



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA MATEMÁTICA

PROYECCIONES ORTOGONALES Y SUS APLICACIONES

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

MATEMÁTICO

AUTOR: CRISTIAN DAVID BUÑAY MARCATOMA

DIRECTOR: MSc. RAMÓN ANTONIO ABANCIN OSPINA

Riobamba – Ecuador

2023

©2023, Cristian David Buñay Marcatoma

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, CRISTIAN DAVID BUÑAY MARCATOMA, declaro que el presente Trabajo de Integración Curricular es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular. El patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 28 de abril de 2023

A handwritten signature in blue ink, appearing to read 'Cristian David Buñay Marcatoma', with a small star symbol at the end of the signature.

Cristian David Buñay Marcatoma

060491459-8

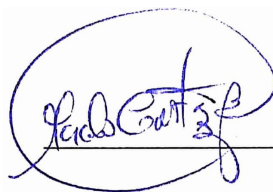
ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA MATEMÁTICA

El Tribunal del Trabajo de Integración Curricular certifica que: el Trabajo de Integración Curricular: Tipo: Proyecto de Investigación, **PROYECCIONES ORTOGONALES Y SUS APLICACIONES**, realizado por el señor: **CRISTIAN DAVID BUÑAY MARCATOMA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

FIRMA

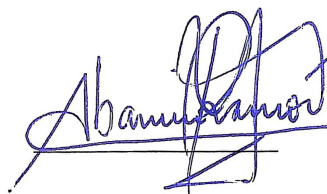
FECHA

Mat. Luis Marcelo Cortez Bonilla
PRESIDENTE DEL TRIBUNAL



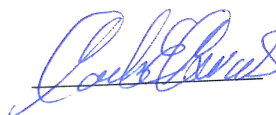
2023-04-28

MSc. Ramón Antonio Abancin Ospina
**DIRECTOR DEL TRABAJO DE
INTEGRACIÓN CURRICULAR**



2023-04-28

MSc. Carlos Eduardo Cova Salaya
**ASESOR DEL TRABAJO DE
INTEGRACIÓN CURRICULAR**



2023-04-28

DEDICATORIA

El resultado de este trabajo de investigación está dedicado a mis padres José Buñay y Teresa Marcatoma quienes me apoyaron en todo momento, con su amor, paciencia, esfuerzo y oración que me han permitido llegar a cumplir hoy un sueño más. Gracias por inculcar en mí el ejemplo de esfuerzo y valentía, por enseñarme a afrontar las dificultades sin perder nunca la fé y la esperanza. A mis hermanos Franklin, Nelly y Jhanneth por su cariño y apoyo incondicional, por estar conmigo durante todo este proceso. Siempre han sido mis mejores guías de vida, por ello les dedico este logro, como una meta más conquistada. Gracias por ser quienes son y por creer en mí.

Cristian

AGRADECIMIENTO

Primeramente quiero expresar mi gratitud al Dios eterno, por ese amor divino y todas las bendiciones derramadas sobre mí y sobre toda mi familia.

Agradezco a mis padres José Buñay y Teresa Marcatoma quienes siempre me han brindado su amor y apoyo incondicional para poder cumplir todos mis objetivos personales y académicos. Estoy infinitamente agradecido porque ellos son los que me han brindado el soporte material y económico para poder concentrarme en los estudios y nunca abandonarlos.

Agradezco a mis hermanos Orlando, Amanda y Jhanneth por apoyarme en todo momento, con sus palabras de ánimo. A mis abuelitos por sus oraciones y a toda mi familia quienes brindaron su apoyo desde un principio.

Le agradezco muy profundamente a mi tutor y a todos los docentes que han sido parte de mi camino universitario. Gracias por su paciencia, por compartir sus conocimientos de manera profesional e invaluable, por su dedicación perseverancia y tolerancia.

Agradezco a mis amigos que se han convertido en mi segunda familia, quienes han sido mis cómplices y hermanos. Agradezco mucho por las horas compartidas, los trabajos realizados en conjunto y las historias vividas de las cuales pude aprender muchas cosas.

Finalmente, quiero expresar mi más grande y sincero agradecimiento a la Dra. Zenaida Castillo PhD., por su esfuerzo, dedicación y paciencia, por sus palabras y correcciones precisas que me ayudaron a lograr y llegar a esta instancia tan anhelada. Muchísimas gracias por su guía y todos sus consejos, donde quiera que vaya, los llevaré conmigo en mi transitar profesional.

Cristian

ÍNDICE DE CONTENIDOS

| | |
|-----------------------------------|----|
| ÍNDICE DE ILUSTRACIONES | ix |
| RESUMEN | x |
| ABSTRACT | xi |
| INTRODUCCIÓN | 1 |

CAPÍTULO I

| | |
|---|---|
| 1. PROBLEMA DE INVESTIGACIÓN | 2 |
| 1.1. Planteamiento del problema | 2 |
| 1.2. Objetivos | 2 |
| 1.2.1. <i>Objetivo general</i> | 2 |
| 1.2.2. <i>Objetivos específicos</i> | 2 |
| 1.3. Justificación | 3 |

CAPÍTULO II

| | |
|---|---|
| 2. MARCO TEÓRICO | 4 |
| 2.1. Antecedentes de la investigación | 4 |
| 2.2. Estado del arte | 5 |
| 2.3. Fundamentación teórica | 7 |

CAPÍTULO III

| | |
|--|----|
| 3. MARCO METODOLÓGICO | 33 |
| 3.1. Descripción del tipo de investigación | 33 |
| 3.2. Descripción del nivel de investigación | 33 |
| 3.3. Descripción del enfoque de la investigación | 34 |

CAPÍTULO IV

| | |
|--|----|
| 4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS | 35 |
| 4.1. Derivación de la técnica de Análisis de Componentes Principales | 35 |
| 4.2. Implementación del algoritmo básico de la técnica del PCA | 50 |

| | | |
|-------------|---|-----------|
| 4.3. | Ejemplo de aplicación de Análisis de Componentes Principales | 53 |
|-------------|---|-----------|

CAPÍTULO V

| | | |
|-------------|---|-----------|
| 5. | CONCLUSIONES Y RECOMENDACIONES | 62 |
| 5.1. | Conclusiones | 62 |
| 5.2. | Recomendaciones | 63 |

BIBLIOGRAFÍA

ÍNDICE DE ILUSTRACIONES

| | |
|--|----|
| Ilustración 1-2: Proyección ortogonal sobre la recta mediante el ángulo entre vectores. . . | 16 |
| Ilustración 2-2: Proyección ortogonal sobre un espacio S de dimensión uno. | 17 |
| Ilustración 3-2: Proyección ortogonal de $v \in \mathbb{R}^3$ sobre el subespacio bidimensional S . . . | 20 |
| Ilustración 4-2: Correlación lineal entre variables. | 31 |
| Ilustración 1-4: Representación del subespacio principal. | 50 |
| Ilustración 2-4: Código básico para la reducción de datos vía PCA. | 54 |
| Ilustración 3-4: Proyección de datos mediante el código básico del PCA. | 55 |
| Ilustración 4-4: Imagen Original de $500 \times 330 \times 3uint8$ | 56 |
| Ilustración 5-4: Reconstrucciones de imágenes con las componentes principales. | 57 |
| Ilustración 6-4: Código básico de reducción de una imagen vía PCA. | 60 |

RESUMEN

El objetivo de esta investigación fue estudiar y describir, desde el punto de vista matemático, la fundamentación teórica que interviene en la reducción de dimensionalidad de datos, mediante la técnica de Análisis de Componentes Principales o PCA (por sus siglas en inglés), para lo cual se revisó detalladamente la bibliografía especializada del tópico con el fin de entender los elementos de la matemática que intervienen en la deducción de esta técnica. La derivación y descripción matemática del Análisis de Componentes Principales se llevó a cabo mediante la descripción en forma detallada del tópico de proyecciones ortogonales de vectores en subespacios de menor dimensión, para obtener una representación o aproximación de los datos que resulte más manejable y sencilla de analizar computacionalmente. El objetivo se logró mediante la indagación exhaustiva y sistemática de la bibliografía existente sobre la técnica de PCA para la obtención del subespacio principal. En esta investigación, el subespacio principal buscado se obtiene minimizando el error cuadrático de reconstrucción promedio, en la norma euclídea, el cual nos conduce al cálculo de los principales autovalores y autovectores de la matriz de covarianzas, los cuales representarán las coordenadas y los vectores óptimos del subespacio de proyección. Se concluye que los elementos de la matemática, en este caso del Álgebra Lineal, como los espacios vectoriales, productos internos, ortogonalidad, proyecciones, autovalores y autovectores, son base fundamental en la descripción del PCA. Adicionalmente, la comprensión de estos conceptos nos permitió elaborar un algoritmo básico para la reducción de dimensionalidad de datos. En el futuro podría realizarse esta derivación desde otra perspectiva, o mediante el uso de una norma distinta a la norma euclídea.

Palabras clave: <MATEMÁTICA>, <ESPACIOS VECTORIALES>, <PRODUCTOS INTERNOS>, <PROYECCIONES ORTOGONALES>, <AUTOVALORES>, <AUTOVECTORES>, <ANÁLISIS DE COMPONENTES PRINCIPALES>, <REDUCCIÓN DE DIMENSIONALIDAD>.

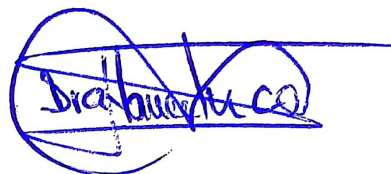
0851-DBRA-UPT-2023



ABSTRACT

The aim of this research was to study and describe from a mathematical point of view, the theoretical basis involved in the reduction of data dimensionality, by means of the Principal Component Analysis or PCA technique; for which the specialized bibliography of the topic was reviewed in detail; in order to understand the elements of mathematics involved in the deduction of this technique. The derivation and mathematical description of the Principal Component Analysis was carried out by means of a detailed description of the topic of orthogonal projections of vectors in subspaces of smaller dimension, to obtain a presentation or approximation of the data that is more manageable and computationally simple to analyze. The aim was achieved by means of the exhaustive and systemic investigation of the existing bibliography on the technique of PCA the obtaining of the principal space. In this research, the principal subspace sought is obtained by minimizing the mean square error of reconstruction, in the Euclidean norm, which leads to the calculation of the principal eigenvalues and eigenvectors of the covariance matrix, which will represent the chordal and optimal vectors of the projection subspace. It is concluded that the elements of mathematics, in this case Linear Algebra, such as vector spaces, inner products, orthogonality, projections, eigenvalues, eigenvectors, are fundamental basis in the description of the PCA. Additionally, the understanding of these notions allowed us to elaborate a basic algorithm for data dimensionality reduction. In the future, this derivation could be performed from another perspective, or by using a norm other than the Euclidean norm.

Keywords: <MATHEMATICS>, <VECTOR SPACES>, <INTERNAL PRODUCTS>, <ORTHOGONAL PROJECTIONS>, <AUTOVALUES>, <AUTOVECTORS>, <PRINCIPAL COMPONENT ANALYSIS>, <DIMENSIONALITY REDUCTION>.



Dra. Nanci M. Inca Ch. Mgs.

0602926719

INTRODUCCIÓN

En los últimos años hemos visto un crecimiento extraordinario en tecnología; y la analítica de datos ha surgido como especialidad de punta para este avance. El procesamiento y control de datos con grandes dimensiones es un reto en la actualidad, y los algoritmos que se emplean deben ser adaptados para hacer eficiente este proceso con las herramientas tecnológicas actuales. El objetivo es extraer información importante de un conjunto de datos que poseen muchas características, ya sean tablas con muchas columnas y/o filas, o imágenes con millones de píxeles, por lo que resulta conveniente disminuir la dimensionalidad de los mismos, que no es más que la reducción del número de variables explicativas. Esto se logra, proyectando la data en un espacio de menor dimensión, y lo ideal es conseguir una proyección que resguarde la información importante de los datos como un todo. En ese sentido, las proyecciones ortogonales han sido de mucha ayuda en problemas como la regresión lineal, o en el procesamiento de imágenes, que pueden modelarse como problemas de mínimos cuadrados, y resolverse eficientemente aprovechando las bondades de la tecnología actual.

La reducción de dimensionalidad se requiere en diferentes aplicaciones, tales como, el reconocimiento de voz, visualización de datos, reducción de ruido o el procesamiento de señales. También se aplica en la predicción del cambio climático, en donde se cuenta con un conjunto de datos; producto de la información recopilada de diferentes fuentes como: los datos de temperatura, humedad, la velocidad del viento y cantidad de lluvia en distintas épocas del año, en un mismo lugar.

Entonces, en este trabajo de investigación se propone realizar un estudio descriptivo sobre las proyecciones ortogonales, su base matemática, y la derivación de la técnica de Análisis de Componentes Principales o PCA (por sus siglas en inglés), para la reducción de dimensionalidad.

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

A continuación se describe el planteamiento del problema de esta investigación, ya que es un punto de partida que da un enfoque a este trabajo.

1.1. Planteamiento del problema

Se propone el análisis, descripción y deducción matemática de la técnica de Análisis de Componentes Principales o PCA (por sus siglas en inglés) usada ampliamente para la reducción de dimensionalidad datos en muchas aplicaciones.

1.2. Objetivos

A continuación se describen los objetivos tanto generales como específicos que se plantean en esta investigación para un buen cumplimiento de la misma.

1.2.1. Objetivo general

Estudiar y describir, desde el punto de vista matemático, las proyecciones ortogonales, y su uso en la reducción de dimensionalidad, según la técnica de Análisis de Componentes Principales (PCA por sus siglas en inglés). Esto se hará a través de la lectura y comprensión de la bibliografía básica y especializada del tópico, y la descripción detallada de esta poderosa herramienta matemática, a fin de que pueda ser entendida por aquellos que se inicien en este tópico del Álgebra Lineal.

1.2.2. Objetivos específicos

1. Seleccionar y organizar la bibliografía especializada del tópico.
2. Entender los fundamentos teóricos, y elementos de la matemática que intervienen en la deducción de la técnica análisis de componentes principales.
3. Investigar los conceptos de Ortogonalidad, Proyecciones Ortogonales, Matriz de covarianza, Autovalores y Autovectores, y la Proyección en espacios de menor dimensión.
4. Derivar y describir matemáticamente la técnica de PCA en la reducción de dimensionalidad.

1.3. Justificación

Se plantea la necesidad de realizar una investigación que genere un documento referencial en el estudio y la descripción del Análisis de Componentes Principales, técnica utilizada ampliamente en la actualidad, para la reducción de dimensionalidad de datos. Algunos conceptos matemáticos, son componentes básicos e importantes en la solución de problemas reales, pero muchas veces pasan desapercibidos, porque no se describe su uso en las distintas aplicaciones, por lo que se describe la técnica de Análisis de Componentes Principales usada ampliamente para el manejo y comprensión de grandes volúmenes de datos. Este tópico es tratado en forma muy general en la bibliografía, y entendemos que a estudiantes de matemática, y a todo aquel que sienta interés por entender esta técnica, desde sus bases, les será de mucha ayuda contar con un documento referencial sobre este tópico.

Este documento servirá de guía a estudiantes de matemática, y a cualquiera que se inicie en este tópico, y le ayudará a consolidar algunos conceptos importantes del Álgebra Lineal, enfatizando su utilidad en aplicaciones reales. También les servirá de motivación hacia otros trabajos similares en los cuales se desentrañe el contenido matemático detrás de otras técnicas reconocidas.

CAPÍTULO II

2. MARCO TEÓRICO

En este capítulo se presentarán los conceptos en los cuales se soporta la investigación; así como también las herramientas matemáticas que permitirán deducir la técnica de análisis de componentes principales.

2.1. Antecedentes de la investigación

El Análisis de Componentes Principales, conocido ampliamente como PCA, es una técnica o método que se utiliza para la extracción, reducción o simplificación de un conjunto de datos, y es ampliamente utilizada en la actualidad en distintas aplicaciones. Fue desarrollada por Pearson (1901) y publicada inicialmente en *Philosophical Magazine*, la cual tuvo como objetivo minimizar la pérdida de información al momento de reducir los datos. En un inicio estaba formulada en un entorno de análisis de regresión lineal, cuyo objetivo fue encontrar la línea o recta que mejor se ajuste a los datos en el plano de mejor representación, dichos ajustes son proyecciones ortogonales que se obtiene mediante la solución del problema de los mínimos cuadrados.

Posteriormente, fueron estudiados por Hotelling (1933), el cual fue publicado en la revista *Journal of Educational Psychology*. Hotelling fue el primero en formular la técnica del análisis de componentes, comenzó con el estudio y la interpretación de las relaciones que existían entre los valores propios o conocido como autovalores de la matriz de covarianza y las cargas de los componentes principales que se obtuvieron gracias a los multiplicadores de Lagrange. Además, mostró las relaciones que se dan entre p variables correlacionadas, es decir, variables que tienen información en común, transformando el conjunto de datos originales en otro conjunto de nuevas variables que no están correlacionadas entre sí, es decir, no hay duplicidad ni redundancia en la información que presenta los datos, este nuevo conjunto de datos que tiene nuevas variables se denomina conjunto de componentes principales. Además Hotelling se centró principalmente en el análisis de componentes que reducen la variabilidad de un sistema de puntos tanto como sea posible (González, N. & Taborda, A., 2015).

Algunos autores como Bro y Smilde (2014) mencionaron la diferencia conceptual del método de Pearson y Hotelling. En el método de Hotelling, las componentes principales están definidas por nuevos ejes. A medida que van en cierta dirección, la cantidad de varianza explicada por

los componentes está en orden descendente; esto significa que el primer componente es el responsable de explicar la mayor parte de la variabilidad, es decir, abarca la mayor parte de los datos, el segundo componente es más variable después del primero, y así sucesivamente (esto se denomina propiedad del eje principal) (González, N. & Taborda, A., 2015). Por otro lado, el método de Pearson define un subespacio proyectado que es más importante que los propios ejes, ya que los ejes solo se especifican como base del subespacio (Restrepo, et al., 2012).

Sin embargo, fue solo con la llegada de las computadoras que esta idea del análisis de componentes principales comenzó a popularizarse. Así, el uso de esta técnica está relacionado con la evolución de las computadoras, Jolliffe (2002) señaló que dado que el algoritmo de solución de análisis de componentes principales requiere mucho poder de cómputo cuando la matriz de datos es grande, principalmente con una gran cantidad de variables, su uso se dio con el aumento de las computadoras, que difundieron esta técnica de reducción de dimensionalidad.

Las primeras aplicaciones desarrolladas mediante la técnica de PCA fueron expuestas, en la década de 1930, en revistas de psicología, tratando de reivindicar la misma importancia de la medida cuantitativa de la inteligencia y personalidad humana, como del interés contemporáneo por los genes, el procesamiento de señales, el análisis de grandes datos meteorológicos, financieros o generados por Internet a gran escala (Trendafilov, 2014).

El Análisis de Componentes Principales es una de las técnicas más conocidas y efectivas para reducir la dimensión de una matriz de datos de alta dimensión la cual es utilizada extensamente en el procesamiento de datos (Jolliffe, 2002). Es una técnica que proyecta observaciones desde un espacio n -dimensional con n variables sobre un espacio q -dimensional donde $q < n$ para abarcar la mayor cantidad de información posible. Según Ligdi González (2020), el Análisis de Componentes Principales es un procedimiento estadístico que transforma ortogonalmente las n dimensiones originales de un conjunto X de datos en un nuevo conjunto de q dimensiones más pequeñas llamadas componentes principales. Esta técnica es ampliamente utilizado en bioestadística, marketing, sociología, informática y muchos otros campos.

2.2. Estado del arte

Durante muchos años se han venido trabajando y mejorando la técnica de Análisis de Componentes Principales para la reducción de dimensionalidad de datos. En 1901, Pearson publicó un artículo sobre el ajuste de sistemas de puntos a líneas planas en múltiples espacios

(Pla, 1986) y Hotelling adoptó el método de Pearson en 1933 y fue el primero en formular el Análisis de Componentes Principales (Del Ángel, 2004). La investigación de Pearson se centró en componentes o combinaciones lineales de las variables originales que generan un plano para el cual el sistema de puntos ajustados debe tener una suma mínima de distancias desde cada punto hasta el plano ajustado (Restrepo, et al., 2012). El objetivo del Análisis de Componentes Principales es reducir grandes cantidades de datos mediante unos cuantos parámetros (Dallas, 2000).

Otro de los aportes mediante el uso de esta técnica de Análisis de Componentes Principales realizada por Pearson se basó en el método de análisis factorial de componentes principales y esto se realizó mediante el análisis de las correlaciones entre variables e identificando estructuras de datos básicas, este método puede extraer factores abstractos que revelan la mayor parte de la información de los datos originales (Ochoa, et al., 2017) y (Solano, et al., 2009), pero en ese entonces no fueron tan reconocidos hasta la aparición de los ordenadores donde estos resultados fueron muy útiles para la reducción de dimensionalidad de datos altamente correlacionados. Según (Almenara, et al., 1998) el Análisis de Componentes Principales es una técnica utilizada para la reducción de dimensionalidad, la cual transforma un conjunto de variables correlacionadas que expresa el conjunto original de variables, en un nuevo conjunto de variables no correlacionadas, este nuevo conjunto no presenta redundancia en la información.

La propiedad no correlacionada de los componentes se destaca por el hecho de que son ortogonales, es decir, en ángulo recto entre sí, lo que significa que los índices miden diferentes dimensiones en los datos (Fujita, M. & Krugman, P., 2004).

