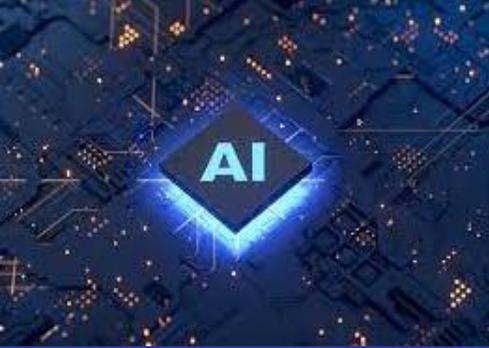


Aprendizaje No Supervisado



Miguel Murillo, M.Sc.



Definición

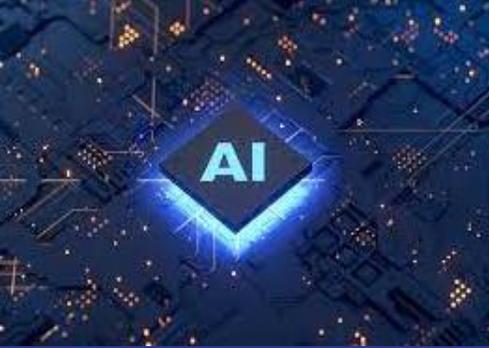
Es una técnica de aprendizaje automático en la que el modelo analiza datos **sin etiquetas predefinidas**, buscando **patrones y estructuras ocultas dentro de ellos**.

En lugar de aprender a partir de ejemplos con respuestas conocidas, el modelo **explora los datos** para **identificar agrupaciones, relaciones o características subyacentes**.

En esta sesión aprenderemos acerca de dos técnicas de aprendizaje automático no supervisado:

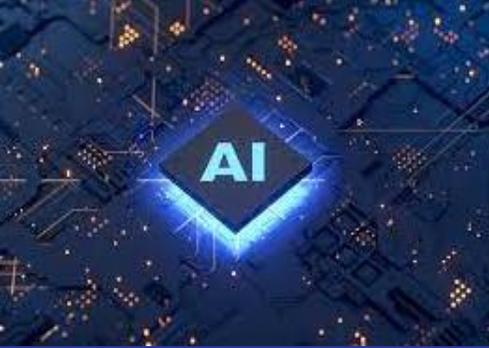
- **Reducción de dimensionalidad**
- **Clustering y detección de anomalías**





Reducción de dimensionalidad





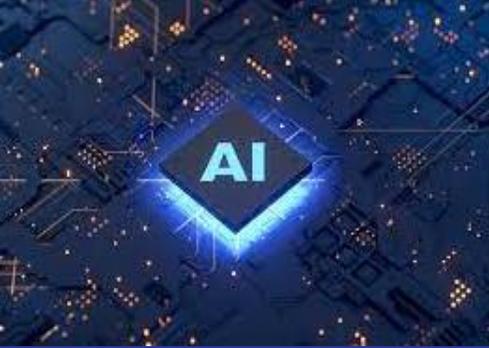
Definición

Es un proceso que transforma datos con **muchas características** en una representación más **compacta con menos características**. El objetivo de esto es:

- **Simplificar los datos:** Reducir la complejidad y hacer que el análisis sea más manejable.
- **Eliminar ruido:** Quitar características menos importantes que no aportan información útil.
- **Mejorar el rendimiento:** Hacer que los algoritmos de aprendizaje automático funcionen de manera más eficiente al reducir el tiempo de procesamiento y el riesgo de sobreajuste (**overfitting**).

En situaciones en las que se tienen muchas características y pocos datos es posible caer en sobreajuste debido al problema de la “maldición de la dimensionalidad” (curse of dimensionality).

