línea horizontal

**Herramientas matemáticas y computacionales para inteligencia artificial**  
Andrés Mauricio González Vargas  
Universidad Autónoma de Occidente

Punto 6 - Informe final

**Samir Hassan - 2190041**

**Gabriel Jeannot - 2185887**

**Carlos Osorio - 2230894**

**Luis Pareja - 2185833**

**Diego Perea - 2185751**

# Elabore un documento en el que se explique cómo se relacionan el álgebra lineal, el cálculo y la estadística con los algoritmos de Regresión Lineal, Regresión Logística, Redes Neuronales, Singular Value Decomposition y Principal Component Analysis.

# Introducción

El álgebra lineal, el cálculo y la estadística son herramientas matemáticas fundamentales para la inteligencia artificial. La IA utiliza técnicas de aprendizaje automático para aprender de los datos y realizar tareas sin ser programada explícitamente para ellas (Goodfellow, Bengio, & Courville, 2016). Estas técnicas dependen en gran medida del análisis y la manipulación de grandes conjuntos de datos, lo que hace que las herramientas matemáticas sean esenciales para la IA.

El álgebra lineal es una rama de las matemáticas que se centra en el estudio de sistemas de ecuaciones lineales y sus propiedades. En la IA, el álgebra lineal se utiliza para resolver sistemas de ecuaciones lineales, realizar cálculos matriciales y realizar transformaciones lineales en datos. Por ejemplo, la regresión lineal utiliza técnicas de álgebra lineal para ajustar una línea recta a un conjunto de datos, lo que permite predecir valores futuros. Además, la descomposición en valores singulares (SVD) es una técnica de álgebra lineal utilizada para la reducción de dimensionalidad, una técnica común en la IA para procesar grandes conjuntos de datos (Strang, 2009).

El cálculo es otra herramienta matemática fundamental en la IA. Se utiliza para calcular las derivadas y las integrales, lo que permite optimizar las funciones objetivo y realizar análisis de sensibilidad en los modelos de aprendizaje automático. Por ejemplo, las redes neuronales utilizan técnicas de cálculo para propagar la señal a través de la red y actualizar los pesos de las conexiones para mejorar la precisión del modelo (Goodfellow, Bengio, & Courville, 2016).

La estadística es la ciencia que se ocupa de recopilar, analizar e interpretar datos. En la IA, la estadística se utiliza para evaluar la precisión y la significancia de los modelos de aprendizaje automático y para realizar pruebas de hipótesis. Además, la estadística se utiliza en la selección de variables, que es un proceso importante en la construcción de modelos de aprendizaje automático. La selección de variables implica identificar las variables más importantes para el modelo y descartar aquellas que no contribuyen significativamente (Hastie, Tibshirani, & Friedman, 2017).

Con base a esto, y previos a realizar el desarrollo de este documento, debemos definir una serie de relaciones pertinentes que nos permitirán abordar la información de manera más adecuada:

Regresión Lineal = Álgebra Lineal

Regresión Logística = Cálculo y Estadística

Redes Neuronales = Álgebra Lineal, Cálculo y Estadística

Singular Value Decomposition = Álgebra Lineal

Principal Component Analysis = Álgebra Lineal y Estadística

Esto significa que cada concepto a analizar, tiene una o varias ramas específicas desde la cuales se le puede relacionar y se puede abordar. Esto teniendo en cuenta que, en términos teóricos y prácticos, todos se relacionan entre sí, pero para optimizar el desarrollo de este documento, se limitará el alcance que cada concepto posee en una profundidad mayor.

# Algoritmos de regresión lineal

# 1. Definición

Habiendo definido previamente la relación directa de este tema con el álgebra lineal, partimos desde su concepto. La regresión lineal es una técnica estadística utilizada en inteligencia artificial para modelar la relación entre una variable dependiente (también conocida como variable de respuesta) y una o más variables independientes (también conocidas como variables predictoras). Los algoritmos de regresión lineal se utilizan para ajustar una línea recta que mejor se adapte a los datos y predecir valores futuros en función de las variables predictoras.