Una de las aplicaciones del Análisis de Componentes Principales en el proceso de purificación de un biofármaco realizada por (Gozá, et al., 2020) el cual se basó en el estudio a datos registrados durante dos (2) años en la fase de purificación de una planta de producción de eritropoyetina recombinante humana basada en múltiples pasos cromatográficos similares al proceso de purificación para uso profiláctico o terapéutico de proteínas recombinantes para vacunas.

Otra de las aplicaciones muy recientes en las Ciencias Técnicas Agropecuarias desarrolladas por (Lucía, et al., 2022) fue acerca de la aplicación de la técnicas multivariadas, con respecto a las bases metodológicas y énfasis en la interpretación de los resultados, además eleva la calidad de las investigaciones científicas en procesos agrarios y afines. En este estudio se desarrolló un ejemplo relacionado con estudios de poscosecha de piñas (variedad Cayena Lisa) donde se incluyeron en el estudio las variables de pérdida de peso en gramos (PP), dureza, índice de color (IC), contenido de sólidos solubles (SSC) y PH, obteniendo como resultado una correlación positiva entre PP, SSC

y PH, mientras que una correlación negativa entre dureza y estas variables fueron las que indican que las mayores PP y PH que se lograron a partir del sexto día y la mayor dureza se logró en los primeros dos días, siguientes.

El Análisis de Componentes Principales se utiliza para identificar patrones de comportamiento en ecosistemas de gran importancia en la agricultura, para explicar deficiencias y alimentación en relación con la morfología animal, factores de producción animal, análisis de pastos y varias aplicaciones como factores ambientales, rebaños, evaluación de la calidad del agua en acuicultura, etc. (Restrepo, et al., 2012). Esta técnica del PCA se viene usando desde hace muchos años; sin embargo, en la actualidad hemos estado frecuentemente en el estudio de la misma, descubriendo y tratando aplicaciones con esta técnica del Análisis de Componentes Principales para la reducción de dimensionalidad de datos.

2.3. Fundamentación teórica

Dentro del Análisis de Componentes Principales (PCA), el Álgebra Lineal es de mucha importancia, ya que manejan muchos conceptos importantes como: espacios vectoriales, bases, transformaciones lineales, autovalores y autovectores, ortogonalidad, proyecciones, matriz de covarianza, entre otros, las cuales son fundamentales para resolver problemas de distinta índole, desde problemas básicos que surgen en la vida real hasta problemas complejos que promueven el avance de la matemática, y de otras ciencias. Tal es el caso del concepto de proyecciones ortogonales, cuyo aporte a la solución de varios problemas teóricos y aplicados es bien reconocido. A continuación revisaremos algunos de estos conceptos o elementos de la matemática que soportan la teoría detrás de la técnica de Análisis de Componentes Principales.

Recordemos que un **espacio vectorial** es una estructura matemática que consta de un conjunto no vacío V de elementos u objetos llamados vectores, en el que están definidas dos operaciones: la suma de vectores y la multiplicación por un escalar, las cuales están sujetas a los axiomas que se dan a continuación:

Suma de vectores

1. Para todo $u, v \in V : u + v \in V$ (Cerradura bajo la suma).
2. Para todo $u, v \in V : u + v = v + u$ (Conmutatividad).
3. Para todo $u, v, w \in V : u + (v + w) = (u + v) + w$ (Asociatividad).
4. Existe el vector e en V tal que para todo $u \in V : u + e = e + u = u$ (Existencia de elemento

neutro).

5. Para todo $u \in V$ existe $u' \in V$ tal que $u + u' = u' + u = e$ (Existencia de elemento inverso).

Multiplicación de un vector por un escalar

1. Si $u \in V$ y $\alpha \in K$ (campo de escalares): $\alpha v \in V$ (Cerradura bajo la multiplicación por un escalar).

2. Para todo $u, v \in V$ y $\alpha \in K$ un escalar: $\alpha(u + v) = \alpha u + \alpha v$.

3. Para todo $u \in V$ y para todo par de escalares α y β en K : $(\alpha + \beta)u = \alpha u + \beta u$.

4. Para todo $u \in V$ y para todo par de escalares α y β en K : $\alpha(\beta u) = (\alpha\beta)u$.

5. Para cada $u \in V$ se tiene que $1 \cdot u = u$ (1 es el elemento neutro en K).

Con frecuencia, estamos interesados en un conjunto particular S de elementos pertenecientes a un espacio vectorial V , y si este subconjunto S , con las operaciones de suma y multiplicación por un escalar, satisface los diez (10) axiomas correspondientes, entonces sería también un espacio vectorial. En este caso, decimos que la estructura $(S, +, \cdot, \alpha)$ mantiene la estructura del espacio vectorial, $(V, +, \cdot, \alpha)$ por lo que S es un subespacio vectorial de V .

Ahora, consideremos un conjunto de vectores $B = \{v_1, v_2, \dots, v_q\}$ cualesquiera de un espacio vectorial V , estos elementos pueden sumarse entre vectores y multiplicarlos por escalares, de esto viene la idea de una combinación lineal. Se dice que el vector x es combinación lineal de los vectores v_1, v_2, \dots, v_q si y solo si existen escalares $\alpha_1, \alpha_2, \dots, \alpha_q$ tal que:

$$x = \sum_{i=1}^q \alpha_i v_i = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_q v_q. \quad (2.1)$$

La idea es ver como es cada vector con respecto a los demás, si son proporcionales o presentan distintas direcciones, ya que esto permitirá más adelante verificar variables o datos que comparten información similar, por lo tanto, es importante verificar la dependencia e independencia lineal de estos vectores.

Entonces, decimos que los vectores del conjunto B son **linealmente dependientes** si existe una combinación lineal de estos vectores que es igual al vector nulo, con escalares $\alpha_1, \alpha_2, \dots, \alpha_q$, no todos iguales a cero, la cual está denotada como:

$$0 = \sum_i^q \alpha_i v_i = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_q v_q \quad \text{para } \alpha_i \neq 0. \quad (2.2)$$

Pero si la única combinación lineal que existe es la trivial, es decir, cuando todos los escalares $\alpha_1, \alpha_2, \dots, \alpha_q$ de esta combinación son iguales a cero, entonces estos vectores son **linealmente**

independientes. Ahora, si consideramos un espacio vectorial V con q vectores v_1, v_2, \dots, v_q linealmente independientes y m combinaciones lineales, entonces cada x_j de espacio vectorial V donde $j = 1, 2, \dots, m$ es una combinación lineal de la siguiente forma:

$$x_j = \sum_{i=1}^q \alpha_{ij} v_i \quad (2.3)$$

Estos q vectores linealmente independientes los podemos representar en términos matriciales, entonces definimos $B = [v_1, v_2, \dots, v_q]$ como la matriz cuyas columnas son los q vectores linealmente independientes, entonces la combinación lineal para x_j se puede escribir como:

$$x_j = B\alpha_j \quad \text{donde } \alpha_j = \begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{qj} \end{bmatrix} \text{ para } j = 1, 2, \dots, m. \quad (2.4)$$

Ahora, si todos los vectores de un espacio vectorial V se pueden expresar como combinación lineal de los vectores de $B_q = \{v_1, v_2, \dots, v_q\}$, siendo éstos linealmente independientes, entonces este conjunto B_q es una **base** de este espacio vectorial.

La cantidad de vectores de una base B_q determinan la **dimensión** del espacio vectorial. Por ejemplo, si $B_{15} = \{v_1, v_2, \dots, v_{15}\}$ es una base del espacio vectorial V , entonces la dimensión de este espacio vectorial es 15, y está denotada como $\text{Dim}(V) = 15$. Por otra parte, un espacio es de dimensión infinita si no existe una base de V formada por un conjunto finito de vectores. Notemos que una base es un conjunto generador mínimo y un conjunto máximo de vectores linealmente independientes, además todo vector x de V es una combinación lineal única de vectores de la base B_q , es decir:

$$x = \sum_{i=1}^q \alpha_i v_i = \sum_{i=1}^q \lambda_i v_i \quad \text{para } i = 1, 2, \dots, q. \quad (2.5)$$

donde los escalares α_i y λ_i resultan ser iguales $\alpha_i = \lambda_i$ con $v_i \in B_q$. Un ejemplo de bases para el espacio vectorial de \mathbb{R}^3 es:

$$B_c = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \quad B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 7 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} \right\}$$

La primera, es conocida como la base canónica de \mathbb{R}^3 y la segunda es una base formada por tres vectores linealmente independientes.

Otro aspecto importante de los espacios vectoriales es el concepto de transformaciones lineales, las cuales tienen aplicaciones importantes en el mundo de la ciencia de datos. Por ejemplo, en las

imágenes, cada vector representa un píxel de dicha imagen y mediante la transformación lineal se puede agrandar o ampliar y girar los píxeles, uno por uno, hasta obtener la imagen que deseamos, y esto se logra mediante esta función conocida como transformaciones o aplicaciones lineales.

Una **transformación lineal** es una función entre dos espacios vectoriales V y W sobre un mismo campo de escalares K , la cual cumple con las siguientes condiciones:

1. $T(u + v) = T(u) + T(v) \quad \forall u, v \in V$
2. $T(\alpha u) = \alpha T(u) \quad \forall u \in V, \forall \alpha \in \mathbb{K}$

Las dos condiciones que definen una transformación lineal, pueden expresarse como una sola, esto es:

$$T(\alpha u + \lambda v) = \alpha T(u) + \lambda T(v) \quad (2.6)$$

Toda transformación lineal tiene una matriz asociada; cuya acción en los elementos del espacio es equivalente a la aplicación de la transformación lineal. Dada una transformación lineal $T : V \rightarrow W$, si la dimensión de V es n , y la dimensión de W es m , entonces la matriz B asociada a esta transformación será de dimensión $m \times n$.

$$\begin{aligned} T : V &\longrightarrow W \\ v &\longmapsto T(v) = Bv \end{aligned} \quad (2.7)$$

Dentro de las características más útiles de los elementos de un espacio vectorial (vectores) encontramos la longitud, la cual puede obtenerse mediante funciones que conocemos como normas, las normas de vectores. A continuación mencionamos la definición de norma y algunos ejemplos de tales funciones.

Una **norma de un espacio vectorial** V no es más que una función $\|\cdot\|$ que asigna a cada vector v su longitud $\|v\|$ de \mathbb{R}^+ , y es definida de la siguiente manera:

$$\begin{aligned} \|\cdot\| : V &\longrightarrow \mathbb{R}^+ \\ v &\longmapsto \|v\| \end{aligned} \quad (2.8)$$

tal que para todo $\alpha \in \mathbb{K}$ y para todo $x, y \in V$ se verifica las siguientes propiedades:

1. $\|v\| \geq 0$ y $\|v\| = 0$ si y solo si $v = 0$.
2. $\|\alpha v\| = |\alpha| \|v\|$
3. $\|u + v\| \leq \|u\| + \|v\|$ Desigualdad triangular

La definición de la norma en (2.8) está dada para cualquier espacio vectorial V sobre un campo \mathbb{K} , en este estudio nos basaremos en normas sobre el campo de los reales. Veamos algunos ejemplos de norma que son mayormente usadas para las medición de distancia entre vectores:

1. **Norma ℓ_1 :** La norma ℓ_1 también conocida como norma de Manhattan en \mathbb{R}^n se define como:

$$\|v\|_1 := \sum_{i=1}^n |v_i| \quad \text{para todo } v \in \mathbb{R}^n, \quad (2.9)$$

donde $|\cdot|$ es el valor absoluto.

2. **Norma ℓ_2 :** La norma ℓ_2 mayormente conocida como la norma euclidiana de $v \in \mathbb{R}^n$ está definida como:

$$\|v\|_2 := \sqrt{\sum_{i=1}^n v_i^2} = \sqrt{v^T v}. \quad (2.10)$$

esta norma calcula la distancia euclidiana de v desde el origen.

3. **Norma ℓ_∞ :** La norma ℓ_∞ conocida como la norma infinita o norma del máximo se define como:

$$\|v\|_\infty = \sup_{1 \leq i \leq n} |v_i| \quad \text{para } v \in \mathbb{R}^n \quad (2.11)$$

Son muchas las aplicaciones que requieren mediciones de distancia entre vectores; y en este caso hablamos de la norma del vector diferencia, usando cualquier norma, o las que nos resulte más apropiada.

El concepto de ángulo de vectores y las proyecciones que se darán más adelante, se explicarán basado en la norma ℓ_2 (2.10) comúnmente conocida como la norma euclídea.

Otro de los conceptos importantes que está estrechamente relacionado con la norma es el producto interno, ya que introduce conceptos geométricos importantes e intuitivos, como la longitud de un vector y el ángulo o la distancia entre dos vectores del espacio vectorial V . Una de las razones importantes de los productos internos es determinar si los vectores son ortogonales entre sí.

Recuerde que el **producto interno** sobre V es una función $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$ que se asocia a cada par ordenado de vectores u, v en V un escalar $\langle u, v \rangle$ en \mathbb{K} tal que para todos los vectores u, v, w en V y todos los escalares α se satisface las siguientes condiciones:

1. $\langle u, v \rangle = \langle v, u \rangle$.
2. $\langle (u + v), w \rangle = \langle u, w \rangle + \langle v, w \rangle$.
3. $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$.
4. $\langle u, u \rangle > 0$ y $\langle u, u \rangle = 0$ si y sólo si $u = 0$.

En la primera condición (1) hay que tener en cuenta el campo donde se trabaja, ya que en el campo de los complejos la primera condición queda expresada de la siguiente forma: $\langle u, v \rangle = \overline{\langle v, u \rangle}$, donde la barra significa la conjugación compleja.

En algunos libros suelen llamar al producto interno como producto escalar $\langle u, v \rangle = u^T v$, pero hay que tener en cuenta que cuando se dice producto interno se refiere a un caso general, es decir, podemos definir un producto interno en cualquier espacio vectorial, ya sea en \mathbb{R}^{100} , en espacio de polinomios, en espacios de matrices, o en otros espacios vectoriales por ejemplo, en el espacio de los polinomios de grado menor e igual a n se puede definir el producto interno $\langle \cdot, \cdot \rangle : \mathbb{P}_n \times \mathbb{P}_n \rightarrow \mathbb{R}$ de la siguiente manera:

$$\langle P, Q \rangle = a_0 b_0 + a_1 b_1 + \cdots + a_n b_n. \quad (2.12)$$

Mientras que, el producto escalar es un caso especial de un producto interno que funciona en el espacio \mathbb{R}^n , en este estudio asumimos que nuestro producto interno es el **producto escalar**, el cual se define como:

$$u^T v = \sum_{i=1}^n u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n, \quad (2.13)$$

donde u, v son elementos de \mathbb{R}^n . Nótese que si $u = v$, y tomamos la raíz cuadrada de $u^T u$ el resultado es la norma ℓ_2 que es la norma euclídea del vector u ; en este caso decimos que el producto interno induce a la norma ℓ_2 .

$$\|u\| = \sqrt{\langle u, u \rangle} = \sqrt{u^T u}. \quad (2.14)$$

En algunos casos, tanto la norma como la distancia entre vectores tienen una interpretación geométrica, por ejemplo en \mathbb{R}^3 la norma de un vector se interpreta como una longitud de ese vector y la distancia como el espacio que separa esos vectores. Pero ¿qué pasa si los elementos del espacio vectorial V son matrices? Es aquí donde se pierde la interpretación geométrica; sin embargo, la utilidad de la medida sigue vigente.

El valor de la norma de un vector v depende del producto interno que se utiliza, si consideramos dos vectores genéricos u, v en \mathbb{R}^2 donde $u = (u_1, u_2)$ y $v = (v_1, v_2)$ podemos definir el producto interno $\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ de la siguiente manera:

$$\langle u, v \rangle = u_1 v_1 - u_1 v_2 - u_2 v_1 + 2u_2 v_2. \quad (2.15)$$

Si $u = (1, 2)$ entonces, la norma de este vector con el producto interno así definido es $\|u\| = \sqrt{\langle u, u \rangle} = \sqrt{5}$, mientras que, si utilizamos como producto interno al producto escalar definido en (2.13) de este mismo vector, entonces obtendremos el siguiente resultado $\|u\| = \sqrt{u^T u} = \sqrt{5}$. Como podemos notar que los valores de la norma de vectores depende del producto interno que se utilice.

Como se expresó anteriormente, la norma está asociada a la distancia entre vectores, si usamos el producto escalar $u^T v$ en \mathbb{R} se obtendrá la distancia euclídea comúnmente conocida.

$$d(u, v) = \|u - v\| = \sqrt{\langle u - v, u - v \rangle} \quad (2.16)$$

Los productos internos además de relacionarse con conceptos de longitud del vector y la distancia entre vectores también está relacionada con otra de las medidas que es el ángulo entre vectores, la cual está estrechamente relacionada con la norma, entonces el ángulo entre u y v se define como el único número real θ tal que cumpla lo siguiente:

$$\cos \theta = \frac{\langle u, v \rangle}{\|u\| \|v\|}. \quad (2.17)$$

Notemos que si alguno de los dos vectores es nulo, no podemos dividir por su norma, por tanto, el ángulo no estaría definido. Ahora, mediante esta definición podemos determinar no solo el ángulo formado entre vectores, sino que también nos permite determinar la orientación de cada una de ellas.

Ortogonalidad entre vectores

La ortogonalidad es una generalización de la noción de perpendicularidad entre dos vectores u, v no nulos de un espacio vectorial V , entonces estos vectores u, v son **ortogonales** si su producto interno es cero, es decir, $\langle u, v \rangle = 0$. Notemos que la ortogonalidad también depende del producto interno que se use, por ejemplo, con el producto escalar, los vectores $u = (3, 1)$ y $v = (-1, 3)$ son ortogonales, mientras que con el producto interno definido en (2.15) estos vectores no son ortogonales, ya que su producto interno es distinto de cero. El único vector que es ortogonal a sí mismo es el vector nulo, esto es consecuencia de la propiedad 4 que define el producto interno.

Ahora, nos centraremos en un caso especial de bases que son: las bases ortogonales y ortonormales, esto ayudará más adelante en la representación de la proyección de forma reducida. Para obtener una base ortogonal B_k , en principio los vectores del espacio vectorial V tienen que ser linealmente independientes y mediante el método de Gram-Schmidt que se encuentran en (Grossman, 2012) y (Lay, 2007), podemos obtener una **base ortogonal** donde los vectores de esta base ortogonal serán ortogonales dos a dos, es decir, $\langle v_i, v_j \rangle = 0$ con $i \neq j$.

Ahora, para normalizarlo sencillamente los vectores deberían ser ortogonales, pero eso no quiere decir que el proceso de normalización se aplique solo cuando se tiene vectores ortogonales. Además, una vez que obtengamos la base ortogonal, podemos normalizar estos vectores obteniendo como resultado una nueva base llamada **base ortonormal**, la cual se define de la siguiente manera:

$$\langle v_i, v_j \rangle = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases} \quad (2.18)$$

Por ejemplo, la base canónica en \mathbb{R}^n es una base ortogonal y cada vector tiene norma uno, por lo que también es base ortonormal. Si consideramos el conjunto $B = \{(2, 3), (6, -4)\}$ en \mathbb{R}^2 este conjunto es base ortogonal.

Una vez comprendido la definición de ortogonalidad entre vectores, lo que nos interesa ahora es sobre los vectores que son ortogonales entre sí es decir, vectores que pertenecen al complemento ortogonal, ya que será de mucha ayuda más adelante.

Recuerde que un **complemento ortogonal** de un subconjunto S cualquiera, es el conjunto S^\perp definido de la siguiente manera:

$$S^\perp = \{v \in V : \langle v, s \rangle = 0, \forall s \in S\}. \quad (2.19)$$

El subconjunto S de V no necesariamente es un subespacio, pero el complemento ortogonal S^\perp si forma un subespacio vectorial de V . Esta definición nos dice que si un vector v es ortogonal a todo vector s de un subconjunto S de V , entonces v es ortogonal a V y el conjunto de todos los vectores v que son ortogonales a S se llama complemento ortogonal de S .

Mediante un subespacio y un complemento ortogonal se puede determinar la descomposición ortogonal del espacio vectorial total.

$$V = S \oplus S^\perp. \quad (2.20)$$

Es decir, cada $v \in V$ la podemos escribir de forma única como suma de vectores.

$$v = \text{Proy}_S(v) + u \quad \text{con} \quad \text{Proy}_S(v) \in S \text{ y } u \in S^\perp. \quad (2.21)$$

Esto implica que el vector $v - \text{Proy}_S(v)$ es un elemento del complemento ortogonal de S , es decir, $v - \text{Proy}_S(v) \in S^\perp$.

Proyecciones Ortogonales

Una de las formas más utilizadas, y eficientes para la reducción de dimensionalidad de datos, está basada en proyecciones ortogonales, ya que se desea proyectar grandes dimensiones de datos en un subespacio de menor dimensión. Estas proyecciones son casos especiales de transformaciones lineales y juegan un papel importante en el proceso de reducción de dimensionalidad mediante esta técnica del PCA, ya que cuando comprimimos o reducimos las grandes dimensiones de datos, perderemos información, y para reducir la pérdida de información, lo ideal es encontrar las dimensiones más representativas en los datos.

Definimos entonces la proyección ortogonal de un vector $v \in \mathbb{R}^n$ sobre un subespacio $S \subseteq \mathbb{R}^n$ con $\dim(S) < n$ como: el vector $Proy_S(v)$ tal que $\langle v, Proj_S(v) \rangle = 0$.

