Máquinas de Vectores de Soporte

Caso Lineal

Luis Norberto Zúñiga Morales

15 de agosto de 2022

Contenido

Motivación

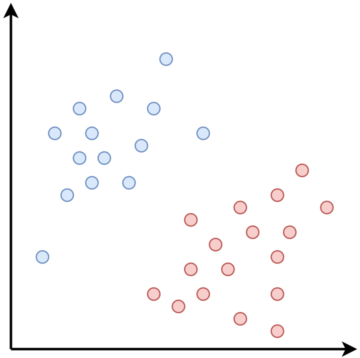

Figura: ¿De cuántas formas se puede separar un conjunto de datos?

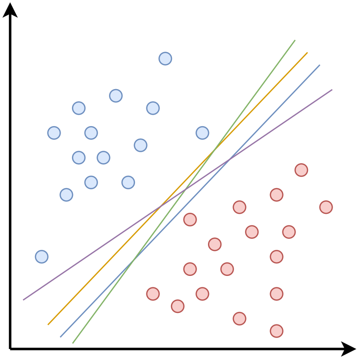

Figura: ¿Cuál de ellas es la mejor?

Iniciemos con algunos supuestos:

- Para formular la idea básica de la MVS es útil comenzar con el caso más sencillo $\rightarrow$ linealmente separable.
- Se tienen L puntos de entrenamiento $\{\boldsymbol{x}_i, y_i\}$.
- $\boldsymbol{x}_i \in \mathbb{R}^D$ representa un vector de entrada de dimensión D.
- Dos clases posibles: $y_i = +1$ o $y_i = -1$

El **hiperplano separador** se puede definir de la siguiente manera:

El **hiperplano separador** se puede definir de la siguiente manera:

$$\boldsymbol{w}^T \cdot \boldsymbol{x} + b = 0 \tag{1}$$

donde:

- $\boldsymbol{w}$ es un vector normal al hiperplano
- $\frac{b}{||\boldsymbol{w}||}$ es la distancia perpendicular del hiperplano al origen

El **hiperplano separador** se puede definir de la siguiente manera:

$$\boldsymbol{w}^T \cdot \boldsymbol{x} + b = 0 \tag{1}$$

donde:

- $\boldsymbol{w}$ es un vector normal al hiperplano
- $\frac{b}{||\boldsymbol{w}||}$ es la distancia perpendicular del hiperplano al origen

Para determinar el hiperplano separador, es posible utilizar a los puntos $\boldsymbol{x}_i$ que se encuentren más cercanos.

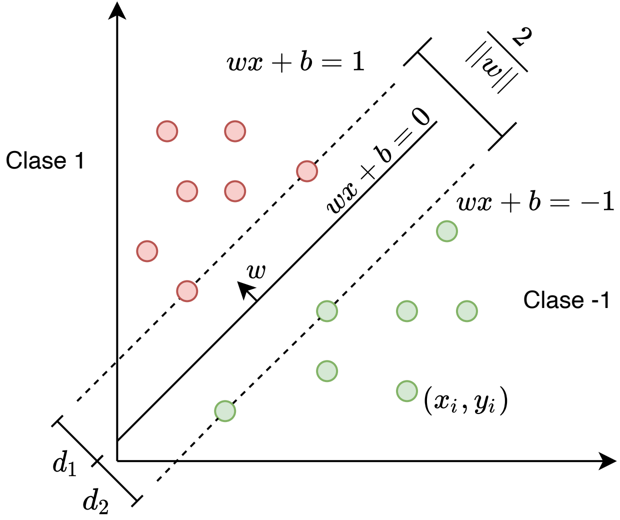

Figura: Idea general de la Máquina de Vectores de Soporte

Para encontrar el hiperplano separador mediante la MVS, se necesita encontrar aquellos $\boldsymbol{w}$ y b tales que

$$\boldsymbol{w}^T \cdot x_i + b \geq +1 \quad \text{para } y_i = +1 \tag{2}$$
$$\boldsymbol{w}^T \cdot x_i + b \leq -1 \quad \text{para } y_i = -1 \tag{3}$$