Según Gareth James et al. (2017), en su libro "An Introduction to Statistical Learning", la regresión lineal es "un enfoque simple pero ampliamente utilizado para modelar la relación entre una variable de respuesta y una o más variables predictoras" (p. 67). El objetivo es encontrar la línea que mejor se ajuste a los datos, minimizando la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos.

Existen dos tipos de regresión lineal: la regresión lineal simple y la regresión lineal múltiple. La regresión lineal simple se utiliza cuando solo hay una variable predictora, mientras que la regresión lineal múltiple se utiliza cuando hay dos o más variables predictoras.

Según Hastie, Tibshirani, & Friedman (2017), La regresión lineal se puede ver como una técnica de interpolación, donde el objetivo es estimar la respuesta media en cada punto de las variables predictoras" (p. 43). Además, mencionan que la regresión lineal es un modelo lineal paramétrico, lo que significa que el modelo asume que la relación entre las variables es lineal y se puede describir mediante un conjunto de parámetros fijos (p. 43).

# 1. Relación con el álgebra lineal

A partir del contexto definido, resulta necesario profundizar en la relación que los algoritmos de regresión lineal tienen con el álgebra lineal. En primera medida, es importante destacar que la regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes. En este sentido, la regresión lineal se basa en la ecuación de una línea recta que se ajusta a los datos para predecir valores futuros. Por su parte, el álgebra lineal se encarga de estudiar las propiedades de los espacios vectoriales y las transformaciones lineales, es decir, las funciones que preservan las operaciones de suma y multiplicación por un escalar.

En la regresión lineal, según Gareth James et al. (2017, p. 70), la ecuación que se utiliza para predecir los valores futuros se puede expresar de manera matricial, donde los coeficientes de la ecuación corresponden a los elementos de la matriz de coeficientes y las variables independientes se agrupan en un vector. De esta forma, la regresión lineal puede verse como una aplicación de las transformaciones lineales estudiadas en el álgebra lineal. Asimismo, los conceptos de vectores y matrices son fundamentales para entender la regresión lineal y sus propiedades. Los algoritmos de regresión lineal están estrechamente relacionados con el álgebra lineal, ya que se basan en la manipulación de vectores y matrices para realizar predicciones. El entendimiento de estos conceptos es fundamental para el desarrollo de modelos de inteligencia artificial basados en regresión lineal.

Con base a la información obtenida en el desarrollo del curso, la regresión lineal se usa a menudo en el aprendizaje automático para predecir valores numéricos en problemas de regresión más simples. Una de las formas más comunes de resolver la regresión lineal es a través de una optimización de mínimos cuadrados que se resuelve utilizando métodos de factorización de matrices de regresión lineal, como una descomposición LU o una descomposición de valor singular, o SVD.

A continuación, se puede ver una representación gráfica de la línea de regresión propia de la aplicación de la regresión lineal en la inteligencia artificial.

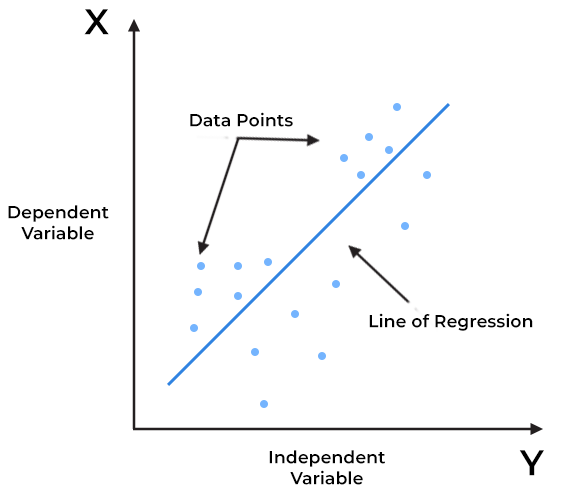


Figura 1: Gráfica de la línea de regresión y los puntos de datos

# Regresión logística

# 1. Definición

La regresión logística se define como un algoritmo de aprendizaje supervisado utilizado en la inteligencia artificial para la clasificación de datos binarios. Este algoritmo se basa en una función sigmoidea que permite predecir la probabilidad de que una observación pertenezca a una de las dos categorías. A diferencia de la regresión lineal, que se utiliza para predecir valores continuos, la regresión logística se enfoca en la predicción de valores discretos. Debido a su capacidad para trabajar con datos binarios y la simplicidad de su modelo, la regresión logística es un algoritmo popular en la inteligencia artificial y una herramienta valiosa para la toma de decisiones en diversos campos. Con base a esta definición, podemos relacionar el concepto de manera directa con el cálculo y la álgebra lineal, de manera independiente para que resulte más entendible.

