

Actividad 9

Ejercicio 1

Investiga sobre dos aplicaciones actuales en donde se aplique la técnica de clústering (agrupamiento) e identifica en cada una de ellas los elementos vistos en el tema 17. Explica por qué es importante el criterio de selección de números de clústeres.

Recordemos que el clústering es una técnica de aprendizaje no supervisado que busca organizar un conjunto de objetos en grupos, llamado clústeres, donde los objetos dentro de un mismo clúster son más similares entre sí que con los otros clústeres. Este proceso se utiliza en una gran variedad de aplicaciones actuales, tales como:

1. Segmentación de clientes: Uno de los usos más habituales del clústering en las empresas es la segmentación de bases de datos de clientes. Se trata de asignar a los clientes a distintos grupos según las características que mejor puedan definirlos. El beneficio de una segmentación altamente precisa es que se podrán enfocar con más eficacia y precisión de comunicación y marketing. Pongamos un ejemplo más práctico. Supongamos en la industria financiera, se planea realizar una estrategia de marketing a clientes y ofrecerles una tarjeta de crédito, sin embargo, hay clientes dentro de la industria con diferentes características, por ende, la tarjeta de crédito debería ser diferentes, considerando sus ingresos mensuales y gastos mensuales, por ende, es mejor dividir a los clientes en grupos de tarjetas de crédito, de ahí la razón por la cual existen productos de tarjeta de crédito de distintas categorías (por ejemplo clásica, oro, platinum, premium, etc). Esto se termina decidiendo con base a distintas características de los clientes, no solo por sus ingresos y egresos mensuales, puede llegar a considerarse la edad, tiempo que ha pasado siendo cliente del banco, estudios, etc.
2. Agrupación de artículos por temáticas: Esta técnica es aplicable en gran variedad de ámbitos. Puede utilizarse para clasificar y organizar contenidos en páginas web, de cara a facilitar la navegabilidad y mejorar la experiencia del usuario. Los sectores en los que se puede aplicar son muy amplios, desde los medios de comunicación hasta las bases de datos científicas, en las que es necesario agrupar los contenidos de formar jerárquica y organizada. Un ejemplo muy común en esta rama es Spotify, usualmente, a final de año, realiza una agrupación a todos los usuarios de la plataforma con el fin de mostrar qué tipo de música o gustos tienen en común todas esas personas, es curioso y al mismo tiempo fascinante notar cómo pueden crearse distintos grupos basándose por las canciones o artistas que más escuche una persona. Basándose en esa misma información, la plataforma inventa nuevas recomendaciones con base a sus gustos musicales.

Ejercicio 2

Identifica en tu centro laboral o entorno cotidiano una situación en la que puedas recopilar datos. Con base en tus datos, determina cuál técnica o modelo del aprendizaje no supervisado te conviene usar para realizar su agrupamiento y justifica tu respuesta.

El desarrollo del programa y las conclusiones están en el archivo anexo a esta entrega o en la ubicación '../notebooks/Tarea9.ipynb'

Ejercicio 3

Investiga cuáles de las diferentes herramientas digitales vistas en clase pueden ayudarte a procesar y extraer información importante de tus datos. Elige tres y clasifícalas por medio de la siguiente tabla

Nombre	Marca	Tipo	Característica	Aplicaciones
SAS	SAS Institute	Nube o Local	Análisis de datos, obtención de datos, preparación de datos para entrenamiento	Bases de datos para modelos
Apache Spark	Apache	Nube o Local	Procesamiento de datos, integración y preparación de datos	Big data, ML, bases para modelos.
Google BigQuery	Google	Nube	Almacenamiento y análisis de datos, consultas SQL, integración con herramientas de Google Cloud	Análisis de datos, creación de modelos, exploración de datos, reconocimiento de imágenes