

EMALCA 2025 - Introduccion al Aprendizaje Estadístico

Tarea #1.

Problema 1. Considera X una variable aleatoria real con segundo momento finito $E[X^2] < \infty$. Encuentra un número fijo c que minimiza:

$$L(c) = E[(X-c)^2].$$

Muestra todos tus pasos con cuidado.

Problema 2.

Considera $Y = f(X) + \epsilon$ con $E(\epsilon) = 0$, $\text{Var}(\epsilon) = 1$
y $f(x) = x$ con $\epsilon_i \stackrel{\text{iid}}{\sim} N(0,1)$

Entrenamos en los puntos $X_1 = 2$
 $X_2 = 2.5$
 $X_3 = 3$

Predicimos (validamos en el punto) $X_* = 2.2$

Consideramos Dos Modelos:

a) 1-NN

b) 3-NN

Cual metodo tiene menor error? Explica lo que encuentras usando el compromiso de sesgo - Varianza.

Problema 1.

$$\begin{aligned}L(c) &= E[(x-c)^2] \\&= E[x^2 - 2cx + c^2] \\&= E[x^2] - 2cE[x] + c^2\end{aligned}$$

$$L'(c) = -2E[x] + 2c = 0$$

$$\boxed{c = E[x]}$$

$$L''(c) = 2 > 0 \Rightarrow \text{Mínimo Global.}$$

una función cuadrática solo tiene un mínimo.

Problema 2.

Estimación

$$1\text{-NN} \quad \hat{f}_{(1)} = Y_1 = 2 + \epsilon_1$$

$$\begin{aligned}3\text{-NN} \quad \hat{f}_{(3)} &= \frac{Y_1 + Y_2 + Y_3}{3} = \frac{(2 + \epsilon_1) + (2.5 + \epsilon_2) + (3 + \epsilon_3)}{3} \\&= 2.5 + \frac{\epsilon_1 + \epsilon_2 + \epsilon_3}{3}\end{aligned}$$

Error Cuadrático Medio (ECM)

1-NN

$$\text{Cesgo: } (2 - 2.2)^2 = 0.04$$

$$\text{Varianza: } \text{Var}(\epsilon_1) = 1$$

$$\text{ECM: } 0.04 + 1 = \boxed{1.04}$$

3-NN

$$\text{Cesgo: } (2.5 - 2.2)^2 = 0.09$$

$$\text{Varianza: } \text{Var}\left(\frac{1}{3}(\epsilon_1 + \epsilon_2 + \epsilon_3)\right) = \frac{1}{3}$$

$$\text{ECM: } 0.09 + 0.33 \approx \boxed{0.42}$$

Miramos que el sesgo del 1-NN es menor que el de 3-NN pero la varianza es mucho mas. En total el metodo de 3-NN es mejor.