Sean u y v vectores no nulos del espacio vectorial V . Entonces la proyección de v sobre u es un vector definido de la siguiente manera:

$$Proy_u v = \frac{u^T v}{\|u\|^2} u. \quad (2.22)$$

Notemos que $\frac{u^T v}{\|u\|}$ es un escalar y representa la componente de v en la dirección del vector u , y además $\frac{u}{\|u\|}$ es un vector unitario en la dirección de u . Si u y v son vectores ortogonales, entonces el producto interno $u^T v = \langle u, v \rangle = 0$, por lo que la proyección de v sobre u es el punto origen de v , es decir, $Proy_u v = 0$. La proyección en sí es una transformación lineal, para verificar esto debemos ver si se cumple con las dos propiedades que define una transformación lineal: Sea v, w dos vectores cualesquiera de \mathbb{R}^n a las cuales se desea proyectar sobre un subespacio S de dimensión menor. Entonces

$$1. \quad Proj_S(v + w) = Proj_S(v) + Proj_S(w).$$

Prueba:

$$\begin{aligned} Proj_S(v + w) &= \frac{u^T (v + w)}{\|u\|^2} u = \frac{u^T v + u^T w}{\|u\|^2} u = \left(\frac{u^T v}{\|u\|^2} + \frac{u^T w}{\|u\|^2} \right) u. \\ &= \frac{u^T v}{\|u\|^2} u + \frac{u^T w}{\|u\|^2} u = Proj_S(v) + Proj_S(w). \end{aligned}$$

$$2. \quad Proj_S(\alpha v) = \alpha Proj_S(v).$$

Prueba:

$$Proj_S(\alpha v) = \frac{u^T (\alpha v)}{\|u\|^2} u = \frac{(u^T \alpha v)}{\|u\|^2} u = \frac{\alpha (u^T v)}{\|u\|^2} u = \alpha Proj_S(v).$$

Hemos comprobado que la proyección, como función entre espacios vectoriales, es una transformación lineal, y por ser una transformación lineal tiene una matriz asociada a esta transformación, a la cual llamaremos P . La proyección de v sobre u entonces puede escribirse como:

$$Proy_u v = \frac{u^T v}{\|u\|^2} u = Pv. \quad (2.23)$$

Considerando la definición del ángulo entre vectores, también podemos expresar la proyección de v sobre u como:

$$Proy_u(v) = \frac{u}{\|u\|} \|v\| \cos \theta, \quad (2.24)$$

donde $\cos(\theta)$ es el coseno entre los vectores u y v . Además, mediante el concepto del complemento ortogonal también se puede definir la proyección ortogonal sobre un subespacio S de V como una transformación lineal $Proj_S : V \rightarrow S$ y esto pasa si:

$$Proj_S u = u, \text{ para todo } u \in S.$$

$$Proj_S w = 0, \text{ para todo } w \in S^\perp.$$

Dado que las proyecciones son transformaciones lineales, entonces por (2.7) tiene una matriz asociada en este caso a la proyección, dicha matriz es conocida como matriz de proyección P la cual cumple que $P^2 = P$.

Estos conceptos del Álgebra Lineal se encuentran al detalle, con ejemplos y ejercicios prácticos en (Grossman, 2012) y (Lay, 2007). A continuación veremos en forma analítica y detallada las proyecciones ortogonales de vectores de dimensión mayor sobre subespacios de dimensión menor, mostrando desde los ejemplos más simples hasta los más complicados, esto nos llevará hacia la reducción de dimensionalidad de datos.

Proyección de \mathbb{R}^2 sobre subespacios unidimensionales

Comencemos mostrando como se calcula y como es la proyección de un vector v de \mathbb{R}^2 sobre un subespacio de dimensión uno (recta). Primero veamos como obtener la proyección ortogonal del vector v sobre el subespacio S de dimensión uno utilizando la definición del ángulo entre vectores y la norma del vector v .

De acuerdo a la figura 1-2, el símbolo θ es el ángulo formado entre el vector v y la recta que es un subespacio S de dimensión uno, y lo que se desea ahora es proyectar el vector v del plano sobre la recta S , la proyección $Proy_S(v)$ es un vector de S por lo que tendrá un vector u unitario del subespacio S en la dirección de S y la norma (longitud) del vector $Proy_S(v)$, es igual la norma del vector v por el coseno del ángulo, es decir, $\|Proy_S(v)\| = \|v\| \cos \theta$ es aquí donde la longitud $v - Proj_S(v)$ es mínima es decir, $v - Proj_S(v)$ es ortogonal al subespacio S .

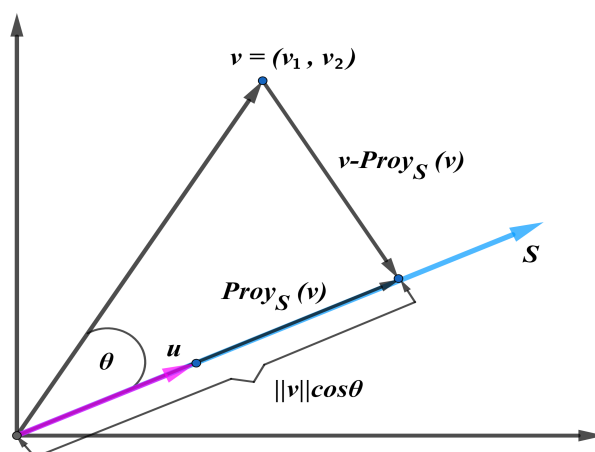


Ilustración 1-2: Proyección ortogonal sobre la recta mediante el ángulo entre vectores.

Realizado por: Buñay, Cristian, 2022.

Por lo tanto, el vector proyección es igual al vector unitario por la norma de la proyección de v sobre S .

$$Proy_S(v) = \frac{u}{\|u\|} \|v\| \cos \theta. \quad (2.25)$$

Ahora, recordando la definición del ángulo entre dos vectores y mediante el despeje de dicha expresión obtendremos lo siguiente:

$$\cos \theta = \frac{\langle u, v \rangle}{\|u\| \|v\|} \implies \|v\| \cos \theta = \frac{\langle u, v \rangle}{\|u\|} \implies \|v\| \cos \theta = \frac{u^T v}{\|u\|}. \quad (2.26)$$

Si reemplazamos la expresión 2.26 en 2.25 vemos que la proyección $Proy_S(v)$ se reduce a la siguiente expresión:

$$Proy_S(v) = \frac{u}{\|u\|} \frac{u^T v}{\|u\|} = \frac{uu^T v}{\|u\|^2} = \frac{uu^T v}{u^T u} = \frac{uu^T}{u^T u} v = u(u^T u)^{-1} u^T v. \quad (2.27)$$

Notemos que $u^T u$ es un número real y uu^T es una matriz cuadrada de orden dos, entonces la $Proy_S(v)$ se reduce a la obtención de la matriz $P = u(u^T u)^{-1} u^T$ y multiplicar por el vector v que se está proyectando.

$$Proy_S(v) = Pv. \quad (2.28)$$

Ahora, veamos como obtener la proyección de un vector $v \in \mathbb{R}^2$ sobre el subespacio de dimensión uno, conociendo los vectores que generan el subespacio S . La gráfica 2-2 nos ayudará a visualizar la proyección del vector v de dimensión dos sobre la recta.

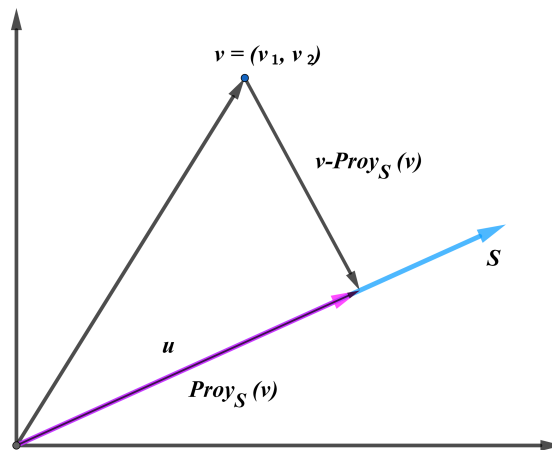


Ilustración 2-2: Proyección ortogonal sobre un espacio S de dimensión uno.

Realizado por: Buñay, Cristian, 2022.

En este caso la recta es un subespacio S de dimensión uno, y la base para este subespacio está formada solo por un vector director $u = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$, la cual genera el subespacio S .

Existen muchos vectores que son base del subespacio S , pero en este caso nos interesa sobre el vector u de la base que coincida con el vector proyección. Como la proyección de v sobre S es un vector de S , entonces por definición, cualquier vector del subespacio S se puede escribir como combinación lineal de los vectores que generan el subespacio S . Por lo tanto, la proyección $Proy_S(v)$ es una combinación lineal del vector u por algún escalar $\alpha \in \mathbb{R}$.

$$Proy_S(v) = \alpha u. \quad (2.29)$$

Como podemos ver en la figura **2-2**, la proyección $Proy_S(v)$ es el vector más cercano al vector, v lo cual significa que la norma $\|v - Proj_S(v)\|$ es mínima, es decir; la distancia entre el vector v y el vector proyección es mínima.

$$d(v, Proj_S(v)) = \min \|v - Proj_S(v)\|.$$

Esto implica que el segmento $v - Proj_S(v)$ es ortogonal al vector u que genera el subespacio S y por ende es ortogonal al subespacio unidimensional S . Recordemos que un vector es ortogonal a otro vector si su producto interno es igual a cero, entonces:

$$\langle u, v - Proj_S(v) \rangle = \langle u, v - \alpha u \rangle = 0. \quad (2.30)$$

Ahora, vamos a determinar las coordenadas de la proyección $Proj_S(v)$ y la matriz de proyección P . Como el producto interno es un producto escalar, es decir, $\langle u, u \rangle = u^T u$ entonces, usando este hecho en (2.30) y mediante las propiedades de la forma bilineal, obtenemos que:

$$\langle u, v - \alpha u \rangle = 0 \implies \langle u, v \rangle - \langle u, \alpha u \rangle \implies u^T v - u^T \alpha u = 0 \implies u^T v = u^T \alpha u.$$

Así, el valor de α queda expresado de la siguiente forma:

$$\alpha = \frac{u^T v}{u^T u}. \quad (2.31)$$

Observe que $u^T v$ y $u^T u$ son productos escalares donde se obtendrá como resultado un número escalar, y además los vectores no pueden ser simplificados, ya que tanto el vector u como el vector v son matrices de 2×1 , además el funcionamiento de la matriz no satisface la ley de multiplicación e intercambio, por lo tanto, no podemos calcular primero con $\frac{1}{u^T}$.

Ahora, conociendo los vectores que generan el subespacio S y con α determinada anteriormente, es posible escribir una expresión para el vector proyección. Esto es;

$$Proj_S(v) = \alpha u = \left(\frac{u^T v}{u^T u} \right) u. \quad (2.32)$$

Esta es la proyección del vector v sobre el subespacio S , lo cual muestra que cuando v se escala, el vector proyección se escala por el mismo múltiplo, mientras que si S escala, el vector $Proj_S(v)$

permanece igual. Recordemos que la multiplicación de un vector por un escalar es conmutativa, es decir, $Proy_S(v) = \alpha u = u\alpha$. Así, la proyección de v sobre S de la ecuación 2.32 se reduce la siguiente expresión.

$$Proy_S(v) = \left(\frac{u^T v}{u^T u} \right) u = u \frac{u^T v}{u^T u} = \frac{(uu^T)v}{u^T u} = \frac{uu^T}{u^T u} v \equiv u(u^T u)^{-1} u^T v. \quad (2.33)$$

Notemos que esta expresión coincide con la expresión 2.27 determinada anteriormente con la definición del ángulo entre vectores. Ahora, observemos que $(u^T u)^{-1}$ es un número real, por lo que en un subespacio de dimensión uno $u(u^T u)^{-1} u^T$ es una matriz de 2×2 definida de la siguiente manera:

$$\frac{uu^T}{u^T u} = u(u^T u)^{-1} u^T = \begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} aa & ab \\ ab & bb \end{bmatrix}.$$

Y la matriz de proyección viene dada como $P = u(u^T u)^{-1} u^T$. Ahora notemos que si $u^T u = \|u\|^2 = 1$ entonces la proyección es $Proy_S(v) = u\alpha = uu^T v$ donde la matriz de proyección es $P = uu^T$ y $\alpha = u^T v$. Esto nos muestra que la proyección de v sobre el subespacio S se obtiene mediante el producto de la matriz de proyección P por el vector v .

$$Proy_S(v) = Pv. \quad (2.34)$$

La matriz de proyección P proyecta cualquier vector v de \mathbb{R}^n sobre el subespacio S de dimensión uno generado por u . La proyección $Proy_S(v)$ es un vector bidimensional, es decir, pertenece a \mathbb{R}^2 pero se encuentra en el subespacio unidimensional y para representar esta proyección solo se necesita una coordenada que se expresa con respecto a la base u del subespacio S .

Proyección de \mathbb{R}^3 sobre subespacios bidimensionales

Supongamos ahora que se desea proyectar un vector de un espacio tridimensional, es decir, un vector $v = (v_1, v_2, v_3)$ de \mathbb{R}^3 sobre el subespacio S de dimensiones dos, es decir, sobre el espacio \mathbb{R}^2 .

Notemos ahora que en la figura 3-2 el vector u de la base del subespacio S , tiene dos componentes u_1 y u_2 de \mathbb{R}^3 las cuales generan el subespacio S de dimensión dos, es decir u_1, u_2 son base del subespacio bidimensional S y además α es el vector de escalares de dos componentes que son α_1 y α_2 tal como se puede observar en la siguiente gráfica.

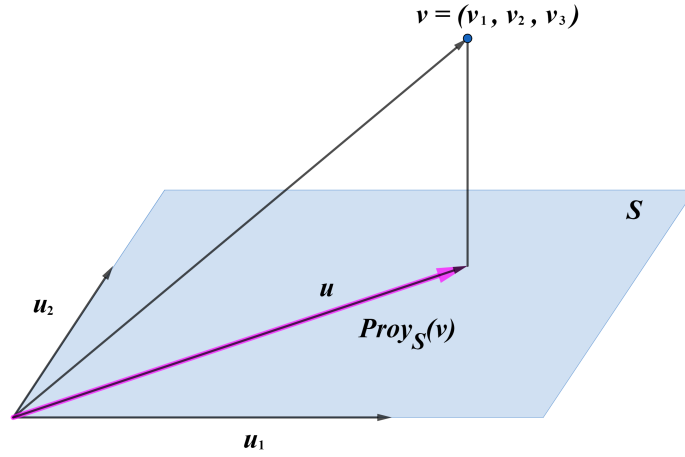


Ilustración 3-2: Proyección ortogonal de $v \in \mathbb{R}^3$ sobre el subespacio bidimensional S .

Realizado por: Buñay, Cristian, 2022.

Entonces, recordemos que cualquier vector del subespacio S se puede escribir como una combinación lineal de los vectores de la base del subespacio S , y como $Proj_S(v)$ es un vector del plano, entonces:

$$Proj_S(v) = u = \alpha_1 u_1 + \alpha_2 u_2. \quad (2.35)$$

El vector proyección $Proj_S(v)$ es el vector único de S más cercano al vector v , esto significa que la norma $\|v - Proj_S(v)\|$ sea mínima es decir;

$$d(v, Proj_S(v)) = \min \|v - Proj_S(v)\|,$$

lo cual implica que el vector $v - Proj_S(v)$ será ortogonal al plano expandido por u_1 y u_2 , lo cual significa que es ortogonal a cada uno de los vectores u_1 y u_2 . Esto se caracteriza de la siguiente manera:

$$\langle v - Proj_S(v), u_1 \rangle = 0 \quad \wedge \quad \langle v - Proj_S(v), u_2 \rangle = 0. \quad (2.36)$$

Sustituyendo la expresión 2.35 en la expresión 2.36 determinada anteriormente, tomando como producto interno al producto escalar y mediante las propiedades de bilinealidad, se obtiene la siguiente deducción:

$$\begin{aligned} \langle v - (\alpha_1 u_1 + \alpha_2 u_2), u_1 \rangle = 0 & \quad \wedge \quad \langle v - (\alpha_1 u_1 + \alpha_2 u_2), u_2 \rangle = 0. \\ \langle v - \alpha_1 u_1 - \alpha_2 u_2, u_1 \rangle = 0 & \quad \wedge \quad \langle v - \alpha_1 u_1 - \alpha_2 u_2, u_2 \rangle = 0. \\ (v - \alpha_1 u_1 - \alpha_2 u_2)^T u_1 = 0 & \quad \wedge \quad (v - \alpha_1 u_1 - \alpha_2 u_2)^T u_2 = 0. \\ v^T u_1 - \alpha_1^T u_1^T u_1 - \alpha_2^T u_2^T u_1 = 0 & \quad \wedge \quad v^T u_2 - \alpha_1^T u_1^T u_2 - \alpha_2^T u_2^T u_2 = 0. \end{aligned}$$

$$v^T u_1 = \alpha_1^T u_1^T u_1 + \alpha_2^T u_2^T u_1 \quad \wedge \quad v^T u_2 = \alpha_1^T u_1^T u_2 + \alpha_2^T u_2^T u_2.$$

Ahora, combinando estas últimas expresiones, tenemos lo siguiente:

$$\begin{aligned} v^T u_1 + v^T u_2 &= \alpha_1^T u_1^T u_1 + \alpha_2^T u_2^T u_1 + \alpha_1^T u_1^T u_2 + \alpha_2^T u_2^T u_2. \\ v^T \underbrace{(u_1, u_2)}_u &= \alpha_1^T \underbrace{u_1^T}_u \underbrace{(u_1, u_2)}_u + \alpha_2^T \underbrace{u_2^T}_u \underbrace{(u_1, u_2)}_u. \\ v^T u &= \alpha_1^T u_1^T u + \alpha_2^T u_2^T u. \\ v^T u &= (\alpha_1^T u_1^T + \alpha_2^T u_2^T) u. \\ v^T u &= \alpha^T u^T u. \end{aligned} \tag{2.37}$$

Notemos que $u^T u$ es una matriz cuadrada de la forma:

$$u^T u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} + c_{11} & a_{12} + b_{12} + c_{12} \\ a_{21} + b_{21} + c_{21} & a_{22} + b_{22} + c_{22} \end{bmatrix}.$$

Además, tiene vectores de la base del subespacio S , por lo que los vectores son linealmente independientes, por lo tanto, la inversa de esta matriz existe. Entonces podemos deducir la expresión 2.37 determinada anteriormente, multiplicando por la inversa de la matriz $u^T u$ para obtener el vector de escalares α .

$$v^T u (u^T u)^{-1} = \alpha^T u^T u (u^T u)^{-1} \implies v^T u (u^T u)^{-1} = \alpha^T I \equiv \alpha = (u^T u)^{-1} u^T v.$$

El vector u se puede ver como una matriz cuya columna son u_1 y u_2 entonces la proyección será el producto de u por el vector α determinada anteriormente, por lo tanto, la proyección representada en la ecuación 2.35 se reduce a la siguiente expresión:

$$Proy_S(v) = u\alpha = u(u^T u)^{-1} u^T v. \tag{2.38}$$

Observemos que P definida de la siguiente manera $P = u(u^T u)^{-1} u^T$ se lo conoce como matriz de proyección. Así, nuevamente obtenemos una expresión similar al cálculo de proyección determinada en (2.25) y (2.32). Por lo tanto, el cálculo de la proyección del vector v de \mathbb{R}^3 sobre el plano se reduce a la multiplicación de la matriz P por el vector v que se está proyectando.

$$Proy_S(v) = Pv. \tag{2.39}$$

Si los vectores de u son ortonormales, entonces $u^T u$ es la matriz identidad, por lo tanto, la matriz de proyección P se reduce a $P = uu^T$ así la $Proy_S(v) = Pv = uu^T v$.

La matriz de proyección P proyecta cualquier vector v de \mathbb{R}^n sobre el subespacio S de dimensión dos generados por u_1, u_2 . Ahora, la proyección $Proy_S(v)$ es un vector tridimensional, es decir, pertenece a \mathbb{R}^3 que se encuentra en el hiperplano y para representar esta proyección solo se necesitan dos coordenadas de α que se expresa con respecto a la base u_1, u_2 del subespacio S .

