

Introducción a Máquinas con Vectores de Soporte - SVM

INTRODUCCIÓN

¿Qué es clasificar?

¿Porqué es importante clasificar?

Ejemplos diarios.

Conceptos matemáticos.

Interpretación geométrica - ejemplo

Aplicaciones

Conceptos matemáticos

- Las SVM son clasificadores derivados de la teoría de aprendizaje estadístico postulada por Vapnik y Chervonenkis
- Las SVM fueron presentadas en 1992 y adquirieron fama cuando dieron resultados muy superiores a las redes neuronales en el reconocimiento de letra manuscrita, usando como entrada pixeles.
- Pretenden predecir a partir de lo ya conocido.

Conceptos matemáticos

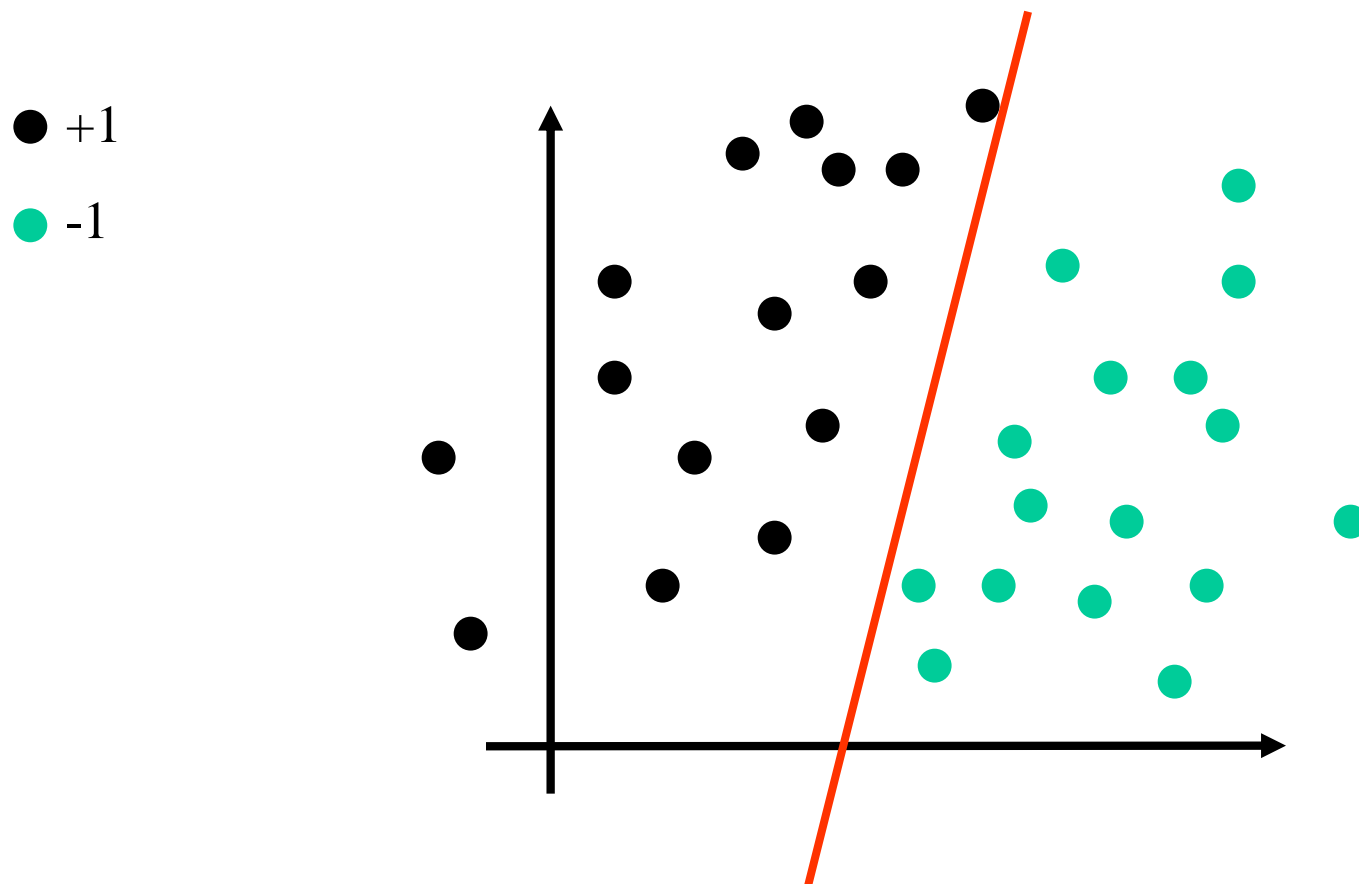
Hay l observaciones y cada una consiste en un par de datos:

un vector $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, l$

una etiqueta $y_i \in \{+1, -1\}$

Supóngase que se tiene un hiperplano que separa las muestras positivas (+1) de las negativas (-1). Los puntos \mathbf{x}_i que están en el hiperplano satisfacen $\mathbf{w} \cdot \mathbf{x} + b = 0$.

Idea inicial de separación



Conceptos matemáticos

\mathbf{w} es normal al hiperplano.

$\frac{|b|}{\|\mathbf{w}\|}$ es la distancia perpendicular del hiperplano al origen.

$\|\mathbf{w}\|$ es la norma euclídea de \mathbf{w}

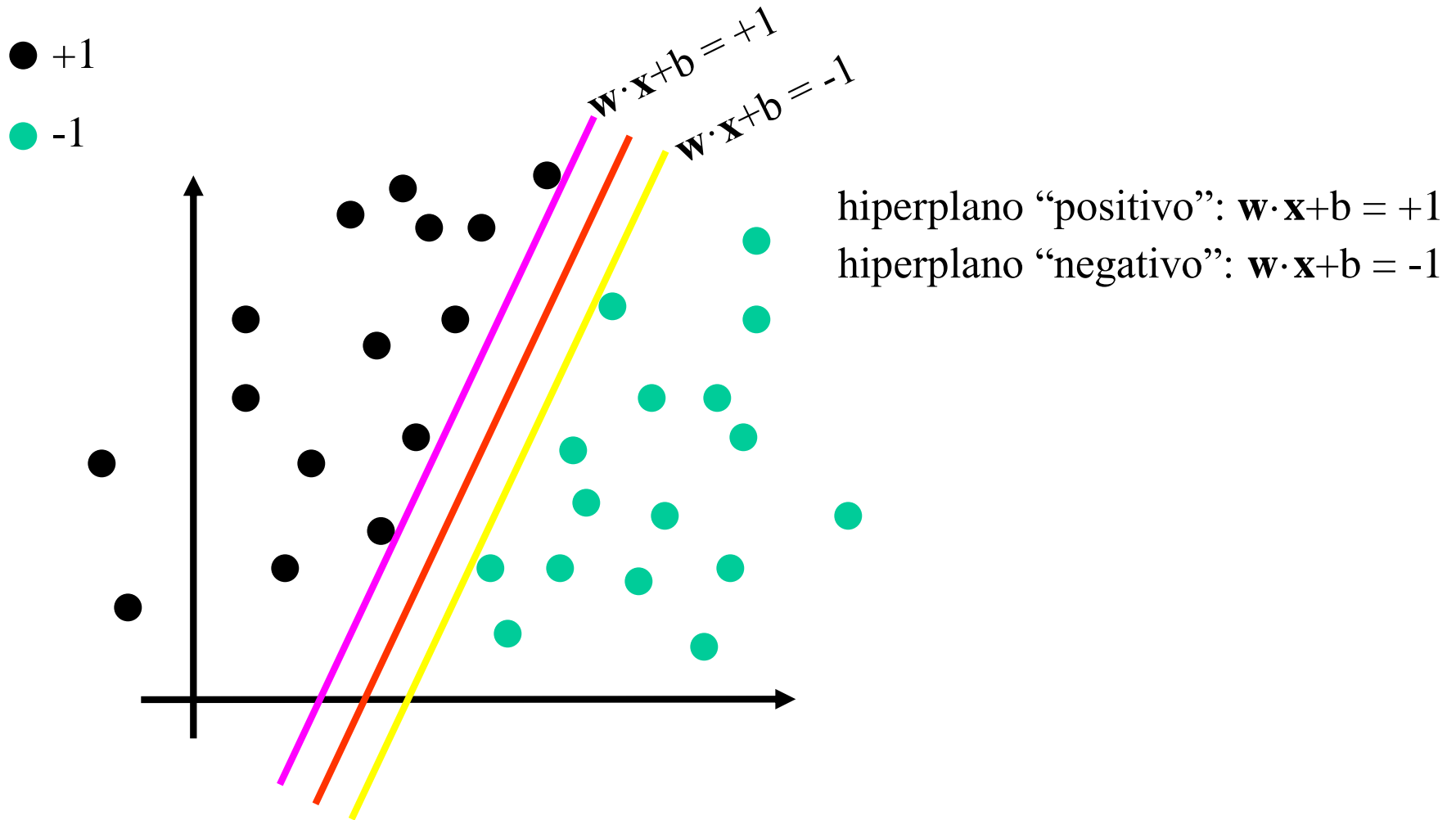
Lo que se quiere es separar los puntos de acuerdo al valor de su etiqueta y_i en dos hiperplanos diferentes:

$\mathbf{w} \cdot \mathbf{x}_i + b = +1$ para $y_i = +1$. (hiperplano “positivo”)

$\mathbf{w} \cdot \mathbf{x} + b = -1$ para $y_i = -1$ (hiperplano “negativo”)

Simplificando: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq +1$

Idea inicial de separación



Conceptos matemáticos

Sea d_+ (d_-) la distancia más corta entre el hiperplano positivo (negativo) y el punto positivo (negativo) más cercano.

Sea el “margen” la distancia entre los hiperplanos “positivo” y “negativo”. El margen es igual a: $\frac{2}{\|\mathbf{w}\|}$

La idea es encontrar un hiperplano con el máximo “margen”. Esto es un problema de optimización:

$$\text{maximizar: } \frac{2}{\|\mathbf{w}\|} \quad \text{sujeto a : } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = +1$$

Conceptos matemáticos

El problema su puede expresar así:

$$\text{minimizar: } \|\mathbf{w}\|^2 \quad \text{sujeto a : } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = +1$$

Pero el problema se puede transformar para que quede más fácil de manejar! Se usan multiplicadores de Lagrange (α_i).

$$L_P \equiv \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) + \sum_{i=1}^l \alpha_i$$

Conceptos matemáticos

Haciendo que los gradientes de L_p respecto a \mathbf{w} y b sean cero, se obtienen las siguientes condiciones:

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \qquad \sum_{i=1}^l \alpha_i y_i = 0$$

Reemplazando en L_p se obtiene el problema dual:

$$L_D = \sum_{i=1}^l \alpha_i + \sum_{i=1, j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

Conceptos matemáticos

La forma para optimizar es:

maximizar:
$$L_D = \sum_{i=1}^l \alpha_i + \sum_{i=1, j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

sujeto a :

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \qquad \sum_{i=1}^l \alpha_i y_i = 0$$

Conceptos matemáticos

Cuando los datos no se pueden separar linealmente se hace un cambio de espacio mediante una función que transforme los datos de manera que se puedan separar linealmente. Tal función se llama ***Kernel***.

También hay métodos para separar los datos (x_i, y_i) directamente aún no siendo separables linealmente, mediante ***funciones polinómicas*** y otro tipo de funciones, las ***Funciones de Base Radial*** (RBF).

Ejemplos de funciones Kernel

- Polinomial con grado d

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^d$$

- Base Radial con ancho σ

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / (2\sigma^2))$$

- Closely related to radial basis function neural networks

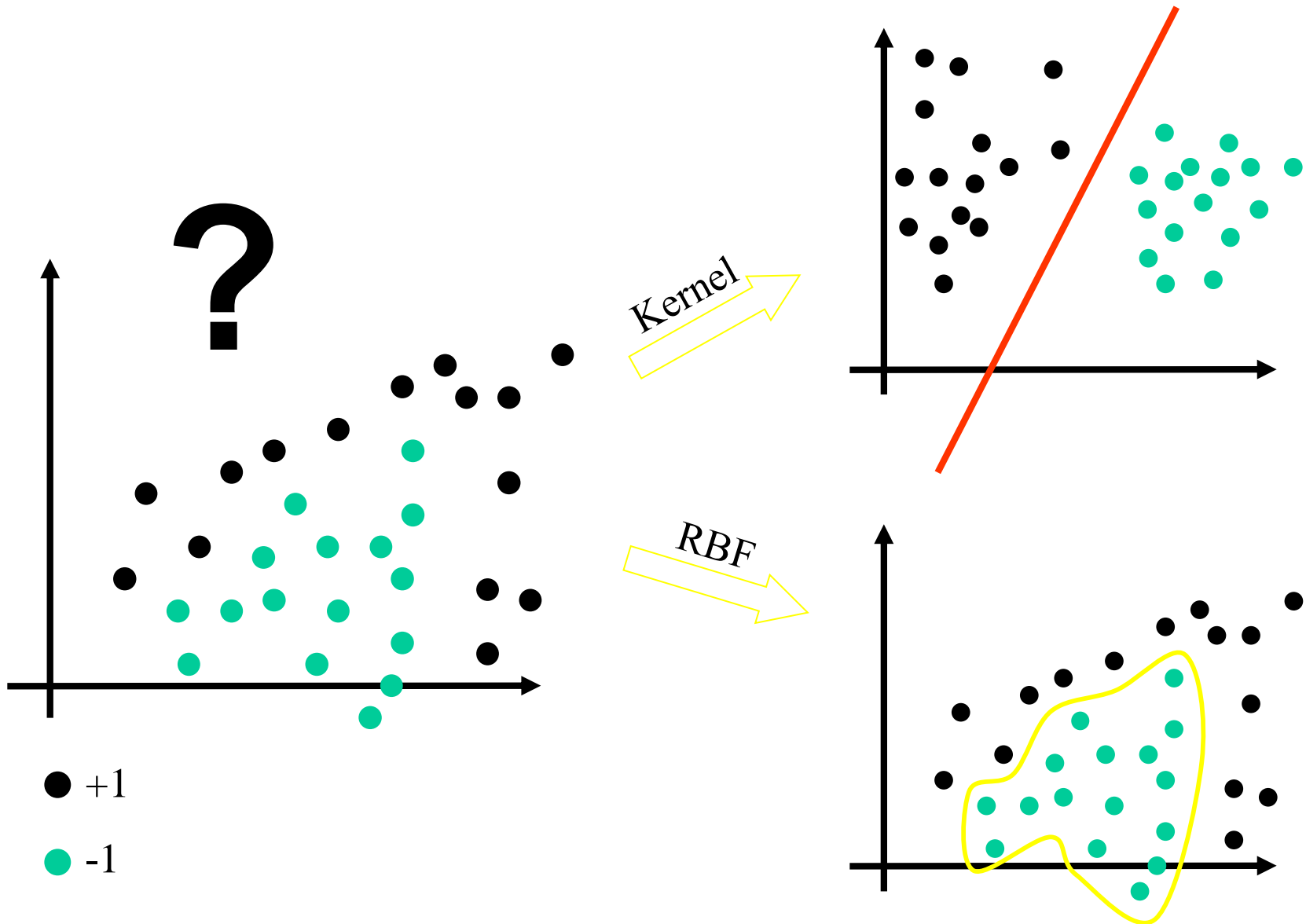
- Sigmoidea con parámetro κ y θ

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\kappa \mathbf{x}^T \mathbf{y} + \theta)$$

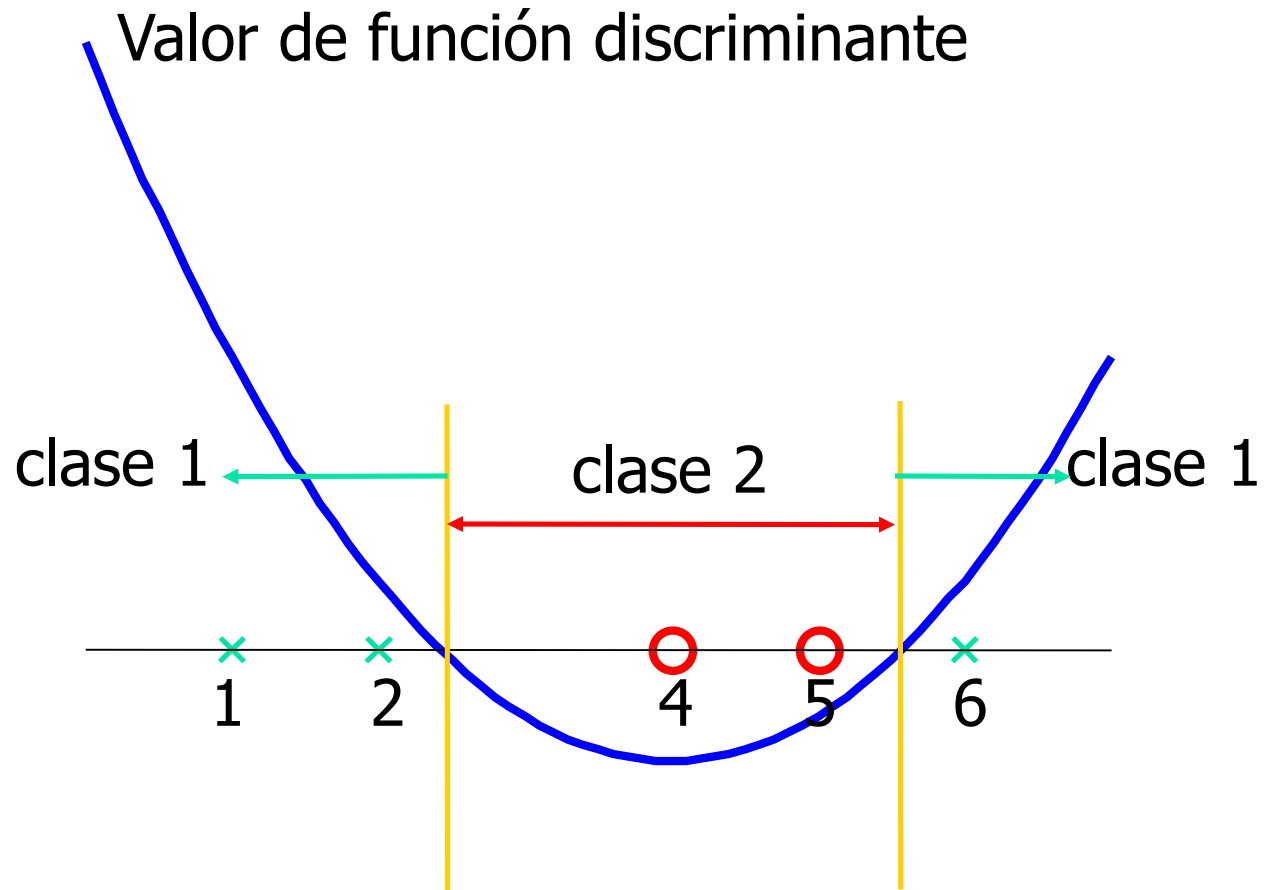
- It does not satisfy the Mercer condition on all κ and θ

