

Ayudantía 9

Support Vector Machines (SVM) y Kernel

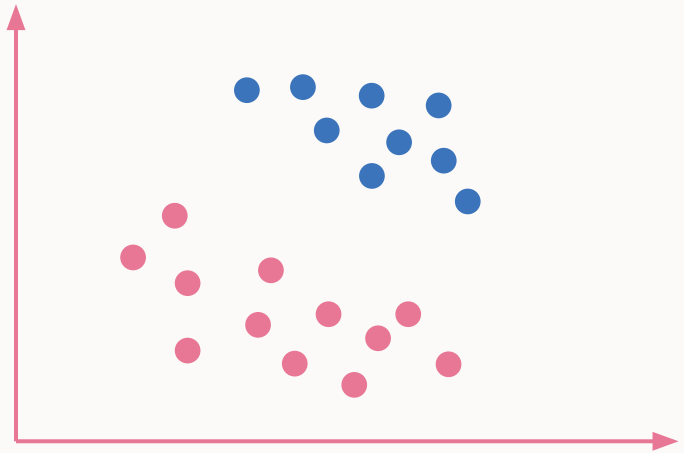
Por Alejandro Held y Celeste Vargas

2 de junio 2024

¿Qué es una SVM?

Es un conjunto de métodos **supervisados** de machine learning diseñados para realizar tareas de **clasificación**.

El objetivo es encontrar una **frontera de decisión óptima** que separe correctamente los datos. Para luego poder **clasificar** nuevos datos.

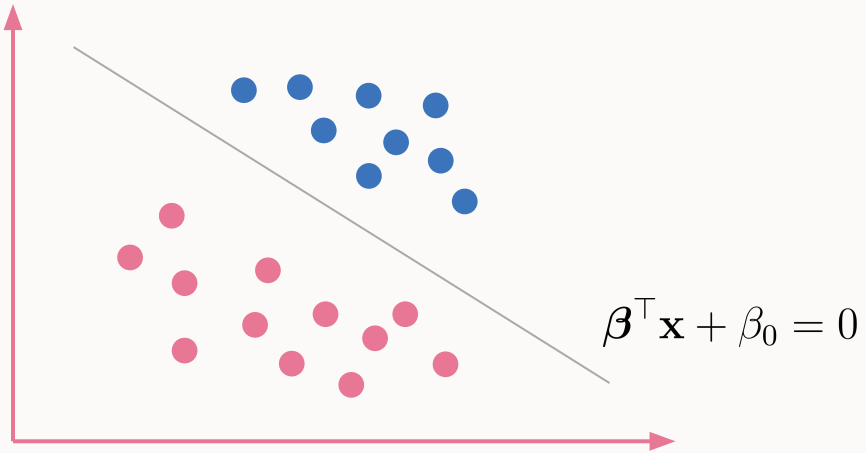


Conceptos importantes

Hiperplano: generalización del concepto de plano a espacios de dimensión arbitraria. Es un subconjunto de un espacio $\mathbb{R}^n$ que tiene dimensión n-1.

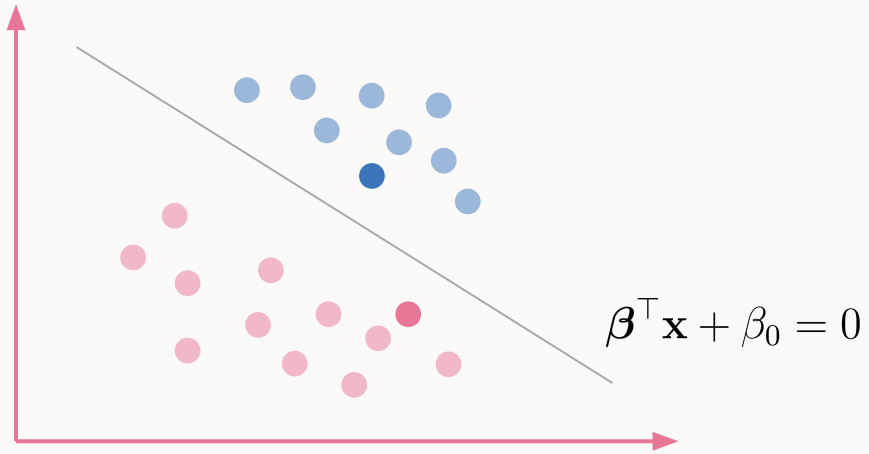
β: vector normal que define la orientación del hiperplano

β_0: define la posición relativa al origen.



