

Curso de IA y Big Data

CPR Badajoz

Francisco Pajuelo Holguera

Sesión 1: Contenidos

- Introducción a la IA
- Anaconda y Jupyter Notebook
- Librerías básicas de IA: matplotlib, pandas, numpy
- Machine Learning. Qué es y tipos

Introducción a la IA

- La **inteligencia artificial** es el conjunto de sistemas o combinación de algoritmos, cuyo propósito es crear máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar conforme la información que recopilan.
- La inteligencia artificial no tiene como finalidad reemplazar a los humanos, sino mejorar significativamente las capacidades y contribuciones humanas.

Introducción a la IA

Stuart J. Russell y Peter Norvig diferencian varios tipos de inteligencia artificial:

- **Sistemas que piensan como humanos:** Estos sistemas tratan de emular el pensamiento humano; por ejemplo las redes neuronales ,actividades como la toma de decisiones, resolución de problemas y aprendizaje.
- **Sistemas que actúan como humanos:** Estos sistemas tratan de actuar como humanos; es decir, imitan el comportamiento humano; por ejemplo la robótica.

Introducción a la IA

- **Sistemas que piensan racionalmente:** Es decir, con lógica (idealmente), tratan de imitar el pensamiento racional del ser humano; por ejemplo, los sistemas expertos, (el estudio de los cálculos que hacen posible percibir, razonar y actuar).
- **Sistemas que actúan racionalmente:** Tratan de emular de forma racional el comportamiento humano; por ejemplo los agentes inteligentes, que está relacionado con conductas inteligentes en artefactos.

Introducción a la IA



Anaconda y Jupyter notebook



Jupyter Notebook es un entorno de trabajo interactivo web que permite desarrollar código en Python de manera dinámica, a la vez que integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes.

Te permite editar y ejecutar documentos de notebook a través de cualquier navegador web, e incluso dentro de Pycharm.

Anaconda y Jupyter notebook



- Windows:
 - Instalación de Python:
<https://www.python.org/downloads/release/python-31010/>
 - Añadir Python al **PATH**:
C:\Users\Francisco\AppData\Local\Programs\Python\Python311\
 - C:\Users\Francisco\AppData\Local\Programs\Python\Python311\Scripts
 - Instalar PIP:
 - Descargamos el archivo: <https://bootstrap.pypa.io/get-pip.py>
 - Abrimos la CMD y ejecutamos -> py get-pip.py
- Linux:
 - Python ya instalado.
 - Sudo apt-get install jupyter-notebook
 - Ejecutar jupyter-notebook

Anaconda y Jupyter notebook



Anaconda es una de las mejores distribuciones de Python. Está orientada a dar soporte integral al desarrollo científico con Python tanto analítico como gráfico. Entre sus características destacan:

- Aplicación gráfica que integra las herramientas y versiones de Python orientadas al desarrollo científico.
- IDE para desarrollar fácilmente aplicaciones científicas, tanto en código como creación de gráficos.
- Gestor de paquetes integrado (Conda).
- Integración avanzada con Jupyter Notebook.

<https://www.anaconda.com/products/distribution>

Anaconda y Jupyter notebook

- Arrancar Anaconda Navigator
- Lanzar Jupyter Notebook

Librerías básicas de IA: NumPy

NumPy es un módulo de Python. El nombre viene de PYthon NUMérico. Es una librería de código abierto muy utilizada para la generación y manejo de matrices multidimensionales. En su mayor parte está escrito en C para optimizar su velocidad de ejecución.

Además, los objetos de Pandas (librería que veremos próximamente) dependen en gran medida de los objetos de NumPy ya que Pandas extiende a NumPy.

Librerías básicas de IA: Pandas

Pandas es un popular paquete de datos de Python porque ofrece unas estructuras de datos potentes y flexibles que facilitan la manipulación y análisis de datos de alto rendimiento. Entre las estructuras más utilizadas se encuentra el DataFrame. Es de código abierto y su nombre viene de PANel DAta.