- La investigación sobre diferentes funciones kernel para diferentes aplicaciones está muy activa

Conceptos matemáticos

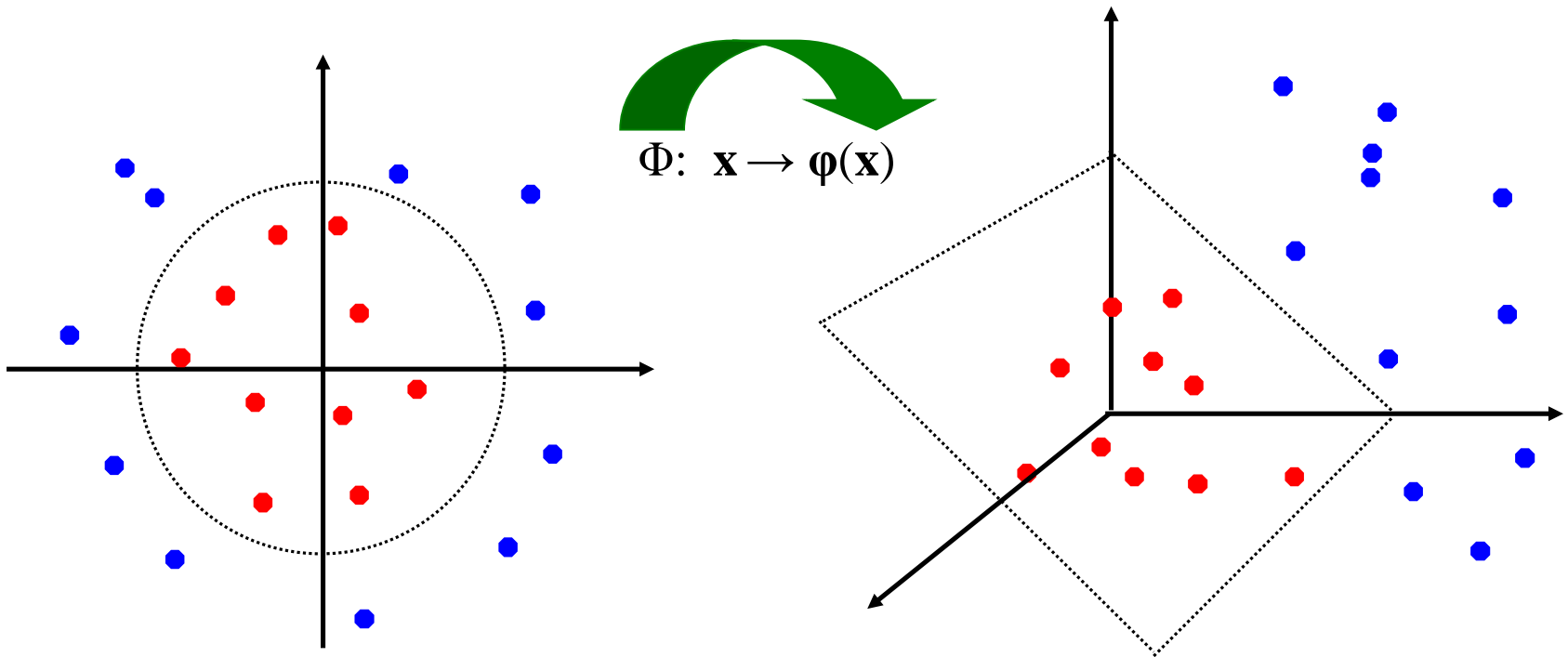


Ejemplo



SVMs No lineales: Espacios Característicos

Idea general: el espacio de características original siempre puede asignarse a un espacio de características de mayor dimensión donde el conjunto de entrenamiento es separable:



Conceptos matemáticos

Algunos problemas con las SVM:

Overtraining: se han aprendido muy bien los datos de entrenamiento pero no se pueden clasificar bien ejemplos no vistos antes. Ej.: un botánico que conoce mucho.

La porción n de los datos no conocidos que será mal calificada, está limitada por:

$$n = \frac{\text{No. vectores de soporte}}{\text{No. de ejemplos de entrenamiento}}$$

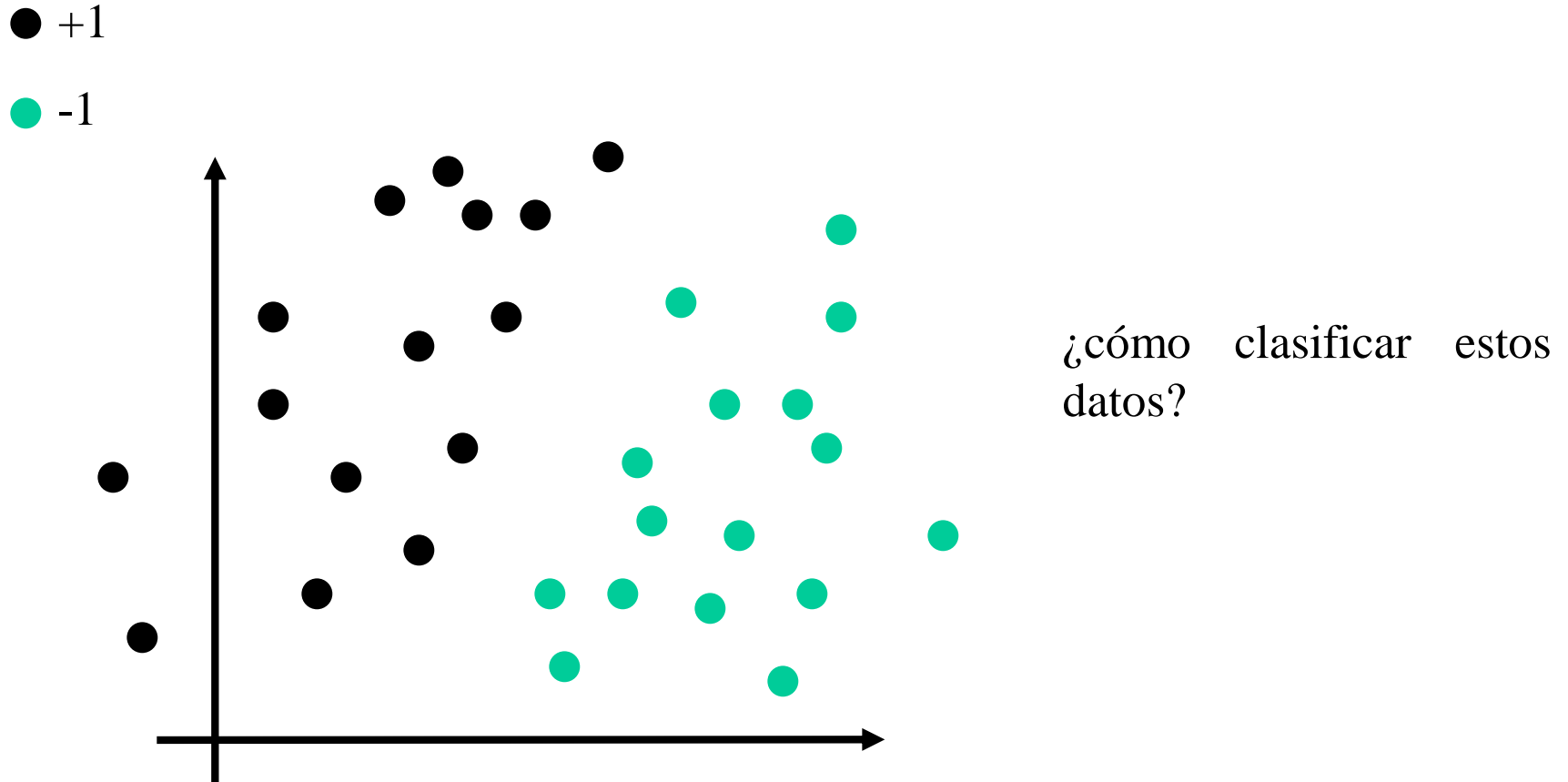
Se aplica el principio de Ockham.