Proyección de \mathbb{R}^4 sobre subespacios tridimensionales

Ahora, veamos que pasa si incrementamos la dimensión, es decir, se desea proyectar un vector $v = \{v_1, v_2, v_3, v_4\}$ de \mathbb{R}^4 sobre el subespacio S de \mathbb{R}^3 . En este caso no podemos visualizar mediante una gráfica como en los casos anteriores, por lo que describiremos de manera analítica esta proyección. Siguiendo la analogía anterior, el vector u de la base que coincide con el vector $Proy_S(v)$ del subespacio S tiene tres componentes u_1, u_2 y u_3 las cuales generan el subespacio tridimensional S . Como la proyección es un vector de S entonces por definición, la proyección se puede escribir como una combinación lineal de los vectores de la base, en este caso de los vectores de u .

$$Proy_S(v) = u = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3. \quad (2.40)$$

Y nuevamente la proyección será el vector de S más cercano al vector v , lo que implica que la distancia del vector v hacia el vector $Proy_S(v)$ será mínima.

$$d(v, Proj_S(v)) = \min \|v - Proj_S(v)\|.$$

Esto significa que el segmento $v - Proj_S(v)$ es ortogonal al subespacio S generado por los vectores u_1, u_2, u_3 , esto implica que el vector $v - Proj_S(v)$ será ortogonal a u_1, u_2 y a u_3 . Esto se caracteriza de la siguiente manera:

$$\langle u_1, v - Proj_S(v) \rangle = 0 \quad y \quad \langle u_2, v - Proj_S(v) \rangle = 0 \quad y \quad \langle u_3, v - Proj_S(v) \rangle = 0. \quad (2.41)$$

Si sustituimos la expresión 2.40 en 2.41 y resolviendo mediante las propiedades del producto interno, se tiene que:

$$\begin{aligned} \langle u_1, v - Proj_S(v) \rangle = 0 & \quad \wedge \quad \langle u_2, v - Proj_S(v) \rangle = 0 & \quad \wedge \quad \langle u_3, v - Proj_S(v) \rangle = 0. \\ u_1^T (v - Proj_S(v)) = 0 & \quad \wedge \quad u_2^T (v - Proj_S(v)) = 0 & \quad \wedge \quad u_3^T (v - Proj_S(v)) = 0. \\ u_1^T v - u_1^T Proj_S(v) = 0 & \quad \wedge \quad u_2^T v - u_2^T Proj_S(v) = 0 & \quad \wedge \quad u_3^T v - u_3^T Proj_S(v) = 0. \\ u_1^T v = u_1^T Proj_S(v) & \quad \wedge \quad u_2^T v = u_2^T Proj_S(v) & \quad \wedge \quad u_3^T v = u_3^T Proj_S(v). \\ u_1^T v = u_1^T (\alpha u) & \quad \wedge \quad u_2^T v = u_2^T (\alpha u) & \quad \wedge \quad u_3^T v = u_3^T (\alpha u). \end{aligned}$$

Sumando estas expresiones se obtiene lo siguiente:

$$[u_1^T, u_2^T, u_3^T]v = [u_1^T, u_2^T, u_3^T]\alpha u.$$

$$u^T v = \alpha u^T u.$$

$$\alpha = \frac{u^T v}{u^T u} = (u^T u)^{-1} u^T v. \quad (2.42)$$

Nuevamente, el vector u se puede ver como una matriz cuya columna son u_1, u_2 y u_3 entonces la proyección será el producto de u por el vector α determinada anteriormente, por lo tanto, la

proyección representada en la ecuación (2.40) se reduce a la siguiente expresión:

$$Proy_S(v) = u(u^T u)^{-1} u^T v. \quad (2.43)$$

Notemos ahora que $u^T u$ es una matriz cuadrada de 3×3 , y sus vectores son linealmente independientes, por lo tanto, tiene inversa. En sí, toda matriz cuadrada cuyos vectores sean linealmente independientes es invertible. Entonces $u(u^T u)^{-1} u^T$ sigue siendo una matriz a la cual se le conoce como matriz de proyección definida como sigue: $P = u(u^T u)^{-1} u^T$. Así, nuevamente obtenemos una expresión similar al cálculo de proyección determinada en (2.25) y (2.32). Por lo tanto, el cálculo de la proyección del vector v de \mathbb{R}^4 sobre el espacio tridimensional se reduce a la multiplicación de la matriz P por el vector v que se está proyectando.

$$Proy_S(v) = Pv.$$

Nuevamente, si los vectores de u son ortonormales, entonces $u^T u$ es la matriz identidad y la matriz de proyección P se reduce a $P = uu^T$ por lo tanto $Proy_S(v) = Pv = uu^T v$. En este caso el vector $Proy_S(v) \in \mathbb{R}^4$ es un vector de cuatro componentes que se encuentra en un espacio de \mathbb{R}^3 , por lo que para representarla solo necesitamos tres coordenadas de α .

Notemos que la matriz de proyección P solamente depende del vector u de la base del subespacio S , entonces podemos generalizar esta idea, y proyectar cualquier vector v de \mathbb{R}^n sobre cualquier subespacio S de dimensión q donde $q < n$.

Entonces, el vector u de la base del subespacio S tendrá q componentes, es decir, $u = (u_1, u_2, \dots, u_q)$ las cuales generan el subespacio S al cual se está proyectando el vector v de \mathbb{R}^n , además el vector de escalares α también tendrá q componentes.

Proyección de \mathbb{R}^n sobre subespacios q -dimensionales ($q < n$)

Recuerde que en el caso anterior el vector u del plano tenía dos componentes u_1, u_2 y cada componente era una matriz de 3×1 entonces el vector u es una matriz de 3×2 , de la misma forma cuando proyectamos un vector de \mathbb{R}^4 sobre el subespacio tridimensional, el vector u de la base de \mathbb{R}^3 está formado por tres componentes u_1, u_2, u_3 donde cada componente es una matriz de 3×1 , entonces el vector u es una matriz cuadrada de 3×3 . Esto nos conduce a la generalización de la proyección mediante matrices.

En **términos matriciales**, podemos representar los vectores u_1, u_2, \dots, u_q de la base B_q del subespacio q -dimensional, en una matriz de la siguiente forma:

$$B = \begin{bmatrix} u_1 & u_2 & \dots & u_q \end{bmatrix},$$

donde cada u_j son vectores columnas, las cuales pertenecen a la base del subespacio S , y el vector de escalares, la cual está definida de la siguiente manera $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_q)$.

Entonces la proyección será una combinación lineal, en este caso de la matriz B por el escalar α , la cual está denotada como sigue:

$$Proy_S(v) = B\alpha. \quad (2.44)$$

Y para que la proyección sea ortogonal, el producto interno de $v - Proj_S(v)$ con cada vector columna u_j , $j = 1, 2, \dots, q$ de la matriz B será igual a cero, es decir:

$$\langle B, v - Proj_S(v) \rangle = 0. \quad (2.45)$$

Si sustituimos la expresión (2.44) en (2.45) y sabiendo que el producto interno es un producto escalar en \mathbb{R}^n , entonces obtendremos lo siguiente:

$$\begin{aligned} \langle B, v - B\alpha \rangle = 0 &\implies B^T(v - B\alpha) = 0. \\ &\implies B^T v - B^T B\alpha = 0. \\ &\implies B^T v = B^T B\alpha. \end{aligned}$$

Notemos que $B^T B$ es una matriz cuadrada y además tiene los vectores de la base, es decir, que sus columnas son linealmente independientes, lo cual significa que la inversa $(B^T B)^{-1}$ existe. Entonces podemos multiplicar $B^T B^{-1}$ por la izquierda a ambos lados de la igualdad de la expresión anterior.

$$\begin{aligned} (B^T B)^{-1} B^T v &= \underbrace{(B^T B)^{-1} B^T B}_{\text{Identidad}} \alpha. \\ \alpha &= (B^T B)^{-1} B^T v. \end{aligned}$$

Una vez obtenida el valor de α , la proyección del vector v de \mathbb{R}^n sobre el subespacio S de dimensión q se define como:

$$Proj_S(v) = B\alpha = B(B^T B)^{-1} B^T v, \quad (2.46)$$

donde $P = B(B^T B)^{-1} B^T$ es la matriz proyección. Si los vectores columnas de la matriz B son ortonormales, es decir, $u_i^T u_j = 0$ si $i \neq j$ y $u_i^T u_j = 1$ si $i = j$, entonces $B^T B$ es la matriz identidad, lo que nos lleva a reducir el cálculo de la proyección a $Proj_S(v) = B\alpha = BB^T v$, donde $P = BB^T$ es la matriz de proyección y el vector de escalares $\alpha = B^T v$.

Con esto en claro, conociendo la base ortonormal $B_q = \{u_1, u_1, \dots, u_q\}$ del subespacio S , podemos proyectar ortogonalmente cualquier vector v de \mathbb{R}^n sobre este subespacio S de dimensión q donde $q \leq n$, la matriz B tendrá como columnas los vectores u_1, u_1, \dots, u_q donde cada vector columna es de n -dimensional, por lo tanto, la matriz B será de $n \times q$ y la proyección estará dada como

$Proy_S(v) = BB^T v$ que será un vector n -dimensional que se encuentra en el subespacio S donde la matriz de proyección es $P = BB^T$ y $\alpha = B^T v$ de q -dimensional. Notemos la importancia de trabajar con vectores ortonormales, ya que reduce grandemente el cálculo de la proyección.

A veces trabajar con un conjunto de vectores linealmente independiente de un espacio vectorial no es suficiente, por lo que en algunos casos es importante obtener bases que sean ortogonales u ortonormales. Si quisiéramos convertir un conjunto de vectores linealmente independientes del espacio vectorial V en una base ortogonal, podemos ortogonalizar a cada uno de sus elementos. Para ello, la primera aplicación que veremos de las proyecciones ortogonales es el denominado el proceso de Gram-Schmidt, que consiste en ortogonalizar un conjunto de vectores linealmente independientes.

Proceso de Gram-Schmidt

Para sacarle mayor provecho a las proyecciones, la matriz del análisis de datos debería tener columnas ortogonales u ortonormales, para ello la primera aplicación que vamos a presentar de uso de las proyecciones es el llamado método de Gram-Schmidt que permite ortogonalizar a partir de un conjunto de vectores linealmente independientes, para luego poder ortonormalizar los vectores ortogonales.

El método de **Gram-Schmidt** es un proceso netamente utilizado para obtener una base ortogonal de un espacio vectorial V cualquiera de \mathbb{R}^n distinto de cero. Este método nos dice que, dado un conjunto no vacío de vectores $\{v_1, v_2, \dots, v_q\}$ linealmente independientes en \mathbb{R}^n devuelve un conjunto de q vectores ortogonales. A continuación veamos como trabaja este proceso:

Paso 1: Fijemos el primer vector v_1 linealmente independiente.

$$u_1 = v_1.$$

Paso 2: Definamos el siguiente vector u_2 como la diferencia entre el vector v_2 con el complemento ortogonal con respecto a v_2 es decir;

$$u_2 = v_2 - \alpha u_1.$$

Lo que se debe hacer es que el vector u_2 sea ortogonal a u_1 , es decir, $u_1^T u_2 = 0$, para ello encontremos el valor del escalar α ;

$$\begin{aligned}
u_1^T u_2 = 0 &\implies u_1^T (v_2 - \alpha u_1) = 0. \\
&\implies u_1^T v_2 - \alpha u_1^T u_1 = 0. \\
&\implies u_1^T v_2 = \alpha u_1^T u_1. \\
&\implies \alpha = \frac{u_1^T v_2}{u_1^T u_1}.
\end{aligned}$$

Sustituyendo el valor de α y mediante la definición 2.24 el vector u_2 ortogonal al vector u_1 queda expresado de la siguiente manera:

$$u_2 = v_2 - \frac{u_1^T v_2}{\|u_1\|^2} u_1 \implies u_2 = v_2 - \text{Proy}_{u_1} v_2.$$

Paso 3: Ahora, definamos el siguiente vector u_3 como:

$$u_3 = v_3 - \alpha_1 u_1 - \alpha_2 u_2.$$

Este vector u_3 debe ser ortogonal a los vectores u_1 y u_2 , para ello encontremos los valores de α_1 y α_2 . Entonces

$$\begin{aligned}
u_1^T u_3 = 0 &\quad \wedge \quad u_2^T u_3 = 0. \\
u_1^T (v_3 - \alpha_1 u_1) = 0 &\quad \wedge \quad u_2^T (v_3 - \alpha_2 u_2) = 0. \\
u_1^T v_3 - \alpha_1 u_1^T u_1 = 0 &\quad \wedge \quad u_2^T v_3 - \alpha_2 u_2^T u_2 = 0. \\
u_1^T v_3 = \alpha_1 u_1^T u_1 &\quad \wedge \quad u_2^T v_3 = \alpha_2 u_2^T u_2. \\
\alpha_1 = \frac{u_1^T v_3}{u_1^T u_1} &\quad \wedge \quad \alpha_2 = \frac{u_2^T v_3}{u_2^T u_2}.
\end{aligned}$$

Así el vector u_3 ortogonal a u_1 y a u_2 se expresa de la siguiente manera:

$$u_3 = v_3 - \frac{u_1^T v_3}{u_1^T u_1} u_1 - \frac{u_2^T v_3}{u_2^T u_2} u_2 \implies u_3 = v_3 - \text{Proy}_{u_1}(v_3) - \text{Proy}_{u_2}(v_3).$$

Paso siguiente: La idea es seguir hasta realizar con todos los vectores $\{v_1, v_2, \dots, v_q\}$ linealmente independientes. Supongamos que se tiene hasta u_q vectores ortogonales, lo que implica que u_q será ortogonal a todos los vectores u_i anteriores, donde cada $\alpha_i = \frac{u_i^T v_q}{u_i^T u_i}$. Así u_q está dada como:

$$\begin{aligned}
u_q &= v_q - \frac{u_1^T v_q}{u_1^T u_1} u_1 - \frac{u_2^T v_q}{u_2^T u_2} u_2 - \dots - \frac{u_{q-1}^T v_q}{u_{q-1}^T u_{q-1}} u_{q-1}. \\
&= v_q - \text{Proy}_{u_1}(v_q) - \text{Proy}_{u_2}(v_q) - \dots - \text{Proy}_{u_{q-1}}(v_q).
\end{aligned}$$

Entonces, para conseguir el vector u_{q+1} definido como:

$$u_{q+1} = v_{q+1} - \alpha_1 u_1 - \alpha_2 u_2 - \dots - \alpha_q u_q,$$

debemos determinar el valor de cada α_i donde $i = 1, 2, \dots, q$. En este caso α_i está dado como:

$$\alpha_i = \frac{u_i^T v_{q+1}}{u_i^T u_i}.$$

Así el vector u_{q+1} es ortogonal a todos los vectores u_i anteriores, el cual está representado de la siguiente manera:

$$\begin{aligned} u_{q+1} &= v_{q+1} - \frac{u_1^T v_{q+1}}{u_1^T u_1} u_1 - \frac{u_2^T v_{q+1}}{u_2^T u_2} u_2 - \dots - \frac{u_q^T v_{q+1}}{u_q^T u_q} u_q. \\ &= v_{q+1} - \text{Proy}_{u_1}(v_{q+1}) - \text{Proy}_{u_2}(v_{q+1}) - \dots - \text{Proy}_{u_q}(v_{q+1}). \end{aligned}$$

El conjunto de vectores $B' = \{u_1, u_2, \dots, u_{q+1}\}$ forman una base ortogonal, ya que todos los vectores son ortogonales entre sí.

Si se tiene una base ortogonal, entonces la podemos ortonormalizar estos vectores, es decir, una base ortonormal $B = \{w_1, w_2, \dots, w_n, w_{q+1}\}$ se obtiene mediante el cálculo de la norma de cada vector de la base ortogonal, donde cada vector w_i de la base ortonormal está representado como:

$$w_i = \frac{u_i}{\|u_i\|}$$

.

Autovalores y Autovectores

En la técnica de Análisis de Componentes Principales PCA, el cálculo de autovalores y autovectores es muy importante. Algunos autores denominan a los Autovectores como vectores propios o Eigenvectors y a los Autovalores como valores propios o Eigenvalue. La definición nos dice que; dado una matriz $\sigma \in \mathbb{R}^{n \times n}$ cuadrada y un vector v no nulo de \mathbb{R}^n tal que:

$$\sigma v = \alpha v \quad \text{con } \alpha \in \mathbb{R}. \quad (2.47)$$

Al escalar α que cumpla esta ecuación se denomina **autovalor** o valor propio de la matriz σ y al vector v que cumpla esta ecuación se denomina **autovector** o vector propio de la matriz σ , asociado a α .

Notemos que el autovalor y autovector está asociados únicamente a matrices cuadradas, además el vector v no nulo debe ser de la misma dimensión de la matriz σ . La ecuación (2.47) no se cumple con cualquier vector, o con cualquier escalar, para encontrar aquel vector y aquel escalar que cumpla la expresión (2.47) empezaremos por determinar los autovalores de una matriz cualquiera, partiendo de la definición:

$$\sigma v = \alpha v \implies \sigma v - \alpha v = 0. \quad (2.48)$$

De esta expresión sacamos el factor común a la derecha al vector v , ya que no podemos tomar como factor común hacia la izquierda, pues nada nos garantiza que σv sea conmutativa, además, cuando se saca como factor común a v no podemos restar la matriz σ con un número real α por lo que se deja la matriz identidad del mismo orden de la matriz σ , así la ecuación (2.48) se reduce a:

$$(\sigma - \alpha I)v = 0. \quad (2.49)$$

La ecuación anterior se denomina como la ecuación matricial de un sistema de ecuaciones lineales homogéneo, donde $\sigma - \alpha v$: es la matriz de los coeficientes del sistema, simplificando la ecuación (2.49) obtenemos que:

$$(\sigma - \alpha I)v = 0 \implies \sigma - \alpha I = 0, \quad (2.50)$$

el 0 es una matriz de términos independientes. Ahora, el siguiente sistema de ecuaciones lineales $\sigma - \alpha I = 0$ es compatible y siempre tiene solución, esto por el hecho de ser homogéneo, por lo tanto, siempre habrá autovalores y autovectores.

Si la solución es compatible determinado existe un único autovector $v = 0$, pero recordemos que en la definición nos dice que el autovector v debe ser distinto del vector nulo. Lo que nos interesa entonces es que la solución sea compatible, indeterminado, esto indicará que siempre existen infinitos autovectores.

Entonces, primero buscamos los autovalores que nos aseguren infinitas soluciones, es decir, para todo α tal que el determinante $Det(\sigma - \alpha v) = 0$, esta ecuación es denominada Ecuación Característica de la matriz. Esto nos permite descartar todos los determinantes distintos de cero. Para mayor entendimiento sobre el cálculo de los autovalores y autovectores se recomienda (Grossman, 2012), (Lay, 2007) y (Datta, 2010) las cuales presentan ejemplos y ejercicios prácticos.

Medidas Estadísticas

A continuación se definen algunos conceptos estadísticos básicos y necesarios para el entendimiento de la técnica de Análisis de Componentes Principales para el proceso de reducción de dimensionalidad.

En estadística siempre se maneja con datos de una población o de una muestra, se entiende por población a un conjunto completo de elementos de estudio que poseen un parámetro común; por ejemplo, los estudiantes de una escuela o los habitantes de una ciudad, pero no tiene que ser necesariamente una población humana, puede ser cualquier colección de datos que posea un parámetro común; por ejemplo el número total de gimnasios en una ciudad. Una muestra es un subconjunto pequeño de toda la población.

Entonces, consideremos un conjunto finito de datos muestrales.

$$X = \{x_1, x_2, \dots, x_N\},$$

donde cada elemento x_i del conjunto posee ciertas características o variables de estudio. El manejo de dichos datos parte de establecer parámetros estadísticos como la media y la varianza que miden la variabilidad de los datos y nos dan información sobre la dispersión de la misma.

La **media** del conjunto es el valor promedio de todos los datos, y lo denotamos con \bar{x} la cual está expresada de la siguiente manera:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + \dots + x_N}{N}. \quad (2.51)$$

donde cada x_i representa una característica o variable de estudio y N es el total de observaciones o el tamaño de la muestra.

La media juega un papel muy importante en la comprensión de la tendencia central de un conjunto de datos; sin embargo, en algunos casos puede ser afectada por valores que están altamente separados o cercanos que el resto de los valores. Cuando deseamos describir una variable, usamos algunas medidas de centralidad y de dispersión, una de ellas es la media mayormente utilizada y la otra es la desviación estándar, pero vemos que cuando la distribución de las observaciones está sesgada, la media no es una buena medida de la posición central, por lo que se recomienda usar la mediana (Solano, et al., 2009). Estas medidas son muy importantes para determinar la dispersión o variabilidad de los datos. Esto conduce a la definición de la **varianza** denotada por s^2 la cual está dada como:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}. \quad (2.52)$$

En algunos casos se divide por $N - 1$ para evitar que el estimador sea sesgado o en algunas para notar que se está trabajando con una muestra, y se divide únicamente por N para notar que se está trabajando con una población. Ahora, mientras mayor sea la dispersión o varianza comúnmente conocida de los datos, mayor es la desviación estándar de la misma y eso nos pudiera dar indicios del comportamiento de los datos para la toma de decisiones. La raíz cuadrada positiva de la varianza es lo que se conoce como la **desviación estándar o típica**.

$$s = + \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (2.53)$$

La desviación típica o estándar es el valor promedio de las desviaciones individuales de cada observación con respecto a la media aritmética, a pesar de que los valores no son iguales, esto permite ver la distancia o medir el grado de dispersión o variabilidad de los datos. Mientras mayor es la desviación estándar, mayor será la dispersión de los datos.

Estos conceptos importantes se deben concebir muy bien para comprender los demás conceptos estadísticos que se describirán más adelante.

Cada variable puede tomar valores cuantitativos o cualitativos; y representa características del objeto de estudio, ya sean características físicas de personas, objetos, o imágenes; de esta manera cuando se requiere tener en cuenta simultáneamente n variables o características de N individuos podemos utilizar una matriz X la cual represente estos datos.

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nN} \end{bmatrix}$$

donde las columnas x_i , $i = 1, 2, \dots, N$ de la matriz X , representan las n características del i -ésimo individuo. Las características de los individuos las podemos ver como vector fila de la matriz, realizando la traspuesta a cada vector columna de la siguiente forma:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix}$$

Es decir, cada vector fila x_i^T representa las características del i -ésimo individuo y para el estudio de estas n características tomaremos las medias \bar{x}_i de cada vector x_i^T y así representarla mediante un vector de medias m , el cual tendrá n componentes de la siguiente manera:

$$m = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{bmatrix}, \text{ donde } \bar{x}_i = \frac{1}{N} x_i^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{x_{i1} + x_{i2} + \dots + x_{in}}{N}$$

donde N es el número total de características de cada x_i^T . Ahora nos interesa saber cómo se comporta una variable en función de lo que hace otra variable. Es decir, se desea saber como varían en conjunto cada par de variables $x = (x_1, x_2, \dots, x_n)$ y $y = (y_1, y_2, \dots, y_n)$ con respecto a sus medias; a esto se le conoce como la **covarianza**, la cual se denota como:

$$\sigma = Cov(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (2.54)$$

Donde \bar{x} es la media muestral de x y \bar{y} es la media muestral de y , además x_i y y_i son los valores de x e y del i -ésimo término de la muestra y N es el número total de registros de una muestra. Si

nuestro conjunto de datos está centrado alrededor de la media, es decir, si la esperanza o el valor esperado $E[X] = 0$, entonces la matriz de covarianza se define como:

$$\sigma = \frac{1}{N} \sum_{i=1}^N x_i x_i^T \in \mathbb{R}^{p \times p} \quad (2.55)$$

Para determinar el tipo de relación que existe entre las variables, bastaría con conocer los signos:

1. Si la $Cov(x,y)$ es positiva, entonces la relación entre variables es directa, es decir, ambas variables varían en la misma dirección.
2. Si la $Cov(x,y)$ es negativa, entonces la relación entre variables es inversa, es decir, la variación es en la dirección opuesta.
3. Si la $Cov(x,y) = 0$ no existe una relación lineal entre las variables.

