Optimización II

Ejercicio para entregar a mano

Francisco Javier Mercader Martínez

2. En este ejercicio estudiaremos una versión sencilla del problema de clasificación con máquinas de vector soporte. Consideremos un conjunto de datos con sólo dos datos

$$X = \{(-1, -, 1; -1), (1, 1; 1)\}$$

donde las dos primeras componentes de cada uno de los vectores anteriores representa sus características, y la tercera componente sirve para clasificar dicho dato dentro de dos clases $y = +1$ e $y = -1$. Supongamos que el hiperplano de separación de las dos clases de datos se escribe como

$$x_1 + x_2 x + x_3 y = 0$$

Se pide:

a. Demuestra que el problema de clasificación con SVD para estos datos se formula como

$$(Primal)\begin{cases} \text{Minimizar en } (x_1, x_2, x_3) & f(x_1, x_2, x_3) = \dfrac{1}{2}(x_2^2 + x_3^2) \\ \text{Sujeto a} & -x_1 + x_2 + x_3 \geq 1 \\ & x_1 + x_2 + x_3 \geq 1 \end{cases}$$

El problema de clasificación con máquinas de vector soporte busca encontrar un hiperplano que separe los puntos de datos de dos clases, maximizando el margen entre las dos clases. En el caso de un conjunto de datos de 2 características (como es en este caso x_1 y x_2 en cada vector de características), el modelo se presenta como un hiperplano en el espacio de características.

En este caso, se nos da un conjunto de datos:

$$X = \{(-1, -1, 1), (1, 1, 1)\}$$

El hiperplano tiene forma general:

$$x_1 + x_2 \cdot x + x_3 \cdot y = 0$$

Lo que significa que el modelo de clasificación tiene tres parámetros: x_1, x_2, y x_3, que son los coeficientes del hiperplano.

- Formulación del problema primal:

El objetivo es encontrar los parámetros x_1, x_2, y x_3 que minimicen una función de costo, generalmente la norma cuadrática de los coeficientes del hiperplano, y cumplir con las restricciones de clasificación correcta para punto.

La función objetivo es la siguiente, que minimiza la norma del hiperplano (lo que es equivalente a maximizar el margen de separación entre las dos clases):

$$f(x, y) = \frac{1}{2}\left(x_2^2 + x_3^2\right)$$

Esto corresponde a minimizar la norma del vector al hiperplano.

Las restricciones que deben cumplir los puntos para estar correctamente clasificados son las siguientes:

- Para el primer punto $(-1, -1, 1)$ (que pertenece a la clase $y=-1$):

$$-x_1 + x_2 + x_3 \geq 1$$

- Para el segundo punto $(1, 1, 1)$ (que pertenece a la clase $y=1$):

$$x_1 + x_2 + x_3 \geq 1$$

Estas restricciones aseguran que los puntos estén correctamente clasificados de acuerdo con el margen de separación.

Por lo tanto, el problema primal se plantea como:

$$(\text{Primal}) \begin{cases} \text{Minimizar en } (x_1, x_2, x_3) & f(x_1, x_2, x_3 = \frac{1}{2}\left(x_2^2 + x_3^2\right) \\ \text{Sujeto a} & -x_1 + x_2 + x_3 \geq 1 \\ & x_1 + x_2 + x_3 \geq 1 \end{cases}$$

Esto demuestra que la formulación primal es correcta.

b. Escribe el problema (Primal) en su forma estándar y demuestra que se satisfacen las condiciones de convexidad adecuadas para que dicho problema sea equivalente a su dual, el cual calcularemos a continuación.

Para escribir el problema primal en su forma estándar, tenemos que identificar claramente la función objetivo y las restricciones.

Su forma estándar del problema de optimización es:

$$(\text{Primal}) \begin{cases} \text{Minimizar en } x & f(x) \\ \text{Sujeto a} & g_i(x) \geq 0, \ i = 1, \dots, m \\ & h_i(x) = 0, \ i = 1, \dots, p \end{cases}$$

En este caso, la función es convexa ya que es una forma cuadrática en x_2 y x_3. Esto se puede verificar porque la matriz de la forma cuadrática (que está dada por $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$) es semidefinida positiva, lo que garantiza que la función es convexa.