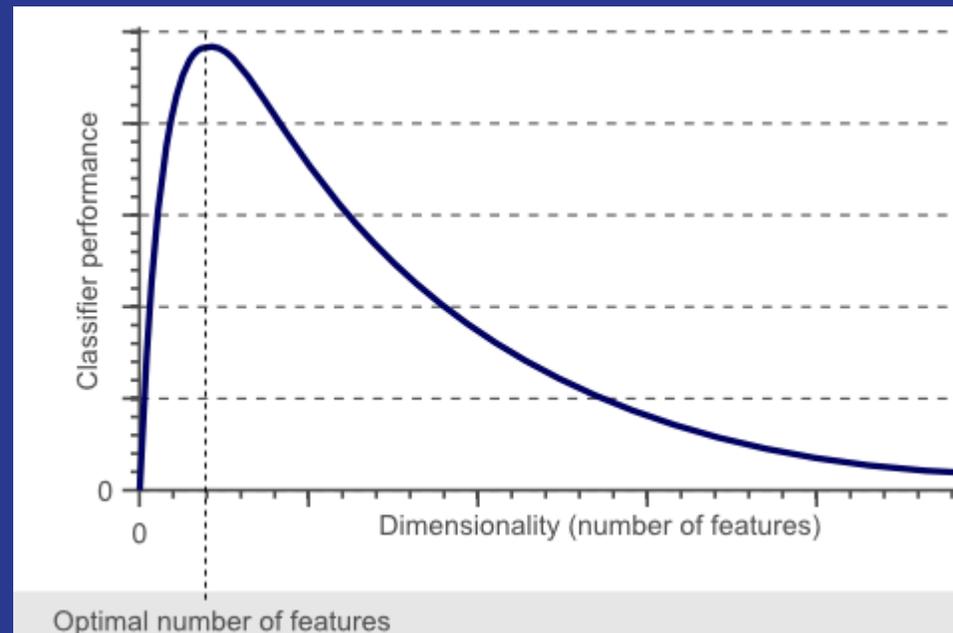


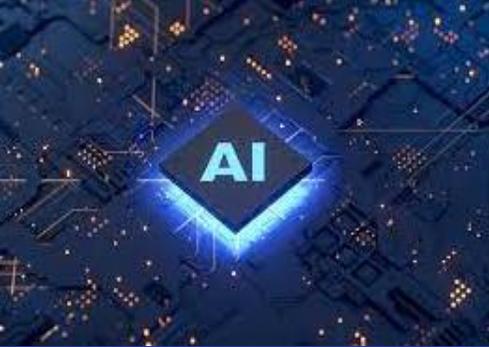


Curse of dimensionality

Se presenta cuando hay un desbalance desproporcionado entre el número de muestras o datos y el número de características, **siendo este último ampliamente superior**.

Esto implica una reducción drástica en el performance de un modelo de aprendizaje automático debido al sobreajuste (overfitting).





Algoritmos de reducción de dimensionalidad

A continuación se ilustran algunos de los algoritmos más famosos:

Análisis de componentes principales (PCA)

Análisis discriminante lineal (LDA)

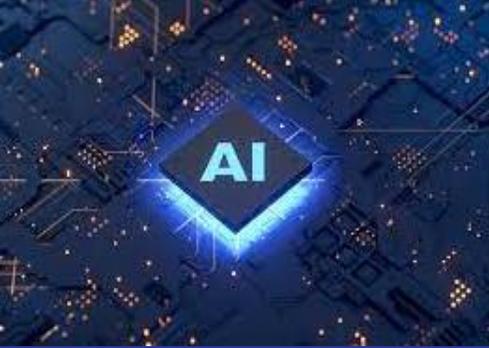
Aproximación y proyección de Manifold uniforme (UMAP)

Autoencoders

Análisis de componentes independientes (ICA)

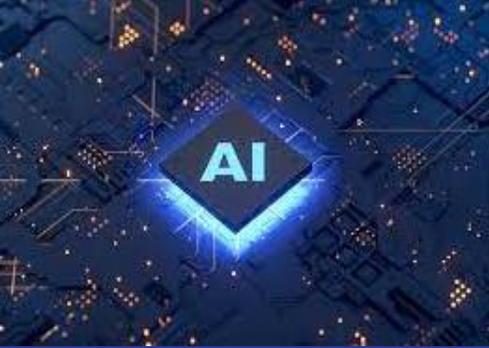
Análisis factorial





Análisis de componentes principales (PCA)

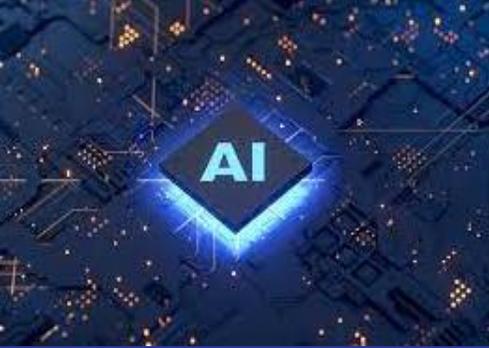




Partiendo del problema

- Hay variables que pueden tener **demasiada correlación**.
 - Varias variables pueden estar midiendo **lo mismo** bajo distintos puntos de vista.
- Cuando hay **mayor varianza** en los datos se considera que existe **mayor información** por parte de ellos.

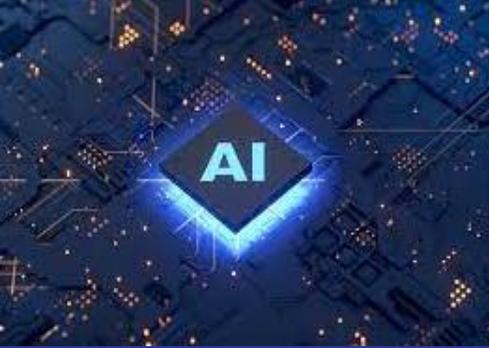




Definición

- Se utiliza para describir un conjunto de datos en términos de nuevas variables (**componentes**) que **no están correlacionadas entre sí**.
- Se calculan las nuevas variables y se ordenan en **modo descendente** en función de su **varianza**.
 - Estas nuevas variables son **combinaciones lineales** de las originales.



