

**Ejercicio I.** La técnica conocida como LIDAR (light detection and ranging) usa la reflexión de luz de láser emitida para detectar compuestos químicos en la atmósfera. Esta técnica ha probado ser una herramienta muy eficiente para el monitoreo de la distribución de diversos elementos polulantes en la atmósfera (Sigrist, 1994).

En el archivo lidar.txt se encuentran datos medidos con a la técnica LIDAR. La variable **range** es la distancia recorrida antes de que la luz sea reflejada de regreso hacia su fuente. La variable **logratio** es el logaritmo del cociente de la luz recibida de dos fuentes de luz láser de distinta frecuencia.

1. A partir de los datos de lidar.txt realice un diagrama de dispersión o scatter plot de **range** (eje x) vs. **logratio** (eje y). Describa la relación entre ambas variables.
2. La función de R **ksmooth** computa el estimador de Nadaraya-Watson a partir de un conjunto de datos  $(x_1, y_1), \dots, (x_n, y_n)$  y lo evalúa en un conjunto de puntos intermedios. Mediante la función de R **ksmooth** estime la función de regresión  $r$  que relaciona a las variables **range** y **logratio**, tomando como variable explicativa a **range**, a partir de los datos dados usando el núcleo normal con ventana  $h = 5$ . Con la función de regresión estimada, obtenga estimaciones de la función de regresión en los valores observados de **range**. Grafique la función de regresión estimada. Repita para valores de la ventana  $h = 10, 30, 50$  y superponga en el mismo plot los puntos correspondientes a las observaciones y el valor estimado de la función de regresión obtenida para  $h = 5, 10, 30, 50$ . Compare los resultados obtenidos con las 4 ventanas.
3. Para cada una de las 4 estimaciones obtenidas en el item anterior compute el Error Cuadrático de Predicción Promediado ( $ECPP(h)$ ). ¿Cuál de las 4 ventanas consideradas da el menor  $ECPP(h)$ ? ¿Cómo se puede justificar lo que está ocurriendo?
4. Halle mediante el criterio de Convalidación Cruzada  $CV(h)$  la ventana óptima. Realice la búsqueda en una grilla para valores de  $h$  entre 3 y 165 con paso 1. Realice un plot de  $h$  vs.  $CV(h)$ .
5. Grafique los puntos observados y la función de regresión estimada por el método de Nadaraya-Watson con la ventana óptima hallada.
6. Compute el Error Cuadrático de Predicción Promediado de la estimación provista por el método de Nadaraya-Watson con la ventana óptima hallada:  $ECPP(h_{\text{opt}})$
7. Compute la Perdida de Convalidación Cruzada asociada al la estimación provista por el método de Nadaraya-Watson con la ventana óptima hallada.  $CV(h_{\text{opt}})$

8. ¿Qué se puede hacer con knn?

**Ejercicio II. Bonus Track** Revisitemos la guía de alturas con este nuevo marco teórico; elija  $h$  y  $k$  con las diferentes propuestas consideradas en clase.