UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Departamento de Matemática

Tarea 8 de Optimización No Lineal (MAT-279)
Segundo semestre de 2022

En el análisis de datos estadísticos, es común encontrar tablas con grandes cantidades de características sobre un individuo. En esta tarea se tratará el problema de la **regresión logística**, donde se utilizará el conjunto de datos descrito en

`https://archive.ics.uci.edu/ml/machine-learning-databases/ionosphere/`

Para introducir la regresión logística, se requiere introducir notación. Sea Y una variable aleatoria con valores en $\{0,1\}$ y sea $\boldsymbol{X}$ un vector aleatorio p-dimensional de covariables. Una forma de modelar Y utilizando la información de $\boldsymbol{X}$ y la naturaleza dicotómica de Y, es mediante la siguiente probabilidad

$$p(x) = \mathbb{P}(Y = 1 | \boldsymbol{X} = x).$$

Es común introducir modelos paramétricos para modelar $p(x)$, digamos $p(x) = p_0(x, \beta)$. El modelo de regresión logística consiste en utilizar la función logística como modelo paramétrico, es decir:

$$p_0(x, \beta) = \frac{1}{1 + \exp\{-x^\top \beta\}}.$$

En la práctica, el vector de parámetros β no se observa y se debe estimar desde las observaciones. En esta tarea procederemos mediante el método de máxima verosimilitud, que consiste en lo siguiente: sean y_i y x_i las observaciones de Y y $\boldsymbol{X}$ del i-ésimo individuo, respectivamente. El estimador de máxima verosimilitud se obtiene mediante la resolución del siguiente problema de minimización:

$$\min_{\beta \in \mathbb{R}^d} \ell(\beta) := \sum_{i=1}^{N} y_i \ln\left(p_0(x_i, \beta)\right) + (1 - y_i) \ln\left(1 - p_0(x_i, \beta)\right). \tag{1}$$

Los métodos numéricos tradicionales suelen fallar cuando el vector de parámetros tiene una dimensión muy grande. Es en este contexto que se suele resolver (1) considerando alguna penalización:

$$\min_{\beta \in \mathbb{R}^d} \ell(\beta) + \varphi_\lambda(\beta), \tag{2}$$

Para esta tarea se pide:

1. Considere una secuencia creciente de parámetros de penalización: $0 < \lambda_1 < \ldots < \lambda_K$.
2. Resuelva (2) considerando $\varphi_\lambda(\beta) = \lambda \|\beta\|_2^2$ (penalización Ridge).
3. Resuelva (2) considerando $\varphi_\lambda(\beta) = \lambda \|\beta\|_1$ (penalización LASSO) usando el método del gradiente proximal; la solución es este caso la denotaremos por $\hat{\beta}^{\lambda_l}$.
4. Realice un gráfico de $\hat{\beta}_i^{\lambda_l}$ v/s $\{\log \lambda_l\}_{l=1}^K$.

Para escoger el parámetro λ en los casos 1 y 2, se suele proceder mediante el método de validación cruzada con 10 grupos. Específicamente para la regresión LASSO, se le pide proceder de la siguiente manera:

a) Genere 10 grupos al azar, obteniendo los conjuntos $\{(y^{(i)}, x^{(i)})\}_{i=1}^{10}$.

b) Considere una secuencia $\{\lambda_l\}_{l=1}^K$. Eliminando al i-ésimo grupo, obtenga el estimador $\hat{\beta}_i^{\lambda_l}$.

c) Para cada λ_l, calcule el error de predicción del i-ésimo grupo $(1/10)\sum_{i=1}^{10}(Y_j - \hat{Y}_j^{(l,i)})^2$, donde

$$\hat{Y}_j^{(l,i)} = \begin{cases} 1 & \text{si } (1 + \exp\{-x_j^\top \hat{\beta}_i^{\lambda_l}\})^{-1} \geq 0.5 \\ 0 & \text{si } (1 + \exp\{-x_j^\top \hat{\beta}_i^{\lambda_l}\})^{-1} < 0.5. \end{cases} \tag{3}$$

d) Escoja el λ_l que minimice el error de predicción promedio.