



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN®



UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE CIENCIAS DE FISICO
MATEMATICAS

Mineria de datos

Resumenes

José Pedro Treviño Hernández

Matricula: 1395501

Contenido

Clustering

Prediccion

Outluers

Patron Secuencial

Reglas de Asociacion

Regresion Lineal

Clustering

El Clustering es una tarea que consiste en agrupar un conjunto de objetos (no etiquetados) en subconjuntos de objetos llamados Clusters. Cada Cluster está formado por una colección de objetos que son similares (o se consideran similares) entre sí, pero que son distintos respecto a los objetos de otros Clusters.

En el campo del ML, el Clustering se enmarca dentro del aprendizaje no supervisado; es decir, que para esta técnica solo disponemos de un conjunto de datos de entrada, sobre los que debemos obtener información sobre la estructura del dominio de salida, que es una información de la cual no se dispone.

Las técnicas de Clustering son apropiadas para los casos de aprendizaje no supervisado en los que se quiere agrupar y tener conocimiento a un alto nivel de cómo se han generado los datos y como están organizados, sin tener conocimiento a priori de estos.

Predicción

La predicción contempla la reacción de los clientes hacia mensajes promocionales, comunicación de nuevos productos, o incluso, poder intervenir en el momento propicio a la pérdida de lealtad del susodicho

El objetivo de las técnicas de predicción no causal (TPNC) es obtener estimaciones o pronósticos de valores futuros de una serie temporal a partir de la información histórica contenida en la serie observada hasta el momento actual. Estas técnicas no requieren la especificación de los factores que determinan el comportamiento de la variable, sino que se basan únicamente en la modelización del comportamiento sistemático de la serie.

Outliers

Un outlier es una observación anormal y extrema en una muestra estadística o serie temporal de datos que puede afectar potencialmente a la estimación de los parámetros del mismo.

En palabras más sencillas un outlier sería una observación dentro de una muestra o una serie temporal de datos que no es consistente con el resto. Imaginemos por ejemplo que estamos midiendo la altura de los alumnos de una clase.

En estadística, un valor atípico es un punto de observación distante de otras observaciones. Un valor atípico puede deberse a variabilidad en la medición o puede indicar error experimental; Estos últimos a veces son excluidos del conjunto de datos. Los valores atípicos pueden ocurrir por casualidad en cualquier distribución, pero a menudo son indicativos de error de medición o que la población tiene una distribución de cola pesada.

Patrón Secuencial

Los patrones secuenciales es un tema de la minería de datos que se ocupa de encontrar patrones estadísticamente relevantes entre ejemplos de datos donde los valores se entregan en una secuencia. Por lo general, se presume que los valores son discretos y, por lo tanto, la minería de series de tiempo está estrechamente relacionada, pero generalmente se considera una actividad diferente. La minería de patrones secuenciales es un caso especial de minería de datos estructurados.

Hay varios problemas computacionales tradicionales clave que se abordan en este campo. Estos incluyen la construcción de bases de datos e índices eficientes para la información de secuencias, la extracción de los patrones que ocurren con frecuencia, la comparación de secuencias en busca de similitudes y la recuperación de miembros de secuencia faltantes.

Reglas de asociación

Las reglas de asociación estudian las ocurrencias de un evento dentro de un conjunto de datos, dadas las diferentes medidas de interés, por ello es importante conocer la estructura en que viene la base de datos, puesto que el método de aprendizaje que presentan las reglas deben ser coherente con el objeto de estudio.

Las reglas de asociación son implicaciones que relacionan la presencia de ítem en las transacciones. Las transacciones son la estructura básica desde donde las reglas de asociación son obtenidas.

Basándose en el concepto de reglas estrictas, Rakesh Agrawal, Tomasz Imieliński y Arun Swami introdujeron reglas de asociación para descubrir regularidades entre productos en datos de transacciones a gran escala registrados por sistemas de punto de venta (POS) en los supermercados.

Regresion Lineal

El modelo de pronóstico de regresión lineal permite hallar el valor esperado de una variable aleatoria ***a*** cuando ***b*** toma un valor específico. La aplicación de este método implica un supuesto de linealidad cuando la demanda presenta un comportamiento creciente o decreciente, por tal razón, se hace indispensable que previo a la selección de este método exista un análisis de regresión que determine la intensidad de las relaciones entre las variables que componen el modelo.

El objetivo de un análisis de **regresión** es determinar la relación que existe entre una variable dependiente y una o más variables independientes. Para poder realizar esta relación, se debe postular una relación funcional entre las variables.

Cuando se trata de una variable independiente, la forma funcional que más se utiliza en la práctica es la relación lineal. El análisis de regresión entonces determina la intensidad entre las variables a través de coeficientes de correlación y determinación.