Reconocimiento de Patrones 2020
Práctica de Laboratorio 1:
Regresión polinomial y validación cruzada

1. Implementar una función para la generación de L sets de datos $\mathcal{D} = \{(x_i, t_i)\}_{i=1,\ldots,N}$. Los x_i serán aleatorios y estarán distribuidos uniformemente dentro cierto intervalo $[a, b]$. Para cada x_i se generará independientemente un valor $t_i = \sin(2\pi x) + r$ correspondiente, donde r será ruido Gaussiano $\mathcal{N}(0, \sigma^2)$. Por ejemplo $L = 200$, $N = 10$, $\sigma = 0.3$ y $[a, b] = [0, 1]$.

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

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

y

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

respectivamente, donde

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

Utilizar la fórmula

$$\mathbf{w}^* = \left(\mathbf{\Phi}^t\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^t\mathbf{t}.$$

para el caso sin término de regularización y

$$\mathbf{w}^* = \left(\lambda\mathbf{I} + \mathbf{\Phi}^t\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^t\mathbf{t}.$$

para el caso con término de regularización. Estimar el valor medio del error y su desviación standard en ambos casos utilizando los L sets de datos generados en el 1er punto, utilizando valores de y de M a seleccionar.

3.
   - (opción A) 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.
   - (opción B) Realizar un análisis sesgo-varianza para distintos valores de λ y un valor fijo de de grado M del polinomio.

     ELEGIR UNA DE ESTAS DOS OPCIONES