

Objetivo :
 Encontrar relaciones dentro un conjunto de transacciones, en concreto, *items* o atributos que tienden a ocurrir de forma conjunta.

SI + ANTECEDENTES + → + CONSECUENCIA

Base de datos transaccional

Una lista Una representación vertical Una representación horizontal

ALGORITMOS:

Apriori Eclat FP-Growth

Objetivo :
 Estimar la precisión de la generalización de un modelo sobre los datos futuros (no vistos/fuera de muestra). Así como la descripción de datos para la toma de decisiones.

Objetivo :
 Búsqueda de patrones que ocurren frecuentemente relacionados con el tiempo o con otras secuencias. El resultado de ésta técnica se presenta como una lista de transacciones.

Análisis de secuencias

Base de datos Secuencia Elemento (Transacción) Evento (Ítem)

ALGORITMOS:

GSP SPADE PrefixSpan

Es el proceso de representación de datos, en formato gráfico, de una manera clara y eficaz. Y es una herramienta poderosa para el análisis e interpretación de datos grandes y complejos.

Tipos de Gráficos

Histogramas Box plot o diagrama de caja Gráfico de dispersión

Gráfico de barras Gráfica de pastel Gráfico de líneas

Sirve para tener escalabilidad con la cantidad de datos. Nos deja encontrar grupos ocultos sobre un conjunto de datos no etiquetado, esto es de gran utilidad para confirmar o descartar alguna teoría asumida de nuestros datos. También puede ayudarnos a descubrir relaciones asombrosas entre conjuntos de datos, que de forma manual, no son reconocibles.

Elementos de K-Medias

Inicialización Asignación Actualización

ALGORITMO:

1er. Seleccionar aleatoriamente un k de los puntos de datos de cada base agrupada.

2do. Asignar a cada punto con la agrupación más cercana, la cual está representada como su centroide

3er. Recalcular la posición del número K de centroides.

4to. Repetir paso 2 y 3 hasta que todos los centroides de las agrupaciones no cambien nunca.

Objetivo :
 Dado un conjunto de elementos de entrenamiento queremos etiquetar los datos en 2 categorías en una base de una línea y entrenar una MSV para construir un modelo que prediga la clase de una nueva muestra futura.

Representación por medio de funciones núcleo o Kernel

Ofrece una solución a este problema, proyectando la información a un espacio de características de mayor dimensión, el cual aumenta la capacidad computacional de la máquinas de aprendizaje lineal.

Es un conjunto de valores observados durante un periodo determinado de tiempo. Este periodo esta dividido en intervalos.

¿Cómo se compone una serie de tiempo?

Tendencia Variación Cíclica Variación Estacional Variación Irregular

Objetivo :
 Es un enfoque a la Regresión lineal donde se realiza el análisis estadístico dentro del contexto de la inferencia Bayesiana.

Tipos de Regresión Bayesiana:

Normal Logístico Probit Poisson Gamma Binomial Negativa

➤ Regresión lineal simple (solo una variable predictora)

➤ Regresión lineal múltiple (más de una variable predictora).

Objetivo :
 Tratar de explicar la relación que existe entre una variable dependiente (variable respuesta) Y un conjunto de variables independientes (variables de predicción).

LIBRERÍAS EN PYTHON:

Sklearn Matplotlib

Objetivo :
 Emplear distintos panoramas y poder seleccionar uno bajo números y propuestas objetivas, evitando la combinación de emociones.

Estructura

Raíz Ramas Nodos Hojas

Determina el número máximo de acuerdo con la una rama.

Brindan los posibles caminos que se tienen de acuerdo con la decisión tomada.

los caminos que se tienen de acuerdo con la decisión tomada.

Raíz: En él se produce la primera división en función de la variable más importante.

Intermedios: Vuelven a dividir el conjunto de datos en función de las variables.

Se ubican en la parte inferior del esquema y su función es indicar la clasificación definitiva.

Pueden afectar considerablemente a los resultados que pueda obtener un modelo de Machine Learning.

Los Outliers pueden significar varias cosas:

Error Límites Punto de interés

Error de carga de datos. En este caso, la detección de outliers nos ayuda a detectar errores.

Valores que se escapan del "grupo medio", pero queremos mantener el dato modificado, para que no perjudique la información.

Puede que sean los casos "anómalos" los que queremos detectar y que sean nuestro objetivo.

