neuronal y estas transmiten los elementos como señales a la siguiente capa, el enlace entre cada capa está afectado por un peso que biológicamente representa el nivel de sinapsis en la conexión. Las salidas de las neuronas de la última capa representan la respuesta de la red neuronal al estímulo inicial. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.37)

Figura 28. ***Arquitectura general de una red neuronal***



Nota; Muestra la arquitectura de una red neuronal de flujo hacia delante. Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.37), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

Los resultados y la discusión en la prueba de cada algoritmo fueron variados en los parámetros, en diferentes intervalos, cada prueba se realizó mediante validación cruzada con $\mathrm{k} = 10$ pliegues. Es decir, el conjunto de instancias disponibles se divide en diez partes iguales usando una para la validación del modelo y el resto para su entrenamiento, el resultado final de la prueba se obtiene al promediar las métricas arrojadas en cada una de las pruebas. La validez de los modelos fue evaluada con las mediciones del porcentaje de predicciones correctas, el área bajo la curva ROC, el índice $Kappa$ y el promedio de la sensibilidad y especificidad de los resultados de la tabla 6. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.37)

Tabla 6.

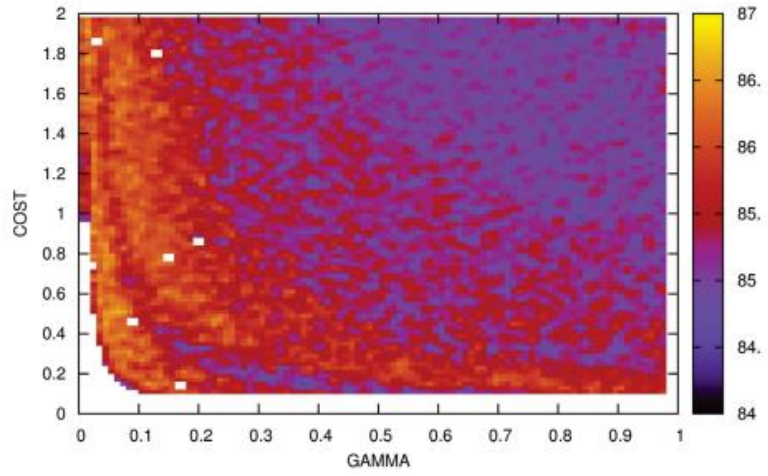
Resultados obtenidos con los algoritmos de clasificación implementados para las variables de desempeño evaluadas.

Algoritmo	Aciertos %	Índice Kappa	Curva ROC	Sensibilidad (%)	Especificidad (%)
MSV	87,0175	0,7349	0,8773	79,07	93,59
AD	87,3684	0,9194	0,7409	80,15	93,08
RN	89,1228	0,9377	0,7909	92,85	87,86

Nota: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.38), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

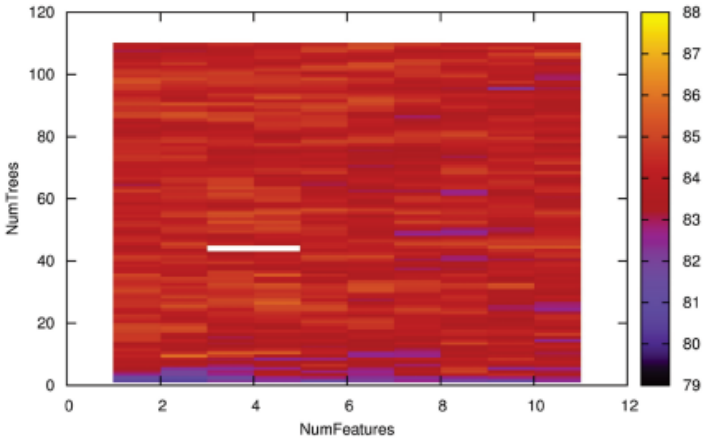
Las máquinas de soporte vectorial fueron implementadas con un núcleo RBF (Función de Base Radial) variando los parámetros γ entre $0\ y\ 1$ con pasos de $0{,}1\ y\ C$ entre $0{,}1\ y\ 2$ con pasos de $0{,}2$; en total fueron realizadas 100 pruebas. Al realizar los experimentos se evidencia la relación entre los parámetros $\gamma\ y\ C$ y los resultados. En la Fig. 29 se puede apreciar que la relación entre estos parámetros no es lineal, pues se forma una curva en la que los valores más altos para las variables de medición del desempeño se dan por la combinación de valores altos de γ y bajos de C y viceversa. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.38)

Figura 29. ***Mapa porcentaje de aciertos MSV.***



Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.38), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

Figura 30. ***Mapa del porcentaje de aciertos para el Random Forest.***

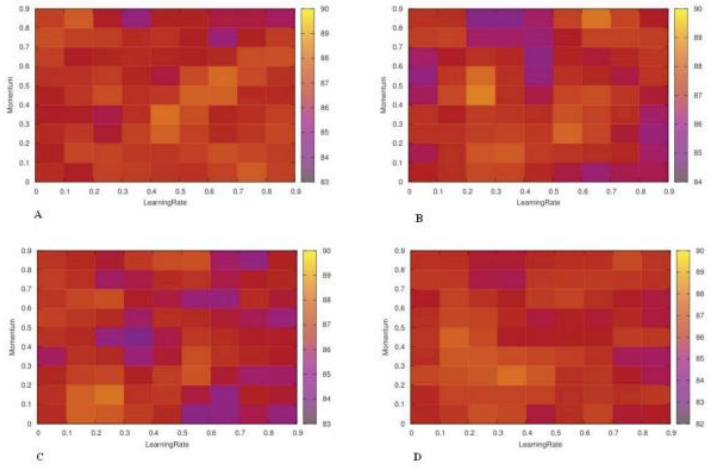


Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.39), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