La figura 4-2 muestra los distintos tipos de relaciones existentes entre variables, las cuales fueron mencionadas anteriormente:

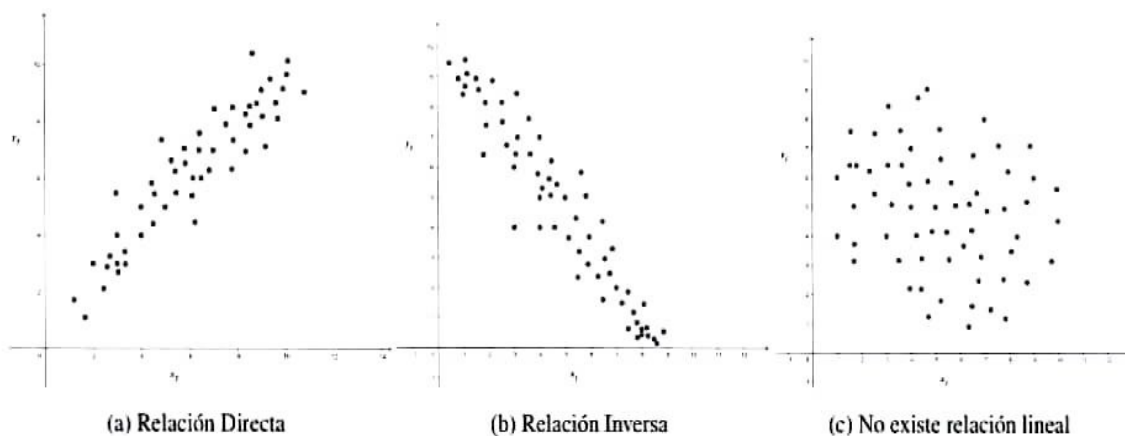


Ilustración 4-2: Correlación lineal entre variables.

Realizado por: Buñay, Cristian, 2022.

La matriz de covarianza también se puede representar mediante una matriz denominada matriz de varianza y covarianza o mayormente conocida como la matriz de covarianza.

$$\sigma = \begin{bmatrix} s_{11}^2 & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22}^2 & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn}^2 \end{bmatrix}$$

Esta matriz tiene en la diagonal principal las varianzas $s_{11}^2, s_{22}^2, \dots, s_{nn}^2$ de cada variable y los elementos que se encuentran fuera de la diagonal principal se le conoce como las covarianzas. Veamos a continuación algunas propiedades de esta la matriz:

1. La matriz σ es simétrica: es decir $\sigma = \sigma^T$.
2. La matriz σ es definida positiva: significa que, para cualquier vector x distinto de cero se tiene que $x^T \sigma x > 0$. Además todos los autovalores de σ son positivos o cero.

Otro de los conceptos estadísticos que estudia la variabilidad de los datos, el cual está relacionado con la covarianza, es el concepto de **coeficiente de correlación lineal**, esto lo podemos visualizar al detalle en (Faraldo, P.; & Pateiro, B., 2013) donde nos muestra que la correlación explica sobre el cambio en una variable la cual varía entre -1 y $+1$. Si $r = -1$ significa que la relación es contraria, es decir, son inversamente proporcionales, ahora, si $r = +1$ significa que las variables están directamente relacionadas, es decir, son proporcionales entre sí, pero si $r = 0$ significa que no existe una relación lineal entre variables, es decir, las variables no están correlacionadas, sin embargo, podría existir otra relación funcional. La correlación entre las variables x y y está definida como:

$$r(x,y) = \frac{Cov(x,y)}{s_x s_y}, \quad (2.56)$$

donde $Cov(x,y)$ es la covarianza de las variables x y y , y s_x y s_y representan la desviación estándar de x y y .

Para una instrucción más detallada de estos conceptos estadísticos que miden la dispersión de los datos y alguna de sus aplicaciones, se recomienda (Faraldo, P.; & Pateiro, B., 2013) y (James, 2013).

CAPÍTULO III

3. MARCO METODOLÓGICO

En este capítulo se describe la metodología de la investigación utilizada para cumplir con los objetivos planteados inicialmente y así obtener los resultados de la investigación de este tópico de Análisis de Componentes Principales vía Proyección Ortogonal.

La metodología utilizada para el desarrollo de este trabajo de investigación es de diseño documental, a nivel descriptivo, con enfoque cualitativo, en donde se emplearon estrategias de búsqueda y organización de la información adecuada, eficiente y apropiada para el tópico de Análisis de Componentes Principales vía Proyecciones Ortogonales, a fin de describir este proceso.

3.1. Descripción del tipo de investigación

La investigación planteada es de diseño documental, ya que la información obtenida fue revisada de manera sistemática la bibliografía especializada en este tópico para la descripción de la técnica de Análisis de Componentes Principales y el desarrollo de un algoritmo básico para la reducción de dimensionalidad mediante la técnica del PCA.

La información que se obtuvo es a través de la indagación exhaustiva, sistemática y rigurosa, utilizando técnica muy precisas de la documentación existente que, directa o indirectamente, aporte la información referente al fenómeno que estudiaremos, esto es, mediante la lectura de documentos como: libros, revistas, papers, tesis, entre otros.

3.2. Descripción del nivel de investigación

Esta investigación es de nivel descriptivo, ya que presenta la descripción de conceptos importantes del Álgebra Lineal como espacios y subespacios vectoriales, transformaciones lineales, ortogonalidad, bases ortogonales y ortonormales, las proyecciones en espacios de menor dimensión, y algunas medidas estadísticas de dispersión como la media, varianza, coeficiente de correlación lineal y la matriz de covarianza las cuales intervienen en el análisis de componentes principales para la reducción de dimensionalidad de datos usada frecuentemente en distintas aplicaciones.

3.3. Descripción del enfoque de la investigación

El enfoque de esta investigación es cualitativo, ya que este documento presenta la descripción de toda la teoría matemática basándose en nuestra percepción, y además se evalúa únicamente la calidad de la herramienta proyección ortogonal dentro de la técnica de Análisis de Componentes Principales en el proceso de reducción de dimensionalidad.

Redacción del documento

Para el cumplimiento satisfactorio del objetivo planteado para el desarrollo de esta investigación, se llevarán a cabo una serie de actividades que se da a conocer a continuación:

1. Selección y organización del material bibliográfico.
2. Estudio y comprensión de las proyecciones ortogonales y sus principales aplicaciones.
3. Entendimiento y descripción del proceso de reducción de dimensionalidad.
4. Derivación de la técnica de componentes principales, para la reducción de dimensionalidad.

Una vez recolectada la información relacionada con el tópico de Análisis de Componentes Principales, se realizó la redacción del documento de manera clara y concisa, donde se detalla toda la teoría matemática que intervienen en la técnica del PCA. Se empezó por escribir una breve introducción y los antecedentes históricos, luego se dio a conocer los fundamentos teóricos donde se describen los conceptos importantes del Álgebra Lineal y algunas medidas estadísticas de dispersión como la media, varianza, matriz covarianza y el coeficiente de correlación lineal. Toda esta teoría matemática fue la base para la descripción y deducción de la técnica de Análisis de Componentes Principales, usada ampliamente en la reducción de dimensionalidad de datos. Finalmente, se logró generar un algoritmo el cual explica cada uno de los pasos para obtener la reducción de dimensionalidad, por último se describen los resultados obtenidos mediante el código básico del PCA. De esta manera se generó el documento “Proyecciones Ortogonales y sus Aplicaciones”.

Uso del Software MATLAB

El software utilizado para la implementación del código es el conocido MATLAB, ya que es un sistema de cómputo numérico que ofrece un entorno de desarrollo integrado con un lenguaje de programación propio. Además es un software muy completo, y tiene una enorme capacidad para resolver problemas del Álgebra Lineal.

CAPÍTULO IV

4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En este capítulo se utilizarán los elementos descritos en el marco teórico para deducir, en forma sistemática, cada paso en el Análisis de Componentes Principales.

El análisis o analítica de datos es el proceso de examinar conjuntos de datos para encontrar tendencias y sacar conclusiones sobre la información que contienen. Esta técnica de Análisis de Componentes Principales para la reducción de dimensionalidad, surge a partir del interés de identificar y eliminar las variables irrelevantes en datos provenientes de distintas aplicaciones, a fin de hacer más eficiente su tratamiento. Particularmente, cuando el problema es tratado computacionalmente, esta reducción conlleva a una mejora del rendimiento computacional y reduce la complejidad, lo que lleva a facilitar el análisis y la comprensión de resultados.

4.1. Derivación de la técnica de Análisis de Componentes Principales

A continuación se presenta la descripción y la derivación de la técnica de Análisis de Componentes Principales; como se dijo inicialmente, el PCA es una de las técnicas más usadas para la reducción de dimensionalidad y manejo de grandes volúmenes de datos. Su objetivo principal es reducir la dimensionalidad de un conjunto de datos originales, en el que se encuentran un gran número de variables correlacionadas. Esta reducción se logra transformando el conjunto de datos originales en un nuevo conjunto de variables no correlacionadas, es decir, en un conjunto en el cual no haya redundancia y mucho menos datos duplicados o que no aporten mucha información, a este conjunto se denomina componentes principales.

Las componentes principales son aquellas que no están correlacionadas, es decir, no existe redundancia en la información, estos componentes están ordenados, ya que las primeras componentes abarcan la mayor parte de la variabilidad de los datos presentes en todas las variables originales. El cálculo de estos componentes principales se reduce a la solución de un problema de autovectores y autovalores para una matriz conocida, la matriz de covarianza, la cual es simétrica y semidefinida positiva. Esto se verá con detalle más adelante.

En sí, la técnica de Análisis de Componentes Principales permite determinar proyecciones \tilde{x}_i o llamados, representaciones o aproximaciones de los datos x_i que sean los más similares posibles a

los datos originales pero en un espacio de menor dimensión.

Notemos que en el estudio de las proyecciones sobre subespacios de dimensión menor se consideraron vectores de \mathbb{R}^n , pero en el estudio del Análisis de Componentes Principales debemos tener en cuenta que cuando se habla de un conjunto X de datos, este conjunto en realidad será una matriz de datos que pertenece a $\mathbb{R}^{m \times n}$ con m y n números reales. Entonces, esta técnica del PCA consiste en el tratamiento de una matriz de datos $(X, \cdot, +)$ como si fuera un espacio vectorial euclídeo que tiene n variables o características. Este conjunto de n variables genera un espacio vectorial, por lo tanto, tienen validez las operaciones de suma y multiplicación por escalares y sus propiedades definidas en el espacio vectorial.

Supongamos que se tiene un conjunto X de datos el cual contiene N vectores x_1, x_2, \dots, x_N , este conjunto X de N vectores es una matriz de datos, es decir, se tiene una muestra de N individuos u objetos de estudio, donde cada x_i con $i = 1, 2, \dots, N$ posee n variables o características, es decir, cada x_i está en un espacio n -dimensional, las cuales representan las variables originales, este conjunto lo podemos visualizar mediante una tabla o matriz de datos de la siguiente manera:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nN} \end{bmatrix}$$

en este caso las filas representan las n características y las columnas representan los objetos de estudio o las observaciones. Para reducir la dimensionalidad del conjunto X de datos se debe proyectar las n variables originales sobre un subespacio q -dimensional, donde $q < n$. Lo que se desea es buscar una representación \tilde{X} que sea lo más parecido o similar posible al conjunto X de datos originales, pero en un subespacio de dimensión menor.

Entonces, consideremos $B_n = \{v_1, v_2, \dots, v_n\}$ como una base ortonormal del espacio \mathbb{R}^n donde cada par de vectores distintos son ortogonales, es decir $v_i^T v_j = 0$ si y solo si $i \neq j$ y cada vector es de norma uno, es decir, $v_i^T v_j = 1$ si y solo si $i = j$. Por definición, cualquier vector x_i , $i = 1, 2, \dots, N$ del conjunto X de datos originales, se puede escribir como una combinación lineal de los vectores de la base, en este caso de la base ortonormal B_n y sea α_i un vector de escalares definido de la siguiente forma $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$, entonces, la combinación lineal de cada dato original x_i de

\mathbb{R}^n viene dada tal como se presenta a continuación:

$$\begin{aligned}x_1 &= \alpha_{11}v_1 + \alpha_{12}v_2 + \cdots + \alpha_{1n}v_n. \\x_2 &= \alpha_{21}v_1 + \alpha_{22}v_2 + \cdots + \alpha_{2n}v_n. \\&\vdots \\x_N &= \alpha_{N1}v_1 + \alpha_{N2}v_2 + \cdots + \alpha_{Nn}v_n.\end{aligned}$$

Para facilitar la notación de la expresión anterior, podemos expresar la combinación lineal para cada x_i , $i = 1, 2, \dots, N$ en términos de sumatoria, la cual viene dada como sigue:

$$x_i = \sum_{j=1}^n \alpha_{ij}v_j = \alpha_{i1}v_1 + \alpha_{i2}v_2 + \cdots + \alpha_{in}v_n, \quad (4.1)$$

La proyección de un dato x_i del conjunto X de datos originales sobre el subespacio unidimensional S generado por un vector v_j , $j = 1, 2, \dots, n$ cualquiera de la base ortonormal B_n , viene expresada tal como se presenta a continuación:

$$Proy_S(x_i) = \alpha_{ij}v_j = v_j\alpha_{ij}. \quad (4.2)$$

Como se está realizando proyección ortogonal, entonces el vector $x_i - Proj_S(x_i)$ será ortogonal a v_j , es decir, $v_j^T(x_i - Proj_S(x_i)) = 0$ reemplazando la expresión 4.2 obtendremos lo siguiente:

$$v_j^T(x_i - \alpha_{ij}v_j) = 0 \implies v_j^T x_i - \alpha_{ij}v_j^T v_j = 0.$$

Como los vectores v_1, v_2, \dots, v_n son ortonormales, entonces $v_i^T v_j = 1$ con $i = j$, es decir, cada vector v_i tiene norma uno, y si determinamos la coordenada α_{ij} de la expresión anterior obtenemos que:

$$\alpha_{ij} = v_j^T x_i = x_i^T v_j. \quad (4.3)$$

Así, α_{ij} con $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, n$ representa o es una aproximación de x_i sobre el espacio unidimensional generado por el vector director v_j .

Ahora, si proyectamos el vector x_i del conjunto de datos originales sobre el subespacio S de dimensión dos, los vectores v_j, v_k de la base ortonormal B_n son los que formarán la base del subespacio S . Entonces la proyección $Proj_S(x_i)$ será una combinación lineal de los vectores v_j y v_k las cuales generan el espacio bidimensional S con escalares α_{ij} y α_{ik} .

$$Proj_S(x_i) = \alpha_{ij}v_j + \alpha_{ik}v_k \quad (4.4)$$

Sabiendo que la proyección ortogonal es aquella que nos dará la menor distancia entre el vector x_i y su proyección, entonces el vector $x_i - Proj_S(x_i)$ será ortogonal al subespacio bidimensional S , esto significa que será ortogonal al vector v_j y a v_k la cual se caracteriza como:

$$(x_i - Proj_S(x_i))^T v_j = 0 \quad \wedge \quad (x_i - Proj_S(x_i))^T v_k = 0.$$

Reemplazando la ecuación 4.4 en la expresión determinada anteriormente obtenemos lo siguiente:

$$(x_i - \alpha_{ij}v_j - \alpha_{ik}v_k)^T v_j = 0 \quad \wedge \quad (x_i - \alpha_{ij}v_j - \alpha_{ik}v_k)^T v_k = 0.$$

mediante la combinación en la expresión anterior obtendremos que $(\alpha_{ij}, \alpha_{ik}) = [v_j, v_k]^T x_i$, la cual es una representación del dato x_i en un espacio de dimensión dos. Ahora, si tomamos tres vectores cualesquiera v_j, v_k y v_h de la base ortonormal B_n como base del subespacio S , entonces obtendremos una representación o aproximación del dato x_i definida como $(\alpha_{ij}, \alpha_{ik}, \alpha_{ih}) = [v_j, v_k, v_h]^T x_i$, donde la proyección será de la forma:

$$Proy_S(x_i) = \alpha_{ij}v_j + \alpha_{ik}v_k + \alpha_{ih}v_h. \quad (4.5)$$

De manera general, si consideramos $q < n$ vectores cualesquiera de la base ortonormal B_n , como vectores de la base del subespacio S de menor dimensión, entonces podemos representarla en términos matriciales, donde esta matriz denotada por B tendrá como columna los q vectores ortonormales, cada vector columna es de n -dimensional, así que la matriz B será de $n \times q$, donde los q vectores generan un subespacio q -dimensional S . Lo que se desea ahora es proyectar la data x_i sobre el subespacio S , entonces la proyección será $Proy_S(x_i) = B\alpha_i = BB^T x_i$ y es un vector n -dimensional que se encuentra en el subespacio S de menor dimensión y el vector de escalares $\alpha_i = B^T x_i$ pertenece a \mathbb{R}^q la cual es una representación de la proyección del dato x_i en un espacio de menor dimensión.

Las proyecciones son aproximaciones o representaciones de los datos originales x_i en un subespacio S de menor dimensión. Estas proyecciones las podemos representar mediante un vector de coordenadas llamado **code**.

$$\alpha_i = \begin{pmatrix} \alpha_{i1} & \alpha_{i2} & \dots & \alpha_{iq} \end{pmatrix} = B^T x_i \in \mathbb{R}^q. \quad (4.6)$$

Para aclarar la idea del **code**, veremos mediante un ejemplo, de como representar la proyección mediante el vector de coordenadas o llamadas **code**, para ello consideremos \mathbb{R}^3 con su base muy conocida que es la base canónica B y \mathbb{R}^2 con su respectiva base canónica. Tomemos como vector $v = (1, 2, 3)^T$ de \mathbb{R}^3 y asumimos que la proyección de v sobre el subespacio S de dimensión dos es $Proy_S(v) = (2, 3)^T$. Entonces podemos escribir la proyección como combinación lineal de los vectores de la base de \mathbb{R}^2 .

$$Proy_S(v) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Donde $\alpha = (2, 3)$ es un vector de coordenadas llamadas code, la cual representa el vector $Proy_S(v)$ sobre el subespacio de dimensión dos.

Mediante esta idea, para representar los datos x_i en un espacio de menor dimensión es suficiente determinar, recordar o almacenar el vector de coordenadas α_i con respecto a la base canónica de un espacio de menor dimensión al cual se está proyectando el vector x_i . Este vector de coordenadas $\alpha_i = B^T x_i$ se le conoce como code.

Sabemos que la proyección $Proy_S(x_i)$ es un vector n -dimensional del subespacio S , es decir, $Proy_S(x_i)$ es un vector que tiene n componentes que se encuentra en el subespacio S de dimensión menor. Por ejemplo, los vectores \tilde{x}_i que pertenecen al subespacio principal S podrían ser vectores de \mathbb{R}^3 en un plano, es decir, la dimensión del plano es dos, pero los vectores todavía tienen tres componentes o coordenadas con respecto a la base estándar de \mathbb{R}^3 . Y la segunda suma es un vector que pertenece al complemento ortogonal de S .

Entonces, el vector proyección por definición se puede expresar como combinación lineal de los vectores de la base ortonormal B_n de \mathbb{R}^n , esto es, existe escalares $\lambda_1, \lambda_2, \dots, \lambda_n$ tal que:

$$Proy_S(x_i) = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n. \quad (4.7)$$

Ahora, podemos seleccionar q vectores de la base ortonormal B_n con $q < n$ y definir la matriz B cuyas columnas son los q vectores seleccionados, estos vectores v_1, v_2, \dots, v_q forman una base B_q y generan un subespacio q -dimensional S . Entonces el vector $Proy_S(x_i)$ de \mathbb{R}^n se puede expresar como combinación lineal de los vectores $v_1, \dots, v_q, v_{q+1}, \dots, v_n$ con escalares $\lambda_1, \lambda_2, \dots, \lambda_q, \lambda_{q+1}, \dots, \lambda_n$ de la siguiente forma:

$$Proy_S(x_i) = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_q v_q + \lambda_{q+1} v_{q+1} + \dots + \lambda_n v_n. \quad (4.8)$$

Lo que buscamos es un subespacio $S \subset \mathbb{R}^n$ de $dim(S) = q$ con $q < n$, donde se encuentran las representaciones del conjunto de datos originales X . Entonces, mediante el concepto de la descomposición ortogonal (2.21), cada vector del espacio n -dimensional se puede escribir de forma única como suma de vectores como sigue:

$$\begin{aligned} Proj_S(x_i) &= \underbrace{\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_q v_q}_{\tilde{x}_i} + \underbrace{\lambda_{q+1} v_{q+1} + \dots + \lambda_n v_n}_{\text{complemento ortogonal}} \\ &= \sum_{k=1}^q \lambda_k v_k + \sum_{k=q+1}^n \lambda_k v_k \\ &= \tilde{x}_i + \sum_{k=q+1}^n \lambda_k v_k. \end{aligned}$$

Como podemos notar, la primera suma \tilde{x}_i es el vector proyección de x_i sobre el subespacio S de dimensión q generado por los vectores v_1, v_2, \dots, v_q de B_q con $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_q)$ la cual se denota

como $\tilde{x}_i = B\lambda$, esto representa la aproximación a los datos x_i en el subespacio q -dimensional S . Notemos que la matriz B tiene como vectores columna todos los vectores de la base B_q .

$$\tilde{x}_i = \sum_{k=1}^q \lambda_k v_k = B\lambda. \quad (4.9)$$

Dentro del Análisis de Componentes Principales se propone descartar o eliminar la segunda suma, es decir, desecha el vector del complemento ortogonal de S , y toma como subespacio principal al subespacio generado por los vectores ortonormales seleccionados v_1, v_2, \dots, v_q de B_n .

