

Curso de posgrado: “Aprendizaje automático: Fundamentos, Herramientas y Aplicaciones”
Diplomado de Telecomunicaciones y Sensado Remoto por Ondas de Radio
Trabajo práctico N° 2
TEMA: Clustering

Aplicando la técnica de Clustering, K-means

Las técnicas de Clustering, intentan descubrir nuevas relaciones entre características (features) o probar o declinar hipótesis que tengamos de nuestros datos.

El conjunto de datos del archivo “mall_customers.csv” (clientes del centro comercial), contiene los registros de 200 clientes de un centro comercial, donde cada perfil de cliente tiene un ID más 4 características. Las columnas que posee el conjunto de datos son:

- identificación del cliente (CustomerID)
- género (gender)
- edad (age)
- ingresos anuales (k \$) (Annual Income)
- puntuación de gasto (1-100) (Spending Score)

Suponiendo que uno es el propietario de un centro comercial y desea comprender a los clientes, como quiénes pueden converger fácilmente (clientes objetivo) para que el equipo de marketing planifique la estrategia de venta.

Fuente de los datos: curso A-Z de aprendizaje automático de Udemy.

La realización del trabajo consiste en:

- I. Aprobar o declinar hipótesis sobre los datos mediante la aplicación de la técnica de clustering K-Means. Algunos ejemplos de hipótesis pueden ser:
 - a) “Los clientes con mayor ingresos anuales son los que más gastan”.
 - b) “Los clientes de mayor edad, con mayor ingreso anual, son los que menos gastan”.

NOTA: es posible plantear todas las hipótesis que se desee para el desarrollo del trabajo.
- II. Mostrar y explicar todos los pasos previos a la aplicación de la técnica, elección del hiperparámetro, resultados obtenidos a partir de las hipótesis que se plantearon.