Para encontrar el hiperplano separador mediante la MVS, se necesita encontrar aquellos $\boldsymbol{w}$ y b tales que

$$\boldsymbol{w}^T \cdot x_i + b \geq +1 \quad \text{para } y_i = +1 \tag{2}$$
$$\boldsymbol{w}^T \cdot x_i + b \leq -1 \quad \text{para } y_i = -1 \tag{3}$$

Ejercicio

¿Cómo se pueden combinar las ecuaciones 2 y 3 en una sola?

Las ecuaciones (2) y (3) se pueden combinar convenientemente en una sola ecuación:

$$y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \qquad\qquad \forall\ i \tag{4}$$

Las ecuaciones (2) y (3) se pueden combinar convenientemente en una sola ecuación:

$$y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \qquad \forall\ i \tag{4}$$

Las ecuaciones de los hiperplanos H_1 y H_2 de soporte (uno para cada clase) se encuentran dadas por:

$$\boldsymbol{w}^T \cdot \boldsymbol{x} + b = +1 \qquad \text{para } H_1 \tag{5}$$

$$\boldsymbol{w}^T \cdot \boldsymbol{x} + b = -1 \qquad \text{para } H_2 \tag{6}$$

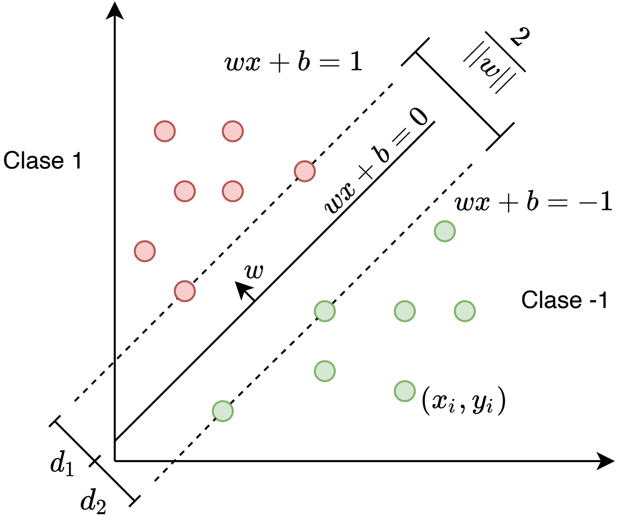

Figura: Idea general de la Máquina de Vectores de Soporte

Idea Principal de la MVS

Hacer el margen de la MVS ($d_1 + d_2$) lo más amplio posible.

Idea Principal de la MVS

Hacer el margen de la MVS ($d_1 + d_2$) lo más amplio posible.

$$\vdots$$

Problema

¿Cuánto mide ese margen?

Ejercicio

Determinar el valor de $d_1 + d_2$ para la MVS.

Tip: ¿Cómo expresar la distancia en términos de $\boldsymbol{w}$ por medio de un vector normal a $\boldsymbol{w}$?

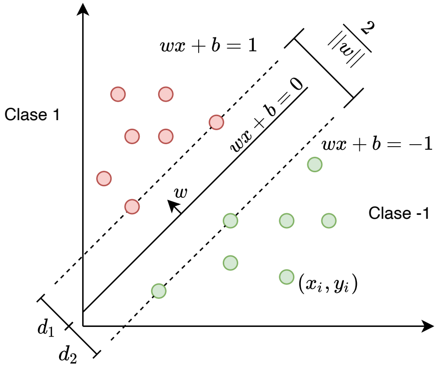

Figura: Idea general de la Máquina de Vectores de Soporte

El margen de la MVS esta dado por

$$d_1 + d_2 = z = \frac{2}{||\boldsymbol{w}||}.$$

El margen de la MVS esta dado por

$$d_1 + d_2 = z = \frac{2}{||\boldsymbol{w}||}.$$

Margen de la MVS

Para hacerlo lo más amplio posible, debemos minimizar $||\boldsymbol{w}||$. El problema de optimización resultante es:

$$\begin{aligned} \text{mín} \quad & ||\boldsymbol{w}|| \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned} \tag{7}$$

¿Problema?