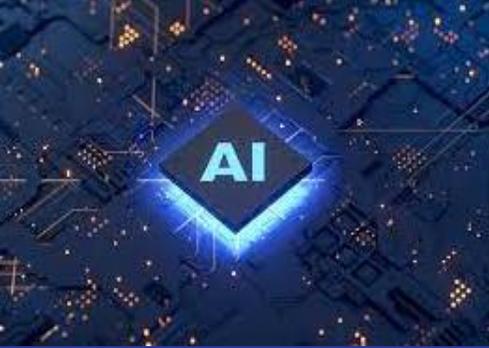


Las nuevas variables están **incorrelacionadas** y sus varianzas van **decreciendo progresivamente**.

$$\{x_1, x_2, \dots, x_m\} \rightarrow \{y_1, y_2, \dots, y_m\}$$
$$\text{Var}(y_1) > \text{Var}(y_2) > \dots > \text{Var}(y_m)$$

El algoritmo para el cálculo de las variables se ilustra a continuación:



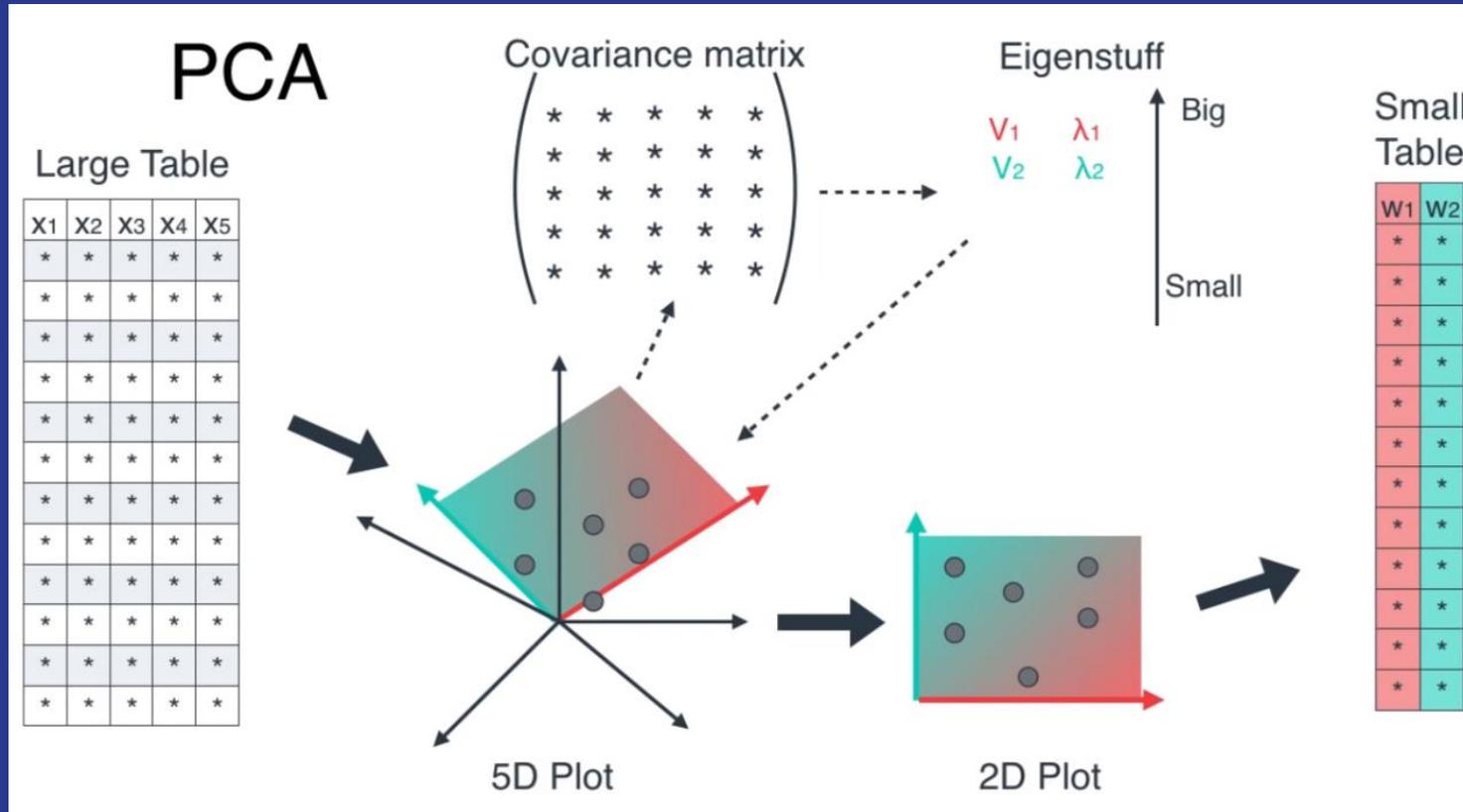
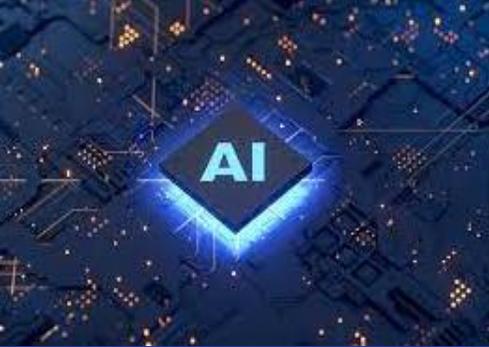