Usando esta librería podemos lograr los 5 pasos típicos en el procesamiento y análisis de datos:

- Cargar datos
- Modelar
- Analizar
- Manipular
- Preparar

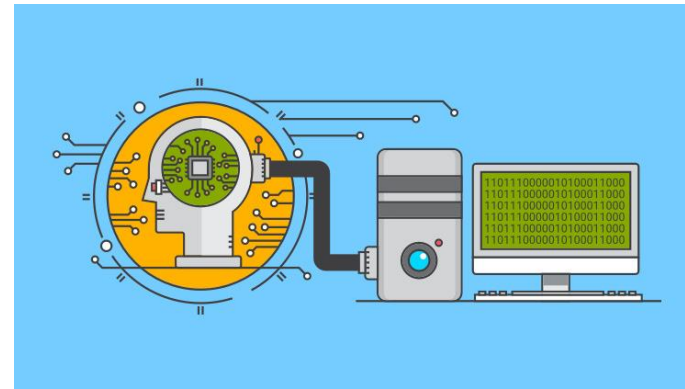
Librerías básicas de IA: Matplotlib

Matplotlib es un popular paquete de visualización de datos de Python.

Es una librería de dibujo de trazos 2D, lo que implica que dibujará gráficos de líneas. Estos son muy útiles tanto para presentar los resultados obtenidos como para empezar a analizar y comprender el conjunto de datos del que partimos.

Es muy sencilla de utilizar: se importa la librería, preparamos unos datos y mostramos por pantalla con la función "plot()".

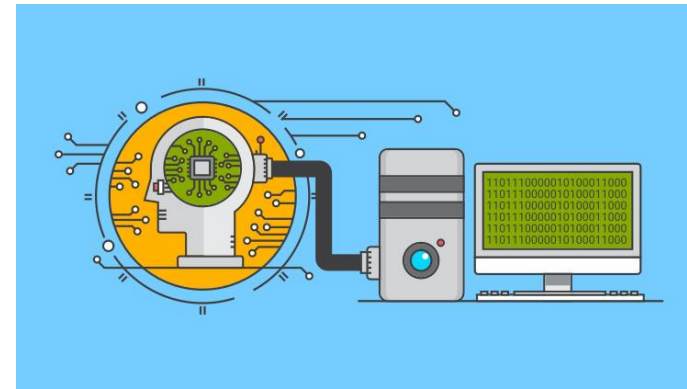
Machine Learning



Se puede definir ML como una disciplina dentro del campo de la informática, muy ligada a la Inteligencia Artificial (IA), mediante la cual se generan sistemas que aprenden de manera automática por sí mismos.

En realidad, el sistema que realmente aprende es un algoritmo que es capaz de valorar todos los datos de entrada y, a partir de ahí, realizar varias acciones, como pueden ser predecir valores o recomendar valores futuros.

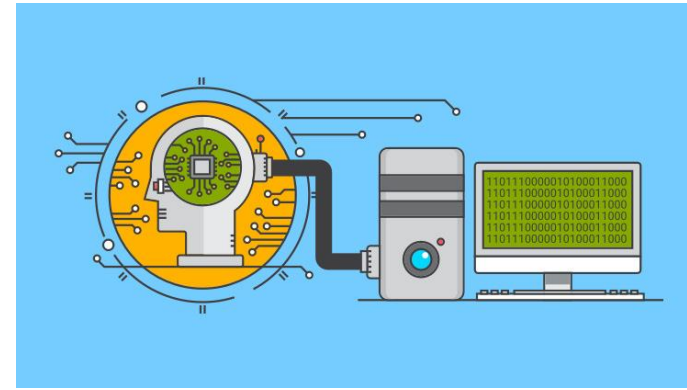
Tipos de ML



- 1. Aprendizaje Supervisado** (SL, Supervised Learning). En este tipo de ML se dispone de un conjunto de datos o dataset llamado de entrenamiento. El algoritmo entrena al sistema proporcionándole una parte de los datos de entrada como datos de entrenamiento, que son datos etiquetados.

Por ejemplo, si tenemos fotos de animales, en concreto de perros y gatos, se etiquetarán estos datos como tales. Una vez que tengamos una gran cantidad de datos, para introducir nuevos datos no hará falta etiquetarlos ya que el sistema, a partir de estos datos de entrenamiento, habrá aprendido qué datos son de perros y cuáles son de gatos.

