

# Clase 17 IMA539

Eduardo Castro Thompson

June 5, 2023

## 1 Clase de Ejercicios

### 1.1 Problema 1

La segmentación de clientes corresponde a agrupar clientes con características similares. Estas características pueden incluir ubicación geográfica, estilos de vida, comportamientos, etc. Uno de los objetivos más conocidos de esta tarea corresponde a la selección de un grupo objetivo para la aplicación de alguna campaña de publicidad.

La base de datos `Mall-customer-data.csv` contiene información de clientes que han realizado compras en una tienda de un centro comercial. Cada muestra del conjunto de datos contiene: - ID. - Edad. - Género. - Ingresos anuales. - Puntuación de gasto.

Se busca determinar cuales son los grupos de clientes que realizan compras. Para esto: 1. Realice un análisis exploratorio del conjunto de datos. Se recomienda: Cantidad de muestras según rango etareo/género, género v/s puntuación de gasto, edad v/s ingreso anual, etc. 2. Entrene un modelo de KMeans en las tres columnas de datos continuos que considere más relevantes. Justifique el número de clusters que eligió. 3. Grafique en 3D los clústers encontrados especificando en cada uno la ubicación del centroide. 4. ¿Es correcta la segmentación encontrada? Concluya.

### 1.2 Problema 2

La competición de Kaggle [California House Prices](#) tiene por finalidad la predicción del precio de casas en ubicadas en el estado de California. La base de datos `callifornia-house-prices.csv` contiene 47.439 muestras y un total de 41 columnas con información relevante a las casas como número de baños, cercanía a escuelas, metros cuadrados, entre otros; además del precio de venta de los inmuebles para el año 2020.

Para resolver este problema, seleccione 5 columnas de datos continuos y entrene dos modelos de regresión (uno lineal y otro no lineal) para la predicción del precio de venta de la edificación, justifique su elección.

Una vez entrenados los modelos reporte al menos  $MSE$  y  $R^2$ .

[ ]: