

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas



Mapa Mental II Técnicas de Minería de Datos

Docente: MC. Mayra Cristina Berrones Reyes

Alumna: Leslie Paola Navarro Sosa

Matrícula: 1728777

Grupo: 003

San Nicolás de los Garza, Nuevo León, miércoles 31 de marzo de 2021

Visualización

Proceso de representación de datos, en formato gráfico, de una manera clara y eficaz.

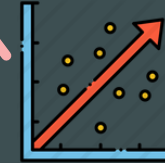


Patrones Secuenciales

Es la búsqueda de patrones que ocurren frecuentemente relacionados con el tiempo o con otras secuencias.

Regresión Lineal

Permite predecir a partir de un muestreo aleatorio.



Regresión Bayesiana

Actualiza la distribución a priori en forma de posteriori.



Outliers

Son los valores que se escapan al rango en donde se concentran la mayoría de las muestras.



Técnicas de Minería de Datos



Máquina Soporte Vectorial

Conjunto de algoritmos de aprendizaje, usados para resolver problemas de clasificación y regresión.



Reglas de Asociación

Permiten contar las combinaciones de artículos o ítems que ocurren con mayor frecuencia en una base de datos.



Métricas de Evaluación

Estima la precisión de la generalización de un modelo sobre los datos futuros.



Series de Tiempo

Conjunto de valores observados durante un periodo de tiempo.



Árboles de Decisión

Modelo de predicción utilizado en el ámbito de la inteligencia artificial.



K Medias

Se utiliza cuando tenemos demasiados datos sin etiquetar para con esto encontrar "k" grupos entre los datos crudos.



Técnicas de Minería de Datos

Árboles de Decisión

Modelo de predicción utilizado en el ámbito de la inteligencia artificial.

- Estructura
 - Nodos raíz
 - Nodos intermedios
 - Hojas
 - Ramas
 - Profundidad del árbol
- Ventajas
 - Fáciles de construir, interpretar y visualizar
 - Selecciona las variables más importantes
 - Categoriza variables numéricas
- Desventajas
 - Tienden al sobreajuste
 - No predicen bien para nuevos casos

Series de Tiempo

Conjunto de valores observados durante un periodo de tiempo.

- Tipos de series
 - Serie Estacionaria
 - Modelos
 - Autoregresivos (AR [p])
 - Medias Móviles (MA [q])
 - Mixto (ARMA [p,q])
 - Criterio de información (AIC pequeño)
 - Serie No Estacionaria
- Componentes
 - Tendencia
 - Variación Cíclica
 - Variación Estacional
 - Variación Irregular

Métricas de Evaluación

Estima la precisión de la generalización de un modelo sobre los datos futuros.

- Clasificación
 - Exactitud
 - Exhaustividad
 - Precisión
 - Especificidad
 - Puntuación F1
 - Predicción
 - Regresión lineal
 - Series de tiempo
 - Regresión bayesiana
 - Asociación
 - Reglas de asociación
 - Patrones secuenciales
 - Agrupamiento
 - K-medias
- Matriz de confusión
 - Error Tipo I
 - Error tipo II

Reglas de Asociación

Permiten contar las combinaciones de artículos o ítems que ocurren con mayor frecuencia en una base de datos.

- Base de datos transaccional
 - Una lista
 - Una representación vertical
 - Una representación horizontal
- Métricas de interés
 - Soporte
 - Confianza
 - Lift

Outliers (Detección de Bordes)

Son los valores que se escapan al rango en donde se concentran la mayoría de las muestras.

- Detección de Outliers
 - Box plot

Visualización

Proceso de representación de datos, en formato gráfico, de una manera clara y eficaz.

- Tipos de gráficas
 - Histograma
 - Gráfico de líneas
 - Gráfico de barras
 - Gráfico de dispersión
 - Box plot
 - Gráfico de pastel
- Ventajas
 - Ayuda en la toma de decisiones
 - Ahorro de tiempo
 - Optimiza la colaboración de la información
 - Funciones de autoservicio para los usuarios
 - Reducción de la carga de trabajo del equipo de tecnología

Patrones Secuenciales

Es la búsqueda de patrones que ocurren frecuentemente relacionados con el tiempo o con otras secuencias.

- Tipos de datos
 - ADN y Proteínas
 - Recorrido de clientes en un supermercado
 - Registros de accesos a una pagina web
- Aplicaciones
 - Música
 - Análisis de mercado
 - Telecomunicaciones
- Ventajas
 - Flexibilidad
 - Eficiencia
- Desventajas
 - Utilización
 - Sesgado por los primeros patrones

Regresión Lineal

- Relación estadística entre dos variables continuas.
 - Tipos
 - Regresión lineal simple
 - Regresión lineal múltiple
 - Métricas
 - Error cuadrático medio
 - Coefficiente de determinación
- Permite predecir a partir de un muestreo aleatorio.

Regresión Bayesiana

Define la distribución a priori para los parámetros, determina la verosimilitud de los datos y actualiza la distribución a priori en forma de posteriori.

- Ventajas
 - Permite ajustar modelos complejos
 - Permite alcanzar estimaciones mas exactas
 - La interpretación de los datos es directa y fácil
 - Permite realizar propagaciones de error
 - Se pueden incluir datos perdidos
- Aplicaciones
 - Descripción de datos
 - Estimación de parámetros
 - Predicción y estimación
 - Control

Máquina Soporte Vectorial

Conjunto de algoritmos de aprendizaje, usados para resolver problemas de clasificación y regresión.

- Usos
 - Regresión
 - Selección de variables
 - Identificación de casos anómalos
 - Clasificación multiclase y binaria
- Características
 - Técnicas de clasificación de datos
 - Trabaja en un espacio de características inducido por la función Kernel

K Medias

Clasificación No supervisada

Se utiliza cuando tenemos demasiados datos sin etiquetar para con esto encontrar "K" grupos entre los datos crudos.

- Elementos
 - Inicialización
 - Asignación
 - Actualización