

TALLER 1: APRENDIZAJE SUPERVISADO VS. NO SUPERVISADO

Angie Caterine Sarmiento

Suponiendo que una técnica particular de extracción de datos se va a utilizar en los siguientes casos, identifique si la tarea requerida es aprendizaje supervisado o no supervisado (descripción). De ser posible determine la tarea específica a realizar: regresión, clasificación, clustering, resumen, modelo de dependencia, regla de asociación, detección de anomalías, etc.

- a. Decidir si otorgar un préstamo a un solicitante en base a datos demográficos y financieros (con referencia a una base de datos de datos similares sobre clientes anteriores).

Aprendizaje supervisado, clasificación:

Se quiere tomar una decisión realizar o no un préstamo con base en datos de clientes anteriores, la variable de respuesta es categórica (Sí, No) por esta razón es una tarea de clasificación como se enuncia en (?, ?)

- b. En una librería en línea, hacer recomendaciones a los clientes sobre artículos adicionales para comprar según los patrones de compra en transacciones anteriores.

Aprendizaje no supervisado, reglas de asociación:

No se cuenta con una variable de respuesta sobre se quiere sugerir a los clientes de la librería cierto tipo de productos descubriendo patrones de asociación de compra basándose en clientes anteriores con compras similares. Además es un problema de canasta de mercado típico de reglas de asociación como se enuncia en (?, ?)

- c. Identificar un paquete de datos de la red como peligroso (virus, ataque de hackers) basado en la comparación con otros paquetes cuyo estado de amenaza es conocido.

Aprendizaje no supervisado:

- d. Identificación de segmentos de clientes similares.

Aprendizaje no supervisado, clustering:

Se identifican subgrupos de clientes basándose en la homogeneidad de sus patrones entre intraclass y en la heterogeneidad interclass como se enuncia en (?, ?)

- e. Predecir si una empresa irá a la quiebra basándose en la comparación de sus datos financieros con los de empresas similares en bancarrota y no en bancarrota.

Aprendizaje supervisado, clasificación:

La variable de respuesta es dicotómica (Quiebra, no quiebra) con base en patrones de empresas que han estado o no en bancarrota, las categorías y la variable de respuesta determinan que es un problema de clasificación como se enuncia en (?, ?)

- f. Estimación del tiempo de reparación requerido para una aeronave en base a un ticket de problema.

Aprendizaje supervisado, regresión:

La variable de respuesta es continua (tiempo de reparación) por tener certeza en lo que se quiere estimar y en que contamos con una variable de respuesta continua es un problema de clasificación, tal como se menciona en (?, ?)

- g. Clasificación automatizada de correo mediante escaneo de código postal.

Aprendizaje supervisado, clasificación

Cada correo tiene un código postal específico correspondiente a su procedencia, estos son extraídos de una base de datos y cada vez que uno es escaneado el sistema reconoce su lugar de origen

- h. Impresión de cupones de descuento personalizados al finalizar la compra de una tienda de comestibles según lo que acaba de comprar y lo que otros han comprado anteriormente.

Aprendizaje no supervisado, regla de asociación:

Los descuentos son asignados por agrupamiento de productos (canasta de mercado) ciertos productos en ciertas cantidades basados en compras anteriores imprimen el descuento correspondiente.

- i. Identificar automáticamente el tema principal abordado por revistas almacenadas en un repositorio en línea.

Aprendizaje no supervisado, clustering

La colección de palabras en cada revista sirven para que se identifiquen patrones de tema de acuerdo al grupo de palabras que contienen, permitiendo así identificar el tema principal.

- j. Desarrollar perfiles de personas según su orientación política y su posición social.

Aprendizaje no supervisado, clustering:

Se identifican patrones en los perfiles de las personas para así agruparlos se espera que exista homogeneidad dentro del grupo y heterogeneidad entre grupos tal y como se menciona en (?, ?)

- k. Encontrar la relación entre la temperatura de un lugar y diversas variables que miden el grado de contaminación presente en el mismo.

Aprendizaje no supervisado:

No se cuenta con una variable de respuesta, clustering identificar patrones mas no relación y las reglas de asociación identifican patrones similares (Mas frecuente en marketing)

- l. Desarrollar una aplicación para teléfonos móviles que sugiera al usuario la trayectoria más adecuada para desplazarse de un punto de la ciudad a otro.

Aprendizaje no supervisado, clustering

Se utilizan datos geográficos que analizan todas las posibles trayectorias y las agrupan de acuerdo a varias especificaciones, como la distancia mas corta el medio de transporte, la ruta mas frecuente entre otras para determinar la trayectoria mas optima.

- m. Desarrollar un dispositivo para el reconocimiento facial de una persona.

Aprendizaje no supervisado

Se carga una imagen al dispositivo y se analizan los píxeles de la imagen para identificar los patrones secuenciales de cada rostro y así identificar a la persona mediante una imagen.

- n. Descubierta un nuevo espécimen de insecto determinar con la mayor precisión a qué género pertenece dicho insecto.

Aprendizaje supervisado, clasificación

La variable de respuesta es el género de insecto y es una variable categorica por lo tanto cumple con las características para ser tareas de clasificación como se menciona en (?, ?)

- o. Predecir vía imágenes satelitales el nivel de dióxido de carbono emitido por un conglomerado industrial.

Aprendizaje supervisado, regresión o clasificación segun como se mida el nivel de dióxido

La variable de respuesta es nivel de dióxido de carbono aunque el nivel se puede ver como categorías (alto, medio, bajo) también se puede ver como una variable continua, si se cumplen estas dos características se habla de una tarea de predicción o de clasificación.

Referencias