Máquinas de Vectores Soporte (SVM)

Norberto Corral, Carlos Carleos

12 de noviembre de 2025

Máquinas de Vectores Soporte (SVM)

Dados $\vec{w} \in \mathbb{R}^p$ y $b \in \mathbb{R}$, fijos, se define una aplicación lineal $f : \mathbb{R}^p \to \mathbb{R}$ como:

$$f(\vec{x}) = \langle \vec{w}, \vec{x} \rangle + b = \vec{w}^\top \vec{x} + b = \sum_{j=1}^{p} w_j \, x_j + b$$

Se llama hiperplano, π, al subconjunto de puntos $\mathbb{R}^p$ que verifican:

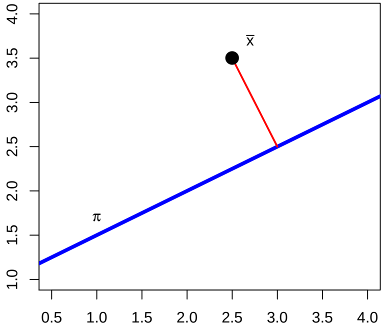
$$\pi = \{ \vec{x} \mid f(\vec{x}) = 0 \}$$

Cualquier hiperplano de $\mathbb{R}^2$ define una recta.

Máquinas de Vectores Soporte (SVM)

El vector $\vec{w}$ es ortogonal a todos los puntos del hiperplano.
La distancia entre un hiperplano π y un punto $\vec{x} \in \mathbb{R}^p$ es:

$$\text{dist}(\pi, \vec{x}) = \frac{f(\vec{x})}{\|\vec{w}\|} \qquad \text{con} \qquad \|\vec{w}\| = \sqrt{\sum_{j=1}^{p} w_j^2}$$



Máquinas de Vectores Soporte (SVM)

Conjunto linealmente separable

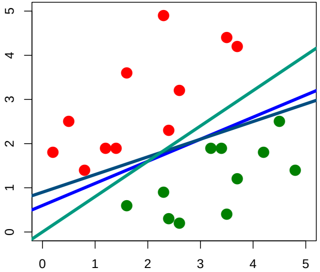
Un conjunto de puntos $(\vec{x}_i, y_i)_{i=1}^{n}$ con $\vec{x}_i \in \mathbb{R}^p$, $y_i \in \{-1, +1\}$, es linealmente separable si existe un hiperplano π tal que:

- $f(\vec{x}_i) = \langle \vec{w}, \vec{x}_i \rangle + b \geq 0 \quad$ si $y_i = +1$
- $f(\vec{x}_i) = \langle \vec{w}, \vec{x}_i \rangle + b \leq 0 \quad$ si $y_i = -1$

Esas dos condiciones son equivalentes a:

$$y_i\, f(\vec{x}_i) = y_i(\langle \vec{w}, \vec{x} \rangle + b) \geq 0 \qquad i = 1, \ldots, n$$

Máquinas de Vectores Soporte (SVM)



Cuando existe un hiperplano de separación un criterio razonable para clasificar un punto $\vec{x}$ es:

$$G(\vec{x}) = \text{signo}\Big(f(\vec{x})\Big)$$

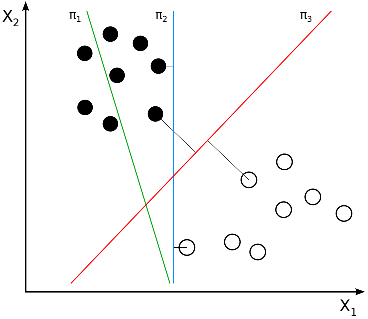
Máquinas de Vectores Soporte (SVM)

Margen de un hiperplano de separación

Dado un conjunto de puntos $(\vec{x}_i)_{i=1}^{n}$ con $\vec{x}_i \in \mathbb{R}^p$, y un hiperplano de separación $\pi = \{\vec{x} \mid \vec{w}^\top \vec{x} + b = 0\}$ se define su margen como el valor M tal que

$$M = \min_{i=1,\dots,n} \operatorname{dist}(\vec{x}_i, \pi)$$

- π_1 no es hiperplano de separación
- π_2 es hiperplano de separación con margen pequeño
- π_3 es el hiperplano de separación con margen máximo



Máquinas de Vectores Soporte (SVM)

Hiperplano óptimo

Es el hiperplano de separación con margen máximo. El problema se puede plantear como la búsqueda del hiperplano óptimo de separación, es decir,

$$\text{Maximizar } M$$

$$y_i \cdot \frac{\langle \vec{w}, \vec{x}_i \rangle + b}{\|\vec{w}\|} \geq M \qquad \forall i = 1, \ldots, n$$

Si

$$y_j \cdot \frac{\langle \vec{w}, \vec{x}_j \rangle + b}{\|\vec{w}\|} = M$$

entonces $\vec{x}_j$ es un **vector soporte**.