Conceptos matemáticos

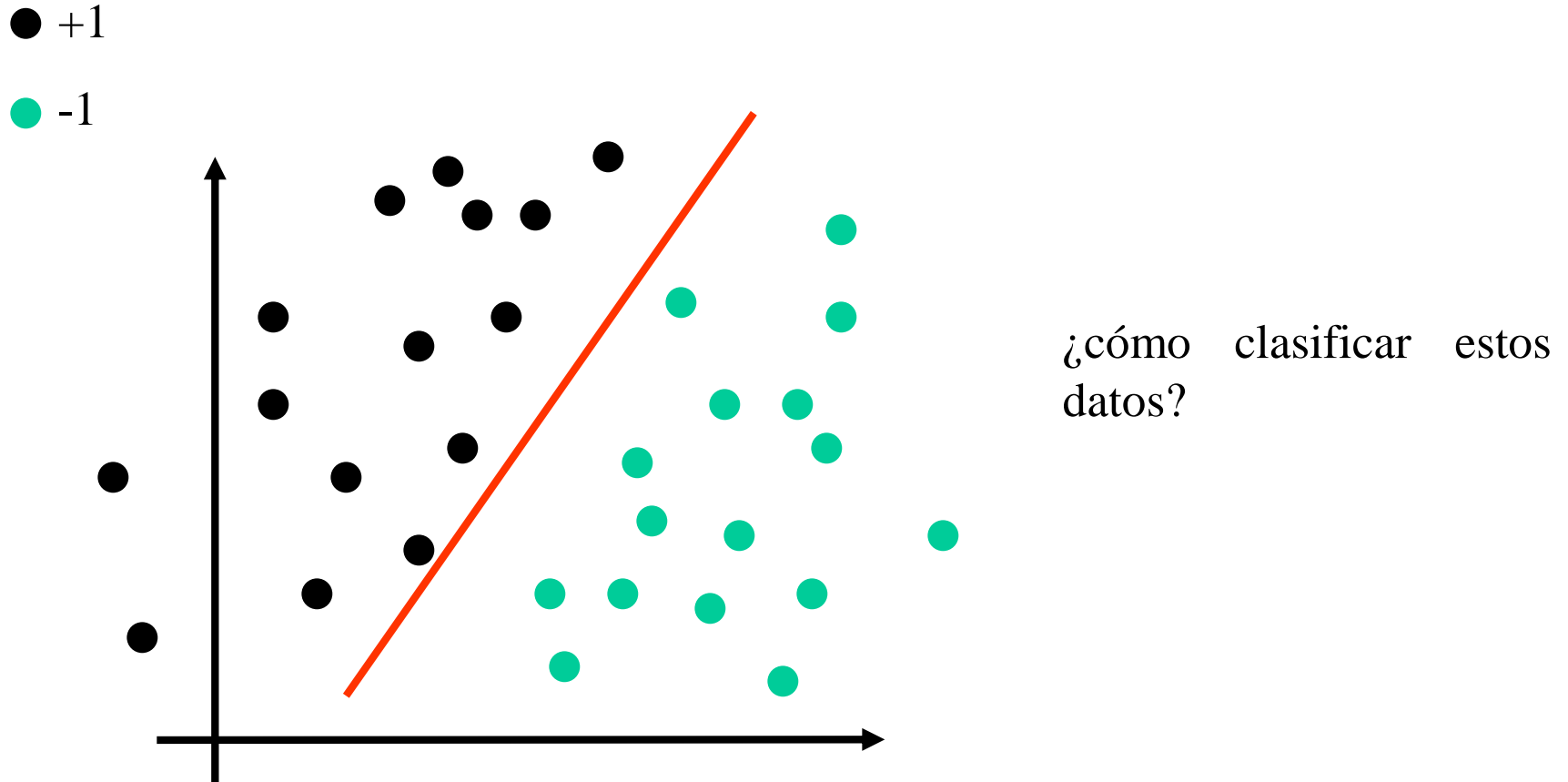
Algunos problemas con las SVM:

Overfitting: no se ha aprendido muy bien la característica de los datos de entrenamiento, por lo que se hace una mala clasificación. Ej.: el hermano del botánico.