$$y_i = \sum_{j=1}^m a_{ij}x_j = A_i \cdot X$$

$$||A_i|| = 1 \forall i = 1, \dots, m$$

- La **primera componente** se calculará usando el vector A_1 para $i = 1$, para este caso se asegura la **máxima varianza** con la condición $||A_1|| = 1$
 - La **segunda componente** se calculará usando A_2 de modo que $i = 2$ está incorrelacionada con $i = 1$ y así sucesivamente.
- Elegir el A_1 que maximice la varianza de $i = 1$ sujeta a la restricción $||A_1|| = 1$ es un problema de optimización que se puede resolver utilizando el método de los **multiplicadores de Lagrange, la matriz de correlaciones, los valores y vectores propios de la misma, así como la matriz de covarianzas.**

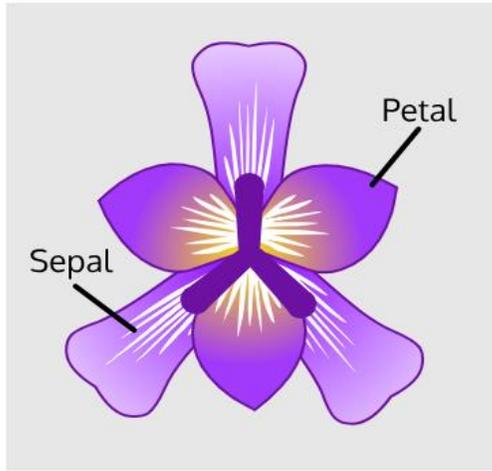




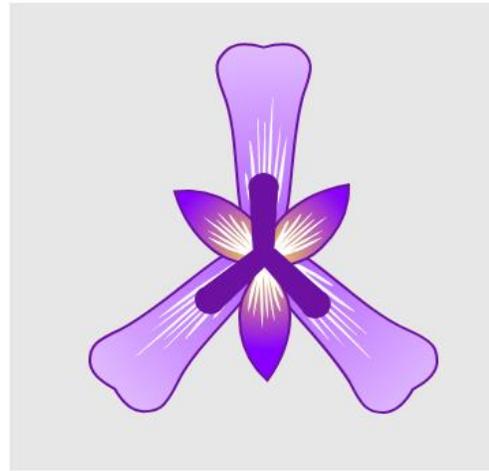
AI

Dato curioso

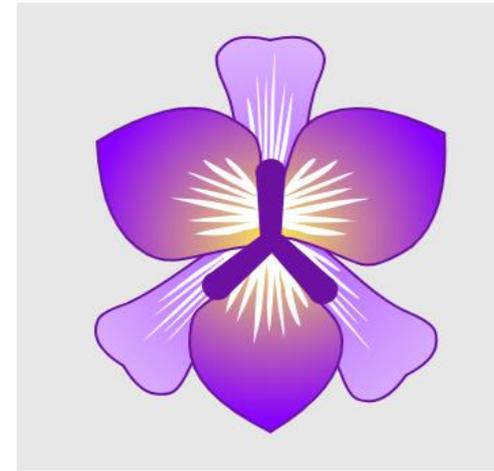
Estos 3 tipos de flores tienen características similares y distintas a la vez



