

Reconocimiento de Patrones 2018

Práctica de Laboratorio 1: Regresión polinomial

1. Implementar una función para la generación de sets de datos $\mathcal{D} = \{(x_i, t_i)\}_{i=1, \dots, N}$ utilizando una función $x(t)$, con t 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. Implementar una función que evalúe el error sin regularización $E_{DRMS} = \sqrt{2E_D/N}$ y con regularización $E_{RMS} = \sqrt{2E/N}$, para un set de datos \mathcal{D} .
4. Realizar un estudio en el que se grafique la influencia del grado del polinomio sobre el error E_{DRMS} . Para esto, se generarán sets de datos de entrenamiento \mathcal{D}_{tr} y de testeo \mathcal{D}_{test} de igual tamaño N , para valores del grado del polinomio desde $M = 0$ hasta $M = N - 1$. Para cada uno de ellos se evaluará varias veces el error E_{DRMS} y se graficarán el valor medio y la desviación standard muestrales para cada valor de M . Indicar cuál es el mejor rango de valores de M .
5. Para el error con término de regularización, dado por

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

se fijará el grado del polinomio M en un valor igual a $N - 1$ y se realizará un estudio en el que se grafique la influencia del grado del polinomio sobre el error E_{RMS} . Para esto, se generarán sets de datos de entrenamiento \mathcal{D}_{tr} y de testeo \mathcal{D}_{test} de igual tamaño N , para diferentes valores de $\ln \lambda$. Para cada uno de ellos se evaluará varias veces el error E_{RMS} y se graficarán el valor medio y la desviación standard muestrales para cada valor de $\ln \lambda$. Indicar cuál es el mejor rango de valores de $\ln \lambda$.