Regresión Simple

Regresión Múltiple

Regresión Polinómica

Clasificación Binaria

Clasificación Multiclases

Clasificación Multi Etiqueta

Cluster Jerárquico

K (# de grupos)

¿Qué librerías son utilizadas?

* Manipulación de datos
* Visualización
* Reportes
* Análisis estadístico

Error Cuadrático Medio

Error Absoluto medio

R2

Regresión Logística

K vecinos más cercanos

Números de vecinos

K-means (Agrupación)

Maquina de vectores de soporte

Hiper Plano

Márgen Máximo

Diferencia entre módulo y librería

* **Regresión Simple**: Es un método que se usa para predecir un valor (como el precio de una casa) basándose en una sola variable (como el tamaño de la casa). Busca encontrar la mejor línea que se ajuste a los datos.
* **Regresión Múltiple:** Similar a la regresión simple, pero aquí utilizamos varias variables para hacer la predicción. Por ejemplo, predecir el precio de una casa usando el tamaño, la ubicación y el número de habitaciones.
* **Regresión Polinómica:** Es una extensión de la regresión que permite ajustar una curva en lugar de una línea recta. Esto es útil cuando los datos tienen una relación más compleja.
* **Clasificación Binaria:** Es un tipo de problema donde se asigna una de dos posibles categorías a un dato. Por ejemplo, clasificar correos electrónicos como "spam" o "no spam".
* **Clasificación Multiclases:** A diferencia de la clasificación binaria, aquí un dato puede pertenecer a más de dos categorías. Por ejemplo, clasificar frutas como "manzana", "naranja" o "plátano".
* **Clasificación Multi Etiqueta:** En este caso, un dato puede pertenecer a múltiples categorías al mismo tiempo. Por ejemplo, una película puede ser clasificada como "acción" y "comedia" a la vez.
* **Cluster Jerárquico:** Es una técnica que agrupa datos en una estructura de árbol, donde los grupos más pequeños se combinan para formar grupos más grandes. Es útil para entender cómo se relacionan los datos entre sí.
* **K (# de grupos):** En técnicas de agrupamiento (como K-means), "K" se refiere al número de grupos que queremos identificar en nuestros datos.
* ¿Qué librerías son utilizadas?: En programación, las librerías son colecciones de funciones y herramientas que facilitan tareas específicas. En Python, por ejemplo, se usan librerías como Pandas para manipulación de datos, Matplotlib para visualización y Scikit-learn para aprendizaje automático.
* **Manipulación de datos:** Se refiere a las operaciones que realizamos sobre los datos, como limpiar, transformar o combinar conjuntos de datos para prepararlos para análisis o modelado.
* **Visualización:** Es el proceso de representar datos gráficamente, como con gráficos o diagramas, para facilitar su comprensión.
* **Reportes:** Son documentos que resumen los hallazgos de un análisis de datos, presentando resultados y conclusiones de manera clara y concisa.
* **Análisis estadístico:** Es el uso de métodos matemáticos y estadísticos para interpretar datos y sacar conclusiones. Incluye técnicas como promedios, medianas y pruebas de hipótesis.
* **Error Cuadrático Medio (MSE):** Es una medida que indica cuánto se desvían las predicciones de un modelo respecto a los valores reales. Se calcula promediando los cuadrados de las diferencias.
* *Error Absoluto Medio (MAE):* Similar al MSE, pero aquí se promedian las diferencias absolutas entre las predicciones y los valores reales. Es menos sensible a los valores extremos.
* **R² (Coeficiente de determinación):** Es una medida que indica qué tan bien se ajusta un modelo a los datos. Un valor de 1 significa que el modelo explica perfectamente los datos.
* Regresión Logística: Es un tipo de regresión que se utiliza para problemas de clasificación binaria. Predice la probabilidad de que un dato pertenezca a una categoría específica.
* **K vecinos más cercanos (K-NN):** Es un algoritmo de clasificación que asigna una categoría a un dato basándose en las categorías de sus "K" vecinos más cercanos en el espacio de características.
* **Números de vecinos:** Se refiere a cuántos vecinos se consideran en el algoritmo K-NN. Este número puede afectar la precisión del modelo.
* **K-means (Agrupació**n): Es un método de agrupamiento que divide los datos en "K" grupos basándose en la similitud. Cada grupo tiene un centroide que representa el promedio de los datos en ese grupo.
* **Máquina de vectores de soporte (SVM)**: Es un algoritmo de clasificación que busca encontrar el mejor límite (hiperplano) que separa diferentes clases en los datos.
* **Hiperplano:** Es un concepto en geometría que se refiere a un límite que separa diferentes grupos en un espacio multidimensional. En 2D sería una línea, y en 3D sería un plano.
* **Márgen Máximo:** En SVM, se refiere a la distancia más grande entre el hiperplano y los puntos de datos más cercanos de cada clase. Un margen más grande generalmente indica un mejor modelo.
* **Diferencia entre módulo y librería**: En programación, un "módulo" es un archivo que contiene código, mientras que una "librería" es una colección de módulos que proporcionan funcionalidades específicas. Las librerías se utilizan para facilitar tareas comunes sin tener que escribir todo el código desde cero.

Una **aplicación típica** es un término que se refiere a programas de software diseñados para realizar tareas comunes o cotidianas en un entorno de computación. Estas aplicaciones están orientadas a resolver necesidades específicas de los usuarios y son utilizadas con frecuencia en diversos campos, como la productividad, el entretenimiento, la comunicación, entre otros. Ejemplos de aplicaciones típicas incluyen:

* **Procesadores de texto** (como Microsoft Word)
* **Hojas de cálculo** (como Excel)
* **Navegadores web** (como Google Chrome)
* **Aplicaciones de correo electrónico** (como Outlook)
* **Programas de mensajería** (como WhatsApp)

Estas aplicaciones suelen ser genéricas, fáciles de usar, y están disponibles para una amplia variedad de usuarios.