# 2. Relación con el álgebra lineal

La regresión logística tiene su base en la optimización de una función de costo que se puede resolver mediante técnicas de optimización matemática, como el método del gradiente descendente. El objetivo es encontrar los coeficientes de la ecuación que mejor ajusten los datos de entrenamiento y permitan predecir correctamente las categorías de los datos de prueba. Como señalan Géron (2019), la regresión logística también puede expresarse de manera matricial, lo que la relaciona con el álgebra lineal.

La regresión logística también puede ampliarse a modelos multiclase mediante la utilización de la función softmax, que se utiliza para estimar las probabilidades de pertenencia a cada una de las categorías. Según Murphy (2012), la función softmax se relaciona con el álgebra lineal debido a que se utiliza para normalizar los valores de una matriz de salida, lo que permite interpretarla como una distribución de probabilidad. Esto ha llevado a la utilización de la regresión logística multiclase en la clasificación de imágenes, texto y otros tipos de datos.

Además de su representación matricial, la regresión logística también se relaciona con el álgebra lineal en la forma en que se actualizan los pesos durante el proceso de entrenamiento. Según Goodfellow, Bengio y Courville (2016), los pesos de la regresión logística se actualizan mediante la multiplicación de la matriz de características por la transpuesta de la matriz de error, lo que se conoce como la regla del producto de la cadena. Esto se asemeja a la forma en que se actualizan los pesos en las redes neuronales, lo que ha llevado a la utilización de la regresión logística como un componente de las mismas.

Por otra parte, el álgebra lineal también es útil para la interpretación de los coeficientes de la regresión logística. Como señala Géron (2019), cada coeficiente se puede interpretar como la pendiente de una recta en el espacio de características, lo que permite analizar su importancia relativa en la predicción de la variable dependiente. Asimismo, la matriz de covarianza de los coeficientes se puede utilizar para analizar su varianza y covarianza, lo que puede ser útil para identificar problemas de multicolinealidad y seleccionar las variables más relevantes en la predicción.

# 3. Relación con el cálculo

La relación con el cálculo se encuentra en la función sigmoidea, que es una función continua y diferenciable utilizada para estimar la probabilidad de pertenecer a cada categoría. La función sigmoidea se utiliza para transformar la variable dependiente en una escala de 0 a 1, lo que permite interpretar la salida como la probabilidad de pertenecer a la categoría positiva. Según Goodfellow, Bengio y Courville (2016), la función sigmoidea también se utiliza en otros algoritmos de aprendizaje automático, como las redes neuronales, y se ha relacionado con la activación de las neuronas en el cerebro.

Este tipo de regresión se basa en la optimización matemática de una función de costo y utiliza la función sigmoidea para estimar las probabilidades de pertenencia a una categoría. Ambas técnicas se relacionan con el álgebra lineal y el cálculo, lo que permite su aplicación en una amplia variedad de problemas de clasificación en la inteligencia artificial. A continuación, se visualiza una función de activación tipo “sigmoid”:

