Trabajo 2 de APR

Vladyslav Mazurkevych

[T4 - 1]: Realizar el desarrollo completo para la obtención de la función de Lagrange primal y dual para las máquinas de vectores soporte con márgenes blandos (separabilidad no lineal) .

Empezaremos la demostración con la Función Discriminante Lineal.

$$\phi : \Re^d \rightarrow \Re : \phi(x;\ \theta) = \theta^t x + \theta_0 \ = \ \sum_{i=1}^{d} \theta_i x_i + \theta_0 \qquad (1)$$

Aquí las θ representan el vector de pesos y el θ_0 representa el umbral. De aquí podemos obtener una función que separe un dataset que es linealmente separable y cumpliría dicha condición:

$$c_n(\theta^t x_n \ + \ \theta_0) \ > \ 0, \quad 1 < n < N \qquad (2)$$

Este caso nos generará muchas soluciones, por lo tanto, para hacerlo más preciso y generar menos soluciones, le añadimos un margen b a la fórmula:

$$b \varepsilon \Re^{\geq 0} : \ \ c_n(\theta^t x_n \ + \ \theta_0) \ \geq \ b \qquad (3)$$

Por otra parte, dado un plano separador y su FDL canónica con respecto a un conjunto S de N puntos $\theta \ \equiv \ (\theta,\ \theta_0)$, la distancia más cercana del vector $x \varepsilon S$ al hiperplano separador H es:

$$r = \frac{1}{||\theta||} \ (4)$$

Así pues el margen de H con respecto a S se define como:

$$2r = \frac{2}{||\theta||} \qquad (5)$$

Que habría que maximizar, pero la ecuación 5 se puede reescribir como:

$$2r = \frac{1}{2} \theta^t \theta \qquad (6)$$

Pero en la vida real los datos no son perfectamente separables, por lo tanto tenemos un caso de no separabilidad lineal. Vamos a introducir una nueva constante que será la C>0, y será la responsable de ponderar la tolerancia de las malas clasificaciones hechas por nuestro modelo, es decir, cuánto fallo quiero permitir que tenga mi modelo. También añadiremos la ζ que será la tolerancia a fallos para cada muestra dada. por lo tanto, el nuevo término a minimizar sería:

$$\text{Minimizar } \frac{1}{2} \theta^t \theta + C \sum_{n=1}^{N} \zeta \qquad (7)$$

Todo esto estaría sujeto a estas nuevas ecuaciones, donde además de que el margen sea 1, le añadimos el ζ, que será nuestro margen extra, 'márgenes blandos':

$$c_n(\theta^t x_n + \theta_0) \geq 1 - \zeta, \quad 1 \leq n \leq N \qquad (8)$$

$$\zeta_n \geq 0 , 1 \leq n \leq N \qquad (9)$$

Pasando la ecuación 8 a un lado, obtendremos lo siguiente:

$$c_n(\theta^t x_n + \theta_0) + \zeta - 1 \qquad (10)$$

Sustituyendo las ecuaciones obtenidas anteriormente en la ecuación correspondiente a la función Lagraniana, obtendremos la primera función pedida por el ejercicio que será la función Lagrange Primal con márgenes blandos para las SVM:
Minimizar:

$$\Lambda(\theta, \theta_0, \zeta, \alpha, \beta) = \frac{1}{2}\theta^t\theta + C\sum_{n=1}^{N}\zeta_n - \sum_{n=1}^{N}\alpha_n(c_n(\theta^t x_n + \theta_0) + \zeta - 1) - \sum_{n=1}^{N}ß_n\zeta_n \quad (11)$$

sujeto a $\alpha_n \geq 0,\ ß_n \geq 0,\ \zeta_n \geq 0$ para $1 \leq n \leq N$.