Asumiendo que el vector \tilde{x}_i representa o es una aproximación a la data x_i en el subespacio S de menor dimensión. Entonces, usaremos exactamente este tipo de representación de \tilde{x} para encontrar las coordenadas óptimas λ y los vectores de la base v_1, v_2, \dots, v_q tal que \tilde{x} sea tan similar como sea posible al punto de dato original x , para lo cual, nuestro objetivo será minimizar la distancia euclidiana $\|x - \tilde{x}\|$.

Para facilitar la notación, consideramos que el conjunto X de datos tiene media o la esperanza cero, al cual se desea representar en el subespacio principal S de q -dimensional, entonces debemos determinar el vector de coordenadas λ y los vectores v_k con $k = 1, 2, \dots, q$ de la base ortonormal B_q tal que al representar los datos en el subespacio q -dimensional, el promedio de la sumatoria de cada error sea mínima, a esto se le conoce como el error cuadrático de reconstrucción promedio y lo denotamos como E . Para medir la similitud tomamos el cuadrado de la distancia euclídea $\|x_i - \tilde{x}_i\|^2$ con el propósito de minimizar el error cuadrático de reconstrucción promedio. Recordemos que la distancia euclídea al cuadrado es igual a su producto interno, es decir:

$$\|x_i - \tilde{x}_i\|^2 = (x_i - \tilde{x}_i)^T (x_i - \tilde{x}_i).$$

Entonces, el problema consiste en minimizar el error cuadrático de reconstrucción promedio definida tal como se presenta a continuación:

$$E = \frac{1}{N} \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2. \quad (4.10)$$

El objetivo es determinar una aproximación $\tilde{x}_i = B\lambda_i$ con $i = 1, 2, \dots, N$, la cual sea similar al dato original x_i , esto considerando que se dispone de una base ortonormal B_n en el espacio de dimensión n donde se encuentran los datos.

Para determinar una aproximación $\tilde{x}_i = B\lambda_i$ similar a la data x_i , debemos hallar el conjunto de escalares λ_i y el conjunto de vectores v_1, v_2, \dots, v_q que sean óptimos para obtener la representación de x_i , por lo que se trata de un problema de optimización multivariable.

Para encontrar las coordenadas λ_i y la base ortonormal del subespacio principal, lo primero que debemos hacer es optimizar las coordenadas λ_i para una base B_q de vectores ortonormales dado (v_1, v_2, \dots, v_q) , luego determinar los vectores (v_1, v_2, \dots, v_q) óptimos de la base ortonormal, las cuales conformarán el subespacio principal S .

Determinación de coordenadas óptimas

Entonces, empecemos determinando las coordenadas óptimas $\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{iq}$, de las proyecciones \tilde{x}_i para $i = 1, 2, \dots, N$. Determinar el vector de coordenadas o llamadas code significa encontrar la representación de la proyección del dato x_i sobre el subespacio S con respecto a los vectores de la base B_q , la cual minimice la distancia entre $x_i - \tilde{x}_i$, y sabemos que si la distancia entre el dato x_i y su proyección es mínima significa que esto será la proyección ortogonal, es decir, el producto interno del vector $x_i - \tilde{x}_i$ con cada vector v_k de B_q será igual a cero.

$$\langle v_k, x_i - \tilde{x}_i \rangle = 0 \implies v_k^T (x_i - \tilde{x}_i) = 0,$$

donde $k = 1, 2, \dots, q$. Para un dato $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ de la matriz de datos X , el conjunto de escalares óptimos para \tilde{x}_i es cuando se anulen las derivadas parciales del error cuadrático de reconstrucción promedio E con respecto a cada coordenada en $\lambda_m = (\lambda_{1i}, \lambda_{2i}, \dots, \lambda_{qi})$, las cuales son funciones de \tilde{x}_i , por lo que requerimos las derivadas parciales en la cual aplicaremos la regla de la cadena para obtener la solución.

$$\frac{\partial E}{\partial \lambda_{ki}} = \frac{\partial E}{\partial \tilde{x}_i} \cdot \frac{\partial \tilde{x}_i}{\partial \lambda_{ki}}. \quad (4.11)$$

Entonces, calculemos la primera derivada parcial, la cual es:

$$\frac{\partial E}{\partial \tilde{x}_i} = \frac{\partial}{\partial \tilde{x}_i} \left(\frac{1}{N} \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2 \right). \quad (4.12)$$

A partir del error cuadrático de reconstrucción promedio definida en la ecuación (4.10) para un dato particular x_i , el término $\|x_i - \tilde{x}_i\|^2$ de la sumatoria se expresa de la siguiente manera:

$$\begin{aligned} \|x_i - \tilde{x}_i\|^2 &= (x_i - \tilde{x}_i)^T (x_i - \tilde{x}_i). \\ &= \begin{bmatrix} (x_{i1} - \tilde{x}_{i1}) & (x_{i2} - \tilde{x}_{i2}) & \dots & (x_{in} - \tilde{x}_{in}) \end{bmatrix}^T \begin{bmatrix} (x_{i1} - \tilde{x}_{i1}) \\ (x_{i2} - \tilde{x}_{i2}) \\ \vdots \\ (x_{in} - \tilde{x}_{in}) \end{bmatrix} \\ &= (x_{i1} - \tilde{x}_{i1})^2 + (x_{i2} - \tilde{x}_{i2})^2 + \dots + (x_{in} - \tilde{x}_{in})^2. \end{aligned}$$

Entonces, el error cuadrático para un dato particular x_i queda expresado como:

$$E = \frac{1}{N} \left[(x_{i1} - \tilde{x}_{i1})^2 + (x_{i2} - \tilde{x}_{i2})^2 + \dots + (x_{in} - \tilde{x}_{in})^2 \right],$$

ahora, si realizamos la derivada parcial de E con respecto a \tilde{x}_i de la expresión anterior, obtendremos como resultado lo siguiente:

$$\begin{aligned}\frac{\partial E}{\partial x_i} &= \frac{1}{N} \left[-2(x_{i1} - \tilde{x}_{i1}) - 2(x_{i2} - \tilde{x}_{i2}) - \cdots - 2(x_{in} - \tilde{x}_{in}) \right]. \\ &= \frac{-2}{N} \left[(x_{i1} - \tilde{x}_{i1}) + (x_{i2} - \tilde{x}_{i2}) + \cdots + (x_{in} - \tilde{x}_{in}) \right],\end{aligned}$$

la derivada parcial de un dato particular x_i que tiene n variables o características recoge como vector fila las derivadas parciales de cada x_{ij} con, $i = 1, 2, \dots, N$, y $j = 1, 2, \dots, n$ entonces:

$$\begin{aligned}\frac{\partial E}{\partial x_i} &= -\frac{2}{N} \left[(x_{i1} - \tilde{x}_{i1}) \quad (x_{i2} - \tilde{x}_{i2}) \cdots (x_{in} - \tilde{x}_{in}) \right]. \\ &= -\frac{2}{N} (x_i - \tilde{x}_i)^T \in \mathbb{R}^{1 \times n}.\end{aligned}\tag{4.13}$$

El vector fila definido en la ecuación (4.13) también se conoce como el gradiente o el jacobiano.

Ahora para obtener $\partial \tilde{x}_i / \partial \lambda_{ki}$ veamos como está definida \tilde{x}_i en la ecuación (4.9). Entonces

$$\begin{aligned}\frac{\partial \tilde{x}_i}{\partial \lambda_{ki}} &= \frac{\partial}{\partial \lambda_{ki}} \left(\sum_{m=1}^q \lambda_{mi} v_m \right). \\ &= \frac{\partial}{\partial \lambda_{ki}} (\lambda_{1i} v_1 + \lambda_{2i} v_2 + \cdots + \lambda_{ki} v_k + \lambda_{(k+1)i} v_{k+1} + \cdots + \lambda_{qi} v_q). \\ &= v_k \quad \text{para } k = 1, 2, \dots, q.\end{aligned}\tag{4.14}$$

Así obtenemos la derivada parcial con respecto a λ_{ki} de la ecuación (4.11) y para obtener la coordenada óptima lo que debemos hacer es establecer esta derivada parcial en cero, esto hace que inmediatamente produzca la coordenada óptima, esto se caracteriza de la siguiente manera:

$$\frac{\partial E}{\partial \lambda_{ki}} = \frac{\partial E}{\partial x_i} \cdot \frac{\partial \tilde{x}_i}{\partial \lambda_{ki}} = -\frac{2}{N} (x_i - \tilde{x}_i)^T v_k = 0,$$

aplicando la propiedad de la traspuesta y la propiedad distributiva obtenemos lo siguiente:

$$\begin{aligned}-\frac{2}{N} (x_i - \tilde{x}_i)^T v_k &= -\frac{2}{N} (x_i^T v_k - \tilde{x}_i^T v_k) = 0. \\ &\implies x_i^T v_k - \tilde{x}_i^T v_k = 0,\end{aligned}$$

mediante el uso de la expresión \tilde{x}_i definida en (4.9) y reemplazando en la expresión anterior se tiene que:

$$\begin{aligned}\implies x_i^T v_k - \tilde{x}_i^T v_k = 0 &\implies x_i^T v_k - \left(\sum_{m=1}^q \lambda_{mi} v_m \right)^T v_k = 0, \\ &\implies x_i^T v_k - \lambda_{ki} v_k^T v_k = 0,\end{aligned}$$

como los vectores v_1, v_2, \dots, v_q son ortonormales, entonces $v_k^T v_k = 1$, así $x_i^T v_k - \lambda_{ki} = 0$, de esta forma obtenemos la coordenada óptima λ_{ki} .

$$\lambda_{ki} = x_i^T v_k, \quad (4.15)$$

donde $k = 1, 2, \dots, q$ y para $i = 1, 2, \dots, N$. Como podemos notar que la coordenada óptima de λ_{ki} coincide con la proyección ortogonal del vector x_i del conjunto de datos originales, sobre el subespacio S de dimensión uno generado por el vector v_k .

Entonces, podemos generalizar esta idea de la siguiente manera: Para cada x_i con $i = 1, 2, \dots, N$ del conjunto de datos originales podemos obtener el valor óptimo λ_{ki} donde $k = 1, 2, \dots, q$. Entonces, la proyección óptima \tilde{x}_i del elemento x_i del conjunto de datos originales X , es la proyección ortogonal y las coordenadas de \tilde{x}_i con respecto a los vectores v_1, v_2, \dots, v_q de la base B_q , son las coordenadas de la proyección ortogonal de x_i sobre el subespacio principal S .

Recordemos que si se tiene una base ortonormal de \mathbb{R}^n entonces la proyección es:

$$\tilde{x} = v_k (v_k^T v_k)^{-1} v_k^T x \implies \tilde{x} = v_k v_k^T x \in \mathbb{R}^n, \quad (4.16)$$

donde $v_k^T x = \lambda_k$ es la coordenada de la proyección ortogonal de x sobre el subespacio definido por v_k . Entonces \tilde{x} es la proyección ortogonal de x sobre el subespacio generado por el vector k -ésimo de la base B_q , y $\lambda_k = v_k^T x$ es la coordenada de esta proyección con respecto al vector de la base v_k . De manera general, si se desea proyectar sobre un subespacio q -dimensional de \mathbb{R}^n , obtenemos la proyección ortogonal de x sobre el subespacio q -dimensional con vectores ortonormales v_1, v_2, \dots, v_q de la base B_q .

$$\tilde{x} = B(B^T B)^{-1} B^T x \implies \tilde{x} = B B^T x. \quad (4.17)$$

En este caso, la coordenada de esta proyección \tilde{x}_i con respecto a la base ortonormal B_q viene dada como $B^T x = \lambda$. Notemos que el vector de coordenadas λ es una representación del vector proyección \tilde{x} . Aunque $\tilde{x} \in \mathbb{R}^n$, es suficiente con determinar las coordenadas $\lambda_1, \lambda_2, \dots, \lambda_q$ para poder representar el vector de proyección \tilde{x} , la cual es similar al dato x_i .

Ahora, las coordenadas que se encuentran en el complemento ortogonal S^\perp con respecto a la base $v_{q+1}, v_{q+2}, \dots, v_n$ son siempre cero. De esta forma hemos demostrado que para una base B_q ortonormal dado, podemos determinar las coordenadas óptimas de \tilde{x} mediante la proyección ortogonal sobre el subespacio principal S .

A continuación pasaremos a demostrar como determinar la mejor base para el subespacio de dimensión menor sobre el cual se encuentran los datos aproximados, es decir, determinaremos los vectores v_1, v_2, \dots, v_q que definen el subespacio principal S .

Determinación de la base de subespacio principal S .

Recordemos que en la ecuación (4.15) se mostró la forma de obtener el valor óptimo λ_{ki} para un determinado vector x_i del conjunto de datos originales X , la cual se calcula como $\lambda_{ki} = x_i^T v_k$ con $k = 1, 2, \dots, q$, y además la proyección \tilde{x}_i la podemos escribir como una combinación lineal de la siguiente manera: $\tilde{x}_i = \lambda_{1i}v_1 + \lambda_{2i}v_2 + \dots + \lambda_{qi}v_q$. Por lo tanto, para determinar los vectores de la base v_1, v_2, \dots, v_q que definen el subespacio principal S , debemos reformular la ecuación (4.10) del error cuadrático de reconstrucción promedio mediante los resultados obtenidos anteriormente, ya que esto hará que sea más fácil determinar los vectores de la base B_q . Ahora, recordemos que:

$$\tilde{x}_i = \sum_{k=1}^q \lambda_{ki}v_k = \sum_{k=1}^q (x_i^T v_k)v_k,$$

si aplicando la propiedad de la simetría del producto escalar a la expresión anterior, obtenemos lo siguiente:

$$\tilde{x}_i = \sum_{k=1}^q (v_k^T x_i)v_k = \sum_{k=1}^q v_k(v_k^T x_i) = \left(\sum_{k=1}^q v_k v_k^T \right) x_i.$$

Notemos en esta parte que el escalar λ_{ki} está dada como $\lambda_{ki} = x_i^T v_k$ y la matriz de proyección P viene representada como:

$$P = \sum_{k=1}^q v_k v_k^T = BB^T. \quad (4.18)$$

Esta matriz de proyección $P = BB^T$ es una matriz cuadrada simétrica de rango q ya que se obtiene a partir de la suma de matrices $v_k v_k^T$ de rango uno que está formada por el producto escalar del k -ésimo vector columna ortonormal del subespacio S .

De la misma forma, el elemento x_i del conjunto X de datos originales se expresa como una combinación lineal de los vectores v_1, v_2, \dots, v_n de la base B_n de la siguiente forma:

$$x_i = \alpha_{1i}v_1 + \alpha_{2i}v_2 + \dots + \alpha_{ni}v_n.$$

En términos de sumatoria se expresa de la siguiente manera:

$$x_i = \sum_{k=1}^n \alpha_{ki}v_k = \sum_{k=1}^n (v_k^T x_i)v_k. \quad (4.19)$$

Esta expresión la podemos representar como la suma de dos sumatorias y aplicando la propiedad de la simetría del producto escalar obtenemos lo siguiente:

$$\begin{aligned} x_i &= \sum_{k=1}^n (v_k^T x_i)v_k = \sum_{k=1}^q (v_k^T x_i)v_k + \sum_{k=q+1}^n (v_k^T x_i)v_k. \\ &= \left(\sum_{k=1}^q v_k v_k^T \right) x_i = \left(\sum_{k=1}^q v_k v_k^T \right) x_i + \left(\sum_{k=q+1}^n v_k v_k^T \right) x_i. \end{aligned}$$

La primera sumatoria representa la proyección \tilde{x}_i en el subespacio principal S y la segunda sumatoria representa el vector en el complemento ortogonal S^\perp . Entonces el término x_i queda expresado como sigue:

$$x_i = \left(\sum_{k=1}^n v_k v_k^T \right) x_i = \tilde{x}_i + \left(\sum_{k=q+1}^n v_k v_k^T \right) x_i.$$

Ahora, mediante estos resultados podemos determinar la distancia que existe entre el dato original x_i y su proyección \tilde{x}_i .

$$\begin{aligned} x_i - \tilde{x}_i &= \tilde{x}_i + \left(\sum_{k=q+1}^n v_k v_k^T \right) x_i - \tilde{x}_i \\ &= \left(\sum_{k=q+1}^n v_k v_k^T \right) x_i = \sum_{k=q+1}^n (v_k^T x_i) v_k. \end{aligned} \quad (4.20)$$

Como podemos notar que la diferencia entre el dato original x_i y su proyección \tilde{x}_i es exactamente la proyección del dato original x_i sobre S^\perp , por lo tanto, el vector $x_i - \tilde{x}_i$ se encuentra en el subespacio S^\perp que es ortogonal al subespacio principal S . Y además la matriz de proyección P de la expresión (4.20) está representada como:

$$P = \sum_{k=q+1}^n v_k v_k^T. \quad (4.21)$$

Tomando la norma al cuadrado a ambos lados de la igualdad de la expresión (4.20) determinada anteriormente, obtenemos lo siguiente:

$$\|x_i - \tilde{x}_i\|^2 = \left\| \sum_{k=q+1}^n (v_k^T x_i) v_k \right\|^2. \quad (4.22)$$

La medida de similitud que se usa es la norma ℓ_2 o norma euclídea comúnmente conocida. Por lo tanto, dentro de la medida establecida, que es la distancia euclídea al cuadrado, el error al aproximar el dato original x_i por su proyección \tilde{x}_i , es igual a la aproximación que ocurriría si usáramos el complemento ortogonal del subespacio principal S , y este hecho se puede usarse de manera equivalente para encontrar la solución óptima que deseamos ver.

Mediante estos resultados vamos a reformular la expresión E del error cuadrático de reconstrucción promedio de la ecuación (4.10). Entonces, reemplazando la expresión (4.22) en la ecuación del error cuadrático $E = \frac{1}{N} \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2$ obtenemos lo siguiente:

$$E = \frac{1}{N} \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2 = \frac{1}{N} \sum_{i=1}^N \left\| \sum_{k=q+1}^n (v_k^T x_i) v_k \right\|^2. \quad (4.23)$$

Ahora calculemos de manera explícita la norma al cuadrado y usamos el hecho de que los vectores v_k forman una base ortonormal. Como la norma euclídea al cuadrado es el producto escalar del vector consigo mismo, entonces produce lo siguiente:

$$\begin{aligned}
 E &= \frac{1}{N} \sum_{i=1}^N \left\| \sum_{k=q+1}^n (v_k^T x_i) v_k \right\|^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n (v_k^T x_i) v_k \right)^T \left(\sum_{k=q+1}^n (v_k^T x_i) v_k \right). \tag{4.24}
 \end{aligned}$$

Ahora, notemos que en esta última expresión el término $x_i^T v_k = \gamma_i$ es un escalar para $k = q+1, q+2, \dots, n$ y los vectores del subespacio del complemento ortogonal son de la forma $v_{q+1}, v_{q+2}, \dots, v_n$, además estos vectores son vectores ortonormales, es decir $v_k^T v_j = 0$ si y solo si $k \neq j$ y $v_k^T v_j = 1$ si y solo si $k = j$ para todo $k, j = q+1, q+2, \dots, n$. Entonces $\sum_{k=q+1}^n (v_k^T x_i) v_k$ es una combinación lineal de la siguiente forma:

$$\sum_{k=q+1}^n (v_k^T x_i) v_k = \gamma_{q+1} v_{q+1} + \gamma_{q+2} v_{q+2} + \dots + \gamma_n v_n. \tag{4.25}$$

Entonces, reemplazando la ecuación anterior en la expresión (4.24) obtenemos lo siguiente:

$$\begin{aligned}
 E &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n (v_k^T x_i) v_k \right)^T \left(\sum_{k=q+1}^n (v_k^T x_i) v_k \right) \\
 &= \frac{1}{N} \sum_{i=1}^N (\gamma_{q+1} v_{q+1} + \gamma_{q+2} v_{q+2} + \dots + \gamma_n v_n)^T (\gamma_{q+1} v_{q+1} + \gamma_{q+2} v_{q+2} + \dots + \gamma_n v_n) \\
 &= \frac{1}{N} \sum_{i=1}^N ((\gamma_{q+1} v_{q+1})^2 + (\gamma_{q+2} v_{q+2})^2 + \dots + (\gamma_n v_n)^2),
 \end{aligned}$$

como los vectores son ortonormales, entonces $v_k^T v_k = 1$ con $k = q+1, q+2, \dots, n$, por lo que la expresión anterior se reduce al siguiente cálculo:

$$E = \frac{1}{N} \sum_{i=1}^N ((\gamma_{q+1})^2 + (\gamma_{q+2})^2 + \dots + (\gamma_n)^2) = \frac{1}{N} \sum_{i=1}^N \sum_{k=q+1}^n \gamma_i^2,$$

pero sabemos que $\gamma_i = x_i^T v_k$, entonces reemplazando en la expresión anterior se obtiene la siguiente expresión:

$$E = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n (v_k^T x_i)^2 \right) = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n (v_k^T x_i)(v_k^T x_i) \right), \tag{4.26}$$

ahora, lo que aplicamos es la propiedad de la simetría del producto interno a la expresión anterior, de la cual obtenemos el siguiente resultado:

$$E = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n (v_k^T x_i)(x_i^T v_k) \right) = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n v_k^T x_i x_i^T v_k \right), \quad (4.27)$$

como el vector x_i no depende de n y el vector v_k no depende de i entonces podemos intercambiar las sumatorias, esto se caracteriza como sigue:

$$E = \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=q+1}^n v_k^T x_i x_i^T v_k \right) = \sum_{k=q+1}^n v_k^T \left(\frac{1}{N} \sum_{i=1}^N x_i x_i^T \right) v_k. \quad (4.28)$$