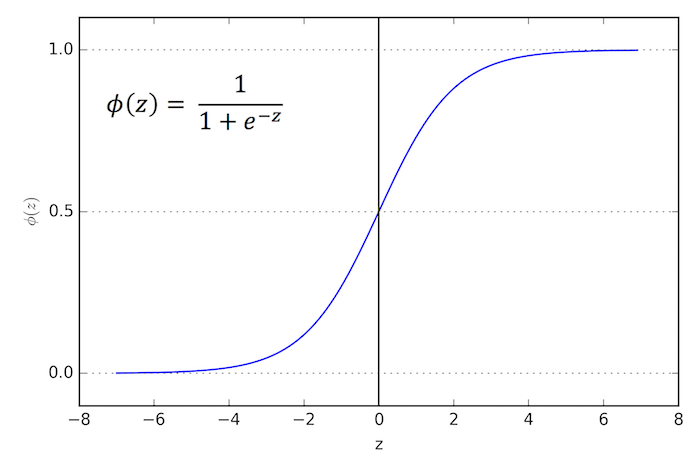


Figura 2: Gráfica de la función de activación sigmoide

Además, la regresión logística también se relaciona con el cálculo a través del método del gradiente descendente utilizado para optimizar la función de costo. Este se basa en la derivada parcial de la función de costo con respecto a los coeficientes de la ecuación, lo que indica la dirección y magnitud del cambio necesario para reducir la función de costo.

# Redes Neuronales

# 1. Definición

Las redes neuronales son un tipo de modelo de aprendizaje automático inspirado en el funcionamiento del cerebro humano. Según Goodfellow, Bengio y Courville (2016), una red neuronal es un conjunto interconectado de unidades de procesamiento llamadas neuronas, que se organizan en capas y procesan la información mediante la aplicación de operaciones matemáticas. Las redes neuronales son utilizadas en una gran variedad de aplicaciones en inteligencia artificial, como la clasificación de imágenes, el procesamiento del lenguaje natural y la predicción de series de tiempo.

Las redes neuronales se entrenan mediante el ajuste de los pesos de las conexiones entre las neuronas, lo que permite que la red pueda aprender a realizar tareas complejas. La retropropagación del error es una técnica comúnmente utilizada para el entrenamiento de redes neuronales, que se basa en la optimización de una función de costo mediante la modificación iterativa de los pesos de las conexiones. Esta técnica se utiliza en la mayoría de los tipos de redes neuronales, como las redes neuronales convolucionales, las redes neuronales recurrentes y las redes neuronales de alimentación directa (feedforward).

# 2. Relación con el álgebra lineal

Las redes neuronales se basan en la utilización de múltiples capas de neuronas, las cuales realizan operaciones matemáticas sobre los datos de entrada para generar una salida. En este sentido, el álgebra lineal es un componente fundamental en el diseño y entrenamiento de las redes neuronales. Según Goodfellow, Bengio y Courville (2016), el proceso de entrenamiento de una red neuronal implica el ajuste de los pesos y sesgos de las neuronas para minimizar una función de costo mediante el uso de técnicas de optimización matemática. Esto se logra mediante la aplicación de operaciones matriciales para realizar la propagación hacia adelante y hacia atrás de los datos en la red.

En particular, la multiplicación de matrices se utiliza en la propagación hacia adelante de los datos en una red neuronal. Cada capa de neuronas utiliza una matriz de pesos para realizar la transformación lineal de los datos de entrada. Según Zhang y LeCun (2015), esta transformación se puede expresar de manera matricial y se puede optimizar mediante el uso de técnicas de descenso de gradiente. La propagación hacia atrás, por su parte, implica la utilización de la matriz de pesos transpuesta para propagar el error en la red y ajustar los pesos y sesgos de las neuronas mediante la actualización de las matrices de pesos.

Además, el álgebra lineal también se utiliza en la representación y manipulación de datos en las redes neuronales. Según Goodfellow, Bengio y Courville (2016), los datos de entrada se suelen representar en forma de matrices, lo que permite su manipulación mediante operaciones matriciales para su procesamiento en la red. Por ejemplo, las imágenes se pueden representar en forma de matrices de píxeles, lo que permite su procesamiento mediante convoluciones y pooling en las redes convolucionales.

Según LeCun, Bengio y Hinton (2015), las redes neuronales convolucionales se basan en el uso de convoluciones, que se pueden representar de manera matricial. Las convoluciones se aplican a la entrada para extraer características importantes y reducir la dimensionalidad antes de pasar los datos a través de múltiples capas de neuronas. La aplicación de convoluciones se realiza utilizando técnicas de álgebra lineal, lo que demuestra la importancia de esta disciplina en el desarrollo de las redes neuronales.

