

# Trabajo #1 -2Do Mod. Introducción a la Analítica. S-02-23.

***Fecha de entrega:*** Nov 1 (Miércoles).

***Indicación:*** Descargue e instale la librería [ISLR2](#) corriendo los siguientes comandos en [R](#).

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

1. (50 %). En este ejercicio, se quiere predecir el número de aplicaciones recibidas en distintos colegios universitarios americanos. utilizando las demás variables en el conjunto de datos [College](#).
  - a) Divida el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba.
  - b) Ajustar un modelo lineal usando mínimos cuadrados en el conjunto de entrenamiento, y informar el error de prueba obtenido.
  - c) Ajustar un modelo de regresión **ridge** en el conjunto de entrenamiento, escogiendo un valor de  $\lambda$  mediante validación cruzada. Informe el error de prueba obtenido.
  - d) Ajuste un modelo **lasso** en el conjunto de entrenamiento, escogiendo el valor de  $\lambda$  mediante validación cruzada. Informe el error de prueba obtenido, junto con el número de estimaciones de coeficientes distintos de cero.
  - e) Ajuste un modelo de PCR en el conjunto de entrenamiento, escogiendo  $M$  mediante validación cruzada. Informe el error de prueba obtenido, junto con el valor. de  $M$  que fué seleccionado.
  - f) Ajuste un modelo PLS en el conjunto de entrenamiento, con  $M$  escogido mediante validación cruzada. Informe el error de prueba obtenido, junto con el valor. de  $M$  que se seleccionó.
  - g) Comente los resultados obtenidos.
    - ¿Con qué precisión se puede predecir la cantidad de solicitudes universitarias recibidas?.
    - ¿Cuanta diferencia hay entre los errores de prueba resultantes de estos cinco enfoques?

2. (50 %). 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.