¿Hay algún problema con la función $||\boldsymbol{w}||$ al momento de diferenciarla?

¿Problema?

¿Hay algún problema con la función $||\boldsymbol{w}||$ al momento de diferenciarla?

El problema de optimización en (7) puede transformarse de la siguiente forma:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned} \tag{8}$$

Recapitulemos...

Hasta este momento hemos hecho lo siguiente:

1. Ya tenemos la idea básica de la MVS.
2. Encontramos cuanto mide el margen amplio.
3. Determinamos cómo maximizar ese margen.

Recapitulemos...

Hasta este momento hemos hecho lo siguiente:

1. Ya tenemos la idea básica de la MVS.
2. Encontramos cuanto mide el margen amplio.
3. Determinamos cómo maximizar ese margen.

¿Qué sigue?

1. Resolver el problema de optimización.
2. Plantear su implementación práctica.

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

El problema de optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned} \tag{11}$$

El problema de optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

Vamos a utilizar multiplicadores de Lagrange α, donde $\alpha_i \geq 0 \ \forall i$.

$$(11)$$

El problema de optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

Vamos a utilizar multiplicadores de Lagrange α, donde $\alpha_i \geq 0 \; \forall i$.

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 - \alpha[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1] \qquad \forall i \tag{9}$$

(11)

El problema de optimización

$$\begin{aligned} \min \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

Vamos a utilizar multiplicadores de Lagrange α, donde $\alpha_i \geq 0 \; \forall i$.

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 - \alpha[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1] \qquad \forall i \tag{9}$$

$$= \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{L} \alpha_i[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1] \tag{10}$$

$$\tag{11}$$

El problema de optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

Vamos a utilizar multiplicadores de Lagrange $\boldsymbol{\alpha}$, donde $\alpha_i \geq 0 \ \forall i$.

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 - \boldsymbol{\alpha}[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1] \qquad \forall i \tag{9}$$

$$= \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{L} \alpha_i[y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) - 1] \tag{10}$$

$$= \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{L} \alpha_i y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) + \sum_{i=1}^{L} \alpha_i \tag{11}$$

Consideraciones al momento de optimizar

Se deben encontrar $\boldsymbol{w}$ y b tales que minimicen

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

Consideraciones al momento de optimizar

Se deben encontrar $\boldsymbol{w}$ y b tales que minimicen

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 \\ \text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) \geq 1 \end{aligned}$$

y los $\alpha_i \geq 0$ que maximicen

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{L} \alpha_i y_i(\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) + \sum_{i=1}^{L} \alpha_i. \tag{12}$$

Función a diferenciar

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

Función a diferenciar

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

Ejercicio

Diferencien $\mathbb{L}_P$ con respecto de $\boldsymbol{w}$ y b, igualando las derivadas parciales a cero:

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

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

Función a diferenciar

$$\mathbb{L}_P = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{L} \alpha_i y_i (\boldsymbol{w}^T \cdot \boldsymbol{x_i} + b) + \sum_{i=1}^{L} \alpha_i \tag{14}$$

Función a diferenciar

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

Diferenciemos $\mathbb{L}_P$ con respecto de $\boldsymbol{w}$ y b, igualando las derivadas parciales a cero:

(14)

Función a diferenciar

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

Diferenciemos $\mathbb{L}_P$ con respecto de $\boldsymbol{w}$ y b, igualando las derivadas parciales a cero:

$$\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{L} \alpha_i y_i x_i = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i \tag{13}$$

$$\tag{14}$$

Función a diferenciar

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

Diferenciemos $\mathbb{L}_P$ con respecto de $\boldsymbol{w}$ y b, igualando las derivadas parciales a cero:

$$\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{L} \alpha_i y_i x_i = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i x_i \tag{13}$$

$$\frac{\partial \mathbb{L}_P}{\partial b} = \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{14}$$

Ecuaciones hasta el momento

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

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

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

Ejercicio

Sustituyan $\frac{\partial \mathbb{L}_P}{\partial \boldsymbol{w}}$ y $\frac{\partial \mathbb{L}_P}{\partial b}$ en $\mathbb{L}_P$.