# 3. Relación con el cálculo

Las redes neuronales tienen una clara relación con el cálculo, en parte, debido a que utilizan funciones de activación no lineales que transforman la entrada y permiten la propagación de la señal a través de las capas ocultas. Según Goodfellow, Bengio y Courville (2016), estas funciones de activación se pueden expresar matemáticamente y se utilizan para calcular la salida de cada neurona y la salida de la red en su conjunto. Además, la propagación hacia atrás de los errores también implica el cálculo de gradientes de la función de pérdida con respecto a los parámetros de la red, lo que requiere la utilización de técnicas de cálculo como la regla de la cadena.

En términos de aprendizaje, las redes neuronales se benefician del uso de técnicas de optimización basadas en el cálculo, como el descenso de gradiente, que permiten ajustar los pesos y sesgos de las neuronas para minimizar la función de pérdida. Según Goodfellow, Bengio y Courville (2016), estos métodos utilizan el cálculo de derivadas parciales de la función de pérdida con respecto a los parámetros de la red, lo que permite actualizar los pesos y sesgos en la dirección que minimiza la pérdida.

Además, el cálculo también se utiliza en la definición de la arquitectura de las redes neuronales. Según Zhang y LeCun (2015), la elección del número de capas, neuronas y funciones de activación se basa en la capacidad de la red para modelar funciones no lineales y para realizar cálculos complejos. Esto ha llevado al desarrollo de redes neuronales profundas, que utilizan múltiples capas para aprender características de alta complejidad y realizar tareas complejas de clasificación y regresión en la inteligencia artificial.

# 4. Relación con la estadística

Las redes neuronales también están estrechamente relacionadas con la estadística en la inteligencia artificial. Según Goodfellow et al. (2016), la estadística es una herramienta importante en la creación de modelos de redes neuronales que se ajusten adecuadamente a los datos y puedan generalizar a datos no vistos. La selección de un modelo adecuado y la validación de su precisión se basan en técnicas estadísticas, como la validación cruzada y la prueba de hipótesis.

Por otra parte, la estadística es esencial en la comprensión de los métodos de regularización utilizados en las redes neuronales. La regularización se utiliza para evitar el sobreajuste, que ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a los datos nuevos. Según Hastie, Tibshirani y Friedman (2009), los métodos de regularización, como la regresión de Lasso y la regresión ridge, se basan en principios estadísticos y se pueden aplicar a redes neuronales para mejorar su capacidad de generalización.

Por último, la estadística también se utiliza en la interpretación de los resultados de las redes neuronales. Las medidas de precisión, como la precisión y el valor F, se basan en principios estadísticos y se utilizan para evaluar el rendimiento del modelo. Además, las pruebas estadísticas, como la prueba t y el análisis de varianza, se pueden utilizar para determinar la significancia estadística de las diferencias en el rendimiento entre diferentes modelos de redes neuronales.

Un ejemplo práctico de la relación entre las redes neuronales y la estadística es el uso de las redes neuronales en la clasificación de imágenes médicas. Por ejemplo, en un estudio realizado por Shen et al. (2017), se utilizó una red neuronal convolucional para clasificar imágenes de cáncer de mama. Los investigadores utilizaron un conjunto de datos de imágenes médicas de cáncer de mama etiquetadas con información clínica, y entrenaron la red neuronal para clasificar imágenes en benignas y malignas. La red neuronal fue capaz de identificar patrones relevantes en las imágenes y mejorar la precisión de la clasificación en comparación con otros métodos de clasificación. El uso de técnicas estadísticas, como el análisis de varianza y la prueba t de Student, también se utilizó para evaluar la efectividad del modelo y determinar la significancia estadística de los resultados.

# Singular Value Decomposition (SVD)

# 1. Definición

La Singular Value Decomposition (SVD) es una técnica matemática utilizada en la inteligencia artificial para la reducción de dimensionalidad y la extracción de características en grandes conjuntos de datos. Según Strang (2009), la SVD es una forma de descomponer una matriz en sus componentes principales, lo que permite la identificación de patrones y estructuras en los datos. La SVD se utiliza en una amplia variedad de aplicaciones, incluyendo el procesamiento de señales, la compresión de imágenes y la recomendación de productos en línea.

