Aprendizaje Estadístico

Máster Interuniversitario en Tecnologías de Análisis de Datos Masivos: Big Data
Prácticas 2-3. Evaluación Continua

Profesores: Jose Ameijeiras Alonso, Manuel F. Mucientes Molina

Método de descenso de gradiente en regresión

En grupos de una a tres personas deberéis resolver un problema de regresión con el método de descenso gradiente. Deberéis entregar el *script* de R (fichero .R) con el que obtuvisteis las respuestas, así como un informe comentando los detalles relevantes. Dicho informe se deberá entregar en formato PDF. Para generar el PDF podéis usar, entre otros, RMarkDown (`https://rmarkdown.rstudio.com/`), LaTex, MS Word, Libre Office, ...

Fecha máxima de entrega: 27 de octubre.

Peso en la evaluación continua: 1.5 punto de los 5 que están asociados al primer bloque.

Considera un modelo de regresión $Y = f(X) + \epsilon$, donde la función f depende de unos ciertos parámetros. Fijando los parámetros del modelo, simula una muestra con n observaciones. Fija una función que quieras minimizar, como la suma de los residuos al cuadrado, y con R obtén los estimadores de los parámetros del modelo mediante el método de descenso gradiente. Si en R ya existe una función que devuelva el estimador de los parámetros, compara tus resultados con los obtenidos con la función correspondiente.

Una opción, y a modo ilustrativo de como se puede resolver este ejercicio, sería considerar el modelo de regresión lineal simple $Y = \beta_0 + \beta_1 X + \epsilon$. Entonces, los pasos a seguir serían los siguientes (véase Sección 2.1 y 2.2 de la práctica de *Modelos de regresión lineal con R*).

1. Simular una muestra de tamaño $n = 100$ de valores de (x_i, y_i), $i \in \{1, \dots, n\}$.

   a) Para ello, se pueden generar 100 observaciones x_i de la uniforme (`runif`).

   b) Los errores del modelo, ϵ_i se pueden generar de la distribución normal de media 0 y varianza σ^2. Por ejemplo, si cogemos $\sigma^2 = 1$, esto se puede hacer con (`rnorm(100,0,1)`).

   c) Los valores de y_i se calcularían como $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

2. Escoger la función a minimizar. En este caso, la suma de los residuos al cuadrado:

$$J(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

3. Obtener las derivadas parciales de la función a minimizar. A partir de estas, aplicad el algoritmo de método de descenso gradiente (véase los apuntes del Tema de *Optimización convexa*). En la Sección 2.1 de la práctica de *Modelos de regresión lineal con R* se dan los pasos a seguir para el caso del modelo de regresión lineal simple.

4. Como en este caso, tenemos disponible la solución óptima a este problema en la función `lm`. Podemos repetir (en un bucle) varias veces los pasos anteriores y hacer una comparativa con los estimadores ($\hat{\beta}_0$ y $\hat{\beta}_1$) que devuelve dicha función. Por ejemplo, podría ser de interés comparar el valor de la función minimizada, $J(\hat{\beta}_0, \hat{\beta}_1)$, con nuestros estimadores y con los obtenidos con la función `lm`.