

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas

Resúmenes de exposiciones

Minería de Datos

Maestra: Mayra Cristina Berrones Reyes

Alumno: Jorge Adrián González Peña

Matricula: 1821695

Grupo: 003

Tema 1: Reglas de Asociación.

Las reglas de asociación es una técnica que establece una regla para asociar entre un conjunto de datos relevantes. Busca asociaciones, patrones comunes entre los datos. Mediante el recuento de soporte y su confianza (que es la probabilidad condicional entre los datos y su asociación) se generan reglas para ajustar los itemset para formar transacciones.

Los conceptos utilizados son: conjunto de elementos, recuento de soporte y confianza.

Reglas de Asociación "A priori"

Utilizamos el principio "A priori" que dice si un conjunto de elementos es frecuente, en sus subconjuntos también lo será.

Reglas de Asociación enfocada en dos pasos

Es generar todos los conjuntos con soporte 2 mínimo

Después generar las reglas con elementos frecuentes y alta confianza.

Tema 2: Clasificación

Este tema fue el que le toco a mi equipo, y me gustó mucho ya que como su nombre lo dice clasifica "ordenas" tomando en cuenta ciertas características de los elementos del conjunto.

Los métodos de clasificación son:

1. Análisis discriminante: método para encontrar una combinación lineal de rasgos.
2. Reglas de clasificación: buscan términos no clasificados de forma periódica.
3. Redes neuronales artificiales: es un modelo de unidades conectadas para transmitir señales.
4. Arboles de decisión: método que a través de una representación facilita la toma de dediciones.

Las características finales que podemos ver son: precisión en la predicción, eficiencia, robustez, escalabilidad e interpretabilidad.

Tema 3: Detección de outliers.

La detección de outliers me parece muy importante ya que estudia el comportamiento de valores externos que difieren del patrón general de una muestra. Prácticamente es la detección de datos atípicos en un conjunto de datos, valores que son muy diferentes a la mayoría.

Las técnicas para la detección de valores atípicos son:

- Prueba de Grubbs
- Prueba de Dixon
- Prueba de Tukey
- Análisis de valores atípicos de Mahalanobis
- Regresión simple

Si se detecta outliers, se puede eliminar o sustituir si se corrobora que los datos atípicos se deben a un error de captura o medición de la variable. Si no es un error de eliminarlo se introduciría un sesgo, afectaría la distribución y varianzas.

Tema 4: Patrones Secuenciales

En minería de datos secuenciales, es la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia. Se enlazan con el paso del tiempo y el orden es importante. Reglas de asociación secuencial expresan patrones secuenciales, esto quiere decir que se dan en instantes distintos en el tiempo.

Análisis de secuencias:

1. Base de Datos
2. Secuencia
3. Elementos de transacción
4. Ítem

Unas de sus características son: El orden importa, La longitud de la secuencia es su cantidad de elementos.

Tema 5: Predicción

La predicción es una técnica que se utiliza para proyectar los tipos de datos que se verán en un futuro o predecir el resultado de un evento. En algunos casos con los datos históricos se puede construir un modelo que explique y trace una predicción sobre el comportamiento futuro.

Algunas de sus aplicaciones son: Revisar los historiales crediticios de los consumidores y las compras pasadas para predecir si serán un resigo creditico en el futuro, predecir el precio de venta de una propiedad, predecir la puntuación de cualquier equipo durante un partido, etc.

Las técnicas de predicción son: Regresión lineal, Regresión lineal multivariante, Regresión no lineal, Regresión no lineal multivariante.

La predicción se basa en ajustar una curva a estos modelos a partir de los datos y así encontrar una relación entre los predictores y los pronosticados.

Tema 6: Regresión

Una regresión es un modelo matemático para determinar el grado de dependencia entre una o más variables, conocer si existe relación entre ellas.

Hay dos tipos de regresión:

- Lineal: Es cuando una variable independiente ejerce fuerza sobre otra variable dependiente
- Múltiple: Es cuando dos o más variables independientes ejercen fuerza sobre una variable dependiente.

Realizamos un análisis de regresión para analizar la relación entre las variables e identificar cuáles son las que tienen mayor impacto en la variable dependiente. Esto nos permite explicar un fenómeno y predecir cosas acerca del futuro, nos ayuda a tomar las decisiones y obtener mejores resultados.

Tema 7: Visualización

La visualización de datos es la representación de información en formato grafico o ilustrado. Al utilizar elementos visuales como cuadros, gráficos o mapas nos proporciona una manera accesible de ver y comprender tendencias, valores atípicos o patrones en los datos.

Tipos de visualización:

- Mapas
- Infografías
- Cuadros de mando
- Gráficos

Podemos tener algunas aplicaciones con la visualización de datos:

- Comprender la información con rapidez
 - Identificar relaciones y patrones
 - Identificar tendencias emergentes.

Tema 8: Clustering

Esta técnica consiste en la división de datos en grupos de objetos similares. Agrupa objetos usando la información que las variables le brindan y una vez hecho esto se coloca clases que son muy similares internamente y a la vez diferente entre los miembros de las demás clases.

Un cluster es una colección de objetos de datos similares entre si dentro del mismo grupo.

Aplicaciones que puede ofrecernos el clustering:

- En el estudio de terremotos
 - En las asegurados
 - En marketing
 - En demografía

Algunos métodos de agrupación son: Jerárquico vs plano, Danos numéricos y/o simbólicos, Determinística vs probabilística, etc..