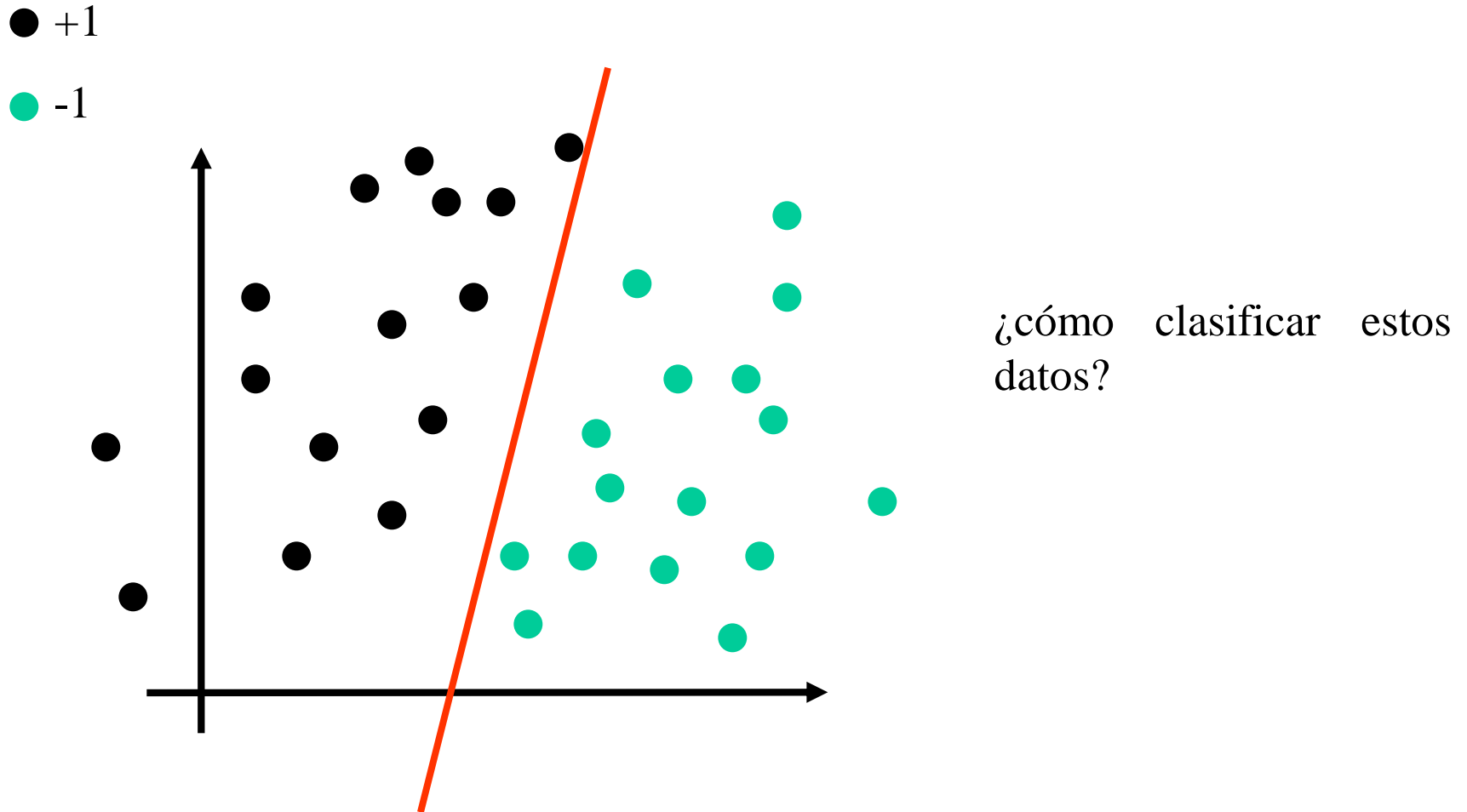
Interpretación geométrica



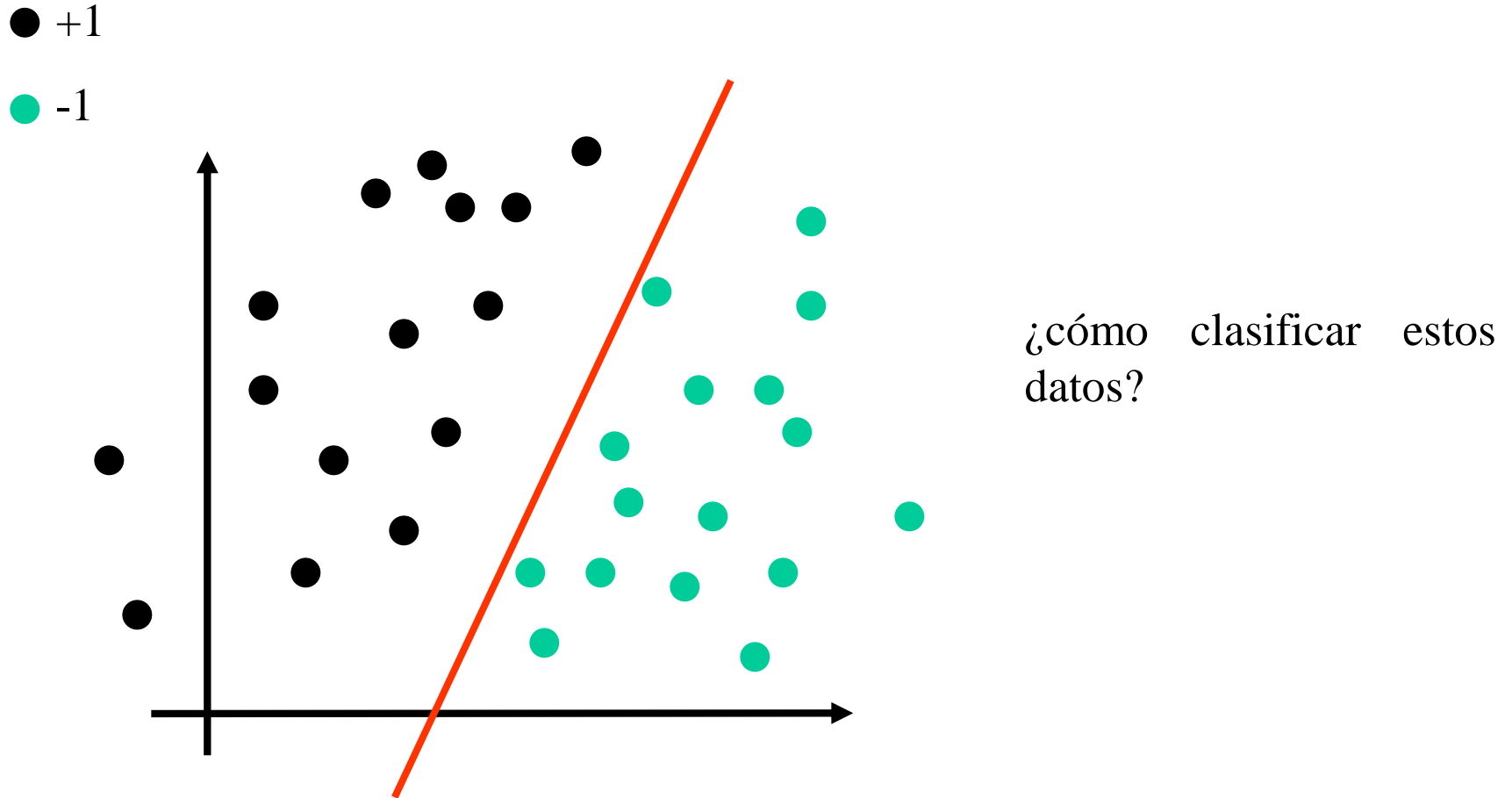
Interpretación geométrica



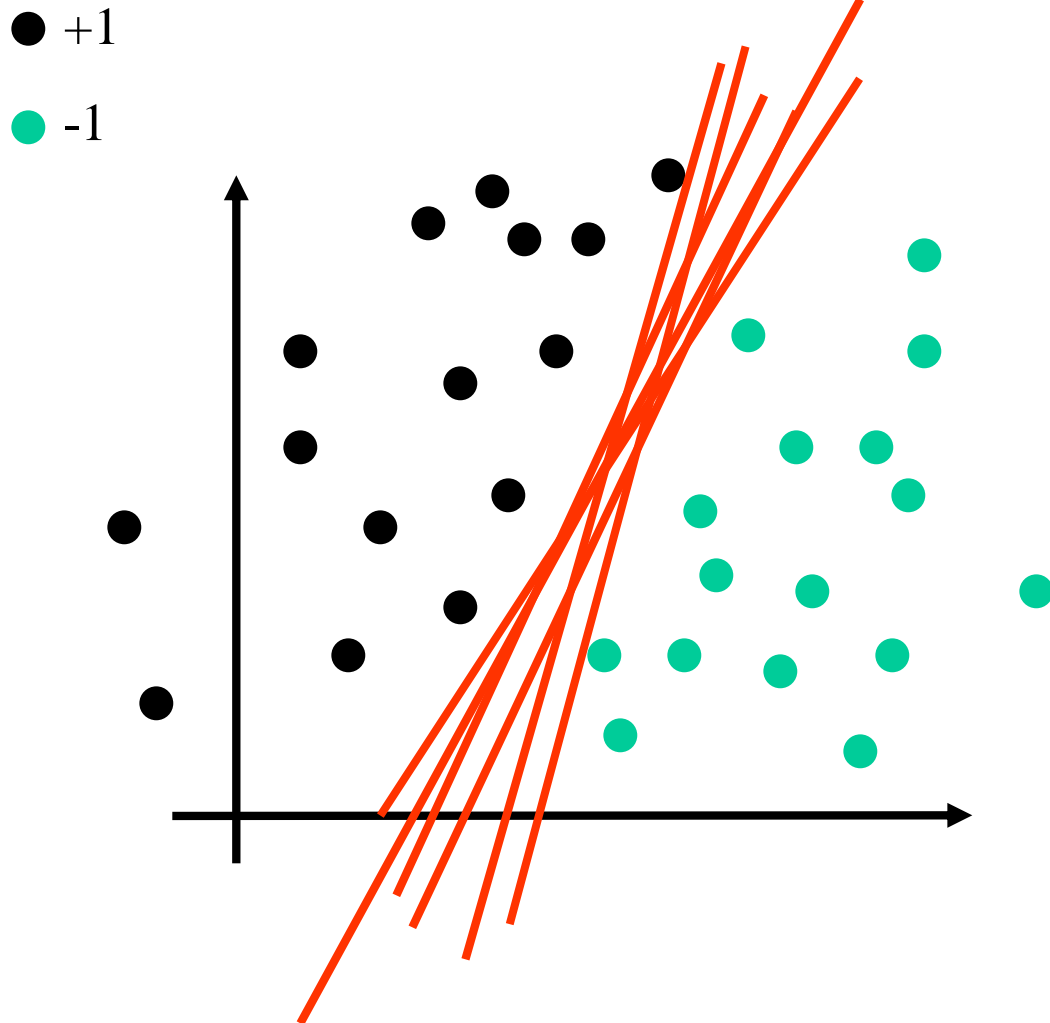