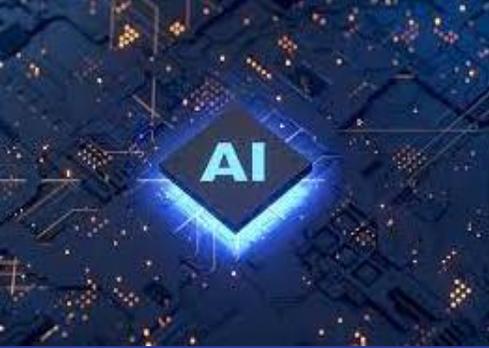
Iris Versicolor



Iris Setosa



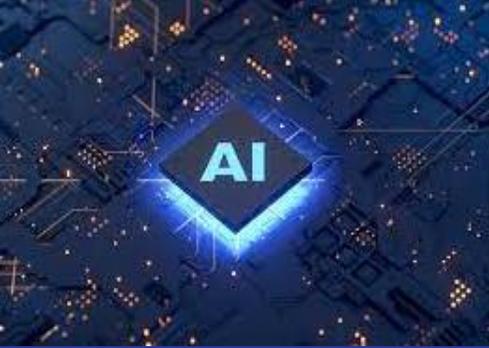
Iris Virginica



Código práctico de uso de PCA

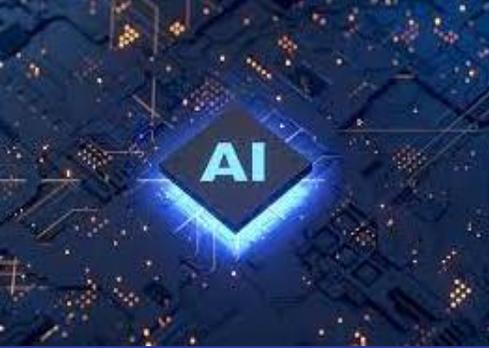
¡Vamos a probar!





Autoencoders



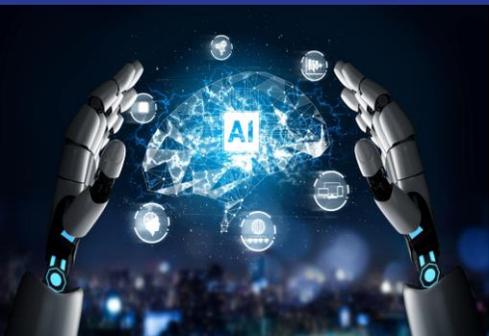
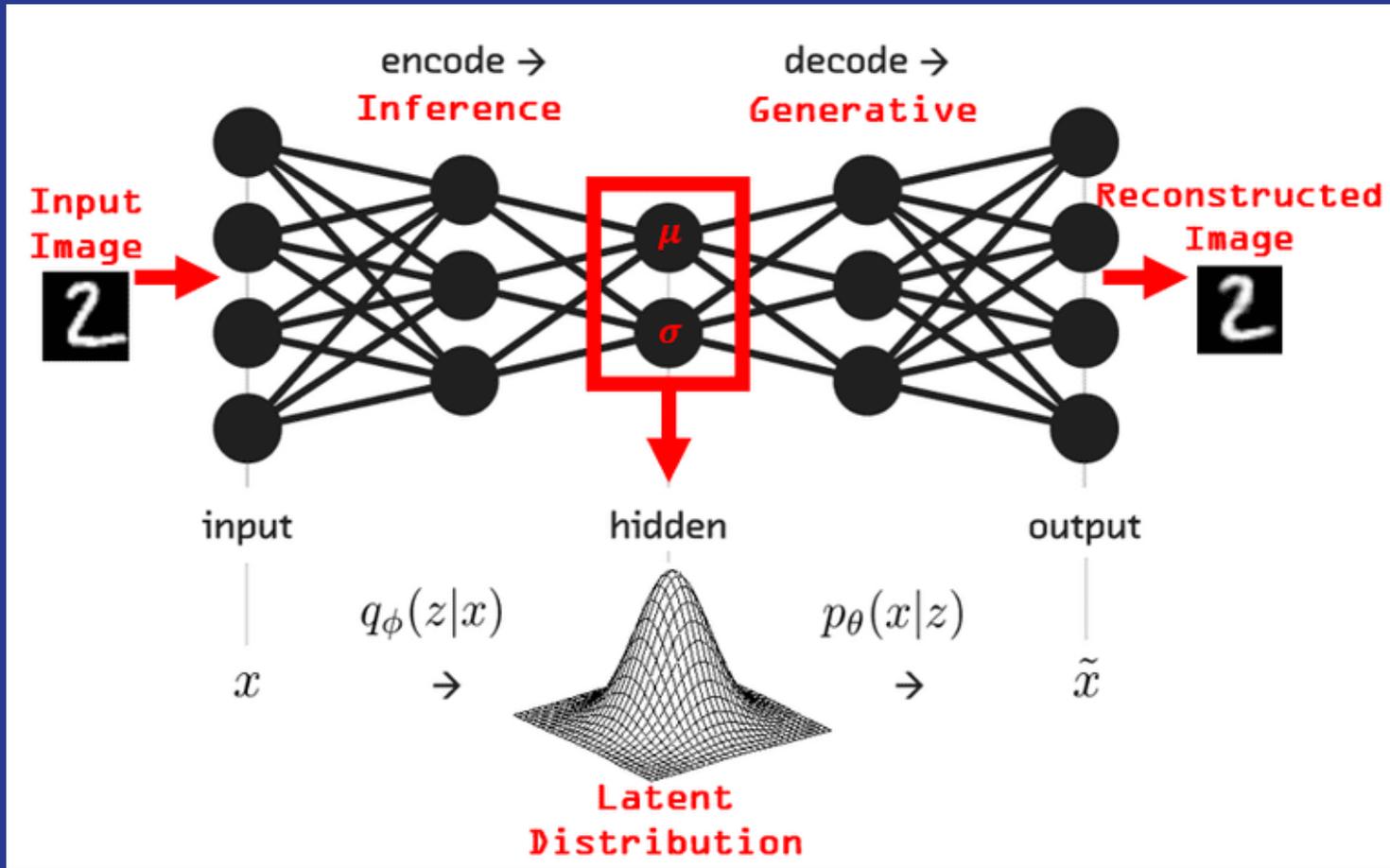
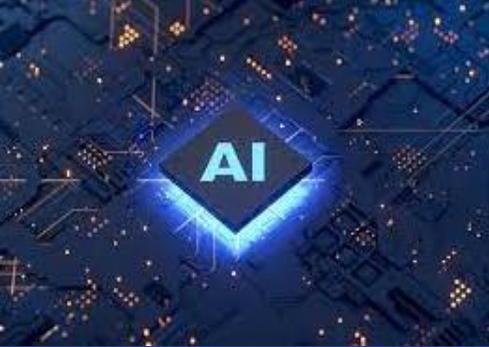