Máquinas de Vectores Soporte (SVM)

Como solamente interviene la dirección de $\vec{w}$ y no su norma, se puede imponer la condición

$$M \cdot \|\vec{w}\| = 1 \Longleftrightarrow M = \frac{1}{\|\vec{w}\|}$$

$$\begin{array}{c} \text{Maximizar } \frac{1}{\|\vec{w}\|} \\ y_i \left[\langle \vec{w}, \vec{x} \rangle + b\right] \geq 1 \end{array} \Longleftrightarrow \begin{array}{c} \text{Minimizar } \|\vec{w}\| \\ y_i \left[\langle \vec{w}, \vec{x} \rangle + b\right] - 1 \geq 0 \end{array}$$

$$\Longleftrightarrow \begin{array}{c} \text{Minimizar } \frac{1}{2}\|\vec{w}\|^2 \\ y_i \left[\langle \vec{w}, \vec{x} \rangle + b\right] - 1 \geq 0 \end{array}$$

Máquinas de Vectores Soporte (SVM)

- Problema de optimización cuadrática con restricciones lineales.
- Se aplican las condiciones de Karush–Kuhn–Tucker (KKT) a un *problema cuadrático convexo* con restricciones lineales.
  - formulación general
  - función lagrangiana
  - condiciones KKT y su interpretación

Máquinas de Vectores Soporte (SVM)

Problema cuadrático convexo con restricciones lineales (general)

Sea el problema

$$\text{minimizar } f(\vec{x}) = \tfrac{1}{2}\,\vec{x}^\top Q\vec{x} + \vec{c}^\top\vec{x},$$

$$\text{sujeto a } A\vec{x} \geq \vec{b},$$

con Q simétrica y semidefinida positiva (convexo), $A \in \mathbb{R}^{m\times n}$.

Objetivo: encontrar $\vec{x}^*$ que minimice f cumpliendo las restricciones.

Máquinas de Vectores Soporte (SVM)

Función lagrangiana

Introducimos multiplicadores $\vec{\lambda} \in \mathbb{R}^m$ (uno por restricción), con $\lambda_i \geq 0$. La lagrangiana es

$$L(\vec{x}, \vec{\lambda}) = \frac{1}{2}\,\vec{x}^\top Q\vec{x} + \vec{c}^\top\vec{x} - \vec{\lambda}^\top(A\vec{x} - \vec{b}).$$

Interpretación: se incorporan las restricciones al objetivo con pesos λ_i.

Máquinas de Vectores Soporte (SVM)

Condiciones KKT

Un par $(\vec{x}^*, \vec{\lambda}^*)$ que satisfaga las siguientes condiciones es óptimo (en problemas convexos estas condiciones son necesarias y suficientes):

- Estacionariedad: $\nabla_{\vec{x}} L(\vec{x}^*, \vec{\lambda}^*) = 0.$
- Factibilidad primal: $A\vec{x}^* \geq \vec{b}.$
- Factibilidad dual: $\vec{\lambda}^* \geq 0.$
- Complementariedad: $\lambda_i^*(A_i\vec{x}^* - b_i) = 0, \quad \forall i.$

Aquí A_i denota la i-ésima fila de A.

Máquinas de Vectores Soporte (SVM)

Interpretación de la estacionariedad

La derivada de L respecto a $\vec{x}$:

$$\nabla_{\vec{x}} L = Q\vec{x} + \vec{c} - A^\top \vec{\lambda}.$$

La estacionariedad exige

$$Q\vec{x}^* + \vec{c} - A^\top \vec{\lambda}^* = \vec{0}.$$

Esta ecuación relaciona la solución primal $\vec{x}^*$ con los multiplicadores duales.

Máquinas de Vectores Soporte (SVM)

Interpretación de la complementariedad

La condición $\lambda_i^*(A_i\vec{x}^* - b_i) = 0$ implica:

- Si $\lambda_i^* > 0$ entonces $A_i\vec{x}^* = b_i$ (la restricción está activa).
- Si $A_i\vec{x}^* > b_i$ entonces $\lambda_i^* = 0$ (la restricción tiene holgura).

Por tanto los λ_i^* son "precios sombra": cuánto cambiaría el objetivo al relajar b_i en una unidad.