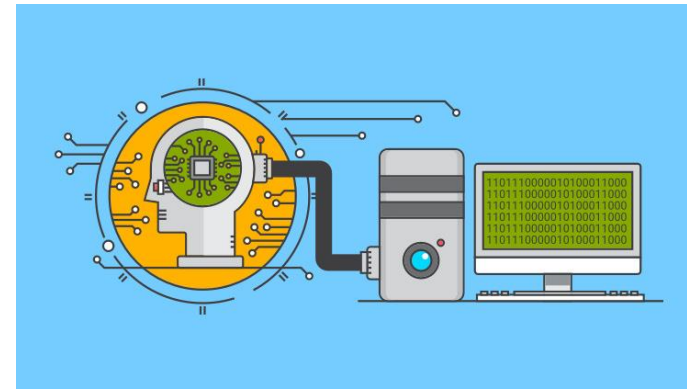
Tipos de ML



2. **Aprendizaje No Supervisado** (NSL, Non-Supervise Learning).

Este tipo de ML no utiliza etiquetas para los datos. También hay datasets, pero no con datos etiquetados. El objetivo que tiene el algoritmo NSL es aprender de manera automática patrones interesantes que puedan existir en los datos de entrada.

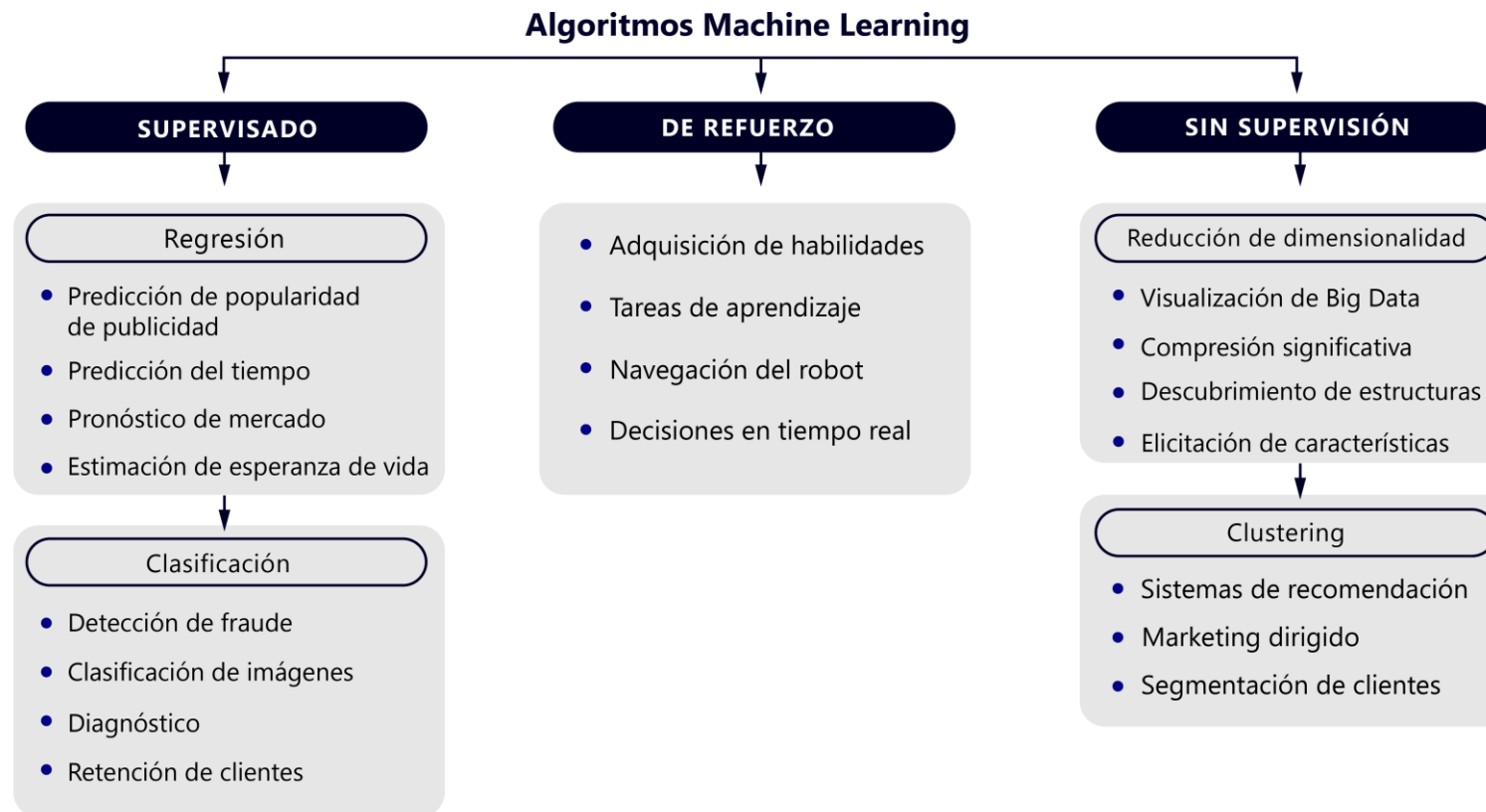
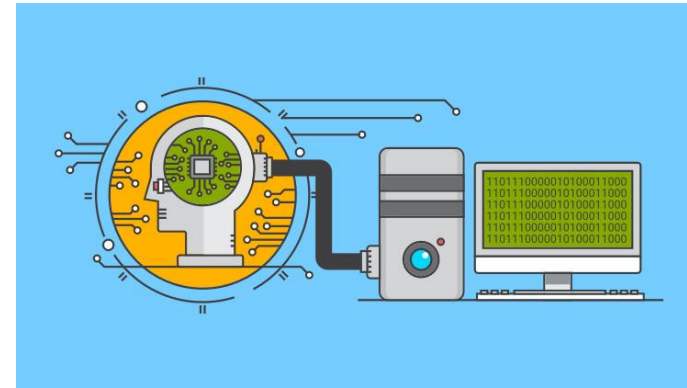
Tipos de ML



