

## Práctica: Validación Cruzada y Bootstrap

1. Lleve adelante un esquema de validación cruzada por K-fold (fije un valor de K) para los datos del archivo **autos.txt**, y en base a ese criterio determine el mejor modelo entre:

- (a)  $Y \approx \beta_0$ .
- (b)  $Y \approx \beta_1 X$ .
- (c)  $Y \approx \beta_0 + \beta_1 X$ .
- (d)  $Y \approx \beta_0 + \beta_1 X + \beta_2 X^2$ .

con  $Y$  =Calidad y  $X$  =Precio. Para el modelo “ganador”, obtenga la regresión y dibuje simultáneamente los datos junto con la curva de ajuste.

2. Genere 100 datos con el siguiente modelo  $Y = 0.5X + \epsilon$  siendo  $X \sim U(-5, 5)$  y  $\epsilon \sim N(0, 1)$ . Proponemos los siguientes modelos competitivos:

- (a)  $Y \approx \beta_0$ .
- (b)  $Y \approx \beta_1 X$ .
- (c)  $Y \approx \beta_0 + \beta_1 X$ .

Mediante un método de K-Fold elija aquel modelo más apropiado. Qué opina de los resultados obtenidos?

3. Con los datos **yield.txt**, determine la mejor relación polinomial existente entre la temperatura y el rinde del cultivo, basándose en un mecanismo de K-fold. Grafique un diagrama de dispersión con la totalidad de los datos y la curva de regresión del mejor modelo obtenido.
4. Con los datos **abalone.txt**, vuelva a tratar de establecer una relación entre el peso total y el diámetro del espécimen, usando una regresión polinomial con respecto al diámetro. Ensayando distintos grados polinomiales distintos, escoja el mejor modelo a través de un mecanismo de K-Fold.
5. Levante la información del archivo **datos1.txt**. Estime el valor medio con las siguientes medidas de posición  $\hat{\alpha}$ :

- (a) Media
- (b) Mediana
- (c) Media 0.1-podada.

El objetivo es estimar el desvío estándar de este estimador. Para eso, se propone el siguiente esquema de Bootstrap:

- (a) Genere una muestra de 100 elementos tomados de datos1.txt elegidos al azar CON reposición.
- (b) Calcule la medida de posición en dicha muestra.
- (c) Repita el procedimiento 1000 veces o más y obtenga 1000 medidas  $\hat{\alpha}_1, \dots, \hat{\alpha}_{1000}$ .
- (d) Calcule el desvío estándar a esas medidas y haga uso de eso como estimador del desvío estándar de la medida de posición.

Repita el procedimiento para cada medida de posición propuesta y determine el de menor desvío estándar.