

# 1 Ejercicio Práctica 5 y 6

**Fecha límite de entrega:** 17 de mayo de 2018 23:59.

Queremos poder establecer la existencia de una relación entre la altura de una estación meteorológica y su temperatura anual promedio. Para tal fin, contamos con el archivo `ortann.csv` que dispone información de diversas estaciones meteorológicas en el estado de Oregon (EEUU). Nos concentraremos en los datos de altitud (elevation) y temperatura anual (tann).

1. Realizar un *scatter-plot* de la elevación y la temperatura anual (eje y la temperatura).
2. Vamos a realizar una regresión no paramétrica con el método de Nadaraya-Watson. Utilizar una implementación propia, casera, de este método. Para elegir el  $h$ , proceda con un método de cross-validation o K-Fold.
3. Plotear los errores obtenidos para cada valor de  $h$  elegido, en base a eso elegir el  $h$  “óptimo”.
4. Considere modelos lineal y cuadrático también, los cuales deben ser escogidos también usando el mismo criterio de cross-validation usado en la regresión no paramétrica con la misma partición. Es decir, en caso de haber usado K-folds deben estar los mismos folds que se usaron para la elección del  $h$ .
5. Determinar así el modelo ganador entre:
  - (a) Modelo no paramétrico con el  $h$  óptimo.
  - (b) Modelo lineal.
  - (c) Modelo cuadrático.
6. Graficar en el *scatter-plot* de elevación vs temperatura las tres curvas correspondientes a cada modelo (no paramétrico con el mejor  $h$ , lineal y cuadrático).

Nota: para las regresiones lineales y cuadráticas se puede usar el comando `lm`, para la regresión no paramétrica se debe usar una implementación propia.