

Practica - día 2

Regresión

Regresión lineal

- Datos físico-químicos del embalse de Cuerda del Pozo de los años 2014 y 2015.
 - Conductividad 2014 vs Salinidad 2014.

- Construye un modelo lineal que calcule los valores de “a” y “b” que minimizan la función de coste.

$$y = f(x) = \alpha_0 + \alpha_1 x$$

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_M^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_M^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_1^{(N)} & x_2^{(N)} & \dots & x_M^{(N)} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_M \end{bmatrix}$$

$$\nabla_{\alpha} \text{Loss} = -X^T (y - X\alpha) = -X^T y + X^T X \alpha = 0$$

$$\alpha = (X^T X)^{-1} X^T y$$

- Compara el resultado con los valores de “a” y “b” calculados a mano:

$$\alpha_1 = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2}$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{x}$$

- y comparar el resultado obtenido con la función “lm” implementada en R:

Regresión no lineal

- Construye un modelo no lineal que calcule los valores de los parámetros que minimizan la función de coste.
 - por ejemplo polinomio de grados 2 y 3

$$y = f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_M x^M$$

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} 1 & x^{(1)} & x^{(1)2} & \dots & x^{(1)M} \\ 1 & x^{(2)} & x^{(2)2} & \dots & x^{(2)M} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x^{(N)} & x^{(N)2} & \dots & x^{(N)M} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_M \end{bmatrix} \quad \alpha = (X^T X)^{-1} X^T y$$

- Compara los modelos no lineales con los obtenidos con la función implementada en R:
 - *"glm(y~poly(x, grado del polinomio))"*
- Comprueba todos los modelos calculando la función de coste con los datos de Conductividad vs Salinidad de 2015.

$$cost = \sum_{i=1}^N \|y^{(i)} - f(x_1^{(i)}, \dots, x_M^{(i)}, \Theta)\|$$