Dualidad fuerte

Definimos el problema dual como

$$\max_{\vec{\lambda} \geq 0} \inf_{\vec{x}} L(\vec{x}, \vec{\lambda}).$$

Para problemas convexos con condiciones de regularidad (p. ej. Slater), existe *dualidad fuerte*: el valor óptimo primal coincide con el valor óptimo dual. Esto permite resolver el problema a través del dual cuando convenga.

Máquinas de Vectores Soporte (SVM)

Formulación primal en SVM (separable)

$$\begin{aligned} &\text{minimizar } \tfrac{1}{2}\|\vec{w}\|^2 \\ &\text{sujeto a } y_i(\langle \vec{w}, \vec{x}_i \rangle + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

Es un problema cuadrático convexo con restricciones lineales en $(\vec{w}, b)$.

Máquinas de Vectores Soporte (SVM)

Función lagrangiana

Multiplicadores: $\alpha_i \geq 0$ (uno por punto)

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{n} \alpha_i \Big[y_i(\langle \vec{w}, \vec{x}_i \rangle + b) - 1 \Big].$$

Máquinas de Vectores Soporte (SVM)

KKT: condiciones explícitas

Estacionariedad:

$$\frac{\partial L}{\partial \vec{w}} = 0 \;\Rightarrow\; \vec{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \vec{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \;\Rightarrow\; \sum_{i=1}^{n} \alpha_i^* y_i = 0$$

Factibilidad primal:

$$y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) \geq 1$$

Factibilidad dual: $\alpha_i^* \geq 0$
Complementariedad:

$$\alpha_i^* \left[y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) - 1 \right] = 0, \quad \forall i$$

Máquinas de Vectores Soporte (SVM)

De las KKT se deduce:

- Si $y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) > 1$ entonces $\alpha_i^* = 0$ (punto estrictamente dentro del margen).
- Si $y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) = 1$ entonces $\alpha_i^* > 0$ (punto en el margen: vector soporte).

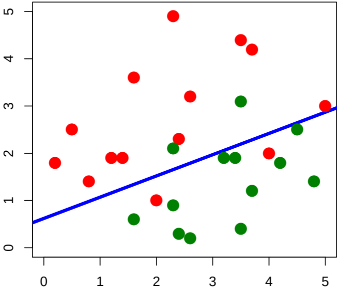
Por tanto la solución se expresa como combinación lineal de *pocos* vectores.

Máquinas de Vectores Soporte (SVM)

En la mayor parte de los problemas reales no existen hiperplanos de separación y se plantea la búsqueda de hiperplanos que verifiquen unas condiciones más suaves:

$$y_i \cdot (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 1 - \xi_i \quad \text{con } \xi_i \geq 0 \;\; i = 1, \ldots, n$$

- Las variables ξ_i se denominan de holgura y permiten la existencia de puntos mal clasificados.
- Cuando $\xi_i = 0$, es el problema anterior.



Máquinas de Vectores Soporte (SVM)

El nuevo problema de optimización viene dado por:

$$L(\vec{w}, \vec{\xi}) = \frac{1}{2}\langle \vec{w}, \vec{w} \rangle + C \sum_{i=1}^{n} \xi_i$$

sujeto a las restricciones

$$y_i \cdot (\langle \vec{w}, \vec{x}_i \rangle + b) + \xi_i - 1 \geq 0; \quad \xi_i \geq 0; \quad i = 1, \ldots, n$$

C representa la penalización de los puntos mal clasificados.

Estos hiperplanos se denominan de margen blando.

Máquinas de Vectores Soporte (SVM)

- La función lagrangiana del hiperplano de margen blando es:

$$L(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \vec{\beta}) =$$

$$= \frac{1}{2}\langle \vec{w}, \vec{w} \rangle + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left[y_i(\langle \vec{w}, \vec{x}_i \rangle + b) + \xi_i - 1 \right] - \sum_{i=1}^{n} \beta_i \xi_i$$

- Aplicando las condiciones KKT se obtiene

$$\forall\ i = 1, \ldots, n : \qquad \frac{\partial L}{\partial \vec{w}} = 0 \iff \vec{w}^* = \sum_{i=1}^{n} \alpha_i^* \cdot y_i \cdot \vec{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \iff \sum_{i=1}^{n} \alpha_i^* \cdot y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \iff C = \alpha_i^* + \beta_i^*$$

$$\alpha_i^* [1 - y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) - \xi_i^*] = 0$$

$$\beta_i^* \, \xi_i^* = 0$$

Máquinas de Vectores Soporte (SVM)