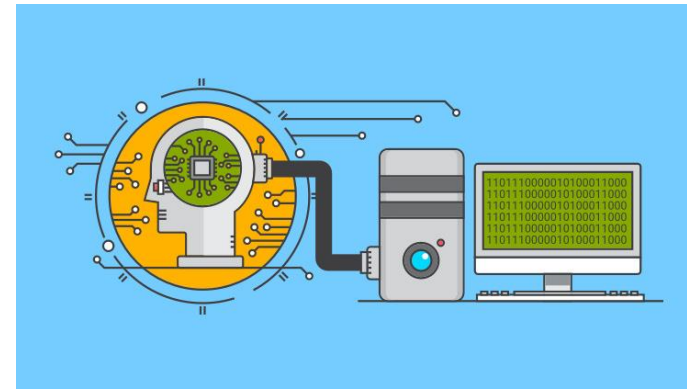
3. Aprendizaje por Refuerzo (RL, Reinforcement Learning). En este tipo de ML los sistemas aprenden a partir de la experiencia o reacción ante un evento.

Este es un método típico que se está aplicando actualmente en el desarrollo de sistemas de conducción autónoma en vehículos. Cuando el coche toma una decisión errónea, se le penalizará, de acuerdo con un registro de valores. De esta forma, el sistema toma mejores decisiones futuras a base de "castigos" y de "premios".

Tipos de ML



Técnicas de SL

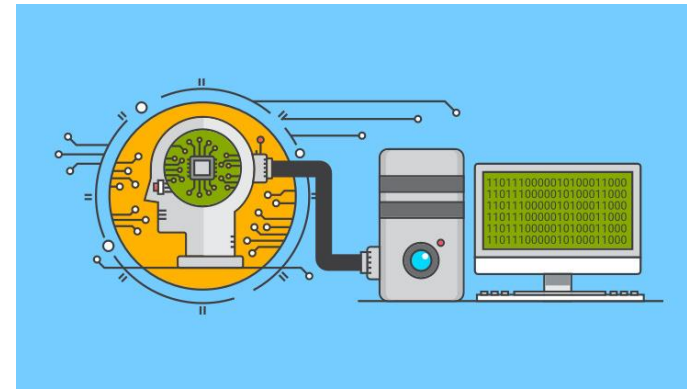


Regresión lineal: Es un algoritmo de aprendizaje supervisado que se utiliza en ML y por lo tanto en IA.

Es uno de los algoritmos más sencillos de entender pues se basa en "dibujar una recta" que nos indica la tendencia de un conjunto de datos continuos (si fueran discretos, utilizaríamos Regresión Logística).