Interpretación geométrica



Interpretación geométrica

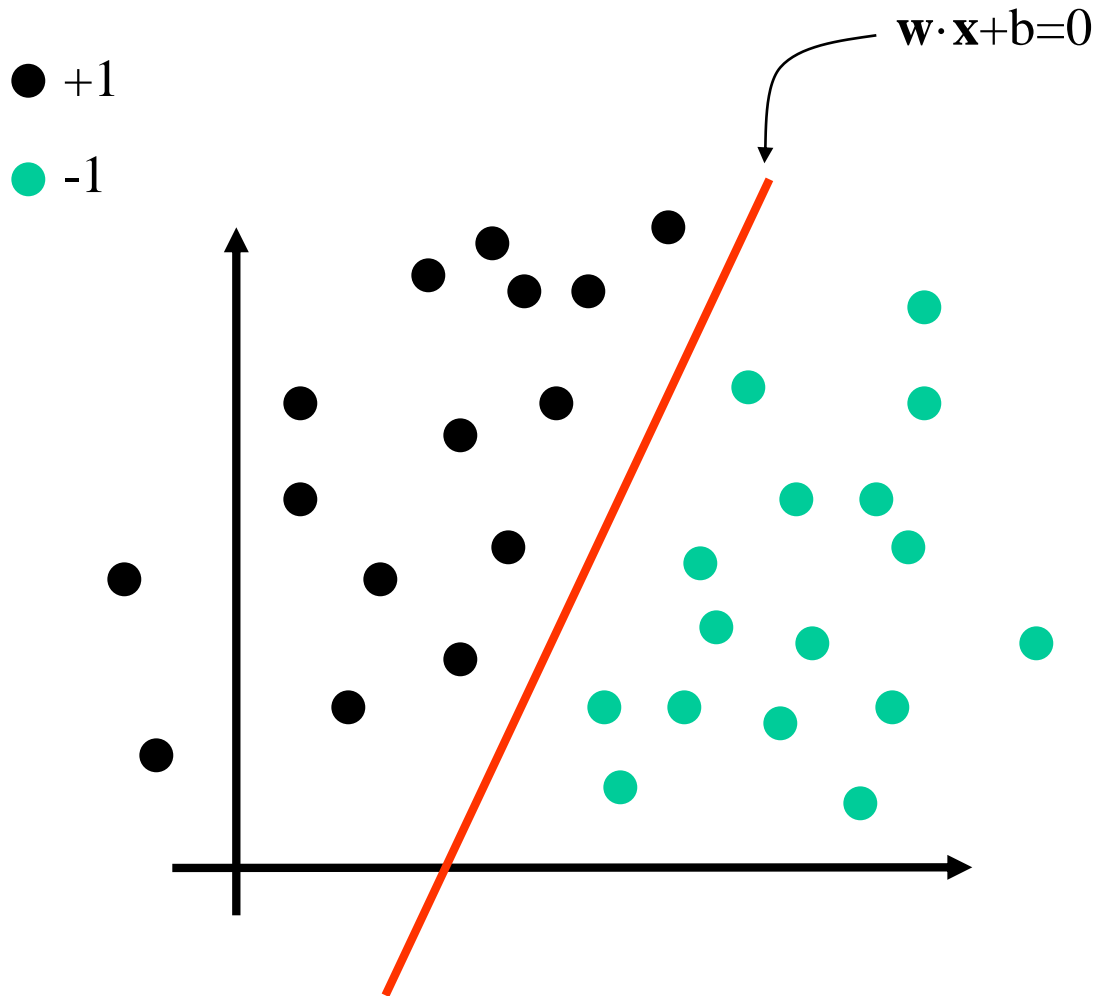


Interpretación geométrica



Cualquiera puede ser buena, ¿pero cuál es la mejor?