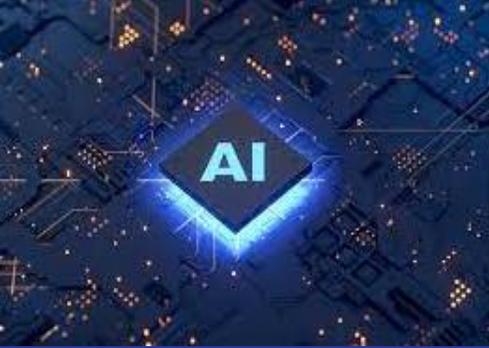
Definición

Es un tipo de red neuronal que se utiliza **principalmente para la reducción de dimensionalidad y la extracción de características**. Está compuesto por dos partes principales:

- **Codificador (Encoder):** Esta parte toma los datos de entrada y **los transforma en una representación de menor dimensión, conocida como el espacio latente o embeddings**. El objetivo del codificador es **aprender una representación compacta y útil de los datos de entrada**.
- **Decodificador (Decoder):** Esta parte toma la representación de menor dimensión generada por el codificador **y trata de reconstruir los datos originales**. La idea es que el decodificador **aprenda a reconstruir los datos de entrada a partir de la representación comprimida, minimizando la diferencia entre los datos originales y los reconstruidos**.

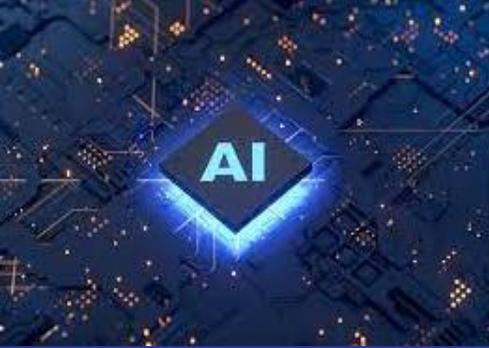






Clustering

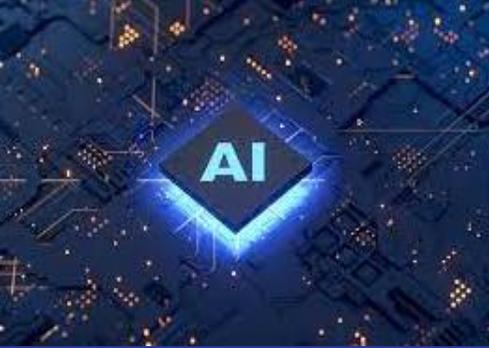




Definición

- Es una técnica de aprendizaje no supervisado que consiste en **agrupar un conjunto de datos en subconjuntos o clusters de tal manera que los objetos dentro de cada cluster sean más similares entre sí que a los de otros clusters.**
- Esta similitud se basa generalmente en una medida de **distancia o de similitud entre los datos.**
- El objetivo del clustering es **identificar una estructura subyacente en los datos, sin tener etiquetas o categorías predefinidas.**





Algoritmos de clustering

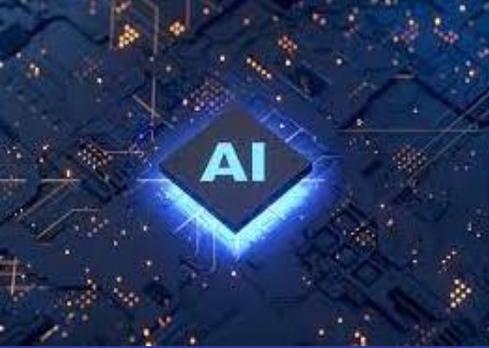
A continuación se ilustran algunos de los algoritmos más famosos:

K-means

DBSCAN (Density-
Based Spatial
Clustering of
Applications with
Noise)

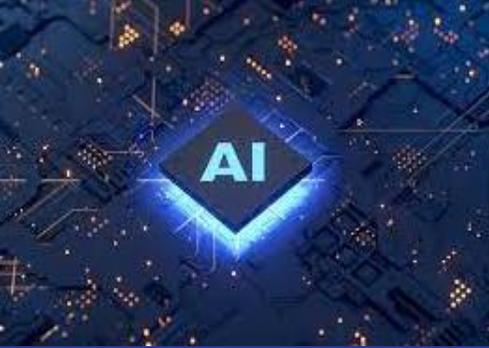
Agglomerative
Hierarchical
Clustering





K-means

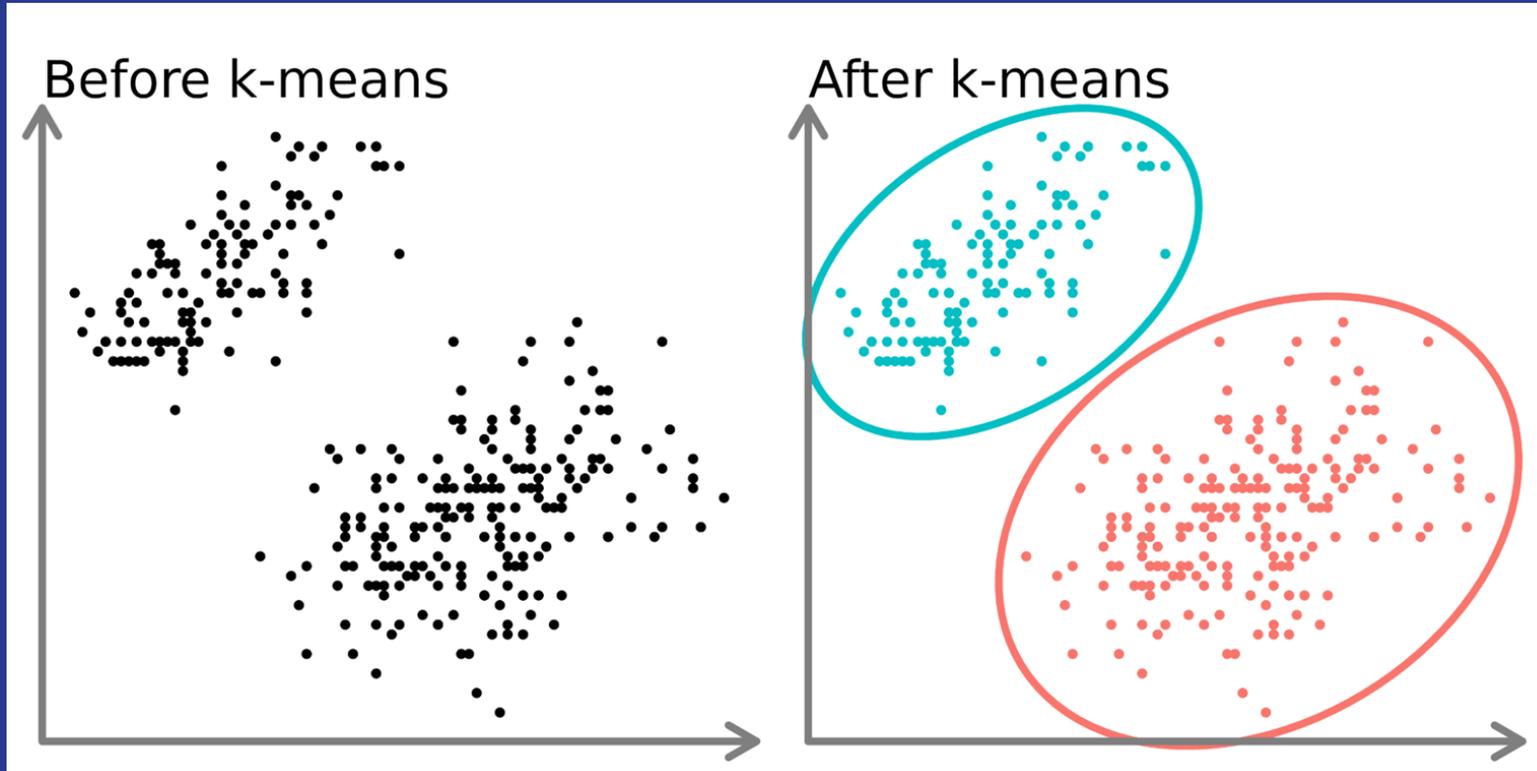
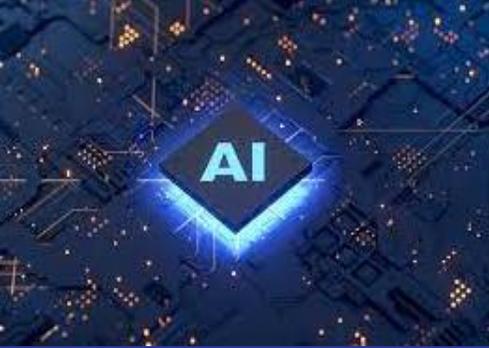