En esta última parte el término $\frac{1}{N} \sum_{i=1}^N x_i x_i^T$ es la matriz de covarianza de la ecuación (2.55), donde los datos están centrados alrededor de la media, es decir, la esperanza o el valor esperado de la variable aleatoria es $E(x) = 0$. Por lo tanto, la expresión del error cuadrático de reconstrucción promedio se reduce al siguiente cálculo:

$$E = \sum_{k=q+1}^n v_k^T \sigma v_k. \quad (4.29)$$

Con este resultado del error cuadrático de reconstrucción promedio E , el problema se transforma en determinar los vectores ortonormales $v_{q+1}, v_{q+2}, \dots, v_n$ del complemento ortogonal S^\perp del subespacio principal S tal que el error cuadrático se minimice, es decir:

$$\text{minimizar } E = \sum_{k=q+1}^n v_k^T \sigma v_k = v_{q+1}^T \sigma v_{q+1} + v_{q+2}^T \sigma v_{q+2} + \dots + v_n^T \sigma v_n. \quad (4.30)$$

Este problema lo podemos resolver mediante el uso del método de los multiplicadores de Lagrange, ya que permite encontrar el máximo o mínimo de una función multivariable cuando presenta alguna restricción o cuando esté sujeta a cierta condición, el cálculo de los multiplicadores de Lagrange la podemos ver a detalle en (Thomas, 2005). En este caso la restricción es que los vectores por ser ortogonales tienen norma uno, a la cual se define como $F := v_k^T v_k = 1$ para $k = q+1, q+2, \dots, n$, entonces la función Lagrangiana queda expresado de la siguiente forma:

$$\mathcal{L}(x, \lambda) = E - \sum_{k=q+1}^n \lambda_k F. \quad (4.31)$$

Reemplazando en esta expresión los valores de E de la expresión (4.31) y la restricción F obtenemos lo siguiente:

$$\mathcal{L}(x, \lambda) = \sum_{k=q+1}^n v_k^T \sigma v_k - \sum_{k=q+1}^n \lambda_k (v_k^T v_k), \quad (4.32)$$

donde la variable λ_k se le conoce como multiplicador de Lagrange. Ahora, $v_k^T v_k = 1$ esto implica que $v_k^T v_k - 1 = 0$. Entonces podemos expresar la función Lagrangiana como:

$$\begin{aligned}\mathcal{L}(x, \lambda) &= v_{q+1}^T \sigma v_{q+1} + \cdots + v_n^T \sigma v_n - \lambda_{q+1}(v_{q+1}^T v_{q+1} - 1) - \cdots - \lambda_n(v_n^T v_n - 1). \\ &= v_{q+1}^T \sigma v_{q+1} + \cdots + v_n^T \sigma v_n + \lambda_{q+1}(1 - v_{q+1}^T v_{q+1}) + \cdots + \lambda_n(1 - v_n^T v_n).\end{aligned}$$

Para determinar la solución debemos realizar el gradiente de la función Lagrangiana $\nabla \mathcal{L}(x, \lambda)$ igual al vector cero, es decir $\nabla \mathcal{L}(x, \lambda) = 0$. Para lo cual se debe realizar la derivada parcial con respecto a cada multiplicador de Lagrange e igualar a cero, de la misma forma la derivada parcial con respecto a cada vector v_k con $k = q+1, q+2, \dots, n$ e igualar a cero, esto se caracteriza de la siguiente manera:

$$\frac{\partial \mathcal{L}}{\partial \lambda_k} = \frac{\partial}{\partial \lambda_k} (v_{q+1}^T \sigma v_{q+1} + \cdots + v_n^T \sigma v_n + \lambda_{q+1}(1 - v_{q+1}^T v_{q+1}) + \cdots + \lambda_n(1 - v_n^T v_n)) = 0.$$

Para $q+1 \leq k \leq n$ obteniendo como resultado

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \lambda_k} &= 0 + (1 - v_n^T v_n) = 0. \\ &= 1 - v_n^T v_n = 0.\end{aligned}\tag{4.33}$$

La expresión $1 - v_n^T v_n = 0$ si y solo si $v_k^T v_k = 1$. De la misma forma para $q+1 \leq k \leq n$ realizamos la derivada parcial de \mathcal{L} con respecto a v_k .

$$\frac{\partial \mathcal{L}}{\partial v_k} = \frac{\partial}{\partial v_k} (v_{q+1}^T \sigma v_{q+1} + \cdots + v_n^T \sigma v_n + \lambda_{q+1}(1 - v_{q+1}^T v_{q+1}) + \cdots + \lambda_n(1 - v_n^T v_n)) = 0.$$

Obteniendo como resultado la siguiente expresión:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial v_k} &= v_k^T \sigma + 0 + \lambda_k(-v_k^T) = 0. \\ &= v_k^T \sigma - \lambda_k v_k^T = 0,\end{aligned}\tag{4.34}$$

esta expresión se cumple si y solo si $v_k^T \sigma = \lambda_k v_k^T$, como la matriz de covarianza σ es una matriz simétrica y por la propiedad de producto interno obtenemos que:

$$\sigma v_k = \lambda_k v_k.\tag{4.35}$$

Notemos que esta expresión es similar a la ecuación (2.47), por lo que los vectores v_k con $k = q+1, q+2, \dots, n$ que optimizan el error cuadrático de reconstrucción promedio E son los autovectores de la matriz de covarianza σ . Ahora, si multiplicamos v_k^T por la derecha a la expresión (4.35) tenemos que:

$$\sigma v_k v_k^T = \lambda_k v_k v_k^T \implies \sigma v_k^T v_k = \lambda_k v_k^T v_k \implies \sigma = \lambda_k.$$

En la expresión anterior se aplicó las propiedades del producto interno y la ortonormalidad de vectores. Con este resultado podemos expresar el error cuadrático de reconstrucción promedio (4.30) como:

$$\begin{aligned}
 E &= v_{q+1}^T \sigma v_{q+1} + v_{q+2}^T \sigma v_{q+2} + \cdots + v_n^T \sigma v_n. \\
 &= v_{q+1}^T \lambda_{q+1} v_{q+1} + v_{q+2}^T \lambda_{q+2} v_{q+2} + \cdots + v_n^T \lambda_n v_n. \\
 &= \lambda_{q+1} v_{q+1}^T v_{q+1} + \lambda_{q+2} v_{q+2}^T v_{q+2} + \cdots + \lambda_n v_n^T v_n.
 \end{aligned}$$

Como los vectores son ortonormales, es decir, $v_i^T v_j = 0$ si y solo si $i \neq j$ y $v_i^T v_j = 1$ si y solo si $i = j$ entonces el error cuadrático de reconstrucción promedio viene dada como:

$$E = \lambda_{q+1} + \lambda_{q+2} + \cdots + \lambda_n. \quad (4.36)$$

Esto significa que el error viene dado por los autovalores de la matriz de covarianza que se encuentran en el complemento ortogonal, pero resulta que, lo que nos interesa son los autovalores asociados a los autovalores de mayor valor, pero si resolvemos el problema de los autovalores con menor valor, automáticamente obtenemos los de mayor valor, y esto sí nos sirve para representar nuestra solución. Por lo tanto, con determinar únicamente los autovalores $\lambda_{q+1}, \lambda_{q+2}, \dots, \lambda_n$ de la matriz de covarianza σ podemos obtener el valor mínimo del error cuadrático de reconstrucción promedio E .

Ahora, como la matriz σ es una matriz cuadrada, simétrica y definida positiva, tiene autovalores reales positivos y los autovectores $v_{q+1}, v_{q+2}, \dots, v_n$ de la matriz de covarianza son vectores que pertenecen al complemento ortogonal S^\perp del subespacio principal S . Entonces, los vectores que se necesitan en la base B_q del subespacio S son los q autovectores de la matriz de covarianza, los cuales corresponden a los q autovalores de mayor valor de la matriz de covarianza, estos q autovectores son los que formarán la base B , los cuales generan el subespacio principal S .

De manera general, la técnica de Análisis de Componentes Principales nos dice que para encontrar un subespacio principal de q -dimensional de \mathbb{R}^n que contenga la mayor cantidad de información posible, debemos elegir los q vectores columnas de la matriz B de la ecuación (4.9) como los q autovectores de la matriz de covarianza σ de datos, estos q autovectores se encuentran asociados con q autovalores de mayor valor.

Por ejemplo, si quisiéramos proyectar ya no un vector tridimensional, sino un conjunto de vectores de este espacio sobre un subespacio bidimensional, es decir, sobre un plano de \mathbb{R}^3 , y para definir el plano necesitamos dos vectores, estos dos vectores serán los dos autovectores de la matriz de

covarianza σ asociados a los autovalores de mayor módulo. Veamos la figura 1-4 donde se puede apreciar la data en un espacio tridimensional, donde el vector resaltado en rojo y el vector resaltado en azul son los autovectores de la matriz de covarianza σ que definen el subespacio principal S .

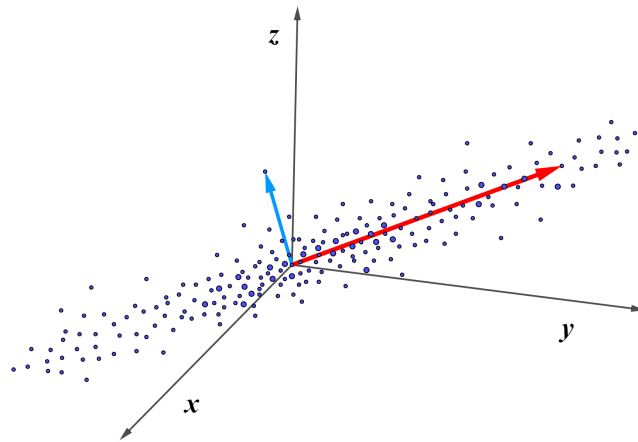


Ilustración 1-4: Representación del subespacio principal S .

Realizado por: Buñay, Cristian, 2022.

El autovector resaltado en la línea roja va en la dirección del mayor de los autovalores y representa la mejor base del subespacio principal S para la proyección, es el primer componente principal el cual representa la mayor cantidad posible de variabilidad en los datos. Ahora, el autovector resaltado en la línea de color azul asociado al autovalor de norma mínima es el segundo componente principal el cual contiene la mayor parte de la variabilidad restante del primer componente principal, este segundo componente será ortogonal al primer componente principal. La serie de datos tomados en este caso de \mathbb{R}^3 , queda representado en un plano de dimensión dos (2) que está definido por los dos autovectores principales asociados a los dos autovalores de mayor módulo, los cuales están representados en la línea marcada en roja y la línea marcada en azul de la figura 1-4, que son las componentes principales.

Cada componente subsiguiente explica la mayor parte de la variabilidad restante de los componentes principales asociados a los autovalores de mayor módulo, es decir, que el conjunto de datos de menor dimensión deberá preservar la varianza de los datos originales y todos los autovectores, además, todas las componentes principales serán ortonormales dos a dos.

4.2. Implementación del algoritmo básico de la técnica del PCA

Como resultado se origina una serie de pasos para obtener el algoritmo básico para la reducción de dimensionalidad, cuyo objetivo es hallar la proyección de los datos originales cuya dimensión

es muy grande sobre el subespacio principal S de dimensión menor.

1. Lo primero es considerar un conjunto de datos o que es lo mismo considerar una matriz de datos donde las filas representan las variables o características y las columnas como observaciones, individuos o cualquier objeto de estudio.

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

2. **Centrar:** Luego, empezaremos por centrar los datos mediante el cálculo de la media μ del conjunto de datos X y restándolo de cada punto x_i , $i = 1, 2, \dots, N$ de datos. Esto asegura que el valor medio del conjunto de datos X sea 0. Esto se realiza para simplificar los cálculos de la derivada y así reducir el riesgo de problemas numéricos.

$$X = x_i - \mu \text{ para } i = 1, 2, \dots, N. \quad (4.37)$$

3. **Estandarización:** Ahora, necesitamos estandarizar los datos originales x_i con $i = 1, 2, \dots, N$ mediante el uso de la media μ_i y la desviación estándar σ_i del conjunto de datos originales en la dimensión i -ésima, respectivamente, entonces x_i^* es la i -ésima componente de $x_i \in \mathbb{R}^n$ para ello dividimos los datos $x_i - \mu_i$ por la desviación estándar σ_i del conjunto de datos para cada dimensión $i = 1, 2, \dots, N$ de modo que:

$$x_i^* \leftarrow \frac{x_i - \mu_i}{\sigma_i} \text{ para } i = 1, 2, \dots, N. \quad (4.38)$$

donde x_i^* es el i -ésimo componente de x_i . Los datos tienen una varianza de 1 a lo largo de cada eje, que se indica mediante las dos flechas en la figura (1-4). Este paso completa la estandarización de los datos.

4. **Matriz de covarianza:** Luego, calculamos la matriz de covarianza de $x_i = \{x_1^*, x_2^*, \dots, x_n^*\}$, la cual es simétrica y definida positiva.

$$Cov(x_i) = \frac{1}{N} \sum_{i=1}^N x_i x_i^T. \quad (4.39)$$

En el caso de grandes dimensiones de datos, calcular los autovalores y los autovectores de la matriz de covarianza σ de dimensión $n \times n$ es muy complicado, ya que se escala cúbicamente en n , esto nos muestra que mientras más alta sea la dimensión resulta computacionalmente muy complicado determinar los autovalores y autovectores. Esto sucede en el caso de que el conjunto de datos es menor a la dimensionalidad de datos, es decir $N < n$.

Lo que se debe hacer entonces es convertir la matriz de covarianza de dimensión $n \times n$ en una nueva matriz de dimensión $N \times N$, esto permitirá calcular de manera más eficiente los autovalores y los autovectores. El cálculo para obtener esta nueva matriz de covarianza de $N \times N$ la podemos revisar de forma detallada en (Deisenroth, 2020).

5. **Identificación de la componente principal:** A partir de la obtención de la matriz de covarianza $Cov(X)$ determinaremos la matriz diagonal E de los autovalores y la matriz V cuyas columnas son los autovectores, las cuales están normalizadas, de modo que:

$$Cov(X)V = VE. \quad (4.40)$$

Dado que la matriz de covarianza es simétrica, mediante la descomposición espectral podemos encontrar una base ortonormal de vectores propios. Además, mediante el método de proyección calculamos los q autovectores asociados a los autovalores de mayor valor, esto en el caso de aplicaciones con grandes dimensiones de datos, pero en este caso se usará la descomposición espectral de la matriz.

6. **Subespacio de proyección:** Ahora, definimos el subespacio principal usando una matriz B cuyas columnas son los q vectores, los cuales son los q autovectores de la matriz de covarianza σ asociados a los q autovalores de mayor valor, que definen el subespacio.

Por ejemplo, si el subespacio lo definen los vectores v_1 y v_2 , entonces $B = [v_1, v_2]$. A partir de ello ya podemos calcular la matriz de proyección $P = BB^T$. mediante el reordenamiento de las columnas de V , los autovalores en E corresponden a los autovectores en las columnas de V .

7. **Proyección:** La proyección de los datos $x_i^* \in \mathbb{R}^n$ sobre el subespacio S está dada como:

$$\tilde{x}_i = Proj_S(x_i^*) = Px_i^* \text{ para } i = 1, 2, \dots, N. \quad (4.41)$$

donde P es la matriz de proyección definida como $P = BB^T$. Realizamos la siguiente observación en esta parte, B es la matriz que tiene como columna a los autovectores asociados a los autovalores más grandes de la matriz de covarianza $Cov(x)$.

Entonces la proyección de la data quedará expresada como:

$$Proj_S(x_i^*) = BB^T x_i^*. \quad (4.42)$$

Estos pasos conducen para generar el código básico para la reducción de dimensionalidad mediante esta técnica de Análisis de Componentes Principales (Deisenroth, 2020).

4.3. Ejemplo de aplicación de Análisis de Componentes Principales

A continuación se presentan dos ejemplos de aplicación de la Técnica de Análisis de Componentes Principales. En el primer ejemplo se realiza la reducción de dimensionalidad proyectando ortogonalmente un conjunto aleatorio de datos bidimensionales sobre un subespacio unidimensional S y en el segundo ejemplo se aplica la técnica del PCA para la reducción de dimensionalidad de una imagen.

Implementación en MATLAB

Para la implementación del algoritmo básico para la reducción de dimensionalidad se utilizó el software llamado MATLAB, ya que en principio es un asistente matemático completo, por lo que podemos gozar de todas las libertades que el software nos ofrece y tiene una enorme capacidad para la solución de problemas del Álgebra lineal. Aunque pudiéramos usar GNU Octave, ya que tiene las mismas prestaciones técnicas que MATLAB, básicamente el comportamiento de ambos software es similar. En particular, los lenguajes de programación, MATLAB y GNU Octave, son dos sistemas tan similares que la mayoría de los programas escritos son compatibles con los compiladores de ambos sistemas. Además, la experiencia del usuario para ambos es muy similar con diseños de ventana casi idénticos.

Por lo tanto, para cualquier estudiante que se está iniciando en el manejo de Octave pudieran realizar sus prácticas sin ningún problema, ya que GNU Octave es una buena elección para esta propuesta de intervención, además permite trabajar a los alumnos en un entorno lo más parecido a MATLAB, y al mismo tiempo es una herramienta que minimiza los costes del centro educativo, porque instalar GNU Octave es gratuito (Erausquin, 2017).

Código básico de reducción de dimensionalidad vía PCA

A continuación en la figura 2-4 se da a conocer un código básico realizado por (Castillo, et al., 2022), este código sigue los pasos del algoritmo básico de reducción de dimensionalidad de grandes volúmenes de datos implementada anteriormente.

Se realiza una práctica para el proceso de reducción de dimensionalidad de datos mediante la técnica de Análisis de Componentes Principales, el conjunto de datos al cual se desea reducir la dimensionalidad son creados de manera aleatoria por este software. Veamos a continuación el código implementada en MATLAB.


```

function [alfa]=PCA(X)
%Reducción de Dimensionalidad vía PCA

N=400
DaT=randn(N,3); % Matriz aleatoria de N filas por 2 columnas.
A=rand(3,3); % Rand crea un arreglo de numeros aleatorios de 2x2.
X=DaT*A;

%Centrando y normalizando
mu1=mean(X(:,1)); s1=std2(X(:,1));
mu2=mean(X(:,2)); s2=std2(X(:,2));
X(:,1)=(X(:,1)-mu1.*ones(N,1))/s1;
X(:,2)=(X(:,2)-mu2.*ones(N,1))/s2;

%Encontrar la matriz de covarianza
CovX=cov(X);

%Identificando el componente principal
[V,E]=eig(CovX);
e=diag(E);
[me,ind]=max(e);

%Definiendo el subespacio de proyección
S=V(:,ind);

%Encontrar los escalares óptimos (code)
alfa=X*S;

%Calculando la proyección (decode)
Proy_S=alfa*S';

%Visualizando los resultados
figure;
scatter(X(:,1), X(:,2), 'k');
scatter(X(:,2), X(:,3), 'b');

hold on;
scatter(Proy_S(:,1), Proy_S(:,2), 'g');
%scatter(Proy_S(:,2), Proy_S(:,3), 'b');
title('PCA: Reducción de dimensionalidad');
xlabel('x');
ylabel('y');
end

```

Ilustración 2-4: Código básico para la reducción de datos vía PCA.

Fuente: Castillo, Zenaida, et al., 2022.

Veamos el resultado obtenido mediante la ejecución del código en la figura 2-4 donde se ejecutó para un conjunto $N = 400$ de datos aleatorios. En este ejemplo seguimos las instrucciones del algoritmo del PCA, ya que dado un conjunto X de datos, lo que hacemos es centrar y normalizar o lo que se conoce como estandarización de los datos, luego se calcula la matriz de covarianza para determinar los autovalores y autovectores, una vez realizada esta acción debemos definir el subespacio de proyección mediante la definición de la matriz de proyección, esto nos ayudará a

determinar las coordenadas o escalares óptimos, con ello ya podemos obtener la proyección de los datos originales en el subespacio unidimensional.

El resultado se refleja en la figura 3-4 donde podemos visualizar que la línea marcada en color verde es una recta que representa la proyección de los $N = 400$ datos aleatorios que están marcadas en color negro, y la relación entre variables es inversa, es decir, la variación es en la dirección opuesta. Este conjunto de datos es proyectado en el espacio unidimensional, el resultado que se obtendrá es un vector de escalar α ya que el subespacio principal al cual se proyecta los datos bidimensionales es de dimensión uno. Esto muestra la reducción de dimensionalidad del espacio de \mathbb{R}^2 hacia el subespacio unidimensional (recta).