La SVD descompone una matriz en tres componentes principales: la matriz de valores singulares, la matriz de vectores singulares izquierdos y la matriz de vectores singulares derechos. Según Geron (2019), los valores singulares representan la importancia relativa de cada dimensión de la matriz, mientras que los vectores singulares representan las direcciones en las que se encuentran los patrones más significativos. La SVD permite la eliminación de dimensiones irrelevantes y la identificación de patrones importantes, lo que puede mejorar la precisión de los modelos de aprendizaje automático y reducir el tiempo de procesamiento en grandes conjuntos de datos.

# 2. Relación con el álgebra lineal

A partir de lo visto en el curso, podemos entender que este es un método de factorización de matrices que se puede usar directamente en aplicaciones como la selección de características, la visualización, la reducción de ruido y más. Los sistemas de recomendación, un subcampo del aprendizaje automático, se ocupan principalmente de los métodos de álgebra lineal. Un ejemplo simple es el cálculo de la similitud entre vectores de comportamiento de clientes dispersos utilizando medidas de distancia como la distancia euclidiana o los productos escalares. Los métodos de factorización matricial, como la descomposición de valores singulares, se utilizan ampliamente en los sistemas de recomendación para destilar datos de usuarios y elementos hasta su esencia para consultas, búsquedas y comparaciones.

La descomposición en valores singulares es un método de descomposición de matrices para reducir una matriz a sus partes constituyentes con el fin de simplificar ciertos cálculos de matrices posteriores.

Donde A es la matriz real n × m que queremos descomponer, U es una matriz m × m, Σ es una matriz diagonal m × n y es la transpuesta V de una matriz n × n

Los valores diagonales en la matriz Σ se conocen como los valores singulares de la matriz original A. Las columnas de la matriz U se denominan vectores singulares a la izquierda de A, y las columnas de V se denominan vectores singulares a la derecha de A. El SVD se calcula mediante métodos numéricos iterativos. El SVD se usa ampliamente tanto en el cálculo de otras operaciones matriciales, como la matriz inversa, como también como un método de reducción de datos en el aprendizaje automático.

Recapitulando y concluyendo, el SVD es un método matemático importante que se utiliza para reducir una matriz a sus partes constituyentes. Este método se usa comúnmente en la inteligencia artificial, especialmente en los sistemas de recomendación. Al aplicar la SVD, se puede simplificar la matriz original y obtener los valores singulares y los vectores singulares izquierdo y derecho, lo que permite simplificar ciertos cálculos matriciales y reducir la cantidad de datos necesarios para realizar ciertas operaciones.

# Principal Component Analysis (PCA)

# 1. Definición

Con base a la información obtenida en el desarrollo del curso, entendemos que, a menudo, un conjunto de datos tiene muchas columnas, quizás decenas, centenas, miles o más. Modelar datos con muchas características es un desafío, y los modelos creados a partir de datos que incluyen características irrelevantes suelen ser menos hábiles que los modelos formados a partir de los datos más relevantes. Los métodos para reducir automáticamente el número de columnas de un conjunto de datos se denominan reducción de dimensionalidad, y quizás el método más popular se denomina análisis de componentes principales, o PCA para abreviar.

Este método se utiliza en el aprendizaje automático para crear proyecciones de datos de alta dimensión tanto para visualización como para modelos de entrenamiento. El núcleo del método PCA es un método de factorización de matrices a partir del álgebra lineal. Se puede usar la descomposición propia y las implementaciones más robustas pueden usar la descomposición en valores singulares, o SVD.

# 2. Relación con el álgebra lineal

La relación entre el análisis de componentes principales (PCA) y el álgebra lineal radica en el hecho de que PCA es una técnica de reducción de dimensionalidad basada en la descomposición de valores singulares (SVD) de una matriz de datos. El SVD es una técnica fundamental en álgebra lineal para el análisis de matrices y se utiliza para calcular la matriz de covarianza en PCA. La matriz de covarianza es una matriz de tipo SVD, por lo tanto, PCA utiliza álgebra lineal para descomponer la matriz de datos en sus componentes principales, que se expresan como combinaciones lineales de las variables originales. Los componentes principales son los vectores propios de la matriz de covarianza, y representan la dirección de máxima variabilidad en los datos.