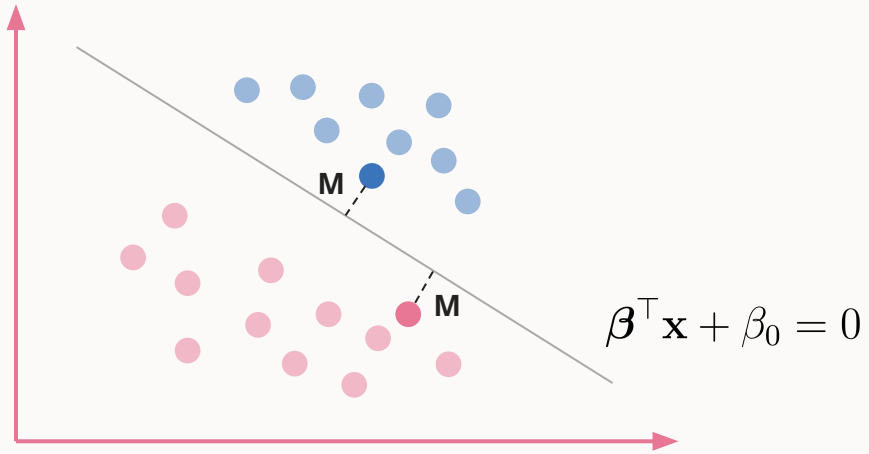
Conceptos importantes

Vectores de soporte: puntos que están más cerca del hiperplano separador.



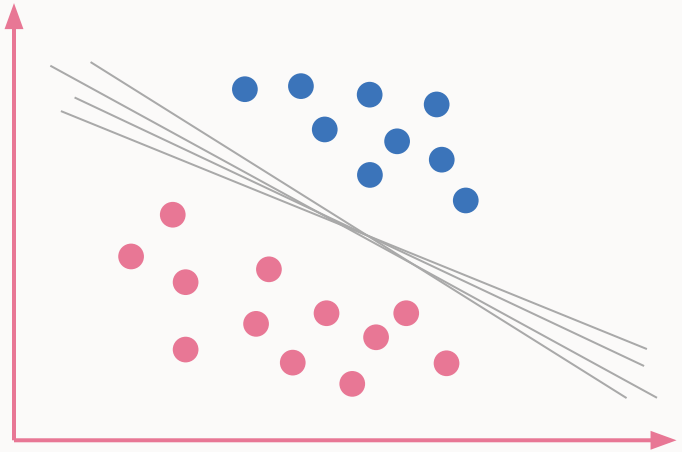
Conceptos importantes

Margen: distancia mínima entre los puntos más cercanos y el hiperplano.



Funcionamiento

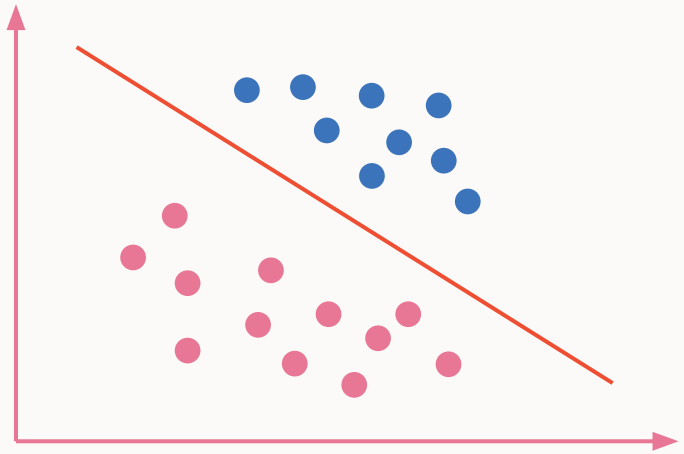
Existen **infinitas** rectas que logran separar los datos.



Funcionamiento

Existen **infinitas** rectas que logran separar los datos.

¿Cómo elegimos la mejor?



Funcionamiento

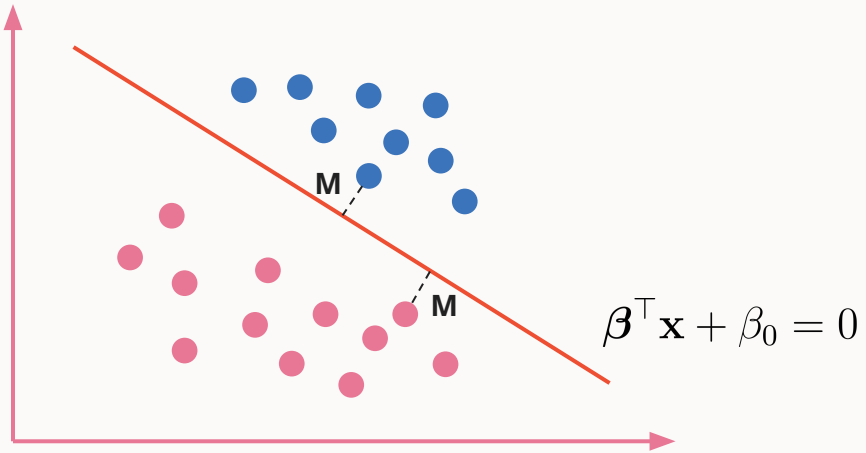
SVM busca encontrar aquel hiperplano que **maximiza** el margen:

$$\max_{\boldsymbol{\beta},\beta_0} \quad M$$

$$\text{s.a.} \quad \frac{1}{\|\boldsymbol{\beta}\|} y_i \left(\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0\right) \geq M, \quad i = 1, \ldots, N$$

$$\min_{\boldsymbol{\beta},\beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2$$

$$\text{s.a.} \quad y_i \left(\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0\right) \geq 1, \quad i = 1, \ldots, N$$



Soft-margin SVM



La restricción impone que el hiperplano clasifique correctamente **todos** los puntos.

Muy sensible al **ruido** o **outliers**: un solo punto mal etiquetado puede hacer que **no exista solución.**

Soft-margin SVM permite que algunos puntos estén **dentro** del margen o sean **mal clasificados.**

Soft-margin SVM

$$\min_{\boldsymbol{\beta},\beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2$$
$$\text{s.a.} \quad y_i\left(\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0\right) \geq 1, \quad i = 1, \ldots, N$$

Hard-margin

$$\min_{\boldsymbol{\beta},\beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum_{i=1}^{N} \xi_i$$
$$\text{s.a.} \quad y_i\left(\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0\right) \geq 1 - \xi_i, \quad i = 1, \ldots, N$$
$$\xi_i \geq 0, \quad i = 1, \ldots, N$$

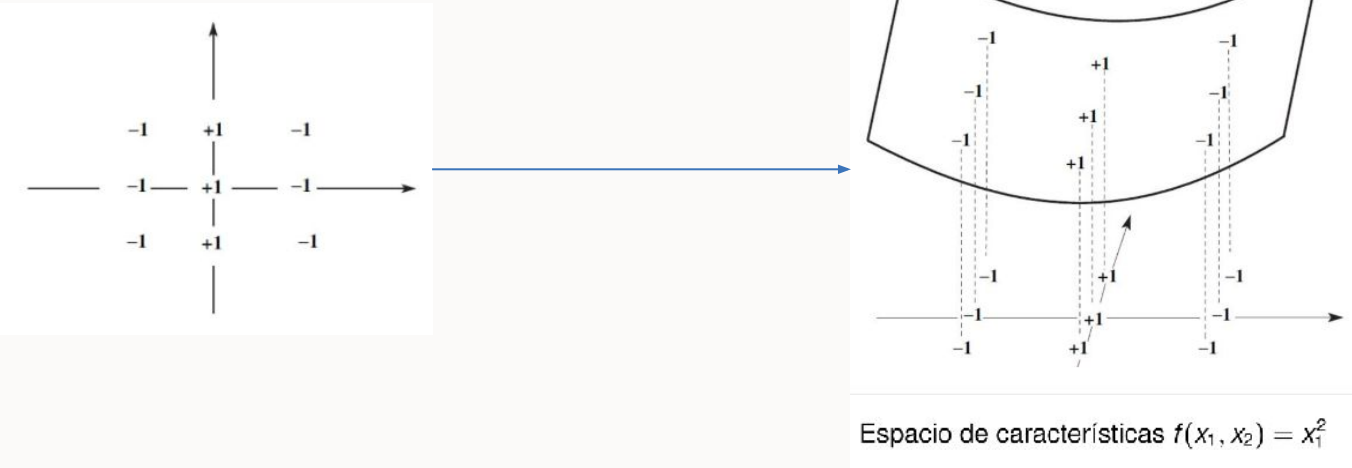
Soft-margin

ξ=0: El punto está bien clasificado y fuera del margen.
0<ξ<1: El punto está bien clasificado, pero dentro del margen.
ξ≥1: El punto está mal clasificado.

Kernel

SVM sigue limitado a ser un **clasificador lineal**

→ **transformación** a las variables para poder representar **más dimensiones** y así poder **clasificar** los datos



¿Cómo puedo transformar los datos a otro espacio?

