

Taller - Quiz: Ajuste Polinomial y Regularización

Objetivo

Practicar conceptos de regresión polinomial, construcción de la matriz de características y análisis de error con datos de entrenamiento, validación y prueba.

1. Lectura de datos

- Lea el fichero `DatosRegresion1D.xlsx`.
- ¿Qué dimensiones tiene el dataset cargado?
- ¿Cuáles son las variables que contiene?

2. Modelo polinomial

Suponga que los datos siguen un modelo polinomial de orden $r = 7$, definido como:

$$h(x) = w_0x^0 + w_1x^1 + w_2x^2 + \cdots + w_rx^r = \sum_{j=0}^r w_jx^j$$

- Construya la matriz de características X de dimensión $n \times d$, con $d = r + 1$.
- ¿Cuál es la forma matricial de X cuando $r = 7$?

3. Ecuación normal

Los coeficientes óptimos se obtienen con la ecuación normal:

$$w^* = (X^T X + n\lambda I_d)^{-1} X^T y$$

donde el coeficiente de regularización de Tikhonov es $\lambda = 0$.

- Entrene el modelo con los primeros 10 clientes ($n_{train} = 10$).
- Calcule w^* .
- Imprima el vector de coeficientes obtenido.

4. Función de coste

La función de coste cuadrática está dada por:

$$J(w) = \frac{1}{n} \sum_{i=1}^n [h(x^{(i)}) - y^{(i)}]^2$$

- Calcule $J(w)$ para distintos órdenes del polinomio $r \in [1, 7]$.
- Evalúe el coste en:
 - Conjunto de entrenamiento ($n_{train} = 10$).
 - Conjunto de prueba ($n_{test} = 990$).

- Dibuje ambas curvas.

5. **Selección de hiperparámetros** Divida el conjunto de datos en:

$$n_{train} = 10, \quad n_{val} = 490, \quad n_{test} = 500$$

- Dibuje el coste de validación en función de $\lambda \in [10^{-4}, 10^4]$.
- Repita para distintos órdenes del polinomio $r \in [1, 7]$.
- ¿Qué combinación de r y λ proporciona el mejor ajuste?