Las restricciones $-x_1 + x_2 + x_3 \geq 1$ y $x_1 + x_2 + x_3 \geq 1$ también son lineales, lo que implica que son restricciones convexas.

Por lo tanto, el problema es convexo, lo que significa que podemos utilizar el enfoque de dualidad para encontrar una solución.

c. Demuestra que el problema dual asociado a (Primal) viene dado por

$$(Dual)\begin{cases}\text{Maximizar} & \Theta(\mu) = -4\mu^2+2\mu \\ \text{Sujeto a} & \mu\geq 0\end{cases}$$

El problema del dual se deriva del problema primal utilizando la técnica de Lagrange para manejar las restricciones.

Para escribir el Lagrangiano del problema primal, asociamos multiplicadores de Lagrange μ_1 y μ_2 con las restricciones:

$$L(x_1, x_2, x_3, \mu_1, \mu_2) = \frac{1}{2}(x_2^2 + x_3^2) + \mu_1(-x_1 + x_2 + x_3 - 1) + \mu_2(x_1 + x_2 + x_3 - 1)$$

Existe el mínimo de L y anula su gradiente.

$$\nabla_{(x_1, x_2, x_3)} = (-\mu_1 + \mu_2,\ x_2 + \mu_1 + \mu_2,\ x_3 + \mu_1 + \mu_2) = (0,0,0)$$

$$\left.\begin{array}{l} -\mu_1 + \mu_2 = 0 \\ x_2 + \mu_1 + \mu_2 = 0 \\ x_3 + \mu_1 + \mu_2 = 0 \end{array}\right\} \begin{array}{l} \longrightarrow \mu_1 = \mu_2 \\ \longrightarrow x_2 = -2\mu_1 \\ \quad\ \ x_3 = -2\mu_1 \end{array}$$

$$L(x_1, \mu_1, \mu_2) = \frac{1}{2}\left((-2\mu_1)^2 + (-2\mu_1)^2\right) + \mu_1\cdot\left(-x_1 + (-2\mu_1) + (-2\mu_1) - 1\right) + \mu_2\left(x_1 + (-2\mu_1) + (-2\mu_1) - 1\right)$$

$$= \frac{1}{2}(4\mu_1^2 + 4\mu_1^2) + \mu_1\cdot(-x_1 - 4\mu_1 - 1) + \mu_2\cdot(x_1 - 4\mu_1 - 1)$$

$$= 4\mu_1^2 - x_1\mu_1 - 4\mu_1^2 - \mu_1 + x_1\mu_2 - 4\mu_1\mu_2 - \mu_2 = \{\mu_1 = \mu_2\}$$

$$= 4\mu_1^2 - \cancel{x_1\mu_1} - 4\mu_1^2 - \mu_1 + \cancel{x_1\mu_1} - 4\mu_1^2 - \mu_1$$

$$= -4\mu_2^2 - 2\mu_1$$

Sujeto a $\mu_1 \geq 0 \longrightarrow \Theta(\mu) = -4\mu^2 + 2\mu$

Finalmente, el problema dual resultante es:

$$(Dual)\begin{cases}\text{Maximizar} & \Theta(\mu) = -4\mu^2 + 2\mu \\ \text{Sujeto a} & \mu \geq 0\end{cases}$$

d. Resuelve el problema dual anterior e infiere de ello que la solución del problema original es $x_1 = 0, x_2 = x_3 = \frac{1}{2}$.

Para resolver el problema dual:

$$\Theta(\mu) = -4\mu^2 + 2\mu$$

tomamos la derivada de $\Theta(\mu)$ con respecto a μ y la igualamos a cero:

$$\Theta'(\mu) = -8\mu + 2 = 0 \rightarrow \mu^* = \frac{1}{4}$$

Sustituyendo $\mu = \frac{1}{4}$ en la función $\Theta(\mu)$:

$$\Theta\left(\frac{1}{4}\right) = -4 \cdot \left(\frac{1}{4}\right)^2 + 2 \cdot \left(\frac{1}{4}\right) = -4 \cdot \left(\frac{1}{16}\right) + \frac{1}{2} = -\frac{1}{4} + \frac{1}{2} = \frac{1}{4}$$

$x_2^* = x_3^* = \left| -2 \cdot \frac{1}{4} \right| = \frac{1}{2}$ y $x_1^* = 0$