Teniendo en cuenta las condiciones que debe cumplir la solución

$$C = \alpha_i^* + \beta_i^*; \quad \beta_i^* \, \xi_i^* = 0, \quad i = 1, \dots, n$$

$$\alpha_i^*[1 - y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) - \xi_i^*] = 0, \quad i = 1, \dots, n$$

se pueden caracterizar los puntos $\vec{x}_i$:

- $\vec{x}_i$ no es separable $\iff \xi_i^* > 0$. Por tanto $\beta_i^* = 0$, $\alpha_i^* = C$ y
$$y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) = 1 + \xi_i^*$$
- $\alpha_i^* = 0 \implies \beta_i^* = C$, $\xi_i^* = 0$ y el punto$\vec{x}_i$ es separable
- $0 < \alpha_i^* < C \iff \beta_i^* \neq 0 \iff \xi_i^* = 0$. Por lo tanto
$$y_i(\langle \vec{w}^*, \vec{x}_i \rangle + b^*) = 1$$
es decir, $\vec{x}_i$ es un vector soporte y permite estimar b^*
$$b^* = \frac{1}{N_{\text{vs}}} \sum_{k=1}^{N_{\text{vs}}} (y_k - \langle \vec{w}^*, \vec{x}_k \rangle)$$

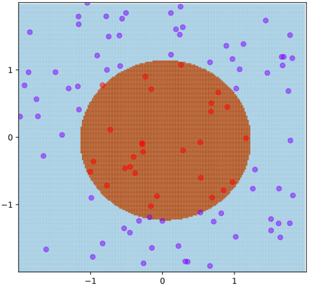
Máquinas de Vectores Soporte (SVM)

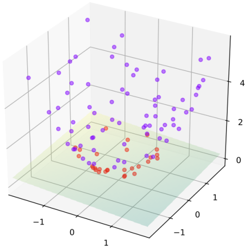
- Cuando las clases no son separables se plantea "pasar" los datos a un espacio de dimensión mayor en el que se cumpla la condición de separabilidad.

$$\begin{aligned} \Phi : \mathbb{X} &\longrightarrow F \\ \vec{x} &\longmapsto \Phi(\vec{x}) = \big(\phi_1(\vec{x}), \ldots, \phi_m(\vec{x})\big) \end{aligned}$$

  donde alguna de las funciones ϕ_j no es lineal.
- El espacio transformado se denomina *espacio de características*.
- El objetivo es buscar el hiperplano óptimo, en el espacio de características, que produce una frontera no lineal en el espacio inicial.

Máquinas de Vectores Soporte (SVM)





Transformación (núcleo): $(a, b) \rightarrow (a, b, a^2 + b^2)$

Máquinas de Vectores Soporte (SVM)

Funciones Núcleo

Sea K una función $K : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$
que verifica las condiciones

1. $K(\vec{x}_1, \vec{x}_2) = K(\vec{x}_2, \vec{x}_1)$ (simétrica)
2. $K(\vec{x}, \vec{x}) \geq 0$ (semidefinida positiva)

Algunos ejemplos de funciones núcleo ($\lambda > 0$, $\gamma > 0$):

- $K(\vec{x}_1, \vec{x}_2) = \langle \vec{x}_1, \vec{x}_2 \rangle$
- $K(\vec{x}_1, \vec{x}_2) = (\lambda \langle \vec{x}_1, \vec{x}_2 \rangle + \gamma)^p$
- $K(\vec{x}_1, \vec{x}_2) = \exp(-\lambda \langle \vec{x}_1 - \vec{x}_2, \vec{x}_1 - \vec{x}_2 \rangle) = \exp(-\lambda \|\vec{x}_1 - \vec{x}_2\|^2)$

Máquinas de Vector Soporte

Teorema de Moore-Aronszajn

Para cualquier función $K : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ simétrica y semidefinida positiva, existe un espacio de Hilbert (normado y completo) y una función $\Phi : \mathbb{X} \rightarrow F$ tal que

$$K(\vec{x}, \vec{x}') = \langle \Phi(\vec{x}), \Phi(\vec{x}') \rangle$$

Este resultado permite calcular el producto escalar $\langle \Phi(\vec{x}), \Phi(\vec{x}') \rangle$ sin necesidad de conocer la función Φ.