De este modo, a través de esta tendencia, es posible la predicción de comportamientos a través de las características de entrada de un determinado modelo.

Técnicas de SL



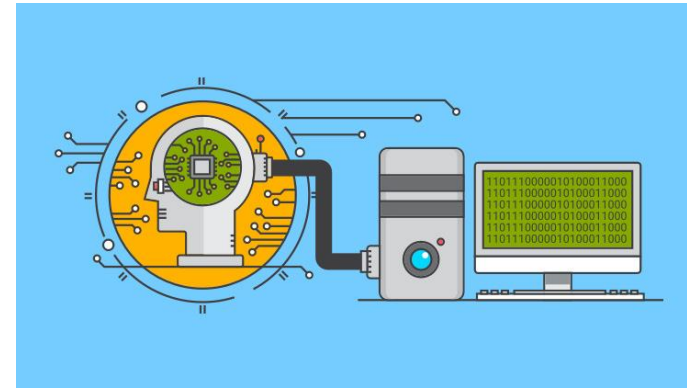
En estadística, regresión lineal es una aproximación para modelar la relación entre una variable escalar dependiente "y" y una o mas variables explicativas nombradas con "X".

Recordemos rápidamente la fórmula de la recta:

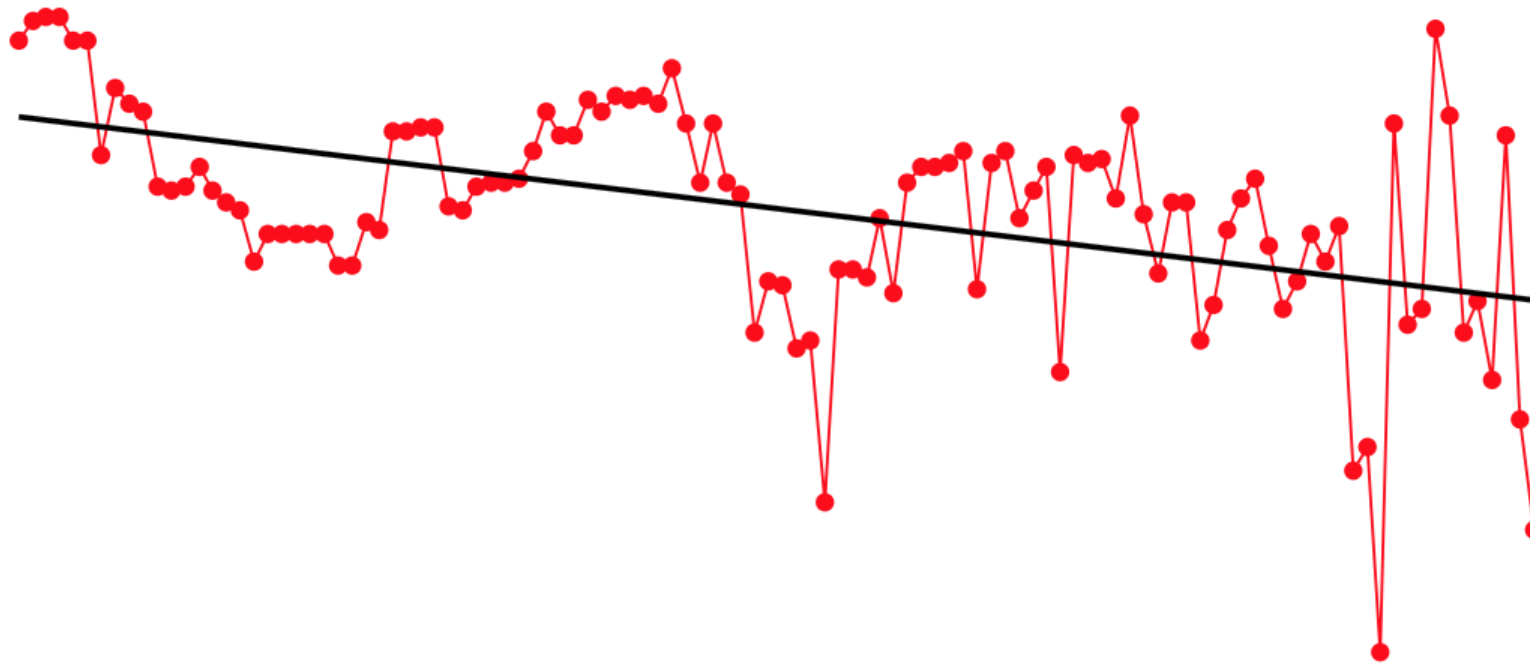
$$Y = mX + b$$

Donde Y es el resultado, X es la variable, m la pendiente (o coeficiente) de la recta y b la constante o también conocida como el "punto de corte con el eje Y" en la gráfica (cuando X=0)

