Aprendizaje automatizado

Algoritmos de aprendizaje

Gibran Fuentes-Pineda

Mayo 2023

Algoritmo de optimización mínima secuencial (SMO)

- Divide el problema de optimización en una serie de subproblemas con 2 multiplicadores de Lagrange (es el mínimo debido a la restricción de desigualdad lineal)

- Es posible optimizar cada subproblema de forma analítica

$$
\begin{aligned}
&0 \le \alpha_1, \alpha_2 \le C \\
&y^{(1)} \cdot \alpha_1 + y^{(2)} \cdot \alpha_2 = k
\end{aligned}
$$

  donde k es el negativo de la suma del resto de los términos de la restricción de igualdad

Algoritmo de descenso por subgradiente (PEGASOS)

- La función bisagra no es diferenciable, pero podemos usar el subgradiente

$$\tilde{\nabla} E(\mathbf{w}, b) = \begin{cases} 0, & y^{(i)} \cdot (\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1 \\ y^{(i)} \cdot \mathbf{x}^{(i)}, & y^{(i)} \cdot (\mathbf{w}^\top \mathbf{x}^{(i)} + b) < 1 \end{cases}$$

- Algoritmo
  1. Inicializamos $\mathbf{w}$ y b a 0
  2. Para $t = 1, \dots, T$ realizar
     2.1 Elige ejemplo $\{x^{(i)}, y^{(i)}\}$ aleatoriamente
     2.2 $\eta^{\{t\}} = \frac{1}{\lambda \cdot t}$
     2.3 Si $y^{(i)} \left(\left(\mathbf{w}^{\{t\}} \right)^\top \mathbf{x}^{(i)} + b \right) < 1$

     $$\mathbf{w}^{\{t+1\}} = (1 - \eta^{\{t\}} \cdot \lambda) \cdot \mathbf{w}^{\{t\}} + \eta^{\{t\}} \cdot y^{(i)} \cdot \mathbf{x}^{(i)}$$

     2.4 En caso contrario

     $$\mathbf{w}^{\{t+1\}} = (1 - \eta^{\{t\}} \cdot \lambda) \cdot \mathbf{w}^{\{t\}}$$

PEGASOS kernelizado

- Es posible entrenar un SVM definido a partir de los parámetros
- Esta versión *kernelizada* es la siguiente:
  1. Inicializa $\alpha_i^{\{0\}}, i = 1, \ldots, n$ a 0
  2. Para $t = 1, \ldots, T$ realizar
     2.1 Elige el índice de un ejemplo aleatoriamente $s \in \{1, \ldots, n\}$
     2.2 $\eta^{\{t\}} = \frac{1}{\lambda \cdot t}$
     2.3 Si $y^{(s)} \cdot \left[\eta^{\{t\}} \cdot \sum_{i=1}^{n} \alpha_i \cdot y^{(i)} \cdot K(\mathbf{x}^{(i)}, \mathbf{x}^{(s)})\right] < 1$, entonces
     $$\alpha_s^{\{t\}} = \alpha_s^{\{t-1\}} + 1$$
     2.4 En caso contrario
     $$\alpha_s^{\{t\}} = \alpha_s^{\{t-1\}}$$