Práctica: Estimación y Regresión no paramétrica

1. Implemente una función **densityEstimator** que reciba los siguientes parámetros:

   - $X_1, \ldots, X_n$ datos observados.
   - h ancho de banda.
   - K, función kernel a utilizar.
   - x punto en donde evaluar la estimación.

   Como resultado, debe devolver la estimación de densidad en el punto x.

2. Genere 100 datos siguiendo una distribución normal estándar, aplique el método anterior sobre diversos anchos de bandas sobre una muestra de mil puntos generados en una grilla uniforme entre -3 y 3. Grafique la densidad obtenida. Elija diversos kernels. Cómo cambia la gráfica en función del núcleo y el ancho de banda?

3. Implemente una función **nwEstimator** que reciba los siguientes parámetros:

   - $X_1, \ldots, X_n$ covariables observadas.
   - $y_1, \ldots, y_n$ variables respuestas observadas.
   - h ancho de banda.
   - K, función kernel a utilizar.
   - x punto en donde evaluar la regresión.

   Como resultado, debe devolver el estimador de Nadaraya-Watson evaluado en el punto x.

4. Considere el dataset **abalone.txt** y construya un estimador para el Peso según el Diámetro. Divida en un training set y en un testing set, regresione los elementos del testing set en base a la estimación del training set, eligiendo algún ancho de banda h y un kernel K. Compare los resultados con los obtenidos por regresión lineal múltiple.

5. Con el mismo dataset **abalone.txt**, tome una secuencia de valores para anchos de banda h y fije un kernel K a gusto. Mediante un mecanismo de K-fold con 5 folds determine el mejor valor h para la regresión.

6. (Simulación).

   (a) Fijando una semilla, considere 100 datos x_i generados a través de una $U(-3, 3)$. Luego, considere $y_i = 0.5 + x_i^2 + 5 * sen(x) + \epsilon_i$ siendo ϵ_i una normal de media cero y desvío estándar 2.

   (b) Grafique los datos obtenidos.

   (c) Mediante un mecanismo de K-fold, elija un valor de h para obtener la "mejor" regresión no paramétrica.

   (d) Grafique en una grilla fina los resultados de su regresión no paramétrica superpuesta con los puntos.

   (e) Grafique conjuntamente las salidas de regresiones que hagan uso de los modelos (nota: es lícita la utilización del comando **lm** para estas regresiones):

   - $y_i \cong \beta_0 + \beta_1 x$.
   - $y_i \cong \beta_0 + \beta_1 x + \beta_2 x^2$.
   - $y_i \cong \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

   (f) Discuta ventajas y desventajas y con qué modelo se quedaría (incluyendo el no paramétrico).