***

Para la segunda parte del ejercicio, tenemos que obtener la función Lagraniana dual, para ello vamos a buscar las soluciones óptimas en función del multiplicador de Lagrange derivando sus componentes e igualandolos a 0:

$$\frac{\partial\Lambda}{\partial\theta} = 0 \Rightarrow \theta^* = \sum_{n=1}^{N} c_n a_n x_n$$

$$\frac{\partial\Lambda}{\partial\theta_0} = 0 \Rightarrow \sum_{n=1}^{N} a_n c_n = 0 \qquad (12)$$

$$\frac{\partial\Lambda}{\partial\zeta} = 0 \Rightarrow \text{CN} - \sum_{n=1}^{N}\alpha_n - \sum_{n=1}^{N}ß_n$$

Una vez hemos obtenido las soluciones óptimas, ahora toca sustituirlas en la ecuación de Lagraniana primal de la ecuación (11)

$$\Lambda(\theta^*, \theta_0^*, \zeta^*, \alpha, \beta)\big|_{\theta^*} =$$

$$= \frac{1}{2}\theta^{*t}\theta^* + C\sum_{n=1}^{N}\zeta_n^* - \sum_{n=1}^{N}\alpha_n(c_n(\theta^{*t}x_n + \theta_0^*) + \zeta^* - 1) - \sum_{n=1}^{N}ß_n\zeta_n^*$$

$$= \frac{1}{2}\sum_{n=1}^{N} c_n a_n x_n \sum_{m=1}^{N} c_m a_m x_m + \text{CN}\cdot\zeta^* - \sum_{n=1}^{N}\alpha_n(c_n(\sum_{m=1}^{N} c_m a_m x_m^t x_n + \theta_0^*) + \zeta^* - 1) - \zeta^*\sum_{n=1}^{N}ß_n$$

$$= \frac{1}{2}\sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m + \text{CN}\cdot\zeta^* - \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m - \theta_0^*\sum_{n=1}^{N}\alpha_n c_n - \zeta^*\sum_{n=1}^{N}\alpha_n +$$

$$+ \sum_{n=1}^{N} \alpha_n - \zeta^* \sum_{n=1}^{N} \beta_n$$

$$= -\frac{1}{2} \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m + \sum_{n=1}^{N} \alpha_n - \zeta^* \sum_{n=1}^{N} \alpha_n - \zeta^* \sum_{n=1}^{N} \beta_n + CN \cdot \zeta^* - \theta_0^* \sum_{n=1}^{N} \alpha_n c_n$$

Sacamos factor común de ζ^*.

$$= \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m + \zeta^* (CN - \sum_{n=1}^{N} \alpha_n - \sum_{n=1}^{N} \beta_n) - \theta_0^* \sum_{n=1}^{N} \alpha_n c_n$$

$$= \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m + \zeta^* (CN - \sum_{n=1}^{N} \alpha_n - \sum_{n=1}^{N} \beta_n) - \theta_0^* \sum_{n=1}^{N} \alpha_n c_n \quad (13)$$

Como la ζ^* tiene que ser la mínima y tiene que ser $\zeta^* \geq 0$, vamos a suponer que la mínima es $\zeta^* = 0$, con lo que la ecuación nos quedaría como:

$$= \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m - \theta_0^* \sum_{n=1}^{N} \alpha_n c_n \quad (14)$$

Como sabemos que está sujeta a $\sum_{n=1}^{N} \alpha_n c_n = 0$; $-\theta_0^* \sum_{n=1}^{N} \alpha_n c_n$ se eliminará dejándonos la ecuación Lagraniana Dual pedida en el ejercicio como la que sigue:

$$= \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n,m=1}^{N} c_n c_m a_n a_m x_n^t x_m \quad (15)$$

Una vez sustituidas la θ en la Lagraniana primal y simplificando, obtenemos la Lagraniana Dual con márgenes blandos para las SVM.

[T4 - 2]: Diseñar el grafo dirigido y acíclico para la clasificación en 5 clases utilizando SVM para dos clases y la estrategia uno-contra-uno.

Para resolverlo, usaremos un DAG.

C = {1, 2, 3, 4, 5}