Universidad de Buenos Aires - Facultad de Ciencias Exactas y Naturales - Depto. de Matemática

Optimización

Primer Cuatrimestre 2023

Trabajo Práctico N° 1: Regularización del modelo lineal

La regresión lineal es un modelo para una variable continua Y como respuesta de otras características o covariables que denotaremos $X_1, X_2, \ldots, X_p \in \mathbb{R}$. La relación viene dada por la fórmula:

$$Y = \sum_{j=1}^{p} \beta_j X_j + \varepsilon,$$

donde ε es un término del error que no depende de las covariables y $\beta_1, \beta_2, \ldots, \beta_p \in \mathbb{R}$ son parámetros desconocidos.

Dadas n observaciones $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^p$, donde $x_j^{(i)}$ representa la j-ésima característica de la observación i y sus respectivas respuestas $y_1, \ldots, y_n \in \mathbb{R}$, se intentará estimar los valores de los parámetros desconocidos. Los criterios usados para determinarlas serán de tal forma que minimicen cierta función que evalúa el desajuste de predecir Y como $\sum_{j=1}^{p} \beta_j X_j$.

Estimador de mínimos cuadrados.

Se busca $\beta \in \mathbb{R}^p$ que minimice la suma residual de cuadrados:

$$f(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_j^{(i)} \right)^2,$$

con $\hat{\beta}$ aquel que minimiza a f.

Ejercicio 1 Calcular su gradiente y matriz hessiana.

Ejercicio 2 Se define la matriz *de diseño* $X \in \mathbb{R}^{n \times p}$ como $X_{ij} = x_j^{(i)}$, es decir, la fila i representa la observación i. Escribir la matriz hessiana en términos de la matriz X y concluir que f resulta convexa en $\mathbb{R}^p$.

Ejercicio 3 Se considera el modelo lineal dado por

$$Y = X_1 + 3X_2 + 5X_3 - 2X_4 + 4X_5 + 9X_6 - 3X_7 + \varepsilon$$

y los archivos `datos1_tp.csv` y `datos2_tp.csv` que contienen 100 observaciones, cada los valores de sus variables explicativas y su respectiva respuesta obtenidas de ese modelo. Para cada tanda de datos, estimar los parámetros β mediante mínimos cuadrados usando el método de descenso por el gradiente combinado con alguna búsqueda lineal. Calcular la cantidad de iteraciones y el error en norma 2 entre β y $\hat{\beta}$ para cada uno y explicar porque difiere dicha cantidad para cada tanda de datos.

Regresión Ridge.

Sea $\alpha > 0$, se busca $\beta \in \mathbb{R}^p$ que minimice la siguiente expresión:

$$f_R(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_j^{(i)} \right)^2 + \alpha \sum_{j=1}^{p} \beta_j^2,$$

con $\hat{\beta}_\alpha$ aquel que minimiza a f.

Ejercicio 4 Calcular su gradiente y matriz hessiana.

Ejercicio 5 Expresar el número de condición de la matriz hessiana de f_R en términos de los autovalores de la de f. ¿Qué sucede cuando $\alpha \to +\infty$?

Ejercicio 6 Para cada tanda de datos y para cada α, estimar los parámetros β minimizando la función de Ridge mediante el método de descenso por el gradiente combinado con la misma búsqueda lineal anterior. Graficar el error en norma 2 entre β y $\hat{\beta}_\alpha$ en función de α y la cantidad de iteraciones en función de α. Explicar el comportamiento del gráfico de la iteraciones.