Interpretación geométrica

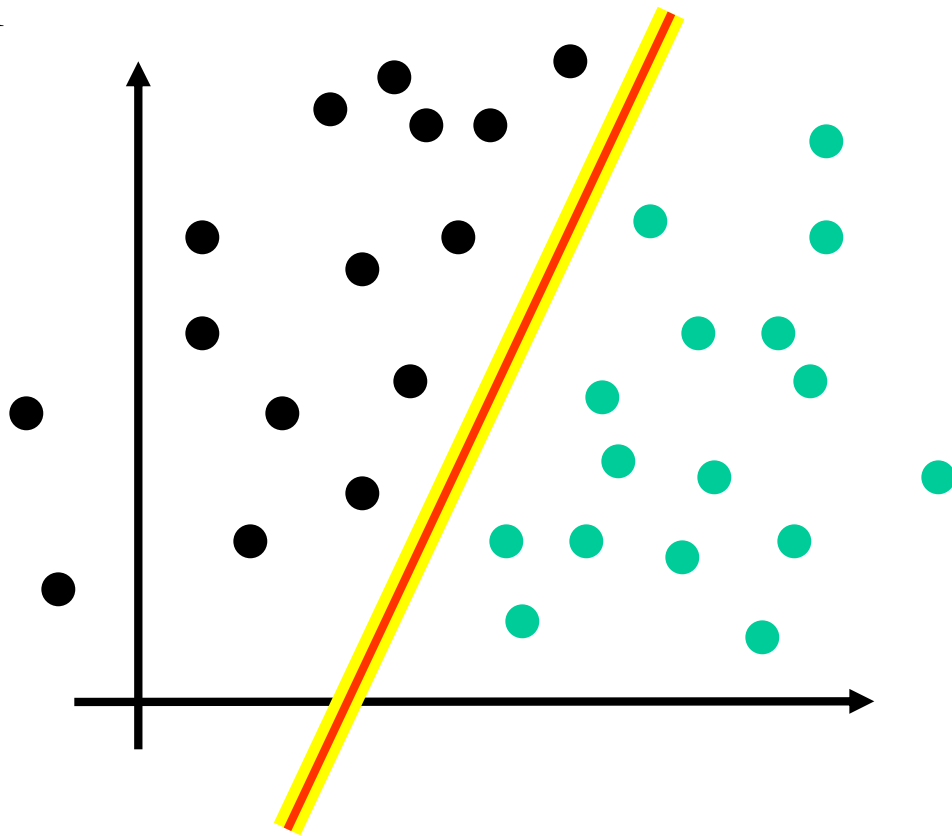


Definimos el hiperplano

Interpretación geométrica

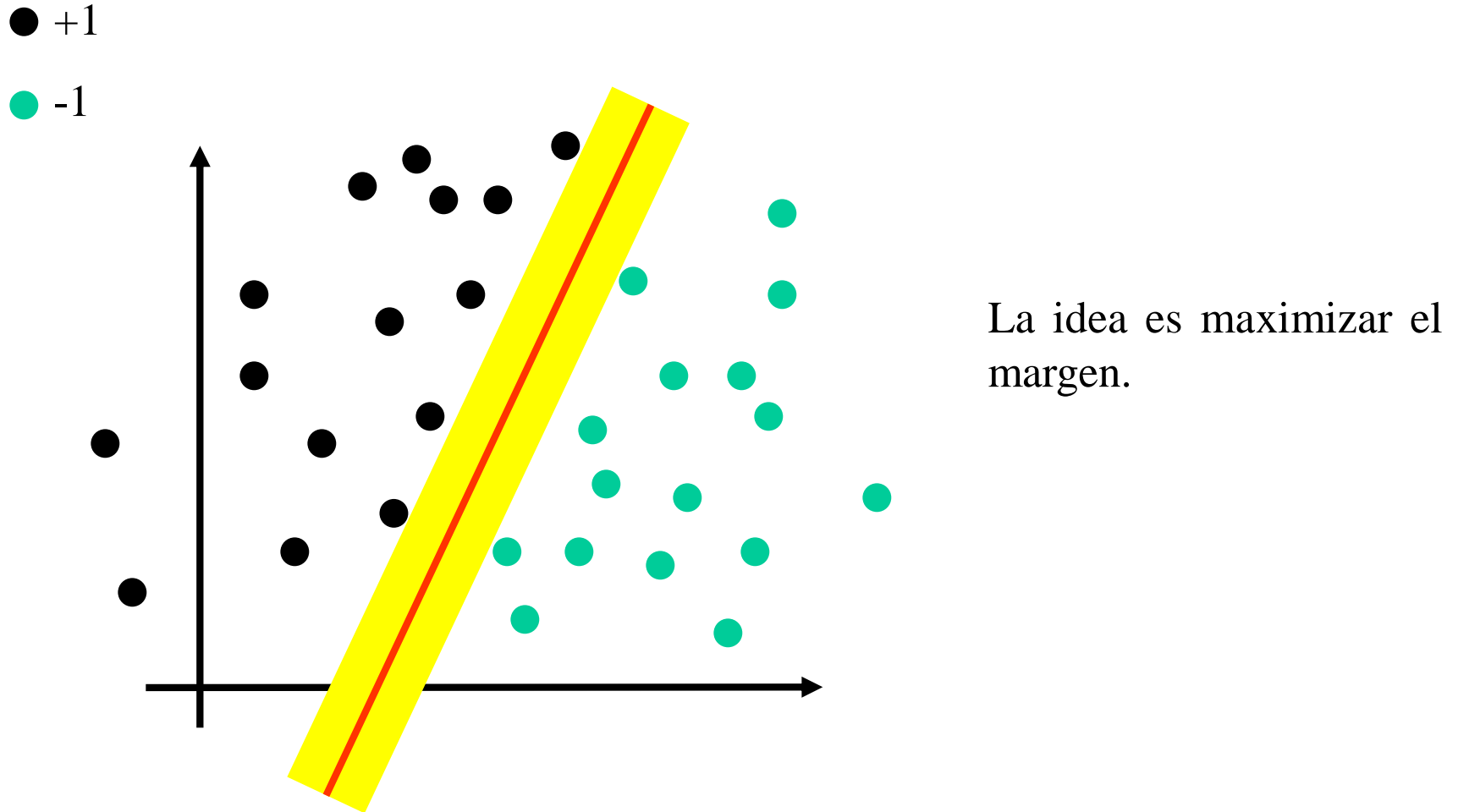
● +1

● -1

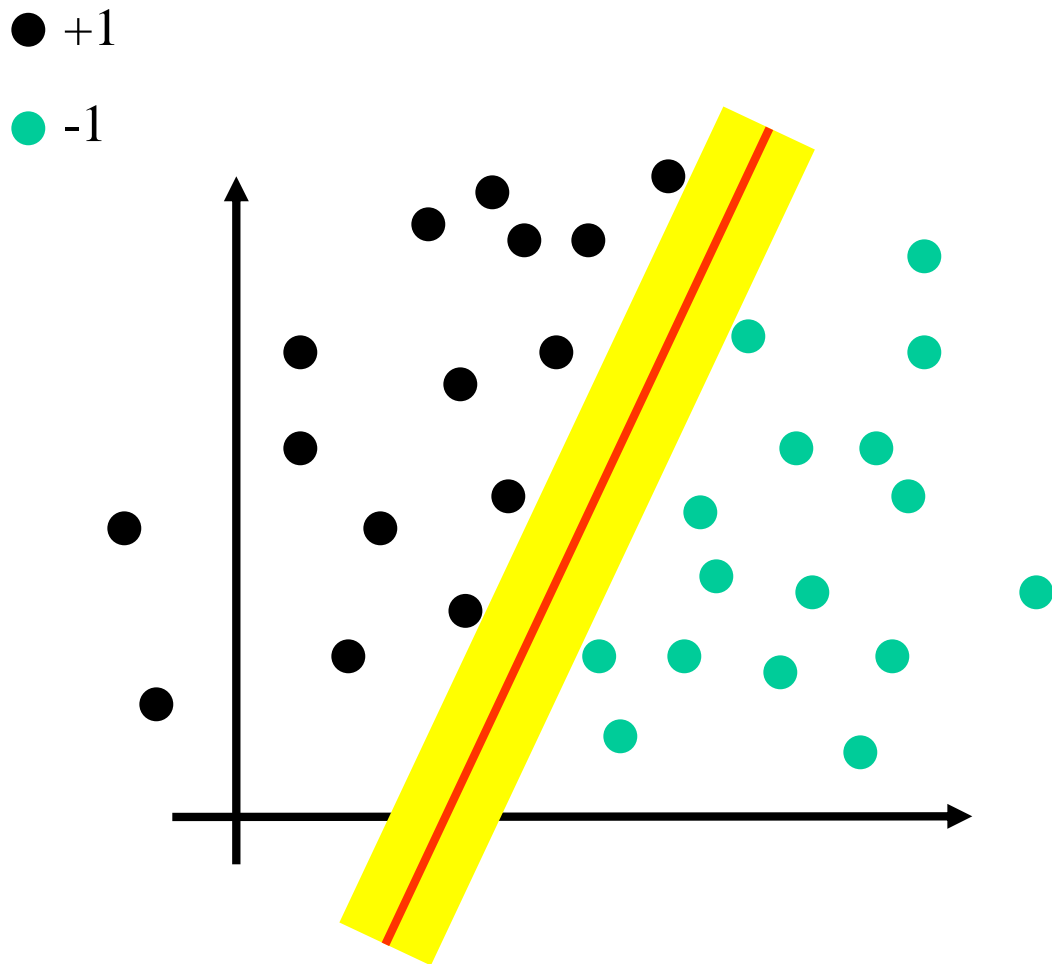


Definimos el margen

Interpretación geométrica



Interpretación geométrica



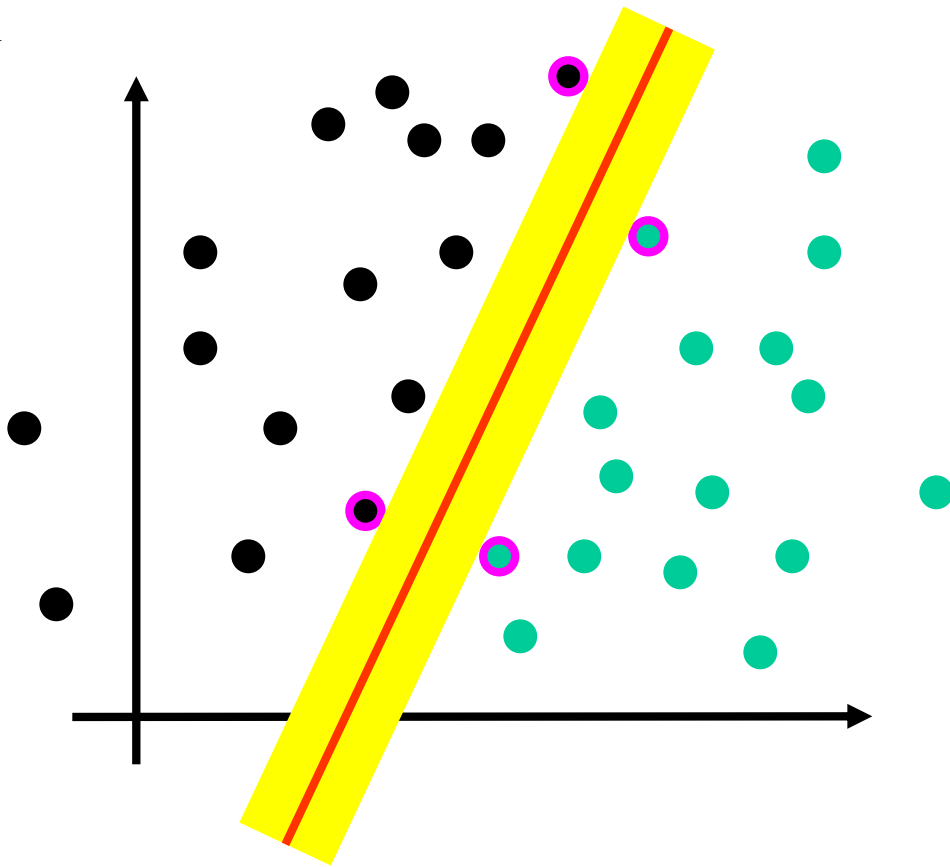
El hiperplano que tenga el mayor margen es el mejor clasificador de los datos.

Esta es la clase más simple de SVM, la LSVM.

