

# Trabajo #1 -2Do Mod. Introducción a la Analítica. S-02-22.

***Fecha de entrega:*** Nov 15.

***Indicación:*** Descargue e instale la librería **ISLR2** en **R**

```
install.packages("ISLR2")  
library(ISLR2)
```

1. Esta pregunta utiliza las variables **dis** (la media ponderada de distancias a cinco centros de empleo de Boston) y **nox** (concentración de óxidos de nitrógeno en partes por 10 millones) del conjunto de datos **Boston**. Vamos a tratar **dis** como predictor y **nox** como respuesta.
  - a) Use la función **poly()** para ajustar una regresión polinomial cúbica y con esta predecir la variable **nox** usando **dis**. Reporte el resultado de la regresión, luego grafique los datos resultantes y los ajustes polinómicos.
  - b) Grafique los ajustes polinómicos para un rango de polinomios de diferentes grados (digamos, de 1 a 10), y reporte la suma de cuadrados de los residuales asociada.
  - c) Realice una validación cruzada o algún otro enfoque para seleccionar el óptimo grado para el polinomio y explique sus resultados.
  - d) Use la función **bs()** para ajustar una spline de regresión para predecir **nox** usando **dis**. Reporte la salida para el ajuste usando cuatro grados de libertad. ¿Cómo ubicó los nodos?. Graique el ajuste resultante.
  - e) Ahora ajuste una spline de regresión para un rango de grados de libertad, y grafique los ajustes resultantes e informe el RSS resultante. Describa los resultados obtenidos.
  - f) Realice una validación cruzada o algún otro enfoque para seleccionar los mejores grados de libertad para una spline de regresión sobre estos datos. Describa sus resultados.
2. En este ejercicio se utilizarán arboles de regresión para predecir los valores de la variable **sales** en la base de datos **Carseats** de la librería **ISLR2**, tratando dicha variable como continua:

- a) Divida el conjunto de observaciones en un conjunto de entrenamiento y un conjunto de prueba. De forma aleatoria. En que proporciones dividió los datos?.
- b) Ajuste un árbol de regresión en el conjunto de entrenamiento. Gráfique el árbol e interprete los resultados. Que valor del MSE **de prueba** obtiene?.
- c) Utilice validación cruzada para determinar el grado óptimo de complijidad del árbol. Consigue la poda del árbol mejorar el EMS **de prueba**?
- d) Utilice el método **bagging** para analizar estos datos. Que valor del MSE obtiene?
  - Use la función `importance()` para determinar cuál de las variables es la más importante.
- e) Ahora utilice un **bosque aleatorio** (*Random-Forest*) para analizar estos datos.
  - Que valor del MSE de **prueba** obtiene?.
  - Use la función `importance()` para determinar cuales variables son las más importantes.
  - Describa el efecto de  $m$  el número de variables consideradas en cada subdivisión, en la tasa de error obtenida.