El estudio con los árboles de decisión fue llevado a cabo mediante el algoritmo Random Forest, variando los parámetros NumTrees (número máximo de árboles) entre 1 y 110 y NumFeatures (número de características) entre 1 y 11. En total se realizaron 1210 pruebas. Este algoritmo logra clasificar correctamente al 88,071% de los espectros si se le permite desarrollar 44 árboles y evaluar cuatro atributos en cada uno. En la figura 29 no se

observa a simple vista una relación entre la variación de los parámetros y las variables de medición; aunque la documentación del Random Forest señala que este es, considerablemente, sensible a la variación del número de características. La red neuronal implementada en WEKA corresponde a la clase MultilayerPerceptron, que es un perceptrón multicapa entrenado usando backpropagation. Se utilizó una capa oculta y se varió la cantidad de neuronas entre 3 y 12 y en cada caso, los parámetros momentum y learnig-rate fueron variados en el intervalo 0-1. Aunque no se percibe que el sistema sea sensible a la variación de los parámetros, al validar el modelo con las medidas de desempeño, se ha obtenido una índice kappa que indica que existe un alto acuerdo entre las predicciones y las clases originales. En cuanto a la sensibilidad y especificidad, los parámetros obtenidos indican que la red neuronal tiene una alta capacidad de detectar correctamente a pacientes sanos y enfermos. La figura 30 corresponde a los mapas del porcentaje de aciertos, variando el número de neuronas.

Figura 31. ***Mapa del porcentaje de aciertos para el MultilayerPerceptron. A) tres neuronas, B) seis, C) nueve y D) doce neuronas.***



Tomada de: "Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa". (p.39), por Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010.

En cuanto a los resultados obtenidos, una vez evaluados cada uno de los modelos de clasificación elegidos, se concluye que una red neuronal de una capa oculta de seis

neuronas con los parámetros momentum y learnig-rate en 0,6 y 0,3 respectivamente, poseen la mayor capacidad de distinguir entre patrones correspondientes a lesiones sanas y malignas. Es decir, existe una alta probabilidad que el modelo clasifique como positivas las instancias positivas; además el alto promedio de sensibilidad y especificidad indican que el modelo tiene un excelente desempeño para diferenciar entre lesiones y tejido sano. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.39)

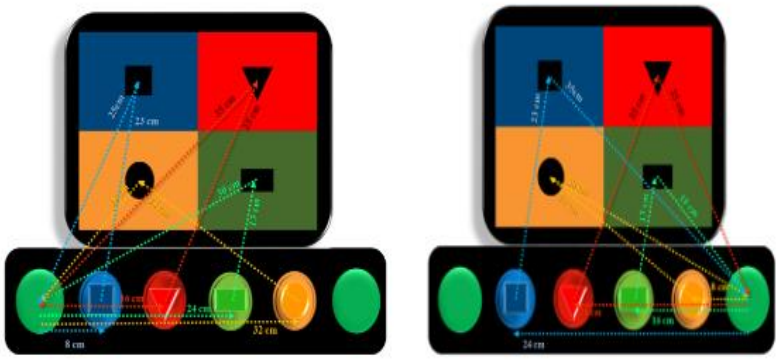
Es evidente que los algoritmos presentados tienen un buen desempeño gracias a las herramientas usadas de la función de clasificación proporcionada por la máquina de soporte vectorial, red neuronal y la arquitectura de los bosques aleatorios. Gracias a la teoría de vectores, matrices, producto punto, transformaciones lineales que proporciona el "álgebra lineal"[8] es posible que los softwares produzcan resultados exactos en el reconocimiento de tejidos malignos y benignos en la piel permitiendo a los especialistas de la medicina diagnosticar para que, puedan intervenir en los pacientes a tiempo y así disminuir el porcentaje de mortalidad. (Guillén, García, Vázquez, Atencio, Ramos, Delgado, 2010, p.40)

Esto nos lleva a resaltar la importancia del "álgebra lineal" para la tecnología de la biomédica en la salud. Así también se destacan, las aplicaciones en diferentes áreas de la salud. Entre otras de las aportaciones se destaca la detención de la parálisis cerebral, estudios realizados en las redes neuronales artificiales para cuantificar motricidad en niños con hemiparesia cerebral; y, podemos diferenciar la parálisis cerebral en función de la parte del cuerpo que se encuentra afectada, teniendo así una clasificación por criterios topográficos. La hemiparesia (centro) se produce cuando la discapacidad se presenta únicamente en la mitad izquierda o derecha del cuerpo. Esto se da, gracias a un sensor inercial (IMU) que se coloca en el dorso de la mano a medir y el Sorting Block Box (SBB),

[8] Conceptos del álgebra lineal estudiados en el capítulo 3, pág. 96, párr.1
Los conceptos del código R software de análisis de datos capítulo 2, pág.53, párr. 1
La función clasificadora capítulo 3, pág. 85, párr.2, pág. 86, párr. 2-3.

ver figura 30, que consta de un tablero con huecos de diferentes formas geométricas (cuadro, triángulo, círculo, rectángulo) y 4 piezas de madera que deben ser colocadas en sus respectivos espacios; dicho tablero tiene por dentro diversos sensores que detectan la fuerza del movimiento y si la pieza está o no en su lugar. (Campos, Urióstegui, González y Roiz, 2020, p.1-2)

Figura 32. ***Tablero Sorting Block Box. Se muestran las distancias (cm) de las trayectorias idealesrecorridas por el brazo izquierdo (izq) y el brazo derecho (der).***



*Tomada de: "*Redes Neuronales Artificiales para Cuantificar la Motricidad en Niños con Hemiparesia Cerebral. In *Memorias del Congreso Nacional de Ingeniería Biomédica". (p.2), por* Campos, Urióstegui, González y Roiz, 2020*.*

4.2. El Diagnóstico del Autismo a través del ADOS-G

La escala ADOS-G es una escala semiestructurada, instrumento basado en la observación de 4 módulos que se gestionan de acuerdo con la edad y las habilidades lingüísticas del niño. Frente al desafío de caracterizar, medir los síntomas y localizar al paciente en el nivel A de funcionamiento, el ADOS -G tiene la ventaja, con su variedad de tareas, para hacer un diagnóstico sobre una base de estudio observacional.

El diagnóstico del autismo requiere del uso de herramientas de diagnóstico, validadas internacionalmente, y utilizadas por los profesionales de la salud, expertos en trastornos del espectro autista; lo cual, requiere de procesamiento de mucha información y un entendimiento técnico de cada una de las áreas evaluadas en ellas. La clasificación del

impacto que tienen cada una de estas áreas, así como la propuesta de un sistema que pueda ayudar a los expertos en el diagnóstico, es una tarea compleja; por lo cual, se presenta una metodología, utilizada para encontrar los elementos más significativos de la herramienta de diagnóstico de autismo ADOS-G a través de redes neuronales artificiales, entrenadas con retropropagación del error. El número de casos para entrenamiento de la red se seleccionó utilizando el método de Taguchi con arreglos ortogonales, reduciendo el tamaño de la muestra de 531,441 a solo 27 casos.

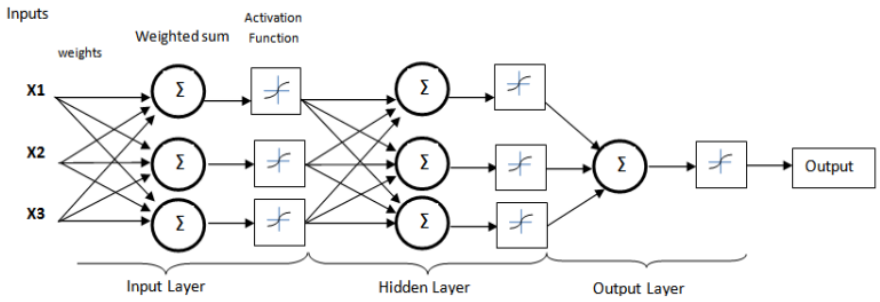
La red entrenada tiene una exactitud del 100% validada con 11 casos diferentes de niños evaluados para diagnóstico de trastorno del espectro autista, con lo que se obtuvo una especificidad y sensibilidad de 1.

La red neuronal artificial se utilizó para clasificar los 12 elementos del algoritmo de la herramienta ADOS-G en tres niveles de impacto: alto, medio y bajo. Se encontró que los elementos “Mostrar”, “Placer compartido durante la interacción” y “Frecuencia de vocalizaciones dirigidas a otros” son las áreas de mayor impacto para el diagnóstico de autismo. La metodología presentada puede ser replicada para diferentes herramientas de diagnóstico de autismo para clasificar sus áreas de otros aportes. (Reyes, Ponce, Grammatikou y Molina, 2014, p.6-7). Este es otro paradigma que presenta las aportaciones de los contenidos del “álgebra lineal” en los reconocimientos y clasificación para el diagnóstico de pruebas que realizan los profesionales de la salud. En donde el álgebra lineal provee de importantes herramientas para el uso de la escala ADOS-G como es la función clasificadora, la cual trabaja mostrando resultados exactos en cada diagnóstico médico.

Las redes neuronales artificiales (RNA) son modelos computacionales basados en un modelo simplificado de versión de redes neuronales biológicas con las que comparten algunas características como: la adaptabilidad aprender, generalización, organización de datos y procesamiento en paralelo. Una RNA está compuesta por capas, una capa de

entrada, una capa de salida y una o varias capas ocultas[9]. Dentro de cada capa hay varias neuronas que están procesando unidades que envían información, a través de señales entre sí y una función de activación determina la salida como se muestra en la figura 33.

Figura 33. ***Red neuronal artificial multicada***



Tomada de: "Metodología para ponderar áreas de evaluación de la herramienta ADOS-G en trastorno del espectro autista con redes neuronales artificiales y método de Taguchi". (p.228), por Ponce, Grammatikou y Molina, 2020.

El método de entrenamiento de retro propagación consiste en minimizar el error, con respecto a los pesos a través del descenso de gradiente. El error es la diferencia entre la salida deseada y la producción real entregada por la RNA. Minimizando el error se logra devolviendo el error desde la capa de salida hasta todos los elementos ocultos involucrados, capas que obtienen una parte proporcional de ese error, cada neurona toma una parte proporcional y vuelve a entrenar el peso para cada entrada durante la próxima época. La función de activación[10] es una función diferenciable de las entradas dada por

$$\mathcal{Y}_k^P = F(\, s_k^{\mathcal{P}})$$

donde,

[9] El perceptrón: sus generalidades. Estudiado en el capítulo 3, pág. 90, párr. 1.
[10] Regla de activación del perceptrón, estudiado en el capítulo 3, pág. 91-92.

$\mathcal{Y}_k^P$ es el valor de salida para cada unidad.

F es la función de activación.

$s_k^{\mathcal{P}}$ es la entrada de la unidad k en el patrón p.

donde,

$$s_k^{\mathcal{P}} = \sum_j w_{jk}\, \mathcal{Y}_j^P + \theta_k \qquad [11]$$

En "álgebra lineal" la expresión es una combinación lineal[12].

Donde, w_{jk} es el factor de ponderación para la entrada j y la salida k. La expresión w_{jk} describe una matriz y θ_k representa el elemento vectorial siendo el umbral.

Donde,

$$j = layer(j)$$

El error se define como el error cuadrático E^p en las unidades de salida para el patrón p entre las salidas deseadas y la salida real. Y puede denotarse como una combinación lineal así,

$$E^p = \frac{1}{2}\sum_{\circ=1}^{N_\circ} (d_\circ^p - \mathcal{Y}_\circ^p)^2$$

La matriz elemental $d_\circ^p$ es la salida deseada para la unidad o en el patrón p.

