Máquinas de Vectores de Soporte

Margen Suave

Luis Norberto Zúñiga Morales

18 de enero de 2022

Contenido

MVS Margen Suave

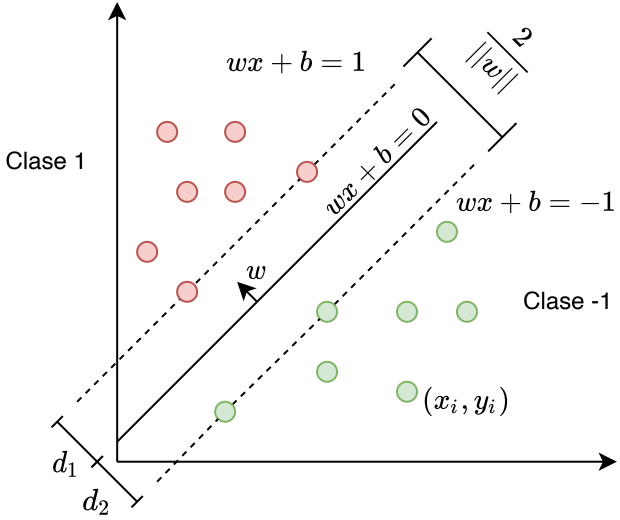


Figura: El caso linealmente separable es un sueño que rara vez se ve en la práctica.

Margen Suave

La idea es relajar las limitantes de H_1 y H_2 para permitir puntos mal clasificados.
Lo anterior se logra introduciendo un variable de holgura positiva ξ_i, $i = 1, \ldots, L$:

$$\boldsymbol{x}_i \cdot \boldsymbol{w} + b \geq +1 - \xi_i \qquad y_i = +1 \tag{1}$$
$$\boldsymbol{x}_i \cdot \boldsymbol{w} + b \leq -1 + \xi_i \qquad y_i = -1 \tag{2}$$
$$\xi_i \geq 0 \qquad \forall i \tag{3}$$

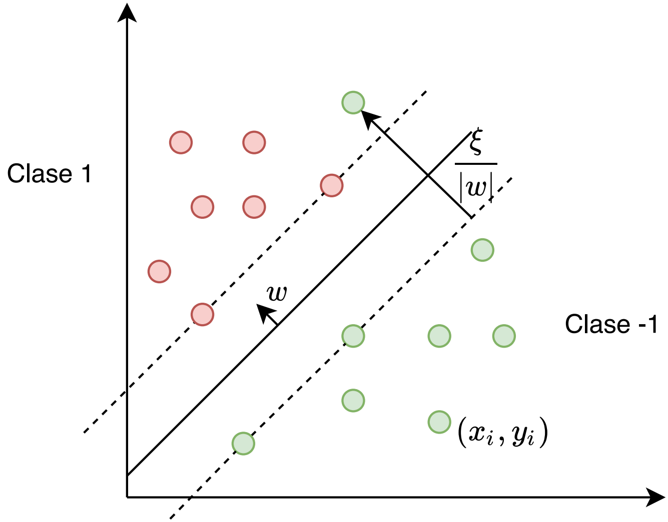


Figura: Idea del margen suave de una MVS.

Ecuaciones con Margen Suave

$$
\begin{aligned}
\mathbf{x}_i \cdot \mathbf{w} + b &\geq +1 - \xi_i & y_i &= +1 \\
\mathbf{x}_i \cdot \mathbf{w} + b &\leq -1 + \xi_i & y_i &= -1 \\
\xi_i &\geq 0 & &\forall i
\end{aligned}
$$

Similar a la ecuación al caso lineal, las ecuaciones (1), (2) y (3) se pueden combinar en una sola.

Ejercicio

Combinar las ecuaciones (1), (2) y (3) en una sola.

Ecuaciones con Margen Suave

$$
\begin{aligned}
\boldsymbol{x}_i \cdot \boldsymbol{w} + b &\geq +1 - \xi_i & y_i &= +1 \\
\boldsymbol{x}_i \cdot \boldsymbol{w} + b &\leq -1 + \xi_i & y_i &= -1 \\
\xi_i &\geq 0 & \forall i
\end{aligned}
$$

Solución

$$
y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \quad \text{donde} \quad \xi_i \geq 0 \; \forall i \tag{4}
$$

Para incluir esta idea en el modelo de la MVS, se adapta el problema de optimización del caso lineal de la siguiente manera:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \quad \forall i \end{aligned} \tag{5}$$

donde el parámetro C controla la razón de intercambio entre la penalización y el tamaño del margen.

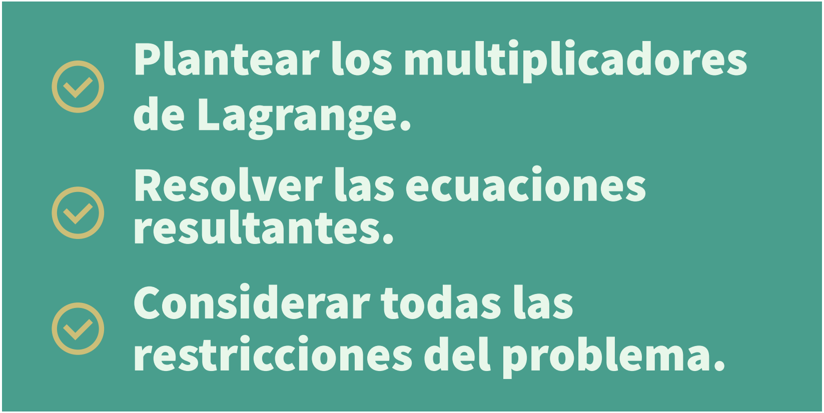
- Se definió el concepto de margen duro .