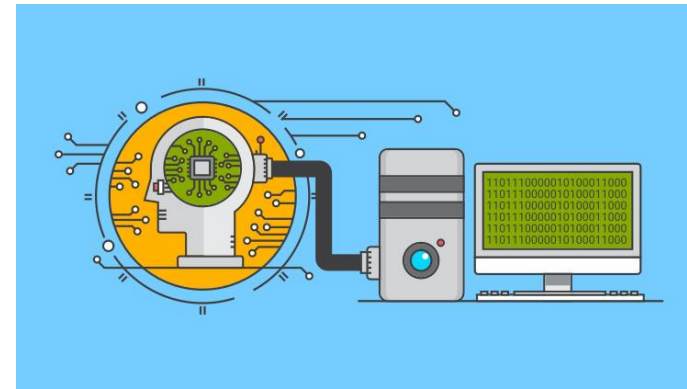
Técnicas de SL



The development in Pizza prices in Denmark from 2009 to 2018



Técnicas de SL

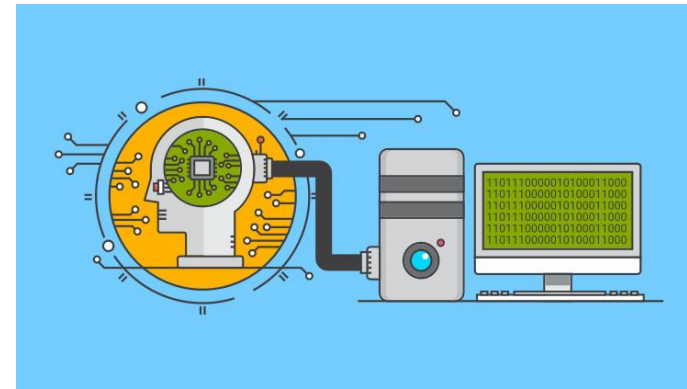


Regresión logística: Es una técnica de análisis de datos que utiliza las matemáticas para encontrar las relaciones entre dos factores de datos. Luego, utiliza esta relación para predecir el valor de uno de esos factores basándose en el otro. Normalmente, la predicción tiene un número finito de resultados, como un sí o un no.

Algunos Ejemplos de Regresión Logística son:

- Clasificar si el correo que llega es Spam o No es Spam
- Dados unos resultados clínicos de un tumor clasificar en "Benigno" o "Maligno".
- El texto de un artículo a analizar es: Entretenimiento, Deportes, Política ó Ciencia
- A partir de historial bancario conceder un crédito o no

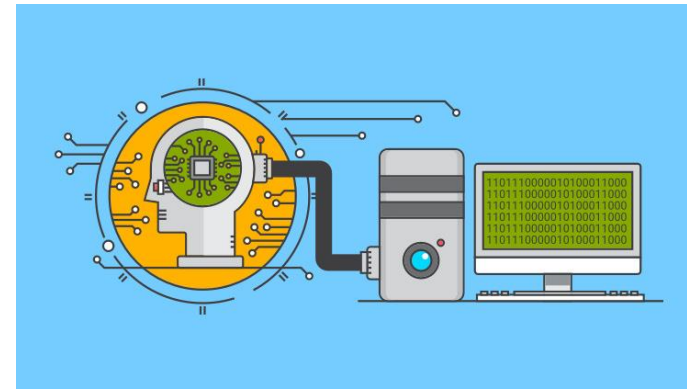
Técnicas de NSL



K-means clustering: Es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características.