Máquinas de Vectores Soporte (SVM)

Derivación del problema dual

Recordando primera condición KKT (estacionariedad):

$$\frac{\partial L}{\partial \vec{w}} = 0 \;\Rightarrow\; \vec{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \vec{x}_i$$

Sustituyendo $\vec{w} = \sum_j \alpha_j y_j \vec{x}_j$ en la lagrangiana y simplificando:

$$L(\vec{\alpha}) = \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle .$$

Problema dual: maximizar $L(\vec{\alpha})$ sujeto a

$$\sum_i \alpha_i y_i = 0, \qquad \alpha_i \geq 0$$

Máquinas de Vectores Soporte (SVM)

$$L(\vec{\alpha}) = \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle.$$

Problema dual: maximizar $L(\vec{\alpha})$ sujeto a

$$\sum_i \alpha_i y_i = 0, \qquad \alpha_i \geq 0$$

Interpretación geométrica y algorítmica:

- El dual depende únicamente de productos escalares $\langle \vec{x}_i, \vec{x}_j \rangle$.
- Esto permite introducir un *kernel* $K(\vec{x}_i, \vec{x}_j)$ y trabajar en espacios de características sin calcular Φ.

Máquinas de Vectores Soporte (SVM)

En el espacio de características se busca la función:

$$f(\vec{x}) = \langle \vec{w}, \Phi(\vec{x}) \rangle + b$$

que en su forma dual se expresa como:

$$f(\vec{x}) = \sum_{i=1}^{n} \alpha_i^* \cdot y_i \cdot K(\vec{x}, \vec{x}_i)$$

Se resuelve mediante el siguiente problema de optimización

$$Max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle$$

$$\sum_{i=1}^{n} \alpha_i \cdot y_i = 0$$

$$0 \leq \alpha_i \leq C \qquad i = 1, \ldots, n$$

Máquinas de Vectores Soporte (SVM)

El problema XOR

Caso	(X_1, X_2)	Y
1	$(+1, +1)$	-1
2	$(-1, -1)$	-1
3	$(+1, -1)$	$+1$
4	$(-1, +1)$	$+1$

Para representar el producto escalar en el espacio de las características se usa el siguiente kernel polinómico:

$$K(\vec{x}, \vec{x}') = (\langle \vec{x}, \vec{x}' \rangle + 1)^2$$

Máquinas de Vectores Soporte (SVM)

El problema dual por resolver es:

$$\text{Maximizar} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot K(\vec{x}, \vec{x}')$$

$$\sum_{i=1}^{n} \alpha_i \cdot y_i = 0$$

$$0 \leq \alpha_i \leq C \qquad i = 1, \ldots, n$$

La solución de α es:

$$\alpha^* = 0{,}125 \quad i = 1, \ldots, 4$$

La función de clasificación que se obtiene es:

$$f(\vec{x}) = 0{,}125 \cdot \sum_{i=1}^{4} y_i \cdot K(\vec{x}, \vec{x}')$$

Máquinas de Vectores Soporte (SVM)

En este caso se pueden obtener las funciones de transformación.

$$\langle \Phi(\vec{x}), \Phi(\vec{x}') \rangle = K(\vec{x}, \vec{x}') = (\langle \vec{x}, \vec{x}' \rangle + 1)^2 =$$

$$1 + x_1^2 (x_1')^2 + x_2^2 (x_2')^2 + 2x_1 x_2 x_1' x_2' + 2x_1 x_1' + 2x_2 x_2'$$

con

$$\Phi = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6)$$

$$f(\vec{x}) = \frac{1}{\sqrt{2}} \phi_4(\vec{x}) = x_1 x_2$$

Máquinas de Vectores Soporte (SVM)

Puntos débiles de las Máquinas de Vector Soporte

- Las MVS son muy sensibles a los parámetros que intervienen en su cáculo. Es aconsejable probar con diferentes valores y analizar los resultados y su estabilidad.
- En los problemas de clasificación es preferible usar un núcleo gausiano y la función objetivo basada en el coste C. En este caso sólo son necesarios los parámetros λ y C. Estrategias posibles de búsqueda:
  - Buscar un C adecuado probando con valores entre 1 y 1000, usando validación cruzada. Con el C seleccionado buscar el λ adecuado.
  - Buscar simultáneamente sobre λ y C, usando una malla.
- Las MVS son sensibles a las unidades de las variables y es conveniente realizar una tipificación previa.

Máquinas de Vectores Soporte (SVM)

La librería **e1071** de R puede usarse para aplicar las máquinas de vector soporte.

```
library (e1071)
mvs_salida <- svm (y ~ x, datos.entrena)
mvs_pred   <- predict (mvs_salida, datos.prueba)
## para optimizar parámetros en una malla:
tune (svm, y ~ x, data = datos.entrena,
      ranges = list (cost=10^(-1:3),
                     gamma=10^(-3:0)))
```

Por omisión, usa núcleo gausiano (radial) donde $\lambda =$ `gamma`.