Interpretación geométrica

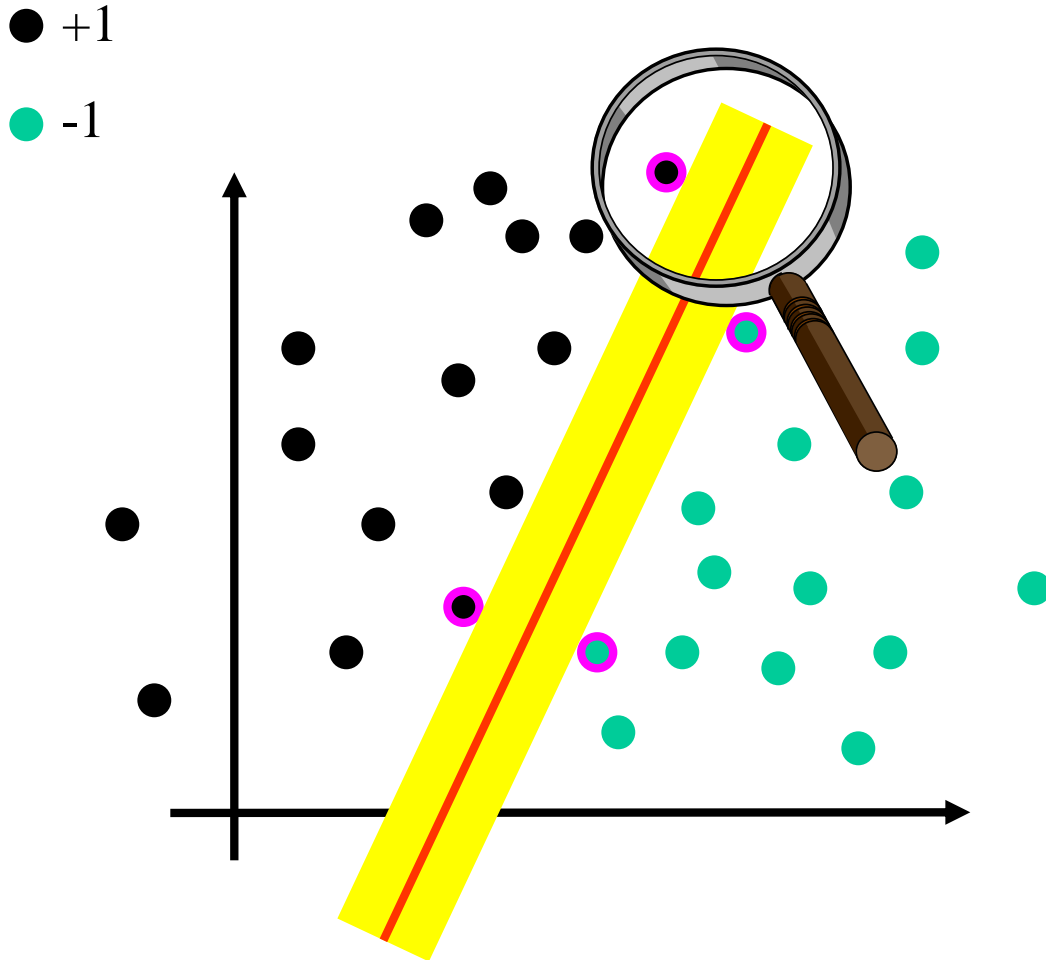
● +1

● -1



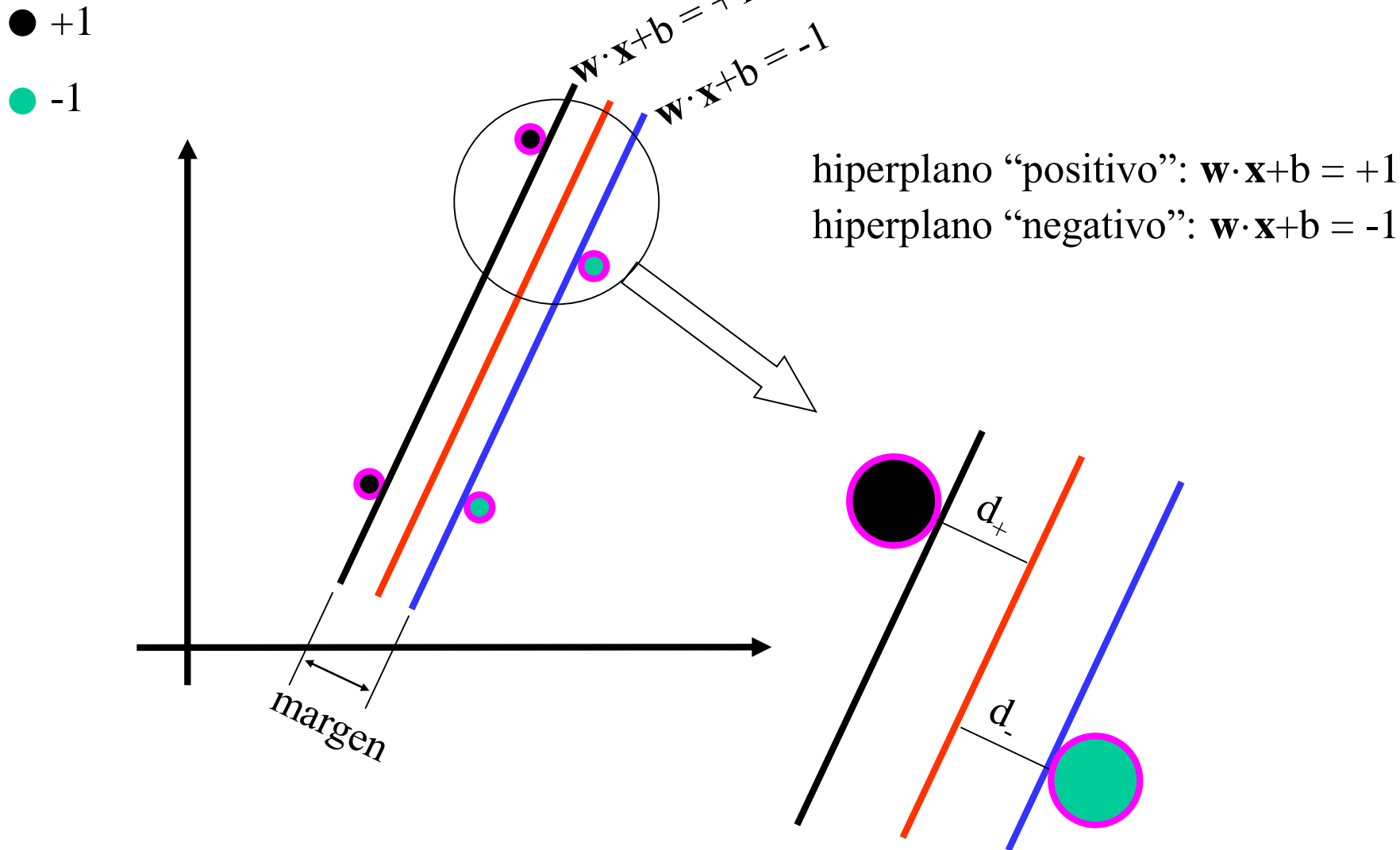
Los vectores de soporte son los puntos que tocan el límite del margen.

Interpretación geométrica

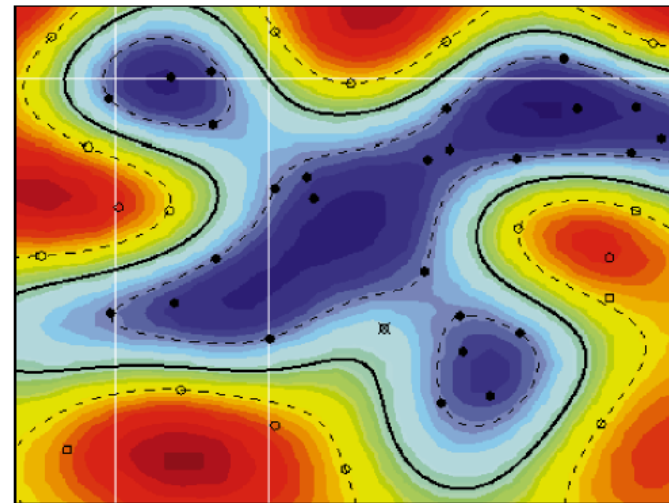
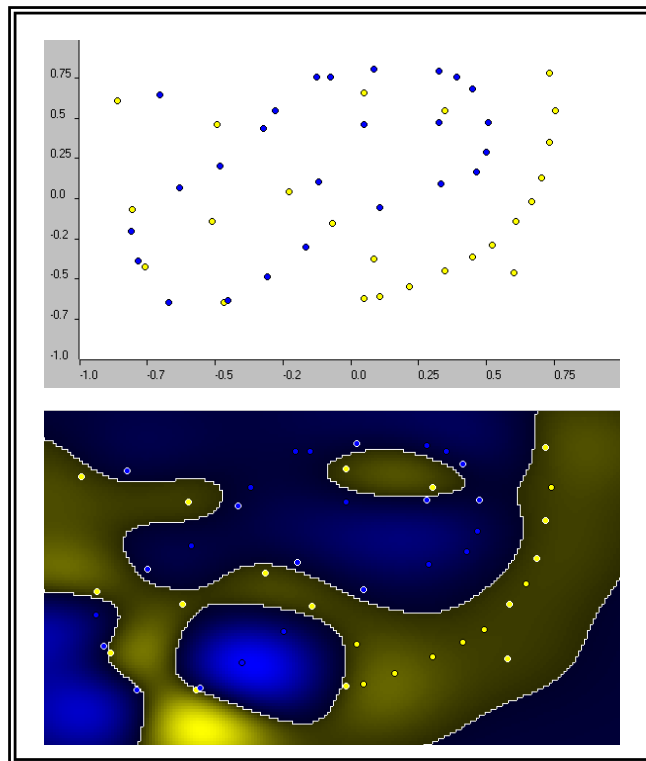
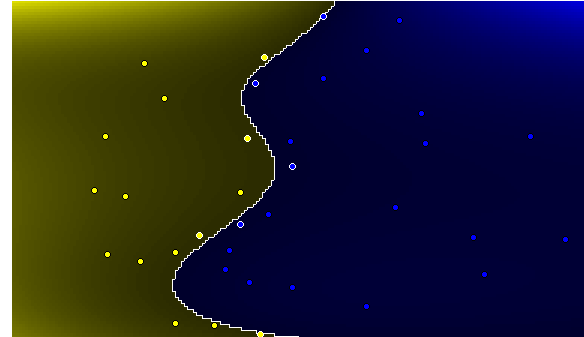
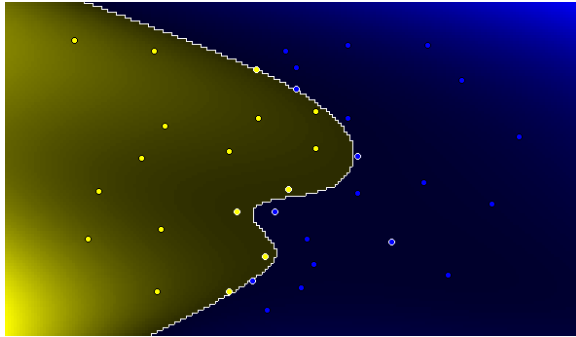


Veamos los hiperplanos
“positivo” y “negativo”

Interpretación geométrica



Separación polinómica y con RBF



Aplicaciones