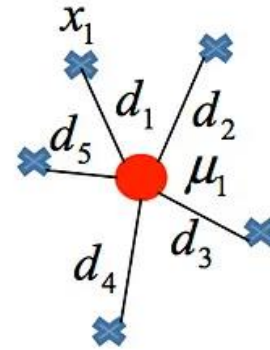
El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.

Técnicas de NSL



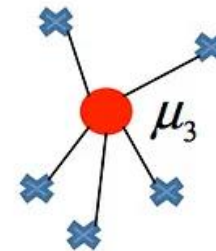
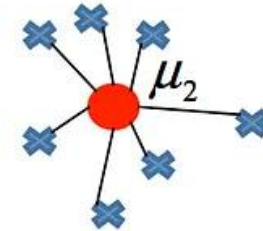
K-means clustering

● Centroid
✕ Sample



$$\sum_{x_j \in S_1} d_j^2 = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

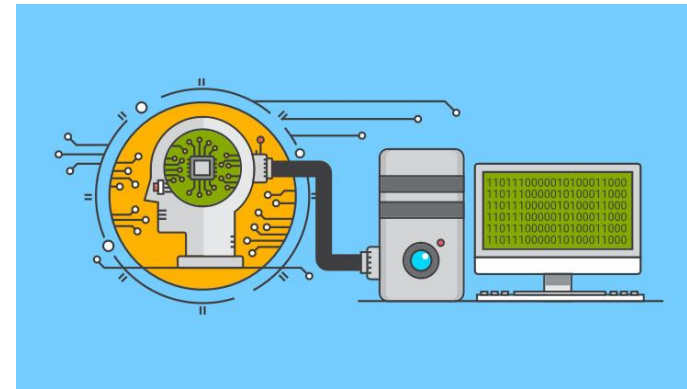
$$\sum_{x_j \in S_2} d_j^2$$



$$\sum_{x_j \in S_3} d_j^2$$

$$\min_S E(\mu_i) = \sum_{x_j \in S_1} d_j^2 + \sum_{x_j \in S_2} d_j^2 + \sum_{x_j \in S_3} d_j^2$$

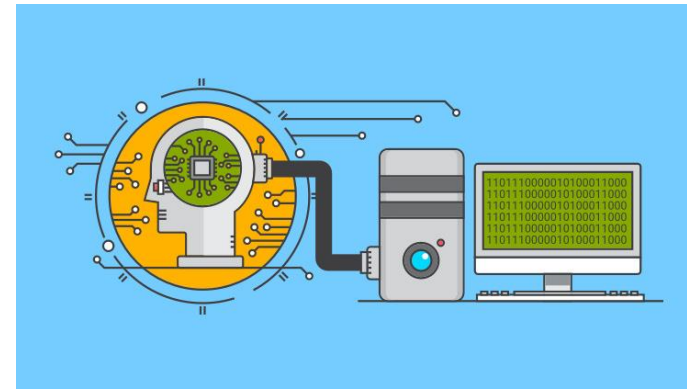
Técnicas de NSL



K-means clustering: Es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características.

El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.

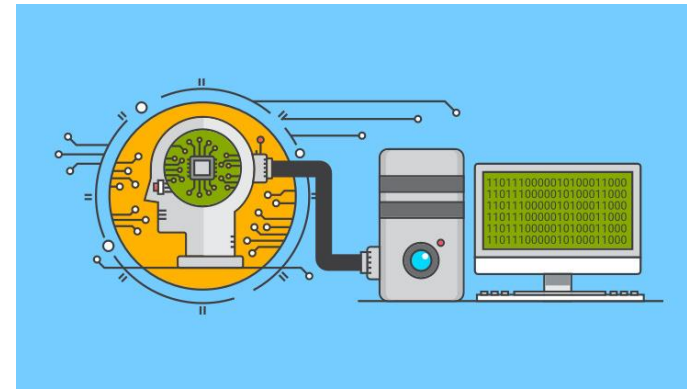
Técnicas de NSL



K-means clustering: El algoritmo consta de tres pasos:

- Inicialización: una vez escogido el número de grupos, k , se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.
- Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.
- Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Técnicas de NSL



K-means clustering:

Gráficamente, podríamos resumir que k-means agrupará personas con las mismas características. En la figura, según sus características; hombres delgados, muy fuertes y otros. En ningún momento etiquetamos personas, como lo haría el Aprendizaje Supervisado.