Al sustituir las ecuaciones (13) y (14) en la ecuación (11), se obtiene una nueva ecuación que depende de α, por lo que hay que maximizar:

$$\mathbb{L}_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j}, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{15}$$

$$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i H_{ij} \alpha_j, \quad H_{ij} = y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j} \tag{16}$$

$$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{H} \boldsymbol{\alpha}, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{L} \alpha_i y_i = 0 \tag{17}$$

Dual y Primal del Problema de Optimización

A la ecuación:

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

se le conoce como el **dual** del **primal**:

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

Hay que maximizar $\mathbb{L}_D$, o bien, minimizar $\mathbb{L}_P$.

El problema de optimización luce de la siguiente forma:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}\boldsymbol{\alpha} \\ \text{sujeto a} \quad & \alpha_i \geq 0 \\ & \sum_{i=1}^{L} \alpha_i y_i = 0 \end{aligned} \tag{18}$$

El problema de optimización luce de la siguiente forma:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}\boldsymbol{\alpha} \\ \text{sujeto a} \quad & \alpha_i \geq 0 \\ & \sum_{i=1}^{L} \alpha_i y_i = 0 \end{aligned} \tag{18}$$

Problema de Optimización Cuadrática

El problema de optimización anterior se le conoce como un problema de **optimización cuadrática.**

Falta encontrar $\boldsymbol{w}$ y b

$\boldsymbol{w}$ se puede encontrar de la ecuación:

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

Ejercicio

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

$$y_s(\boldsymbol{x}_s \cdot \boldsymbol{w} + b) = 1 \tag{19}$$

Sustituyan

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

en la ecuación anterior.
Nota: Recuerden las restricciones sujetas al problema de optimización del dual. ($\alpha_i > 0$)

Ecuaciones

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

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

Al sustituir, se obtiene que

$$y_s(\sum_{k \in S} \alpha_k y_k \boldsymbol{x}_k \cdot \boldsymbol{x_s} + b) = 1 \quad (20)$$

donde S denota el conjunto de índices de los vectores de soporte, el cual se construye al encontrar los índices i tales que $\alpha_i > 0$.

Ejercicio

A partir de

$$y_s(\sum_{k\in S} \alpha_k y_k \boldsymbol{x}_k \cdot \boldsymbol{x_s} + b) = 1 \tag{21}$$

calcular el valor de b.

Ejercicio

A partir de

$$y_s\left(\sum_{k\in S}\alpha_k y_k \mathbf{x}_k \cdot \mathbf{x_s} + b\right) = 1 \tag{21}$$

calcular el valor de b.

Multiplicando por y_s, observando que $y_s^2 = 1$:

$$\begin{aligned} & y_s^2\left(\sum_{k\in S}\alpha_k y_k \mathbf{x}_k \cdot \mathbf{x_s} + b\right) = y_s \\ \Rightarrow \quad & b = y_s - \sum_{k\in S}\alpha_k y_k \mathbf{x}_k \cdot \mathbf{x_s} \end{aligned} \tag{22}$$

Ahora, surge la pregunta sobre cuál o cuáles vectores de soporte $\boldsymbol{x}_s$ utilizar para calcular b.

En la práctica, se utiliza un promedio de todos los vectores en S:

$$b = \frac{1}{N_S} \sum_{s \in S} (y_s - \sum_{k \in S} \alpha_k y_k \boldsymbol{x}_k \cdot \boldsymbol{x}_{\boldsymbol{s}}) \tag{23}$$

Valores Finales

El hiperplano separador óptimo es

$$\boldsymbol{w}^T \cdot \boldsymbol{x} + b = 0$$

donde

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

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