El error cuadrático sumado es entonces E dado por

[11] Especificación de la Función, estudiado en el capítulo 3 pág. 116.
[12] Representación de transformaciones por matrices, estudiado en el capítulo 2, pág. 36.

$$E = \sum_{p} E^{p}$$

donde E^{p} es el error en el patrón p. Aplicando la cadena de reglas

$$\frac{\delta E^{p}}{\delta w_{jk}} = \frac{\delta E^{p}}{\delta s_{k}^{\mathcal{P}}} \frac{\delta s_{k}^{\mathcal{P}}}{\delta w_{jk}}$$ [13]

Siendo el módulo de una razón. Donde,

$$\frac{\delta s_{k}^{\mathcal{P}}}{\delta w_{jk}} = \mathcal{Y}_{j}^{P}$$

Definición de una regla de activación

$$\Delta_{p} w_{jk} = \gamma \delta_{k}^{p} \mathcal{Y}_{j}^{p}$$

Donde,

$$\delta_{k}^{p} = \frac{\delta E^{p}}{\delta s_{k}^{\mathcal{P}}}$$

A aplicar la regla de la cadena para el cambio en el error en función de la salida y el cambio en la salida, en función de los cambios del aporte,

$$\delta_{k}^{p} = -\frac{\delta E^{p}}{\delta s_{k}^{\mathcal{P}}} = -\frac{\delta E^{p}}{\delta \mathcal{Y}_{k}^{\mathcal{P}}} \frac{\delta \mathcal{Y}_{k}^{\mathcal{P}}}{\delta s_{k}^{\mathcal{P}}}$$

A usar la regla de la cadena, cuando k es una unidad oculta, se llama h

$$\frac{\delta E^{p}}{\delta \mathcal{Y}_{k}^{\mathcal{P}}} = \sum_{\circ=1}^{N_{\circ}} \frac{\delta E^{p}}{\delta s_{k}^{\mathcal{P}}} \frac{\delta s_{k}^{\mathcal{P}}}{\delta \mathcal{Y}_{k}^{\mathcal{P}}} = -\sum_{\circ=1}^{N_{\circ}} \delta_{\circ}^{p} \; w_{h\circ}$$

Estos rendimientos a

[13] Espacio de producto interno, estudiado en el capítulo 2, pág. 47, párr. 3.

$$\delta_h^p = F'(s_h^p) \sum_{\circ=1}^{N_\circ} \delta_\circ^p w_{h\circ} \quad [14]$$

Aumentar el número de neuronas ocultas puede evitar caer en un mínimo local y disminuir el error, pero puede consistir en un largo proceso de entrenamiento.

El “álgebra lineal” es una herramienta que aporta propiedades que permiten enlazar con diferentes teorías como: diferenciación, las series u otras empleadas en la matemática moderna facilitando la compresión de la ingeniería biomédica.

Es de importancia también señalar las aportaciones del “álgebra lineal” en los Nanorobots como es en el caso de estimación de la velocidad de la sangre. Siendo un importante paradigma que evidencia la presencia de los contenidos del “álgebra lineal” para el desarrollo, progreso y funcionamiento de la innovadora ingeniería de la biomédica.

4.3. Estimación de la Velocidad de la Sangre basada en un Observador Óptimo para la Navegación Magnética de Nanorobots

Los nanorobots magnéticos se utilizan recientemente en un sistema de administración de fármacos, con el fin de realizar procedimientos médicos mínimamente invasivos. Una de las principales dificultades para controlar los nanorobots es el complicado comportamiento de la fuerza de arrastre hidrodinámica y los factores inciertos de su dinámica hidrodinámica y los factores inciertos de su dinámica.

La fuerza de arrastre depende de la velocidad de la sangre, pero por desventaja en anteriores diseños de sistemas de control por retroalimentación, se suponía que la velocidad de la sangre era constante. Esta suposición no reflejaba la realidad y reducía la viabilidad de los sistemas de control. Teniendo en cuenta esta eficaz observación se propone para estimar la velocidad de la sangre. La única información necesaria para el

[14] Transfomaciones Lineales, estudiado en el capítulo 2, pág. 35.

observador propuesto y el sistema de control global es la posición de los nanorobots. Las características, la estabilidad y las reglas de ajuste de ganancia del observador óptimo propuesto.

Los resultados de la simulación se realizan en MATLAB/Simulink para demostrar la eficacia del observador óptimo propuesto, así como del sistema de control de retroalimentación global para la navegación de Nanorobots en vasos sanguíneos. (Do, 2017, p. 3)

4.3.1. Formulación del Problema sobre la Navegación de Nanobots en Vasos Sanguíneos

La modelización dinámica del espacio de estados de los nanorobots en arterias es el siguiente: **sistema de ecuaciones lineales**

$$\begin{cases} \dot{x}_1 = \dot{x}_2 \\ \dot{x}_2 = f(x_2, v_1) + \dfrac{t_m M}{\rho}\, u \\ y = x_1 \end{cases} \qquad (1)$$

donde $x_1\ y\ x_2$ son la posición y la velocidad agregadas respectivamente, u es el gradiente de campo magnético considerado como entrada de control, v_1 es la velocidad de la sangre, t_m es la relación ferromagnética, ρ es la densidad del agregado, M es la

M es la magnetización, e y es la salida, que es la posición medida por $MRI\ o\ MPI$, respectivamente. (Do, 2017, p. 2)

En “álgebra Lineal” $\dot{x}_1 = \dot{x}_2$ son elementos vectoriales de n-dimensiones, $y = x_1$ representa la ecuación lineal.