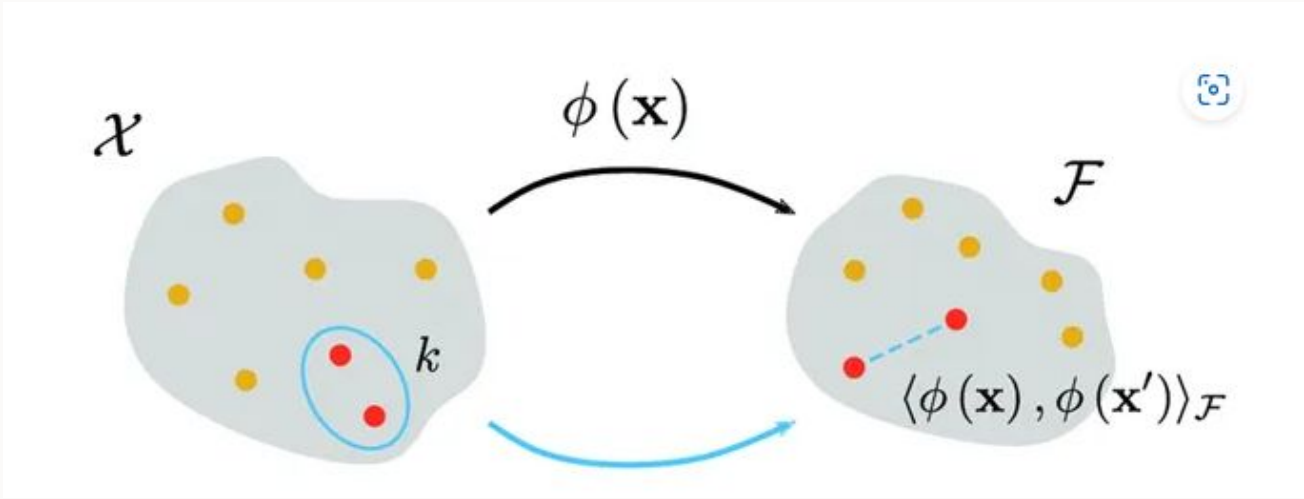


La idea nace de ver SVM dual

$$\max_{\alpha} \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} y^{(i)} y^{(j)} \alpha_i \alpha_j x^{(i)} . x^{(j)}$$

$$\text{subject to } 0 \leq \alpha_i \leq C \text{ for all } i, \text{ and } \sum_{i=1}^{M} \alpha_i y^{(i)} = 0.$$

¿Cómo puedo transformar los datos a otro espacio?



¿Cómo puedo transformar los datos a otro espacio?



$$K(x, x') = \langle \phi(x), \phi(x') \rangle$$

Lineal

$$K(x, x') = x^T x'$$

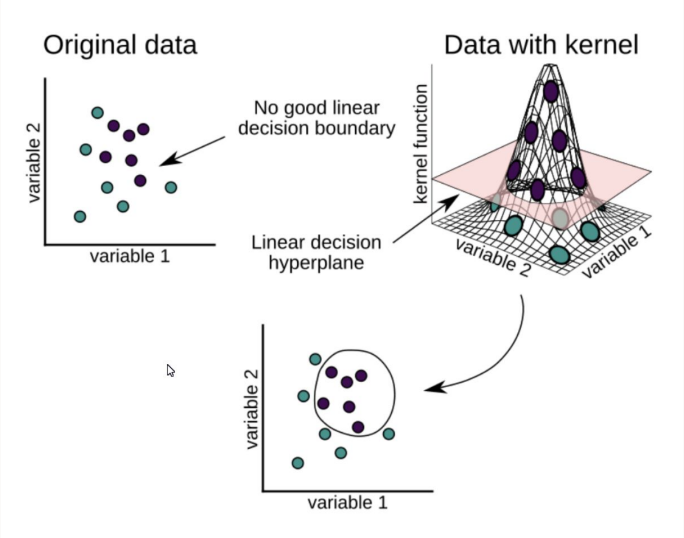
Polinomial

$$K(x, x') = (x^T x' + c)^d$$

RBF

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right)$$

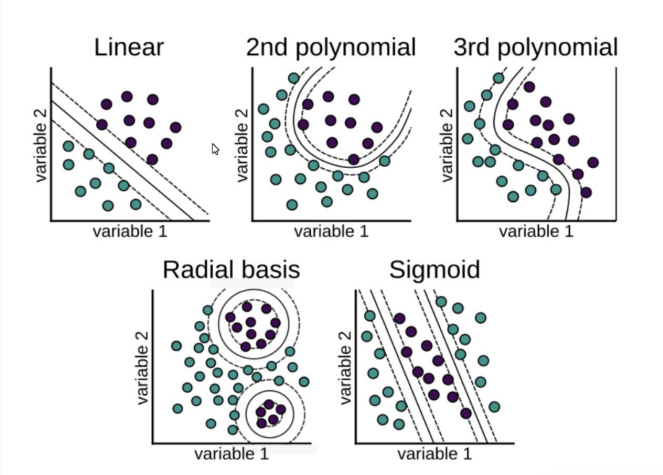
Función de Kernel



Conjunto de funciones que permite **transformar** el espacio de características con el que trabajamos

Entonces… ¿Qué es una función de Kernel?

Función que cambian el “punto de vista” de los datos para encontrar similitudes



Existen varios tipos de Kernel para **distintas aplicaciones**