- Se definió el concepto de margen duro.
- Se introdujo la idea del margen suave para considerar errores de clasificación.

MVS Margen Suave

Resumen hasta el momento

- Se definió el concepto de margen duro.
- Se introdujo la idea del margen suave para considerar errores de clasificación.
- Se formuló el problema de optimización ajustado a la idea del margen suave.



Plantear los multiplicadores de Lagrange.
Resolver las ecuaciones resultantes.
Considerar todas las restricciones del problema.

Problema de Optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \quad \forall i \end{aligned}$$

Problema de Optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \quad \forall i \end{aligned}$$

Similar al caso lineal, es momento de formular el Lagrangiano:

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i - \sum_{i=1}^{L}\alpha_i[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{L}\mu_i\xi_i \quad (6)$$

Consideraciones al momento de resolver el problema de optimización

De manera similar a la MVS Lineal, debemos encontrar $\boldsymbol{w}$, b y ahora ξ_i que minimicen

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \quad \forall i \end{aligned}$$

y $\boldsymbol{\alpha}$ y $\boldsymbol{\mu}$ que maximicen

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i - \sum_{i=1}^{L}\alpha_i[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{L}\mu_i\xi_i$$

Ejercicio

Dado el Lagrangiano

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i - \sum_{i=1}^{L}\alpha_i[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{L}\mu_i\xi_i$$

encontrar

$$\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}} = 0 \quad (?)$$
$$\frac{\partial \mathbb{L}_P}{\partial b} = 0 \quad (?)$$
$$\frac{\partial \mathbb{L}_P}{\partial \xi_i} = 0 \quad (?)$$

Solución

Diferenciando con respecto a $\boldsymbol{w}$, b y ξ_i e igualando a cero:

$$\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{L} \alpha_i y_i x_i = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i \tag{7}$$

$$\frac{\partial \mathbb{L}_P}{\partial b} = \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{8}$$

$$\frac{\partial \mathbb{L}_P}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \Rightarrow C = \alpha_i + \mu_i \tag{9}$$

Ejercicio

Sustituir

$$\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{L} \alpha_i y_i x_i = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i$$

$$\frac{\partial \mathbb{L}_P}{\partial b} = \sum_{i=1}^{L} \alpha_i y_i = 0$$

$$\frac{\partial \mathbb{L}_P}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \Rightarrow C = \alpha_i + \mu_i$$

en la ecuación del Lagrangiano

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L} \xi_i - \sum_{i=1}^{L} \alpha_i [y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{L} \mu_i \xi_i$$

Solución

Ya que $\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}}$ y $\frac{\partial \mathbb{L}_P}{\partial b}$ son las mismas que en el caso lineal, obtenemos

$$\mathbb{L}_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j}, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{10}$$

$$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i H_{ij} \alpha_j, \quad H_{ij} = y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j} \tag{11}$$

$$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H} \boldsymbol{\alpha}, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{12}$$

Sin embargo, al considerar la ecuación

$$C = \alpha_i + \mu_i$$

y el hecho que $\mu_i \geq 0 \; \forall i$, se sigue que $\alpha > C$. Por lo tanto, se debe encontrar:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}\boldsymbol{\alpha} \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq C \;\; \forall_i \\ & \sum_{i=1}^{L} \alpha_i y_i = 0 \end{aligned} \tag{13}$$

Ejercicio

Para encontrar b, cualquier punto x_s que funcione como vector de soporte satisface la ecuación

$$y_s(\boldsymbol{x}_s \cdot \boldsymbol{w} + b) = 1 \tag{14}$$

Considerando que

$$\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i$$

obtendríamos el mismo valor de b para el caso de margen duro. ¿Qué consideración debemos añadir el caso del margen suave?

Solución

Para calcular b, se sigue la misma idea que en el caso lineal dado que

$$\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i$$

son la misma para el caso de margen duro y suave. En este caso, los vectores de soporte son aquellos cuyos índices i cumplan la desigualdad $0 \leq \alpha_i \leq C$.

MVS Margen Suave

Implementación Práctica

1. Crear $\boldsymbol{H}$, donde $H_{ij} = y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$.

1. Crear $\boldsymbol{H}$, donde $H_{ij} = y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$.
2. Elegir el valor del parámetro C, el cual permitirá penalizar clasificaciones erróneas.

1. Crear $\boldsymbol{H}$, donde $H_{ij} = y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$.
2. Elegir el valor del parámetro C, el cual permitirá penalizar clasificaciones erróneas.
3. Encontrar las α_i que maximicen

$$\sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}\boldsymbol{\alpha}$$

sujeto a las restricciones

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{L} \alpha_i y_i = 0.$$

mediante un programa para resolver problemas de optimización cuadrática.

4. Calcular $\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i \boldsymbol{x}_i$.

4. Calcular $\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i \boldsymbol{x}_i$.
5. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.

4. Calcular $\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i \boldsymbol{x}_i$.
5. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.
6. Calcular el valor de b mediante la ecuación

$$b = \frac{1}{N_S} \sum_{s \in S} (y_s - \sum_{k \in S} \alpha_k y_k \boldsymbol{x}_k \cdot \boldsymbol{x}_s).$$

4. Calcular $\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i \boldsymbol{x}_i$.
5. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.
6. Calcular el valor de b mediante la ecuación

$$b = \frac{1}{N_S} \sum_{s \in S} (y_s - \sum_{k \in S} \alpha_k y_k \boldsymbol{x}_k \cdot \boldsymbol{x}_s).$$

7. Cada elemento del conjunto de prueba x_t se clasifica evaluando

$$y_t = sgn(\boldsymbol{w}^T \cdot \boldsymbol{x}_t + b).$$