El sistema (1) se apoya con propiedades importantes del Álgebra Lineal como, el símbolo $\dot{x}$, representa un vector en n-dimensiones (el espacio vectorial escalar) así, se representa la velocidad vectorial del flujo en la posición espacial n-dimensional (Zamora,

Ibargüen, 2012, p. 50). Apoyada en la teoría de espacio vectorial en los sistemas de espacios vectoriales y funciones que se pueden describir finalmente en una transformación lineal como es el caso de $f(x_2, v_1) + \frac{t_m M}{\rho}\ u$ donde, $f(x_2, v_1)$ es el vector producto de las matrices elementales y $\frac{t_m M}{\rho}$ la razón vectorial, sistema utiliza importantes contenidos del "álgebra lineal" para el estudio de la estabilidad o inestabilidad de la velocidad en el flujo sanguíneo por medio de la función diferencial. El "álgebra lineal" contribuye a estos sistemas de transformaciones lineales por las propiedades de matrices, vectores, espacios vectoriales, espacios de funciones, ordenadores y combinaciones lineales[15].

La función f se expresa como

$$f(x_2, v_1) = \frac{1}{m}\left[-sgn(x_2 - v_1)\ \left((a|x_2 - v_1|) + b(x_2 - v_1)^2 + c\frac{(x_2 - v_1)^2}{1 + d\sqrt{|x_2 - v_1|}} + V\left(\rho_f - \rho\right) g\right)\right] \quad (2)$$

Donde

$$a = \frac{6\pi\eta r}{\beta},\ b = \frac{0.2\ \rho f \pi r^2}{\beta^2},\ \ c = \frac{3\ \rho f \pi r^2}{\beta^2},\ \ d = \sqrt{\frac{2r\rho_f}{\beta\eta}} \quad (3)$$

$$sgn(z) = \begin{cases} -1\ if\ z\ < 0 \\ 0\ if\ z = 0 \\ -1\ if\ z\ > 0 \end{cases}$$

[15] Conceptos definidos en el capítulo 2 Matrices (pág. 23), vectores - espacios vectoriales (pág.24), espacios de funciones, vector ortogonal (pág. 51), transformaciones (pág. 35) -combinaciones lineales (pág. 30).

En (3), r es el radio agregado, β es una relación adimensional relacionada con el efecto de pared causado por la oclusión por un agregado esférico de radio $r, \eta\ y\ \rho_f$ son la viscosidad y la densidad de la sangre, respectivamente.

Obsérvese, que la función f en $(1)\ y\ (2)$ incluye la expresión de la fuerza de arrastre y el peso (el último término del paréntesis es el peso y los demás son la fuerza de arrastre) mientras que la entrada de control u es proporcional a la fuerza magnética.

Además, en (1), sólo se puede medir la posición x_1 del nanorobot, la velocidad del nanorobot y la velocidad de la sangre son necesarias para ser calculado y estimado. Con el fin de estimar la velocidad de la sangre, su dinámica, la velocidad pulsátil de la sangre $v_1$puede definirse como una serie de Fourier truncada de enésimo orden

$$\begin{cases} \dot{v_1} = v_2 \quad producto\ vectorial \\ v_2 = -\omega^{\dot{2}}(v_1 - v_3) \quad matrices\ de\ aplicaciones\ lineales \\ \vdots \qquad asusibas\ n - dimensiones \\ \dot{v}_{2k-1} = v_{2k}\ \ espacio\ vectorial \\ \dot{v}_{2k} = -\omega^2(k^2 v_{2k-1} - v_{2k+1})\ matrices\ de\ aplicaciones\ lineales \\ \vdots \\ \dot{v}_{2n+1} = 0 \quad vector\ ortogonal \end{cases} \tag{4}$$

Donde ω es la pulsación del flujo sanguíneo que se supone conocido y v_{2n+1}es el valor medio de la velocidad de la sangre. (Do, 2017, p. 2)

Es interesante resaltar que elementos como: el sistema de espacio vectorial, transformación lineal y la forma cuadrática de Jordan ayudan a visualizar la presencia de los contenidos del “álgebra lineal” en la ingeniería biomédica, los cuales son de gran importancia para el desarrollo tecnológico.

El sistema (4) muestra aplicaciones teóricas del doble de una matriz n-dimensiones lo que constituye las características de un vector que en su conjunto forman una

combinación lineal. Es relevante apreciar que, así como la teoría del "álgebra lineal" contribuye a estimar la velocidad de la sangre; también, la teoría diferencial con la serie de Fourier truncada de enésimo orden del cálculo es oportuno aportan importantes teorías a la ingeniería biomédica.

4.4. Diseño del Control de Realimentación

4.4.1. Diseño de Control Adaptativo en Modo Deslizante

Consideremos la siguiente ley ASMC para el sistema (1).

$$\begin{cases} \sigma = e_1 + \alpha e_2 \quad combinación\ lineal \\ u = -\frac{1}{\alpha}\frac{\rho}{\tau_m M}(\alpha f - \quad x_{1r} - \alpha \dot{x}_{1r} + \theta sgn(\sigma))\ transformación\ lineal \\ \dot{\theta} = \frac{|\sigma|}{\gamma} \quad ortogonal \end{cases} \tag{5}$$

donde $e_1 = x_1 - x_{1r}$ y $e_2 = x_2 - \dot{x}_{1r}$ son los errores, x_{1r} es el comando de referencia de posición de los nanorobots, $\alpha\ y\ \gamma$ son las ganancias de control. Nanorobots. En (5), la ganancia de control adaptativa θ se estima en línea, mediante una ley para evitar ganancias de control elevadas, así como con altos esfuerzos de control. La ley SMC en (5) puede incluir chattering, debido al uso de una función de signo. Existen varias formas de reducir la vibración, sin embargo, un método sencillo es la función de signo aproximada mediante la siguiente función aproximada:

$$sing(\sigma) = \frac{\sigma}{|\sigma| + \varepsilon} \tag{6}$$

donde ε es un número positivo muy pequeño[16]. (Do, 2017, p. 2)

[16] Conceptos definidos en el capítulo 2, pág. 46.

