Técnicas de optimización

Marcos Esteve Casademunt

Octubre 2018

Demostrar que en cualquier problema de clasificación en C clases, la estimación de máxima verosimilitud de la probabilidad a priori de la clase c, $1 \leq c \leq C$ es $p_c = n_c/N$ donde $N = n_1 + ... + n_c$ es el número total de datos observados y n_c es el número de datos de la clase c.

Modelo:

$$P(c = 1) = p_1, ..., P(c = C) = p_c$$

$$\sum_{c=1}^{C} p_c = 1$$

$$\theta = (p_1, ..., p_c)$$

Verosimilitud y logaritmo de la verosimilitud:

$$P(S|\theta) = \prod_{c=1}^{C} \prod_{i=1}^{n_c} p_c = p_1^{n_1} * ... * p_c^{n_c}$$

$$q_s(\theta) = L_s(\theta) = \log(P(S|\theta)) = \sum_{c=1}^{C} n_c \log(p_c)$$

$$\theta^* = \underset{\sum_{c=1}^{C} p_c = 1}{\operatorname{argmax}} \sum_{c=1}^{C} n_c \log(p_c)$$

Lagrangiana:

$$\Lambda(p_1, ..., p_c, \beta) = \sum_{c=1}^{C} n_c \log(p_c) + \beta(1 - \sum_{c=1}^{C} p_c)$$

Soluciones óptimas en función del multiplicador de Lagrange:

$$\frac{\delta\Lambda}{\delta p_c} = \frac{n_c}{p_c} - \beta = 0$$

$$P_c(\beta) = \tfrac{n_c}{\beta}$$

Función dual de Lagrange: Aplicaremos propiedades de los logarítmos...

$$\Lambda_D(\beta) = \sum_{c=1}^{C} n_c \log(\frac{n_c}{\beta}) + \beta(1 - \sum_{c=1}^{C} \frac{n_c}{\beta})$$

$$= n_1 \log(n_1) - n_1 \log(\beta) + ... + n_c \log(n_c) - n_c \log(\beta) + \beta - N$$

$$= \ -N \log(\beta) + \sum_{c=1}^{C} n_c \log(n_c) + \beta - N$$

Valor óptimo del multiplicador de Lagrange:

$$\frac{\delta\Lambda_D}{\delta\beta} = \frac{-N}{\beta} + 1 = 0 \Rightarrow \beta^* = N$$

Solución final:

$$p_1^* = \frac{n_1}{N} \quad ... \quad p_c^* = \frac{n_c}{N}$$

Existe una variante de la función de Widrow-Hoff que incluye un término de regularización con el objetivo de que los pesos no se hagan demasiado grandes:

$$q_S(\boldsymbol{\theta}) = \frac{1}{2}\sum_{n=1}^{N}\left(\boldsymbol{\theta}^t\boldsymbol{x}_n - y_n\right)^2 + \frac{\boldsymbol{\theta}^t\boldsymbol{\theta}}{2}$$

Aplicando la técnica de descenso por gradiente, obtener la correspondiente variante del algoritmo de Widrow-Hoff y la correspondiente versión muestra a muestra.

$$\bigtriangledown q_s(\theta) = \sum_{n=1}^{N}(\theta^t x_n - y_n)x_n + \theta$$

Suponiendo que:

$$v = (x, y, z) \qquad \frac{\delta v^t v}{\delta v} = \frac{\delta(x^2, y^2, z^2)}{\delta v} = 2v$$

Versión del algoritmo:

$$\theta(1) = arbitrario$$

$$\theta(k+1) = \theta(k) + p_k\sum_{n=1}^{N}(y_n - \theta(k)^t x_n)x_n - \theta(k)$$

Versión del algoritmo muestra a muestra:

$$\theta(1) = arbitrario$$

$$\theta(k+1) = \theta(k) + p_k(y(k) - \theta(k)^t x(k))x(k) - \theta(k)$$