

# Trabajo #2 -2Do Mod. Introducción a la Analítica.

## S-02-22.

***Fecha de entrega: Nov 29.***

1. Este ejercicio utiliza el conjunto de datos `OJ` el cual es parte de la librería `ISLR`
  - a) Cree un conjunto de entrenamiento con una muestra aleatoria de 800 observaciones y un conjunto de prueba que conste del resto de observaciones.
  - b) Ajuste un **clasificador de soporte vectorial** utilizando `cost = 0.1`, con `Purchase` como la variable respuesta y las demás como predictores.
    - Utilice la función `summary()` para obtener un resumen de estadísticas y describa los resultados obtenidos.
  - c) Que tasas de error de entrenamiento y de prueba obtiene?.
  - d) Utilice la función `tune()` para obtener un valor óptimo del parámetro `cost`. Considere valores en el rango de 0.01 a 10.
  - e) Calcule nuevamente las tasas de error de entrenamiento y de prueba usando el valor óptimo obtenido de `cost`.
  - f) Repita items de (b) hasta (e) ajustando esta vez una **máquina de soporte vectorial (svm)** con un nucleo **radial**. Utilizando el valor de default para  $\gamma$ .
  - g) Repita items (b) hasta (e) utilizando nuevamente una máquina de soporte vectorial pero esta vez con un nucleo **polinomial**, usando `degree = 2`.
  - h) En general cuál método parece proporcionar los mejores resultados en estos datos?.
2. Considere el conjunto de datos `USArrests`. En este ejercicio se agruparán los estados en `USArrests` con agrupamiento jerárquico
  - a) Utilice agrupación jerárquica con enlace completo y distancia euclidiana, para agrupar los estados.
  - b) Corte el **dendrograma** a una altura que dé como resultado **tres** clusters. ¿Qué estados pertenecen a qué cluster?
  - c) Agrupe jerárquicamente los estados utilizando un enlace completo y distancia euclidiana, después de escalar las variables para tener una desviación estándar uno.

- d)* ¿Qué efecto tiene el escalado de las variables en la estructura jerárquica del agrupamiento obtenido? En su opinión, ¿deberían las variables ser escaladas antes de que se calculen las disimilitudes entre observaciones? Proporcione una justificación para su respuesta.