4.4.2. Cálculo de la Velocidad de los Nanorobots

Dado que la velocidad del nanorobot es la derivada temporal de la de la posición del nanorobot, puede calcularse directamente a partir de la posición del nanorobot, mediante la siguiente fórmula en tiempo discreto

$$x_2 = \frac{\varphi}{T+\varphi}\, x_2(k-1) + \frac{1}{T+\varphi}[x_1(\mathrm{k}) - x_1(\mathrm{k}-1)] \qquad (7)$$

donde $k\ y\ k-1$ denotan dos instantes de muestreo consecutivos, T es el período de muestreo y φ es una constante temporal suficientemente pequeña. (Do, 2017, p. 2)

La teoría de las transformaciones lineales aporta importantes propiedades matemáticas para la programación de los softwares utilizados en la ingeniería biomédica, para que los resultados sean exactos como en el caso de controlar la velocidad de la sangre como en (5) y (7).

4.4.3. Diseño del Observador de la Velocidad de la Sangre

Combinando (1) y (4), obtenemos las siguientes ecuaciones dinámicas para estimar la velocidad de la sangre,

$$\begin{cases} \dot{x}_1 = x_2 \ \ producto\ vectorial \\ \dot{x}_2 = -ax_2 + \dot{a}x_3 + f'(x_2, x_3, u)\ \ combinación\ lineal \\ \dot{x}_3 = x_4\ producto\ vectorial \\ \dot{x}_4 = -\omega^2(x_3 - x_5)\ Matrices\ de\ aplicaciones\ lineales \\ \vdots \quad asucibas\ n-\ dimensiones \\ \dot{x}_{2k+1} = x_{2k+2}\ \ espacio\ vectorial \\ \dot{x}_{2k+2} = -\omega^2(k^2 x_{2+1} - v_{2k+3})\ \ combinación\ lineal\ de\ un\ espacio\ vectorial \\ \vdots \\ \dot{x}_{2n+3} = 0\ vector\ ortogonal \\ y = x_1\ \ ecuación\ lineal \end{cases}$$

Donde, $x_{i+2} = v_i \forall i\ \in \{1,2,\cdots n\}$ y

$$f'(x_2, v_1, u) = \frac{1}{m}\left[-sgn(x_2 - v_1)\left(b(x_2 - v_1)^2 + c\frac{(x_2 - v_1)^2}{1 + d\sqrt{|x_2 - v_1|}}\right.\right.$$

$$\left.\left. + V(\rho_f - \rho)g\right) + \frac{t_m M}{\rho}u\right]$$

La expresión matemática muestra que, a través de las transformaciones lineales y los espacios de funciones se puede integrar para el desarrollo de la ingeniería biomédica otras propiedades en el área del cálculo superior, en el cálculo diferencial de la doble derivada para continuar con futuras investigaciones en este tema innovador.

La ecuación (8) puede reescribirse en el siguiente espacio de estados, en la siguiente aplicación lineal

$$\begin{cases} \dot{x} = Ax + Bf' \\ y = x_1 \end{cases} \tag{9}$$

Donde, $x = [x_1 \ \cdots \ x_{2n+3}]^T$, $B = [0 \quad 1 \quad 0 \cdots 0]^T$,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & & 0 & 0 & 0 \\ 0 & -a & a & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 & 0 \\ 0 & 0 & -\omega^2 & 0 & & 0 & 0 & 0 \\ & & & & \ddots & & & \\ 0 & 0 & 0 & 0 & & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & & -n^2\omega^2 & 0 & \omega^2 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 0 \end{bmatrix}$$

(Do, 2017, p. 3)

La forma canónica de Jordan es la forma de la matriz de un endomorfismo de un espacio vectorial en cierta base asociada a la descomposición en suma directa de subespacio invariantes bajo dicho endomorfismo. Dicha forma canónica consistirá en que la matriz estará formada por "bloques de Jordan" en la diagonal y bloques de ceros fuera de ella.

Entonces, el modelo de observador óptimo para estimar la velocidad de la sangre puede representarse como,

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Ly - LC\hat{x} + Bf' \\ y = Cx \\ \hat{x}_3 = C_T\hat{x} \end{cases} \quad (10)$$

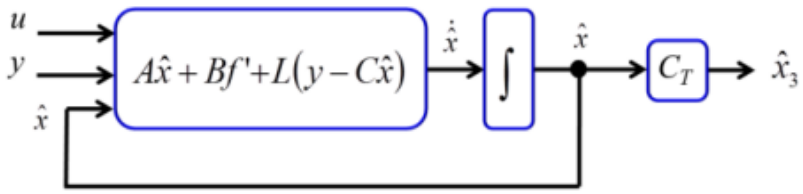
Donde, $C = [1 \quad 0 \cdots \quad 0]^T, C_T = [0 \quad 0 \quad 1 \quad 0 \quad \cdots 0]^T, \hat{x}$ es el vector estimado de x, y L es el vector de ganancia del observador del observador. (Do, 2017, p. 3)

En este sistema de aplicaciones lineales se pueden observar elementos de los espacios vectoriales Euclidiano, sistema lineal y el vector estimado de una transformación lineal.

La Figura 34 ilustra el diagrama de bloques propuesto de la velocidad de la sangre.

El diseño observado sobre la velocidad de la sangre presentando en (8,9,10) se puede designar como un sistema de espacios vectoriales diferencial, siendo así una trasformación lineal, integrada con derivada, integral, transformada de Laplace referente al cálculo superior.

Figura 34. ***Diagrama de bloques del observador óptimo de la velocidad***



Tomada de: "Blood Velocity Estimation Based on an Optimal Observer for Magnetic Nanorobots Navigation." (p.3), por Do, T. D, 2017.

La dinámica de error del observador de la corriente de carga puede obtenerse como sigue:

$$\dot{\hat{x}} = (A - LC)\tilde{x} \quad (11)$$

Donde, $\tilde{x} = x - \hat{x}$

La ecuación vectorial figura 34 expresada es una transformación lineal[17] .Es curioso notar que contiene parte diferencial e inferencial como elementos del cálculo superior. Esto resalta que el "álgebra Lineal" se aplica en los contenidos de otras áreas de la matemática como: cálculo diferencial e inferencial, ecuaciones diferenciales, series de Fourier, álgebra y otros.

La matemática forma parte de nuestra vida en todas las áreas y así como el "álgebra lineal" contiene importantes teorías que se aplican en la ingeniería biomédica; también, existen otras áreas matemáticas como: ecuaciones diferenciales, series de Fourier u otras que tienen contenido importante que aportar a la ingeniería biomédica permitiendo nuevos temas de investigación documentales.

Teorema 11. Consideremos la siguiente ecuación algebraica de $Riccati$ (ARE)

$$AP + PA^T - PC^T - PC^T R^{-1} CP + Q = 0 \qquad (12)$$

$donde, Q \in R^{(2n+3)x(2n+3)}$es una matriz simétrica semidefinida, $\mathbb{R}$ un número positivo, y $P \in R^{(2n+3)x\,1}$ es un vector solución. Y la matriz de ganancia de

L dada por,

$$L = PC^T R^{-1}$$

Entonces, el error de estimación converge exponencialmente a cero. (Do, 2017, p. 3)

La ecuación (12) contiene importantes propiedades del "álgebra lineal" como, la matriz simétrica, vectores, espacios vectoriales, transformaciones lineales u otras que aportan a la estimación de la velocidad de la sangre basada en un observador óptimo para

[17] Capítulo 2, pág. 35.

la navegación magnética de Nanorobots. Esta investigación muestra que se puede estimar con exactitud la velocidad del flujo sanguíneo, a través de nanorobots con los fundamentos teóricos del “álgebra lineal”.

Los estudios e investigaciones contemporáneas expuestos en el anterior capítulo sobre alguno de los equipos implementados en la ingeniería biomédica para diagnosticar: tejidos cancerígenos, el autismo desde temprana edad y regular la velocidad sanguínea por nanorobots son equipos que contienen la presencia de los contenidos importantes del “álgebra lineal”. Esto refuerza que la implementación de la teoría del “álgebra lineal” en el funcionamiento de los equipos biomédicos ha permitido el avance y desarrollo tecnológico de los mismos. Gracias a ellos, la ingeniería biomédica cuenta con equipos modernos y actualizados con mejores funciones que favorecen a las personas y por ende mejoran su condición de salud.

En los ejemplos mostrados la teoría del “álgebra lineal” se constituye en una herramienta que permite enlazar diferentes aplicaciones lineales como: serie de Fourier, ecuaciones diferenciales con las derivadas parciales, el teorema de Laplace u otros contribuyendo a explicar los conceptos matemáticos utilizados en la ingeniería biomédica.

Lo anteriormente expuesto, nos lleva a concluir que la teoría de “álgebra lineal” proporciona importantes conceptos para el desarrollo de la ingeniería biomédica; además, contribuye a que otros contenidos matemáticos: algoritmos y teoremas se puedan enlazar para poder dar soluciones a las investigaciones que se desarrollan en la ingeniería biomédica; con lo cual, se ha avanzado en la exactitud de los diagnósticos.

CONCLUSIONES

En esta investigación, se ha abordado el tema de la teoría del “álgebra lineal” utilizada en la programación de los equipos modernos que la ingeniería biomédica ha implementado en la detección de diversas enfermedades. Y después de haber realizado este amplio estudio se ha llegado a las siguientes conclusiones:

- Las investigaciones realizadas sobre equipos biomédicos del siglo XXI evidencian que en su implementación, soporte y funcionamiento se han empleado teorías dadas por el “álgebra lineal”. Algunos teoremas, algoritmos y definiciones de esta teoría contribuyen a la programación y avances de estos equipos. Datos de esta investigación nos ha llevado a conocer que en el tema de diagnosticar células cancerígenas sirve de herramienta el Código R. Este código se utiliza en el aprendizaje de máquina para la programación de los equipos; de allí, que se apoye en teorías del “álgebra lineal” como: la teoría de soporte de vectores y matrices. Éstas permiten el ingreso preciso de los datos con lo cual se facilita la detección de los tejidos cancerígenos y se logra un diagnóstico en menor tiempo. Esto les brinda a los especialistas la oportunidad de tratar a tiempo la enfermedad y lograr reducir las estadísticas de mortalidad generadas por el cáncer.
- Los temas planteados en este estudio, actualmente, se utilizan en muchos de los equipos biomédicos. Así, tenemos el caso de: la teoría de vectores en donde se apoya el perceptrón o aprendizaje de máquina, la teoría del hiperespacio que se aplica en la teoría de las transformaciones lineales y en los espacios vectoriales, la teoría de las funciones lineales vectoriales que permiten resolver operaciones en n-dimensiones. Las nociones de los contenidos de estas teorías permiten que esta tecnología continúe progresando porque facilita el enlace con otros campos de la matemática. Así es el caso del cálculo diferencial e inferencial, en donde el “álgebra

lineal" le aporta propiedades que permiten operar algoritmos para mejorar la ejecución de los equipos. Los estudios demuestran que se necesitan de los contenidos del "álgebra lineal" para los avances tecnológicos de los equipos biomédicos. Así también, es importante señalar que el "álgebra lineal" proporciona herramientas de enlace para trabajar en el campo diferencial, en las derivadas parciales, las series de Furrier y otras, despertando el interés por estudio de campo; el cual, forja el camino para próximos trabajos investigativos con este tema innovador.