Además, el PCA también puede interpretarse como una transformación lineal de los datos originales a un nuevo espacio de características en el que las variables se han decorrelacionado. Esta transformación se realiza mediante una matriz de rotación ortogonal, que se puede calcular a partir de la matriz de covarianza. Como señala Abdi y Williams (2010), "el análisis de componentes principales es una técnica que utiliza la descomposición de valores singulares (SVD) para rotar los ejes de un espacio de datos y decorrelacionar las variables" (p. 433). Esta transformación puede ser útil en aplicaciones de aprendizaje automático, ya que puede reducir la cantidad de variables necesarias para describir los datos y mejorar la eficiencia computacional de los modelos.

# 3. Relación con la estadística

El PCA se utiliza ampliamente en el aprendizaje automático y la estadística, en particular en la exploración y visualización de datos de alta dimensión. En un estudio realizado por Jolliffe (2002), el PCA se describe como "una técnica estadística de reducción de datos que permite la representación de datos multivariantes complejos en un espacio de menor dimensión". La reducción de dimensiones permite una visualización y comprensión más efectiva de los datos. Además, el PCA también se utiliza en la regresión lineal y en la clasificación de datos para reducir la cantidad de variables y mejorar la precisión del modelo.

Otra aplicación del PCA en la estadística es su uso en el análisis de componentes independientes (ICA), que es un método para separar señales mixtas en sus fuentes originales. En un estudio realizado por Hyvärinen et al. (2001), se describe el ICA como "una técnica de descomposición ciega de señales en componentes no gaussianos independientes". El PCA se utiliza como una técnica de preprocesamiento para ICA, ya que reduce la dimensión de los datos y mejora la capacidad del modelo para separar las señales mixtas.

Para concluir y resumir, el PCA es una técnica estadística de reducción de dimensiones que se utiliza ampliamente en el aprendizaje automático y la exploración de datos. La reducción de dimensiones permite una visualización y comprensión más efectiva de los datos, así como una mejora en la precisión del modelo en la regresión y la clasificación de datos. Además, el PCA se utiliza como una técnica de preprocesamiento para ICA, una técnica para separar señales mixtas en sus fuentes originales.

# Referencias

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. Recuperado de: <https://www.deeplearningbook.org/contents/convnets.html>

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning: data mining, inference, and prediction. Springer. Recuperado de: <https://hastie.su.domains/Papers/ESLII.pdf>

Strang, G. (2009). Linear algebra and its applications. Cengage Learning. Recuperado de:  
<https://www.academia.edu/36871773/Linear_Algebra_and_Its_Applications_Fourth_Edition>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An introduction to statistical learning: with applications in R. Springer. <https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf>

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc. Recuperado de: <http://powerunit-ju.com/wp-content/uploads/2021/04/Aurelien-Geron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-OReilly-Media-2019.pdf>

Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press. Recuperado de: <https://books.google.com.co/books?hl=es&lr=&id=RC43AgAAQBAJ&oi=fnd&pg=PR7&dq=Murphy,+K.+P.+(2012).+Machine+learning:+a+probabilistic+perspective.+MIT+press.&ots=ummDexRv07&sig=MDCjH6QvBV9Yp5SXxUu0891DJkM#v=onepage&q=Murphy%2C%20K.%20P.%20(2012).%20Machine%20learning%3A%20a%20probabilistic%20perspective.%20MIT%20press.&f=false>

Zhang, Y., & LeCun, Y. (2015). Text understanding from scratch. arXiv preprint arXiv:1502.01710. Recuperado de: <https://arxiv.org/pdf/1502.01710.pdf>

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19, 221-248. Recuperado de: <https://doi.org/10.1146/annurev-bioeng-071516-044442>

Abdi, H. (2010). Singular value decomposition (SVD) and generalized singular value decomposition (GSVD). In Encyclopedia of research design (pp. 1-7). Sage. Recuperado de: <https://personal.utdallas.edu/~herve/Abdi-SVD2007-pretty.pdf>

Jolliffe, I. T. (2002). Principal component analysis. John Wiley & Sons, Ltd. Recuperado de: <http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf>

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). Independent component analysis. John Wiley & Sons, Ltd. Recuperado de: <https://www.cs.helsinki.fi/u/ahyvarin/papers/bookfinal_ICA.pdf>