

Reconocimiento de Patrones 2020

Práctica de Laboratorio 1:

Regresión polinomial y validación cruzada

1. Implementar una función para la generación de sets de datos $\mathcal{D} = \{(x_i, t_i)\}_{i=1, \dots, N}$ utilizando una función $t(x)$, con los x_i aleatorios distribuidos dentro cierto intervalo $[a, b]$ adicionándole ruido independiente a cada valor de t_i . Por ejemplo $N = 10$, $t = \sin(2\pi x)$ y $t \in [0, 1]$, ruido $\mathcal{N}(0, \sigma^2)$.
2. Implementar una función que devuelva el vector de pesos óptimo \mathbf{w}^* para un set de datos \mathcal{D} , un grado de polinomio M y error cuadrático sin término de regularización $E_D(\mathbf{w})$ y con término de regularización $E(\mathbf{w})$ dados por

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (y(x_i, \mathbf{w}) - t_i)^2,$$

y

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (y(x_i, \mathbf{w}) - t_i)^2 + \frac{\lambda}{2} \sum_{j=0}^M |w_j|^2,$$

respectivamente, donde

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M = \sum_{j=0}^M w_j x^j,$$

3. Determinar la mejor combinación de valores para los hiperparámetros λ y M mediante validación cruzada (cross-validation). Considerar 3 valores para λ y 3 valores para M .