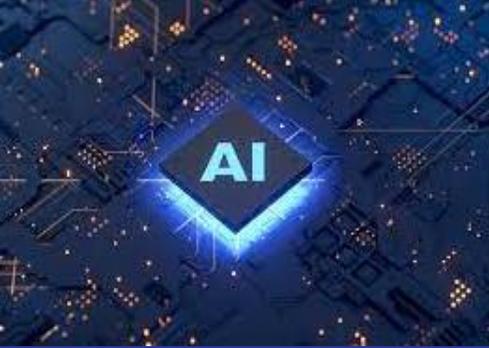


Definición

- Es un algoritmo de clustering que agrupa datos en un número predefinido de clusters (K).
- Funciona asignando cada dato al cluster cuyo centroide (punto medio) esté más cercano.
- Luego, **recalcula** los centroides de cada cluster y **repite el proceso hasta que las asignaciones no cambien o se alcance un criterio de convergencia.**
 - **El objetivo es minimizar la variación dentro de cada cluster.**



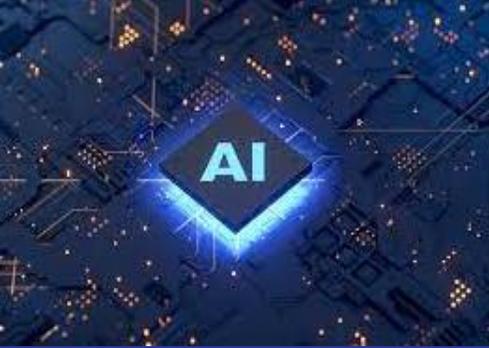




Código práctico de uso de clustering

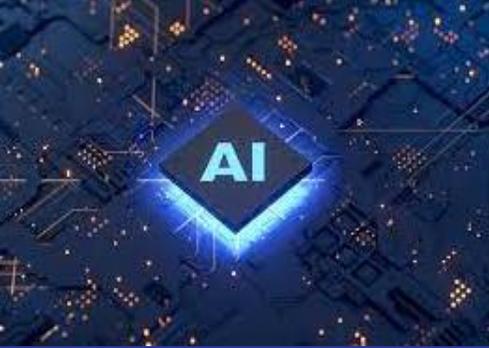
¡Vamos a probar!





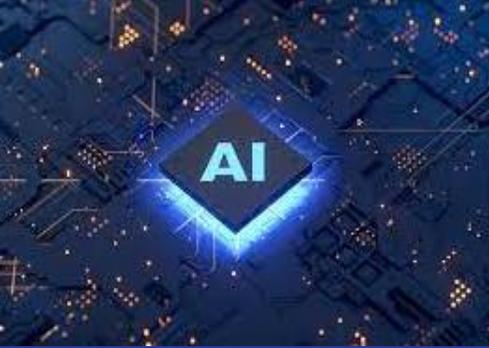
Detección de anomalías

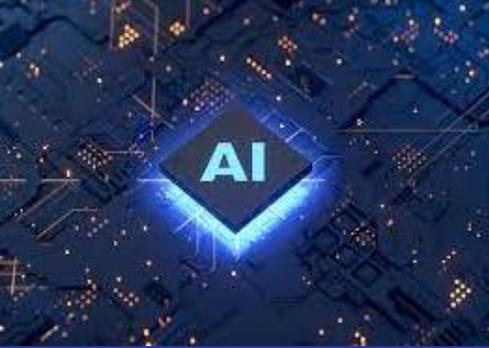




Definición



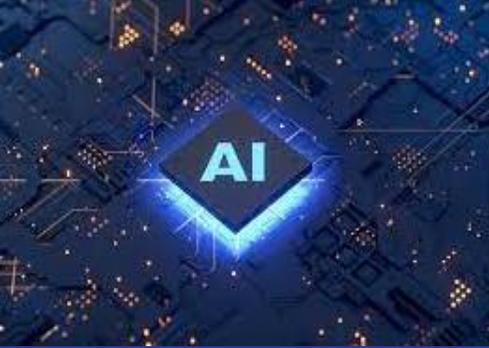




Código práctico de uso de detección de anomalías

¡Vamos a probar!





Gracias por su atención
¿Dudas?