- Se han definido importantes elementos del "álgebra lineal" que se emplean, en la actualidad, para garantizar la programación y ejecución de los equipos que la ingeniería biomédica utiliza para detectar cualquier problema de salud. Se demostró que el "álgebra lineal" contribuye a que estos equipos sean programados y den resultados en corto tiempo con diagnósticos exactos; para que el especialista en medicina pueda intervenir, inmediatamente, y mejorar la salud del paciente. Las teorías del "álgebra lineal" son esenciales para poder explicar y desarrollar los avances de los equipos biomédicos, por ende, son fundamentales los temas conceptuales de: el producto punto y la longitud, vectores unitarios, espacio y sub espacio de vectores, la desigualdad triangular, distancia entre vectores, ángulos, ortogonalidad de vectores, álgebra de vectores, ecuaciones de lineales, determinantes, matrices simétricas, entre otros.
- Los equipos implementados en la ingeniería biomédica se han utilizado para: diagnosticar tejidos cancerígenos, diagnosticar el autismo desde temprana edad y regular la velocidad sanguínea por nanorobots. Estos equipos, son programados con el contenido del "álgebra lineal" los cuales le permiten su funcionamiento. Dicho instrumental, se está investigando e innovando de manera continua; y, esto se debe al aporte de la teoría del "álgebra lineal" que permite el avance y desarrollo tecnológico de los mismos.

- Uno de los equipos más utilizados por los cirujanos al momento de regular la velocidad sanguínea, es el que utiliza un nanorobot. Este equipo para la ejecución del diseño del programa necesita algunos contenidos del “álgebra lineal” como la teoría de la transformación lineal de vectores, que es la que permite que se pueda programar para mantener la velocidad sanguínea estable. También los nanorobots en la robótica utilizan soporte de vectores para que al momento de operar un paciente se pueda controlar el movimiento con exactitud, esto gracias al uso de la teoría del vector normal y la geometría implicada en el mismo.

- Es importante destacar, que, así como la teoría del “álgebra lineal” proporciona importantes conceptos para el desarrollo de la ingeniería biomédica, para apoyar a los funcionarios de la salud; las investigaciones presentadas también muestran otros importantes contenidos de las matemáticas que favorecen el avance y desarrollo de los equipos de esta ingeniería. Entre estos se destacan: las series de Fourier, ecuaciones diferenciales con las derivadas parciales, el teorema de Laplace u otros. El “álgebra lineal” contribuye a que otros contenidos matemáticos en el campo diferencial, como los ya mencionados, se puedan enlazar para poder dar soluciones a las investigaciones que se desarrollan en la ingeniería biomédica.

REFERENCIAS BIBLIOGRÁFICAS

Babini, N. (2020) Las antecesoras de la computadora: las primeras máquinas de calcular. *Saber y tiempo*, 63.

Belke, C. H., & Paik, J. (2017). Mori: a modular origami robot. *IEEE/ASME Transactions on Mechatronics*, *22*(5), 2153-2164.

Campos, L. A., Urióstegui, I. Q., González, Y. Q., & Roiz, V. B. (2020, Octubre). Redes Neuronales Artifciales para Cuantificar la Motricidad en Niños con Hemiparesia Cerebral. In *Memorias del Congreso Nacional de Ingeniería Biomédica* (Vol. 7, No. 1, pp. 5-12).

Do, T. D. (2017). Blood Velocity Estimation Based on an Optimal Observer for Magnetic Nanorobots Navigation. In *MATEC Web of Conferences* (Vol. 108, p. 05002). EDP Sciences.

Guillén, E. O., Garcia, G. I., Vázquez-Montiel, S., Atencio, J. A. D., Ramos, J. C., & Delgado, F. G. (2010). Métodos de clasificación para identificar lesiones en piel a partir de espectros de reflexión difusa. *Revista Ingeniería Biomédica*, *4*(8), 34-39.

Mora, Javier. (2022). Proyecciones de la ciencia de datos en la cirugía cardíaca. *Revista Médica Clínica Las Condes*, *33*(3), 294-306.

Obi Tayo. (16 de enero de 2022). *Matemáticas para principiantes en ciencia de datos: análisis de componentes principales. Medium.* https://benjaminobi.medium.com/math-for-data-science-beginners-principal-component- analysis-a2c0058f33a2.

Reyes, M., Ponce, P., Grammatikou, D., & Molina, A. (2014). Metodología para ponderar áreas de evaluación de la herramienta ADOS-G en trastorno del espectro autista con redes neuronales artificiales y método de Taguchi. *Revista mexicana de ingeniería biomédica*, *35*(3), 223-240.

Vargas, C. (2012). Alan Turing: Máquinas e inteligencia. en conmemoración de los 100 años de su nacimiento. *Revista Aporía*, *4*, 43-63.

Wayne Richards, (2022) *"Importancia Del Álgebra Lineal En La Ciencia De Datos"*, creado el 8 de agosto 2022. Marioh.

Zamora-Vélez, J. M., & Ibargüen-Ocampo, F. J. (2012). Una introducción conceptual a la teoría de contracción. *Revista de Investigaciones Universidad del Quindío*, *23*(1), 48-55.

BIBLIOGRAFÍA

Burgos, J. D. (1997). Álgebra Lineal. Ed.

Corso, C. L. (2009). *Aplicación de algoritmos de clasificación supervisada usando Weka.* Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba.11

Corrales, Martín y Solanilla. (2016). *Estructuras de Álgebra: Tópicos de Álgebra Lineal.* Editora Novo Art. (1), 19-120.

Hoffman, K., Kunze, R., & Finsterbusch, H. E. (1973). *Álgebra lineal*. Prentice-Hall Internacional.

Lay, D. C. (2007). *Álgebra lineal y sus aplicaciones*. Pearson educación.

Ossa Jorge, (2016). *Elementos del álgebra lineal en el aprendizaje de máquina.* [Tesis de Maestría], Universidad Tecnológica De Pereira]. https://repositorio.utp.edu.co/server/api/core/bitstreams/74d1feb5-84f2-40ea-880f-f02751bac8e5/content

Paitán, H. Ñ., Mejía, E. M., Ramírez, E. N., & Paucar, A. V. (2014). *Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis*. Ediciones de la U.

Pereyra, L. E. (Ed.). (2022). *Metodología de la investigación*. Klik.

Sepúlveda, S., & Farfán, E. (2014). El arte de programar en R: un lenguaje para la estadística.