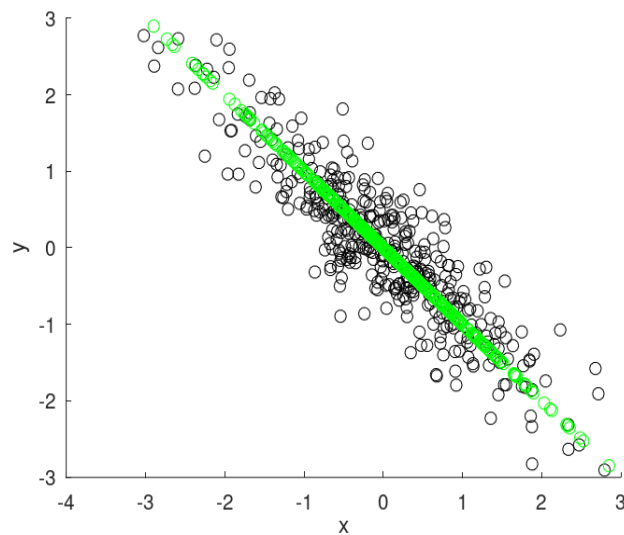


Ilustración 3-4: Resultado de la proyección de datos mediante el código básico del PCA.

Realizado por: Buñay, Cristian, 2023.

La implementación que se hizo es una idea básica de la cual podemos generalizar siguiendo la analogía del algoritmo. Por ejemplo, si quisiéramos reducir un conjunto de datos aleatorios que se encuentran en el espacio tridimensional, proyectaremos estos datos a un subespacio de dimensión dos, donde obtendremos como salida el vector de escalares α , el cual tendrá dos coordenadas, ya que el subespacio al cual se está proyectando es bidimensional, estas dos coordenadas son los dos autovalores asociados a los dos autovectores de mayor valor de la matriz de covarianza.

En el caso general, si consideramos un conjunto $X = \{x_1, x_2, \dots, x_N\}$ de N datos donde cada dato tenga n características o variables al cual se desea proyectar en un subespacio S de dimensión q ,

$q \leq n$, entonces obtendremos como salida el vector de escalares α que tiene como coordenadas los q autovalores de mayor módulo y los vectores que definen el subespacio de proyección serán los q autovectores de la matriz de covarianza σ asociados a los autovalores del módulo máximo, esto muestra la reducción de dimensionalidad de datos.

En el caso de que la matriz de covarianza σ es simétrica y deficiente en rango, lo que se recomienda para la solución es hallar la descomposición en valores singulares truncada de la matriz de datos X . Estos la podemos ver en detalle en referencias recientes como (Deisenroth, 2020) y (Trefethen, 1997).

Reducción de dimensionalidad de una imagen vía PCA

En el siguiente ejemplo, aplicaremos el algoritmo básico de reducción de dimensionalidad mediante la técnica de Análisis de Componentes Principales al conjunto de datos de una imagen a color de (iStock, 2016) cuyo tamaño es de $500 \times 330 \times 3 \text{unit8}$, esta imagen contiene 165,000 píxeles para que podamos interpretar cada imagen en este conjunto de datos como un vector $x \in \mathbb{R}^{165,000}$, es decir, es un punto de datos en un espacio de 165,000 dimensiones, donde cada píxel responde a tres dimensiones, una para cada canal o capa de color rojo, verde y azul o conocido como RGB por sus sigas en inglés (Red, Green, Blue).

Para el desarrollo de este ejemplo se usó el software MATLAB, ya que posee muchas opciones que facilitan el manejo y es una interfaz amigable. A continuación en la figura 4-4 se presenta la imagen original.



Ilustración 4-4: Imagen Original de $500 \times 330 \times 3 \text{unit8}$.

Fuente: Imagen realizado por iStock, 2016.

Para obtener la reducción de dimensionalidad de esta imagen, seguimos las instrucciones del algoritmo del PCA, establecido anteriormente. Esto menciona en resumen que primero, se debe centrar los datos mediante el cálculo de la media y la estandarización los datos de entrada, lo que se conoce como normalización de los datos, luego calculamos la matriz de covarianza de los datos para obtener los autovectores y autovalores de la matriz de covarianza, ahora debemos ordenar los autovalores de mayor a menor y elegir los q autovectores que se correspondan con los autovectores q más grandes (donde q es el número de dimensiones del nuevo subespacio de proyección).

Mediante esto pasamos a definir la matriz de proyección S con los q autovectores seleccionados. Una vez definida el subespacio de proyección S , debemos proyectar los datos originales sobre el subespacio definido por los q vectores seleccionados. Luego transformamos el conjunto de datos originales X (estandarizado) vía S para obtener las nuevas características q -dimensionales.



(a) PCs:1



(b) PCs:5



(c) PCs:10



(d) PCs:25



(e) PCs:50



(f) PCs:100

Ilustración 5-4: Reconstrucciones de imágenes con las componentes principales.

Realizado por: Buñay, Cristian, 2023.

El resultado se puede apreciar en la figura 5-4, donde podemos ver las reconstrucciones de la imagen original mediante el uso de los subespacios principales de dimensiones 1, 5, 10, 25, 50 y 100 respectivamente. Observemos que con un subespacio principal unidimensional, es decir, con una sola componente principal obtenemos un resultado muy deficiente y no podemos identificar el objeto de la cual se trata, por ende no se puede hacer el análisis con dicho resultado, ahora veamos ¿que ocurre si incrementamos la dimensión del subespacio de proyección?, entonces tomamos un subespacio de dimensión cinco, es decir, con las cinco primeras componentes y tampoco podemos apreciar con claridad, pero cada vez que incrementamos la dimensión del subespacio principal, es decir, el número de componentes principales, se incrementa también la presentación en este caso de la imagen.

Observemos que con tan solo las diez primeras componentes principales obtenemos una reconstrucción decente de la imagen original donde ya podemos identificar de lo que se trata la imagen, aunque sigue siendo borrosa.

Así, a medida que vamos incrementando el número de las componentes principales se van tomando en cuenta más detalles y obtenemos una mejor reconstrucción la cual mejora la presentación y hace que la calidad de la imagen sea cada vez más clara y nítida. Notemos que con tan solo 100 componentes principales, obtenemos efectivamente una reconstrucción casi perfecta de la imagen, esto implica que si eligiéramos el número máximo de las componentes principales, en este caso, 330 PCs, recuperaríamos la imagen exacta sin pérdida de compresión.

Pero ¿cuál es el número óptimo de Componentes Principales para un análisis efectivo? (Bagnato, 2018) menciona que: dado un conjunto de datos con N observaciones y n variables o características, el número de componentes principales que se puede formar de este conjunto es $\min(N - 1, n)$. Sin embargo, no existe un método ideal que permite tomar cierto número de las componentes principales para un análisis eficiente de los datos, esto depende del problema y del juicio del analista. Generalmente, no se eligen más de tres componentes principales para facilitar la representación gráfica y la interpretación si proporcionan suficiente variabilidad, ya que el objetivo es visualizar los datos. Para realizar un análisis eficiente de los datos se debe considerar las componentes principales que abarquen el 85 % de la variabilidad total de los datos. En algunos casos puede pasar que solo una componente principal abarque más del 85 % de la variabilidad de los datos, por lo que depende mucho del problema en cuestión.

A continuación en la figura 6-4 se presenta el código básico implementado en MATLAB para la reducción de dimensionalidad de una imagen mediante el uso de la técnica de Análisis de Componentes Principales.

```

% Reducción de Dimensionalidad vía PCA
% Perspectiva Proyección

imagen_original = imread('flor2.jpg');
figure;
imshow(imagen_original);
title('Imagen Original');

% Descomposicion de imagen por tonos (RGB)
rojo = imagen_original(:,:,1); %capa_roja
verde = imagen_original(:,:,2); %capa_verde
azul = imagen_original(:,:,3); %capa_azul

[N,M]=size(rojo); %numero de filas y columnas de la matriz

%preparando las matrices para su procesamiento
A_red = double(rojo);
A_green = double(verde);
A_blue = double(azul);

%Centrando y normalizando
mu1=mean(A_red); s1=std2(A_red);
mu2=mean(A_green); s2=std2(A_green);
mu3=mean(A_blue); s3=std2(A_blue);

N1=(A_red(:,1)- mu1.*ones(N,1))/s1;
N2=(A_green(:,2)-mu2.*ones(N,1))/s2;
N3=(A_blue(:,1)-mu3.*ones(N,1))/s3;

%Encontrar la matriz de covarianza
CovR=cov(A_red);
CovG=cov(A_green);
CovB=cov(A_blue);

%Identificando el componente principal
[U,E1]=eig(CovR);
e1=diag(E1);
[me1,ind1]=max(e1);

[V,E2]=eig(CovG);
e2=diag(E2);
[me2,ind2]=max(e2);

[W,E3]=eig(CovB);
e3=diag(E3);
[me3,ind3]=max(e3);

% Definiendo el subespacio de proyección
k = input('Introducir el valor de k = ');
S1=U(:,M-k:M);
S2=V(:,M-k:M);

```

```

S3=W(:,M-k:M);

%Encontrar los escalares óptimos (code)
alfa1=A_red*S1;
alfa2=A_green*S2;
alfa3=A_blue*S3;

%Calculando la proyección (decode)
Proy_S1=alfa1*S1';
Proy_S2=alfa2*S2';
Proy_S3=alfa3*S3';

%Visualizando los resultados
Proy_S1 = uint8(floor(Proy_S1));
Proy_S2 = uint8(floor(Proy_S2));
Proy_S3 = uint8(floor(Proy_S3));

% Concatenando las imagenes de los tres colores (RGB)
Imagen_aprox = cat(3, Proy_S1, Proy_S2, Proy_S3);

% Mostrando la solución
figure;
imshow(Imagen_aprox);
title('Imagen Comprimida');

```

Ilustración 6-4: Código básico de reducción de una imagen vía PCA.

Realizado por: Buñay, Cristian, 2023.

El PCA presenta muchas relaciones con otras técnicas, en algunos textos sobre análisis multivariado, el PCA es tratado como si fuera parte del análisis factorial. Aunque el análisis factorial y el PCA tienen objetivos similares, en realidad son técnicas bastante diferentes. Sin embargo, hay algunas formas en las que PCA se puede utilizar en el análisis factorial y se describen brevemente. El uso de PCA para ortogonalizar un problema de regresión, al reemplazar un conjunto de variables altamente correlacionadas por sus componentes principales, es bastante conocido. El Análisis de Componentes Principales se utiliza a veces como un paso preliminar o en conjunto con otras técnicas estadísticas, siendo el ejemplo obvio en regresión (Jolliffe, 2002).

Una de las mayores ventajas que tiene esta técnica de Análisis de Componentes Principales es que cuando reduce la gran dimensionalidad de datos, la primera componente principal abarca o conserva la mayor variabilidad de los datos y la segunda componente principal abarca la mayor variabilidad de los datos restantes del primer componente principal, esto hace que con algunas pocas componentes principales podemos realizar el análisis efectivo de los datos. Además, es una de las técnicas que reduce el tiempo al máximo al momento de la ejecución.

El Análisis de Componentes Principales es una técnica no supervisada, por lo tanto, presenta la desventaja de no presentar resultados satisfactorios en muchos problemas supervisados de clasificación o regresión. Otra de las desventajas que tiene el PCA es que solo se puede realizar

combinaciones lineales de los datos originales, esto hace que en ciertos casos pierda mucha información. Existen algunas otras técnicas que reducen la dimensionalidad de los datos, las cuales hacen competencia con la técnica del PCA, tales como: Análisis de factores, Análisis de discriminante lineal, Análisis de componente independiente, Descomposición en valores singulares.

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

En este capítulo se describen las conclusiones y las recomendaciones obtenidas a lo largo de este proceso de investigación, esto con el fin de dar la continuidad en este tópico de Análisis de Componentes Principales vía proyecciones ortogonales.

5.1. Conclusiones

El objetivo de esta investigación fue estudiar y describir, desde el punto de vista matemático, las proyecciones ortogonales y su uso en la reducción de dimensionalidad mediante la técnica de Análisis de Componentes Principales o PCA.

Mediante el uso de conceptos importantes del Álgebra Lineal tales como: espacios y subespacios vectoriales, bases ortogonales y ortonormales, autovalores y autovectores, normas, proyección ortogonal y otros que están relacionados, se obtuvo la descripción en forma detallada de la técnica de Análisis de Componentes Principales que es ampliamente conocida por sus distintas aplicaciones en la reducción de dimensionalidad. Finalmente, se ordenan los pasos a seguir para la reducción de dimensionalidad con esta técnica; los cuales permiten elaborar un código que se utiliza para mostrar la utilidad de la técnica.

Durante muchos años se han venido trabajando con el Análisis de Componentes Principales en distintas ramas de la ciencia con el propósito de aportar en los avances científicos y tecnológicos, existen muchas referencias en relación a este tópico, muchas de ellas requiere un alto nivel de conocimiento. Una de las aplicaciones más relevantes en la actualidad mediante el uso de esta técnica es el aprendizaje automático o machine learning.

Existen otras derivaciones de esta técnica del PCA, tales como: la maximización de la varianza, y la perspectiva de variables latentes, las cuales se basan en modelos estadísticos. En esta investigación hemos usado las herramientas básicas de la matemática que yacen en el Álgebra Lineal, y que permiten la derivación inicial de la técnica, con el fin de crear conocimiento que permita, no solo su entendimiento, sino también su uso.

5.2. Recomendaciones

Esta investigación muestra todos los detalles de la teoría matemática que intervienen en la derivación de la técnica de Análisis de Componentes Principales, por lo cual se recomienda su uso como base para el entendimiento y futuras investigaciones en la reducción de dimensionalidad de datos.

Adicionalmente, para la derivación del algoritmo de la técnica del PCA, se usó la norma euclídea, ampliamente conocida, la cual está relacionada con el producto escalar, pero se hizo mención de otras normas relacionadas con el producto interno, por lo que se recomienda a futuros estudiantes que tengan interés en el tópico realizar el estudio de la derivación de la técnica del PCA basado en normas diferentes a la norma euclídea o desde una perspectiva distinta a la de proyección ortogonal. Otro trabajo, un poco más de carácter computacional, sería calcular solo los autovalores del módulo máximo.

Se recomienda que se dé a conocer a los estudiantes de la carrera de matemática como uno de los temas a tratar dentro de la asignatura del Álgebra Lineal para que así puedan ir adquiriendo conocimiento de la parte aplicada de la matemática.

BIBLIOGRAFÍA

CASTILLO, Zenaida; et al. "Detalles de la derivación e implementación de algoritmo básico para la reducción de dimensionalidad con PCA". *Polo del Conocimiento*, vol. 7, n° 2 (2022), pp. 1884-1907.

DEISENROTH, Marc; et al. *Mathematics for Machine Learning*. Cambridge University Press, 2020. pp. 317-338.

GROSSMAN, S. I. *Álgebra lineal*. 7ª ed. México: McGraw Hill Educación, 2012. ISBN: 978-607-15-0760-0, pp. 1-765.

LAY, D. C. *Álgebra lineal y sus aplicaciones*. 3ª ed. México: Pearson educación, 2007. ISBN: 978-970-26-0906-3, pp. 1-584.

PEARSON, K. *On lines and planes of closest fit to systems of points in space*. *Philosophical Magazine*, 1901. pp.559-572. [Consulta:23 octubre 2022] Disponible en: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj61oK19ur-AhV3RzABHSdwALkQFnoECBoQAQ&url=https%3A%2F%2Fzenodo.org%2Frecord%2F1430636%2Ffiles%2Farticle.pdf&usg=A0vVaw3oUQS_9JmIKHKU_m83xJLR.

JOLLIFFE, I. T. *Principal Component Analysis*[en línea]. 2ª ed. *Encyclopedia of Statistics in Behavioral Science*, 2002. pp. 1-518. [Consulta:23 octubre 2022] Disponible en: [http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf).

DATTA, B. N. *Numerical Linear Algebra and Applications*. Society for Industrial and Applied Mathematics, 2ª ed. 2010. pp. 1-128.

THOMAS, Jr; GEORGE, B. *Cálculo, Varias Variables*. 12^a ed. México. Pearson Education, 2005. ISBN: 978-607-32-0209-1, pp. 747-828.

JERÓNIMO, Gabriela; et al. *Álgebra Lineal*. Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 2008. ISSN 1851-1317, pp. 1-281.

GALLURDO, Mario & MARTÍNEZ, Francisco. *Apuntes de Álgebra Lineal*. Universidad Politécnica de Madrid, 1977. pp. 1-155.

SaALEM, N; & HUSSEIN, S. "Data dimensional reduction and principal components analysis". *Procedia Computer Science*, 2019, vol. 163, pp.292-299.

KHERIF, F; & LATYPOVA, A. "Principal Component Analysis". En *Machine Learning*. Academic Press, 2020. pp. 209-225.

DEMMELE, J. W. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics. 1997. pp. 1-388.

SALAZAR, C.; & DEL CASTILLO, S. *Fundamentos Básicos de Estadística* [en línea]. 1^a ed. 2018. pp. 49-75 [Consulta: 04 octubre 2022]. Disponible en: <http://up-rid2.up.ac.pa:8080/xmlui/bitstream/handle/123456789/1371/EstadísticaFundamentos%20Básicos%20de%20-Libro.pdf?sequence=1&isAllowed=y>.

JAMES, Gareth; et al. *An introduction to statistical learning*. New York: springer. 2013. Vol. 112, pp. 373-413.

SMOLA, A.; & VISHWANATHAN, S. *Introduction to machine learning*. Cambridge University, UK, 32(34), 2008. pp. 91-139.

RENCHER, A.; & BRUCE G. *Linear models in statistics*. 6^a ed. Hoboken, N.J: Wiley-Interscience, 2008.

TREFETHEN, L.; & BAU, D. *Numerical linear algebra* [en línea], 1997. pp. 1-234. [Consulta: 02 noviembre 2022]. Disponible en: https://cdn.bc-pf.org/resources/math/algebra/linear_algebra/Trefethen_Bau-Numerical_Linear_Algebra.pdf

ALMENARA, J. et al. ¿Qué es el análisis de Componentes principales? *Jano-1268*, 1998. pp.58-60.

RESTREPO, Luis; et al. *Aplicación del análisis por componentes principales en la evaluación de tres variedades de pasto*. *Revista Colombiana de Ciencias Pecuarias*, 2012, vol. 25, n° 2, pp. 258-266.

PLA LE. *Análisis Multivariado: Método de Componentes Principales*. Washington: Organización de Estados Americanos; 1986.

DEL ÁNGEL NV. *Una aplicación del álgebra lineal a la estadística (componentes principales)*. 2004.

DALLAS, J. E. *Métodos multivariados aplicados al análisis de datos*. México: Thomson Paraninfo S.A. 2000.

LUCÍA, et al. “Análisis de componentes principales, una herramienta eficaz en las Ciencias Técnicas Agropecuarias”. *Revista Ciencias Técnicas Agropecuarias* [en línea], 2022, vol. 31. [Consulta: 14 diciembre 2022]. Disponible en: <http://ref.scielo.org/rd2j69>.

TRENDAFILOV, N. T. “From simple structure to sparse components: A review Computational Statistics”, 2014. pp. 431-454.

OCHOA, J. et al. “Propuesta de un modelo de medición de la competitividad mediante análisis factorial”. *Contaduría y administración*, 2017. pp. 775-791.

GIL C. *ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)* [blog]. Junio, 2018. [Consulta: 17 diciembre 2022]. Disponible en: https://rpubs.com/Cristina_Gil/PCA.

BAGNATO, J. I. *Aprende Machine Learning* [blog]. Octubre, 2018. [Consulta: 17 diciembre 2022]. Disponible en: <https://www.aprendemachinellearning.com/comprende-principal-component-analysis/>.

FUJITA, M.; & KRUGMAN, P. “The new economic geography: Past, present and the future”. *Papers in regional science*, 2004. pp.139-164.

GOZÁ, O.; et al. “Aplicación del Análisis de Componentes Principales en el proceso de purificación de un biofármaco”. *Vaccimonitor*, 2020. vol. 29, n° 1, pp.5-13.

SOLANO, E.; et al. “Análisis de componentes principales en la Competitividad en Colombia”. *Revista Ibérica de Sistemas e Tecnologías de Información* [en línea], 2009. pp. 260-271. [Consulta: 25 noviembre 2022]. Disponible en: <https://repositorio.cuc.edu.co/handle/11323/5951>.

FARALDO, P.; & PATEIRO, B. *Estadística y la metodología de la investigación*. Universidad de Santiago de Compostela. 2012-2013. pp.1-15.

GONZÁLES, N.; & TABORDA, A. *ANÁLISIS DE COMPONENTES PRINCIPALES SPARSE Formulación, algoritmo e implicaciones en el análisis de datos*. Recuperado de: https://gredos.usal.es/bitstream/handle/10366/126046/TFM_MAADM_Gonz%E11ez_Taborda.pdf?sequence=4.

ERAUSQUIN, M. *Introducción a la programación numérica con GNU Octave para Matemáticas de 2º de bachillerato*, 2017. Recuperado de: <https://reunir.unir.net/>

handle/123456789/4705.

ISTOCK. *Flor libre de Derechos*, 2016. Tipo Imagen jpg. [Consulta: 25 enero 2023]. Disponible en: <https://www.istockphoto.com/es/foto/blanco-solo-hermosa-rosa-con-fondo-negro-gm517518010-89524925>.





epoch

Dirección de Bibliotecas y
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 16 / 06 / 2023

| |
|--|
| INFORMACIÓN DEL AUTOR/A (S) |
| Nombres – Apellidos: Cristian David Buñay Marcatoma |
| INFORMACIÓN INSTITUCIONAL |
| Facultad: Ciencias |
| Carrera: Matemática |
| Título a optar: Matemático |
| f. Analista de Biblioteca responsable: Ing. Rafael Inty Salto Hidalgo |

0851-DBRA-UPT-2023

