**Tarea Numero 1**

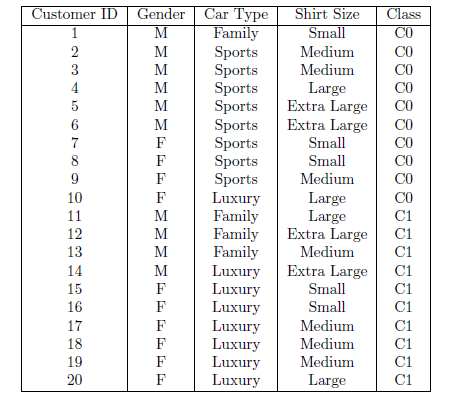
**Mineria de Datos**

**Octubre 2017-2018**

Ejercicio 1

Considere los ejemplos de entrenamiento de la siguiente tabla. Para un problema de clasificación binario

1. Obtenga el Gini index para toda la colección de ejemplos de entrenamiento
2. Obtenga el Gini index para el atributo Customer ID
3. Obtenga el Gini index para el atributo genero
4. Obtenga el Gini index para el tipo de carro usando un multiway Split
5. Cual atributo es mejor Genero, tipo de carro, o tamaño de la camiseta
6. Explique el porque curstomer id no puede ser usado como atributo prueba a pesar de que tiene el fini mas bajo.



Ejercicio 2

Considere los ejemplos de entrenamiento mostrados en la siguiente tabla para un problema de clasificacion binario.

1. Cual es la entropía de esta colección de datos de entrenamiento con respecto a la clase positiva
2. Cual es el information gain o ganancia de la información de a1 y a2 relativa a estos ejemplos de entrenamiento
3. Para a3 , el cual es un atributo continuo, obtenga el information gain para todos sus posibles splits.
4. Cual es el mejor Split entre (a1, a2, y a3) de acuerdo a information gain
5. Cual es el mejor Split (entre a1 y a2) de acuerdo a la tasa de error
6. Cual es el mejor Split (entre a1 y a2) de acuerdo al Gini index

