Máquinas con Vectores de Soporte - SVM

Misael López Ramírez

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:

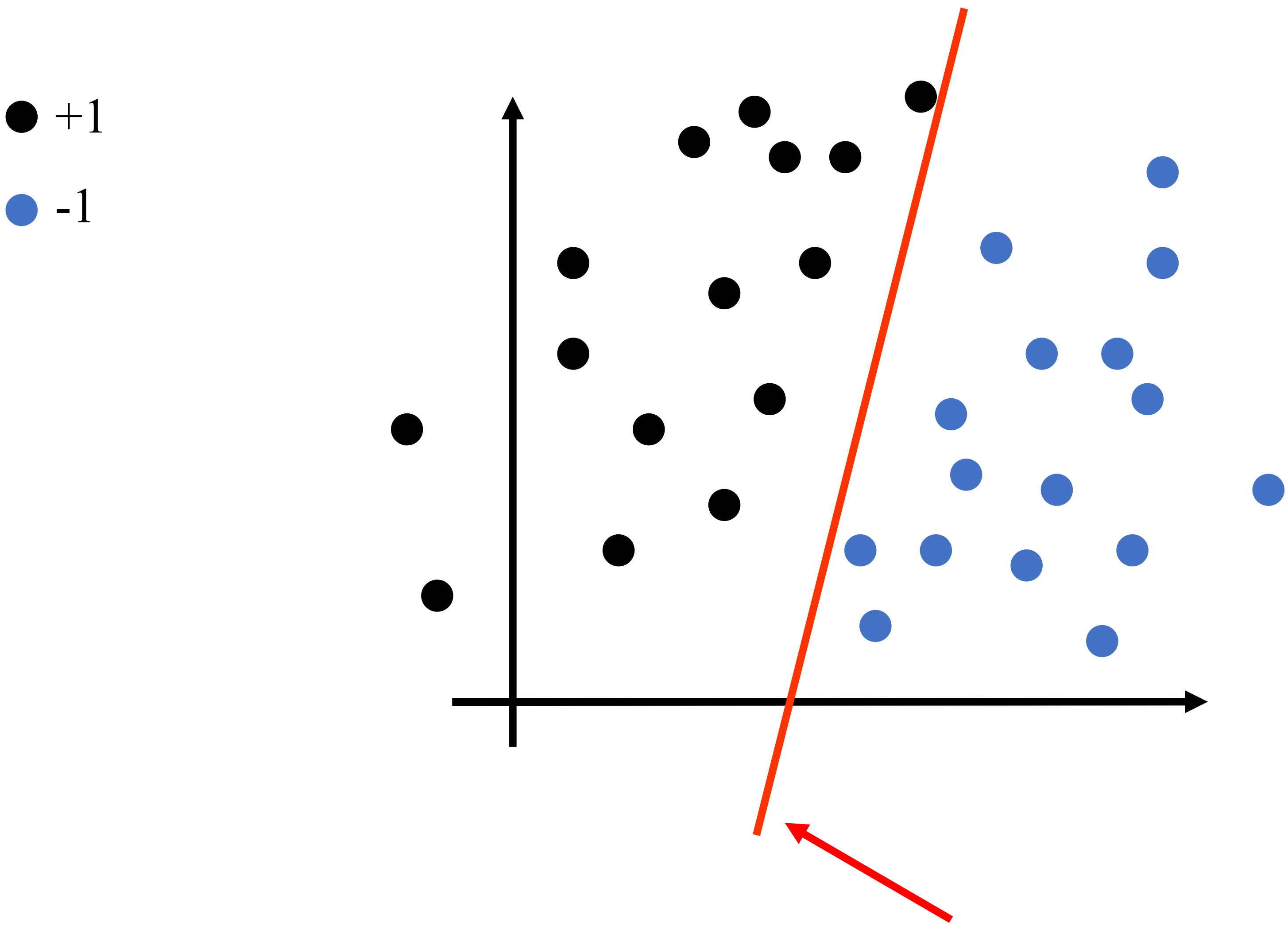
un vector $x_i \in R^n, i = 1,\ldots,l$

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 $\mathbf{w}\cdot\mathbf{x}+b=0$.

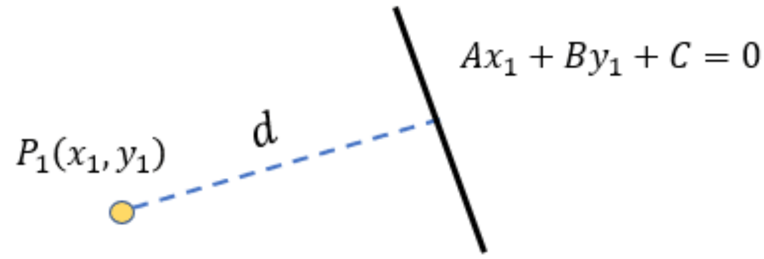
Ecuación de la línea recta

Idea inicial de separación



$\mathbf{w}\cdot\mathbf{x}+b=0$

Conceptos matemáticos



$$d = \frac{|Ax_1 + By_1 + C|}{\sqrt{A^2 + B^2}}$$

w es es normal al hiperplano.

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

$\|w\|$ es la norma euclídea de **w**

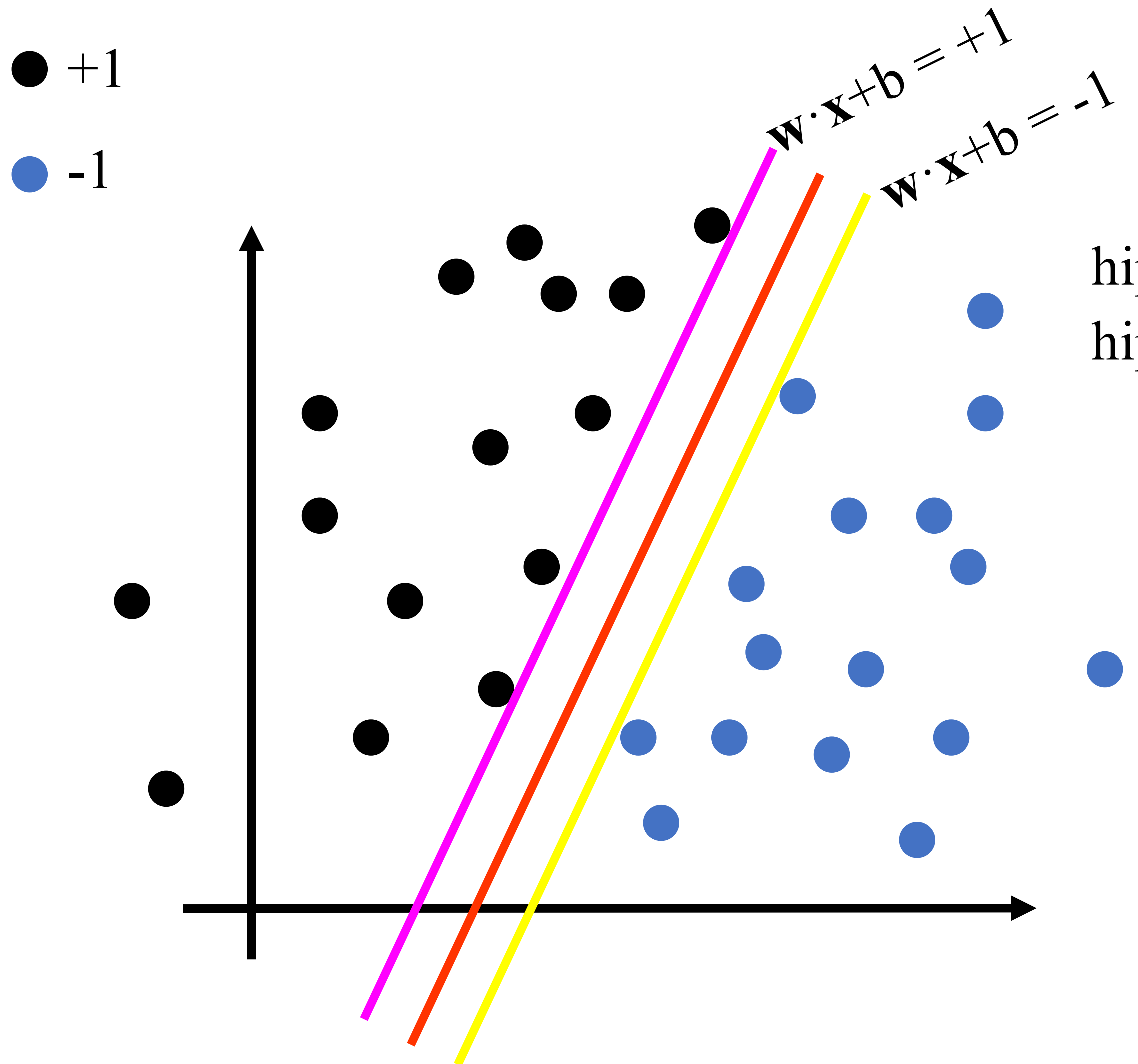
Lo que se quiere es separar los puntos de acuerdo al valor de su etiqueta y_i en dos hiperplanos diferentes:

$\mathbf{w}\cdot\mathbf{x_i}+b\geq+1$ para y_i=+1. (hiperplano "positivo")

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

Simplificando: $y_i(\mathbf{w}\cdot\mathbf{x_i}+b)\geq 0$

Idea inicial de separación



hiperplano "positivo": $\mathbf{w}\cdot\mathbf{x}+b = +1$

hiperplano "negativo": $\mathbf{w}\cdot\mathbf{x}+b = -1$

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: $\frac{2}{\|W\|}$

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

maximizar: $\frac{2}{\|W\|}$ sujeto a : $y_i(W \cdot X_i + b) \geq +1$

Ó minimizar

Conceptos matemáticos

El problema su puede expresar así:

$$\text{minimizar: } \frac{1}{2}\|W\|^2 \quad \text{sujeto a : } y_i(W \cdot X_i + b) \geq 1$$

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

 # de observaciones ó muestras

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

Multiplicaodres de lagrange: https://www.youtube.com/watch?v=P1YDelGxRTc

Conceptos matemáticos

Multiplicadores de Lagrange

minimizar: $\frac{1}{2}\|W\|^2$ sujeto a : $y_i(W \cdot X_i + b) \geq 1$

Teorema de Lagrange: $\boxed{\nabla f(x,y,z) = \lambda \nabla g(x,y,z)}$

Metodo de los multiplicadores de Lagrange:

1) $\frac{df}{dx}(x,y,z) = \lambda \frac{dg}{dx}(x,y,z)$ **3)** $\frac{df}{dz}(x,y,z) = \lambda \frac{dg}{dz}(x,y,z)$

2) $\frac{df}{dy}(x,y,z) = \lambda \frac{dg}{dy}(x,y,z)$ **4)** $g(x,y,z) = c$

$$L_P = f(x,y,z) - \lambda g(x,y,z)$$

Función de lagrange

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

Conceptos matemáticos

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

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

Reemplazando en L_p se obtiene el problema dual:

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

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

Hay penalización por error de clasificación

Conceptos matemáticos

La forma para optimizar es:

$$\min\ f(x) = \mathbf{c}^T \mathbf{x} + \tfrac{1}{2}\mathbf{x}^T \mathbf{Q}\,\mathbf{x}$$
$$s.a.\ \mathbf{A}\mathbf{x} \le \mathbf{b}$$
$$\mathbf{x} \ge 0$$

maximizar:

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

*L_D mayores a cero son nombrados vectores de Soporte (NV).

sujeto a :

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

El problema se reduce a encontrar el máximo de L_D con las restricciones anteriores

*PROGRAMACIÓN CUADRATICA

*Se puede resolver con el método del punto interior

Conceptos matemáticos

La forma para optimizar es:

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

*L_D mayores a cero son nombrados vectores de Soporte (NV).

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

Descripción

Solver para funciones objetivas cuadráticas con restricciones lineales.

encuentra un mínimo para un problema especificado porquadprog

$$\min_x \frac{1}{2} x^T H x + f^T x \text{ such that } \begin{cases} A \cdot x \leq b, \\ Aeq \cdot x = beq, \\ lb \leq x \leq ub. \end{cases}$$

, y son matrices, y,,,, y son vectores.$HAAeqfbbeqlbubx$

Conceptos matemáticos

Al encontrar el vector L_D máximo calculamos los valores de pendiente W y b con las formulas optimos:

$$W = \sum_{i=1}^{L} \alpha_i y_i x_i \qquad b = \frac{1}{NV}\sum_{i=1}^{NV}(y_i - W * x_i)$$

*NV numero de vectores de soporte.

Sustituimos W y b en la ecuación de la línea recta de la forma

$$W \cdot x + b = -1 \qquad W \cdot x + b = +1$$

*Por lo tanto clasifica entre dos clases únicamente con el signo de tal forma nos queda como:

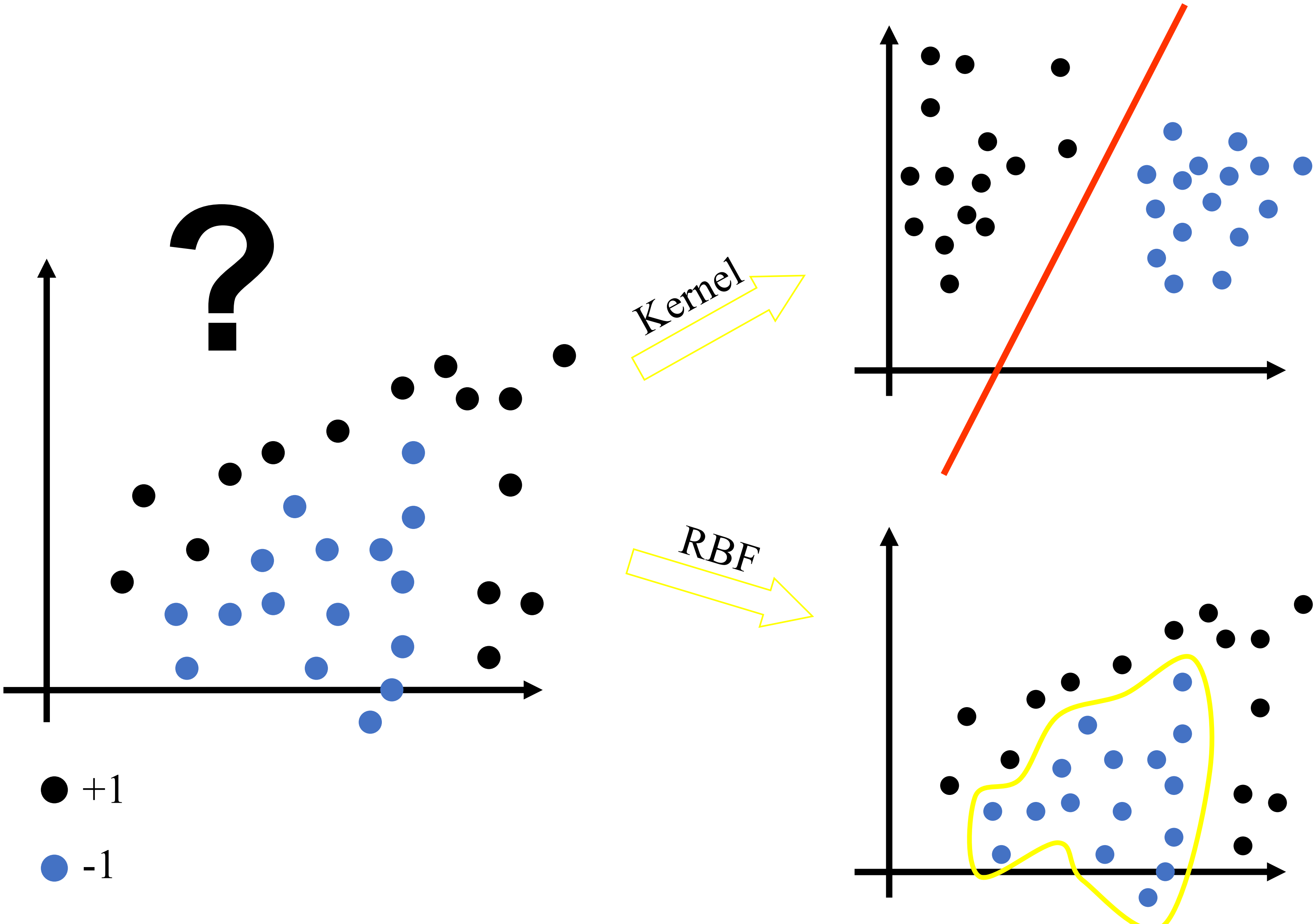
$$f(x) = sign\,(W \cdot x + b)$$

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.***

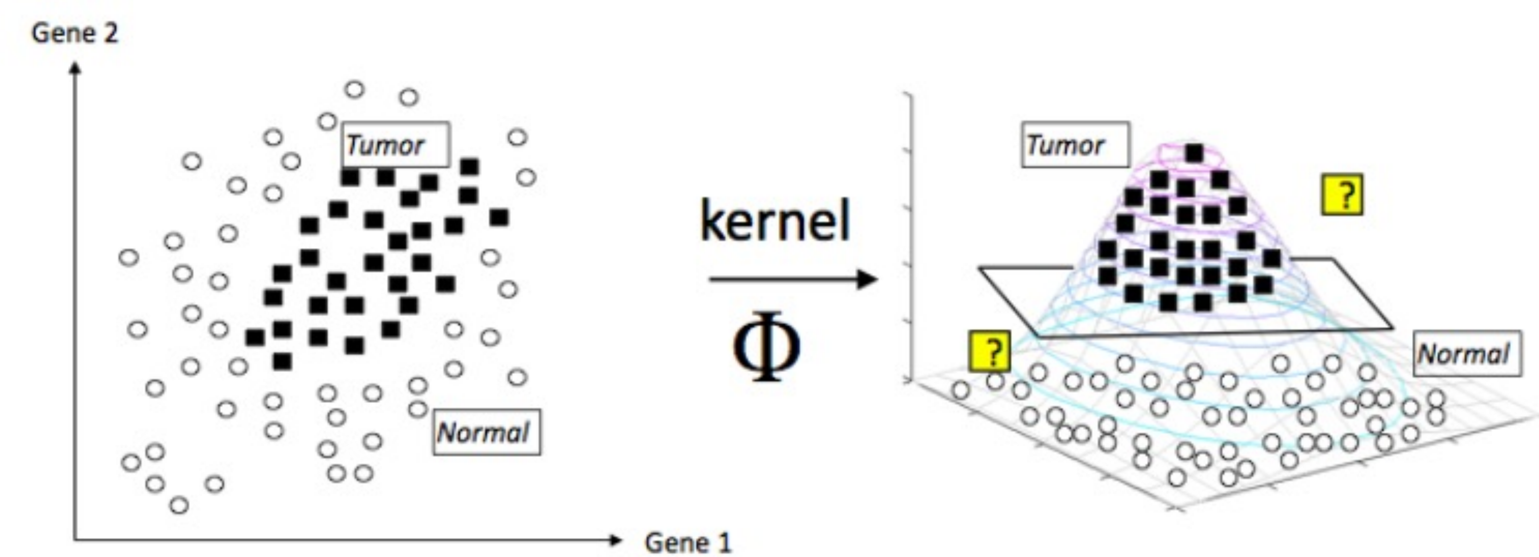
También hay métodos para separar los datos (xi,yi) directamente aún no siendo separables linealmente, mediante ***funciones polinómicas*** y otro tipo de funciones, las ***Funciones de Base Radial*** (RBF).

Conceptos matemáticos



Conceptos matemáticos

La forma del kernel nos puede ayudar a obtener una mejor discriminación entre clases como lo muestra ala siguiente figura:



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:

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

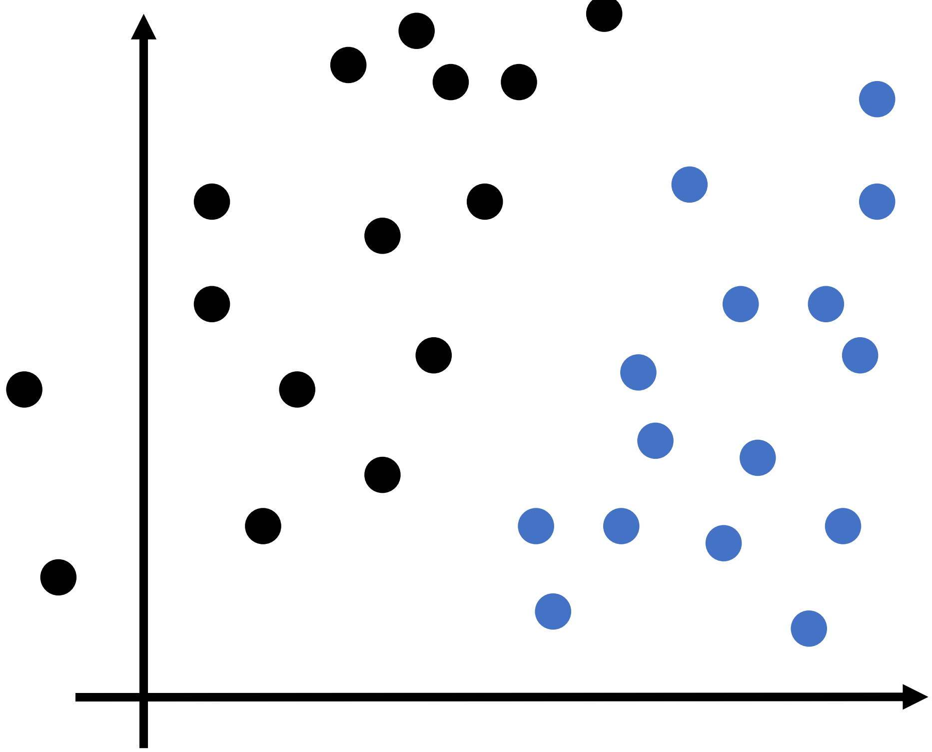
Se aplica el principio de Ockham.

Interpretación geométrica

● +1

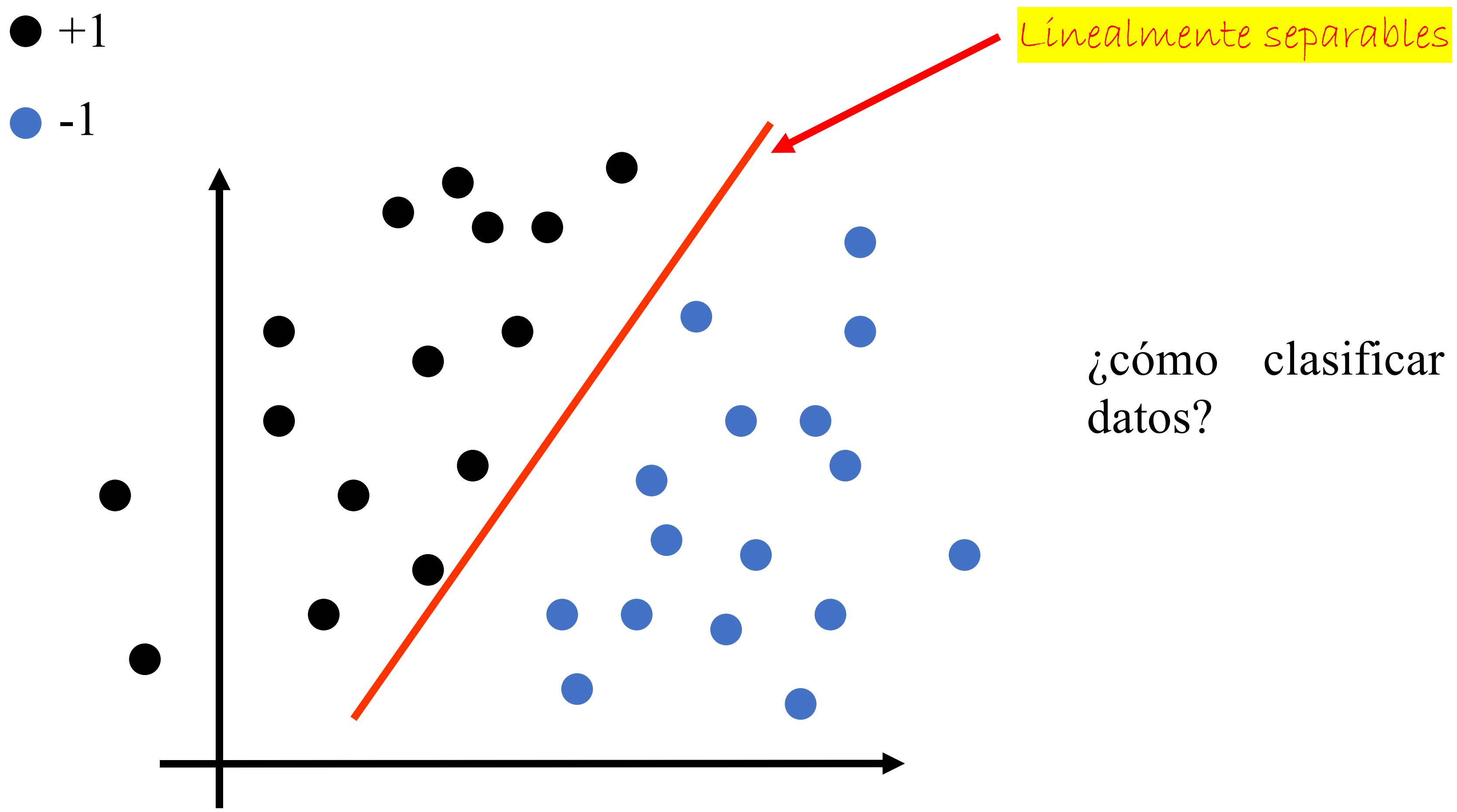
● -1

2 clases a clasificar



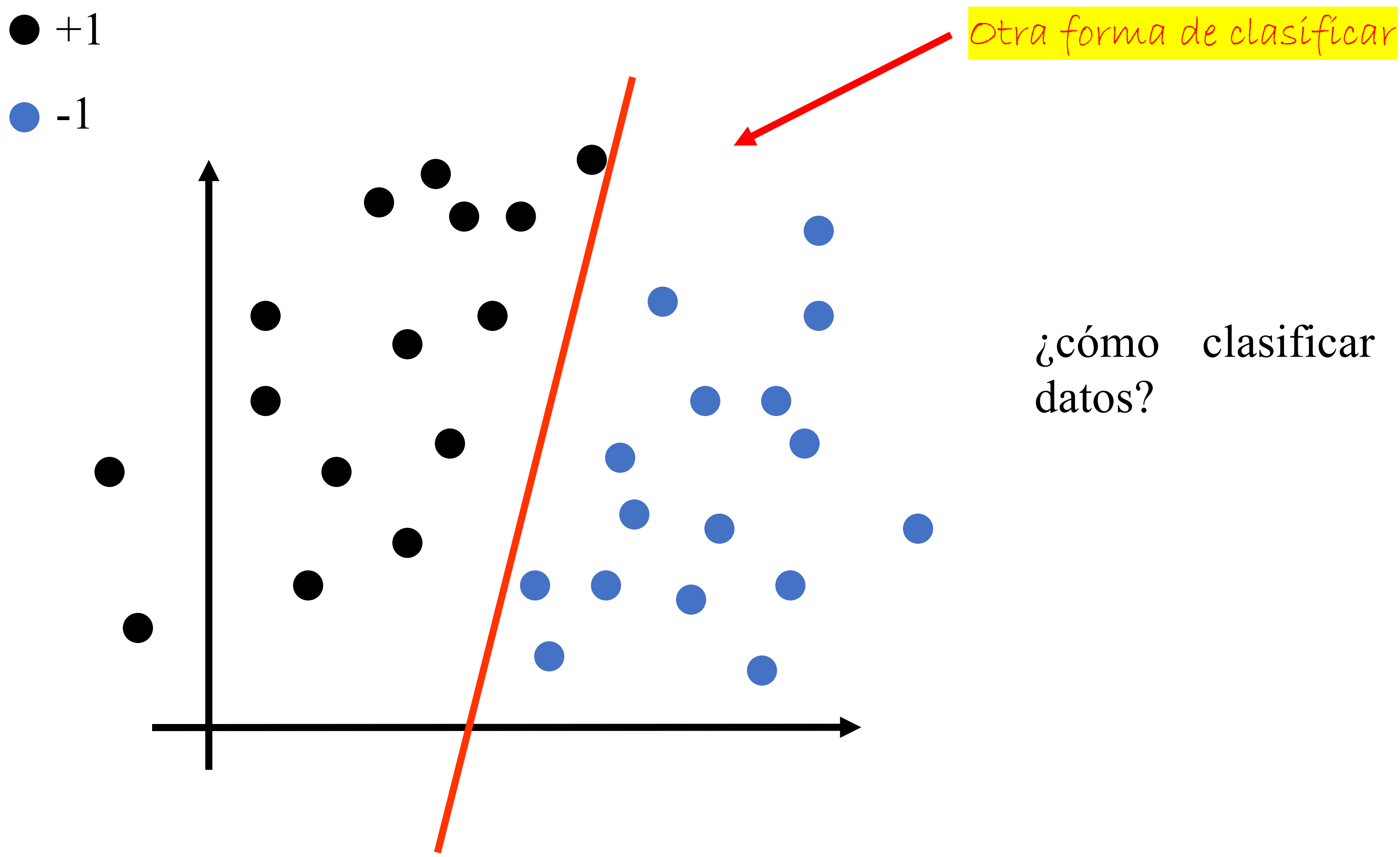
¿cómo clasificar estos datos?

Interpretación geométrica



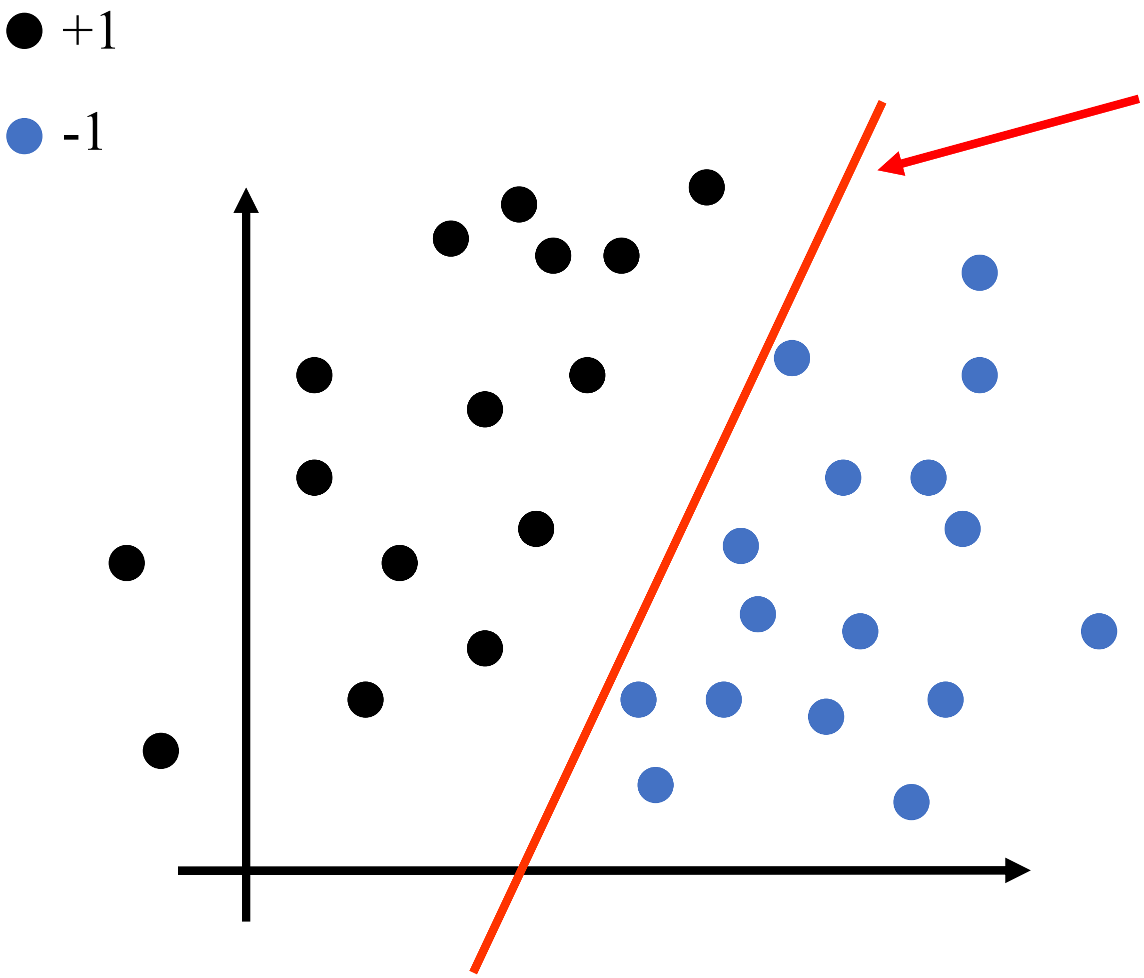
¿cómo clasificar estos datos?

Interpretación geométrica



¿cómo clasificar estos datos?

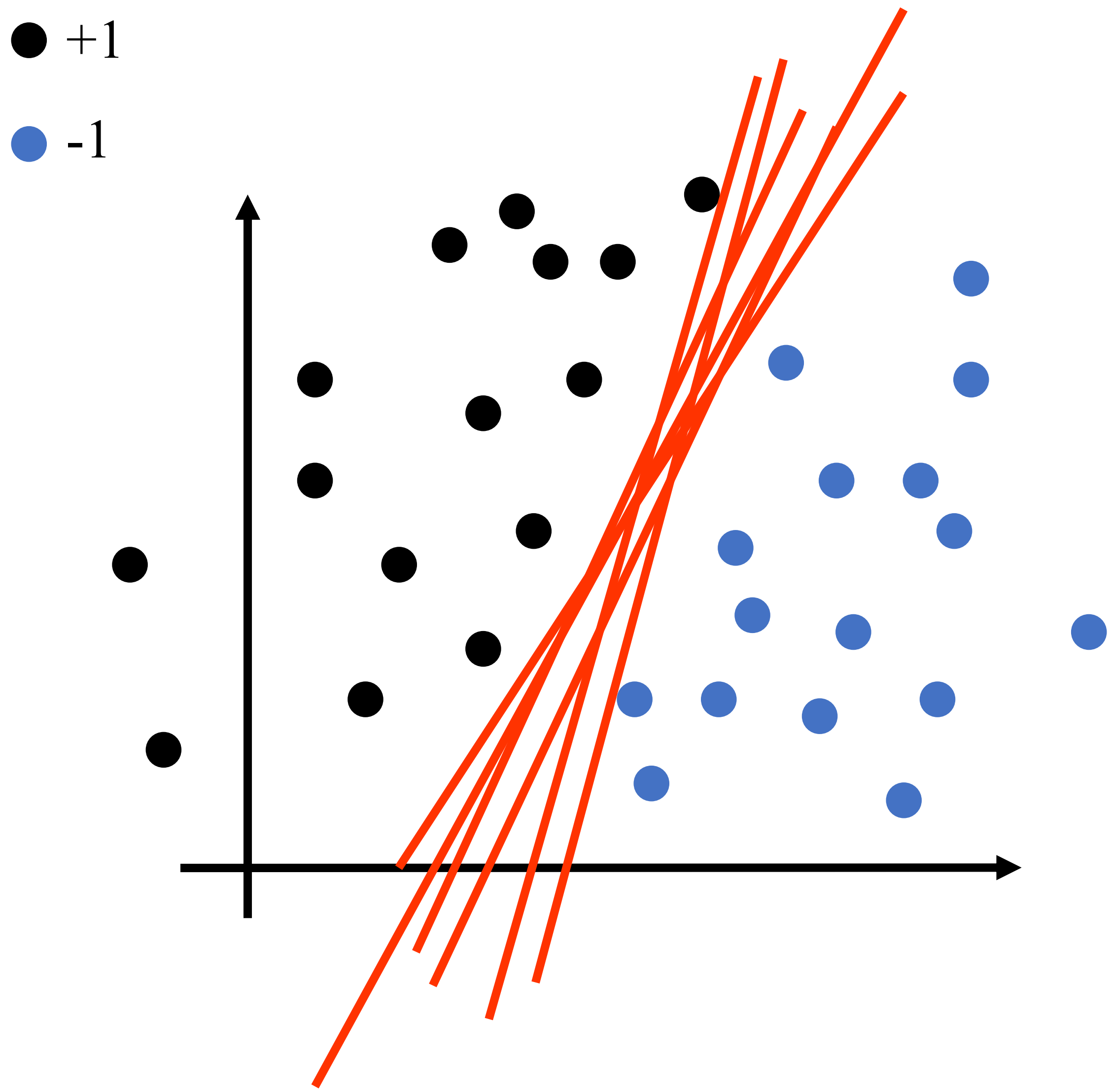
Interpretación geométrica



Podemos ver que puede tener un numero infinito de rectas

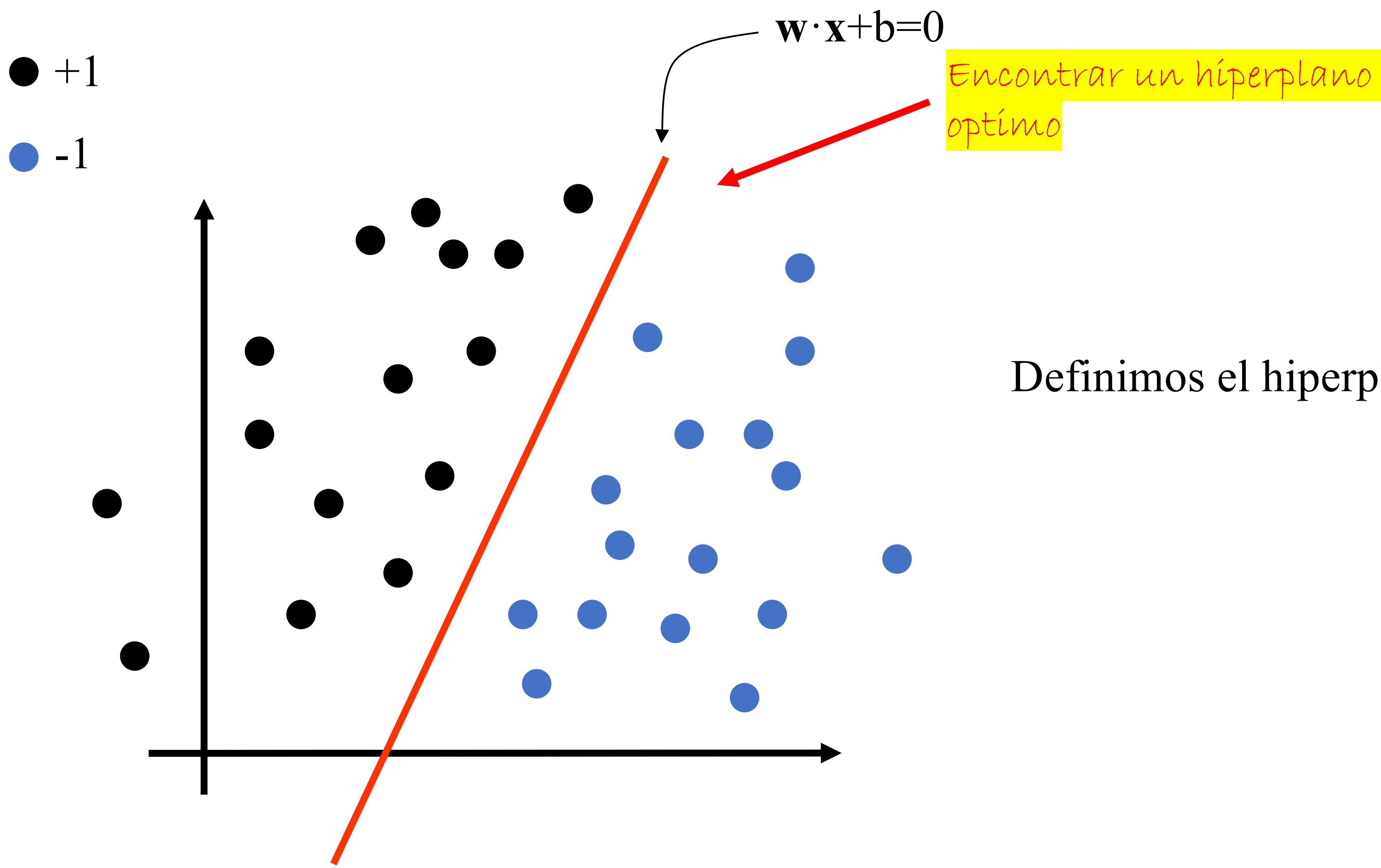
¿cómo clasificar estos datos?

Interpretación geométrica



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

Interpretación geométrica

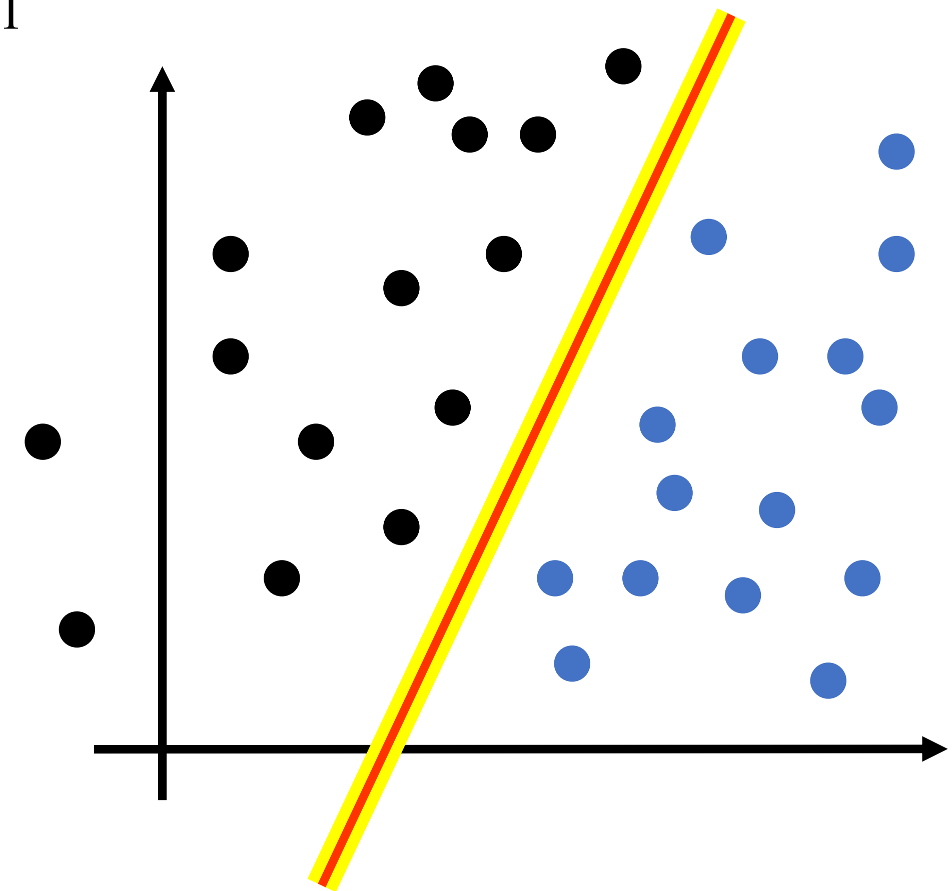


Definimos el hiperplano

Interpretación geométrica

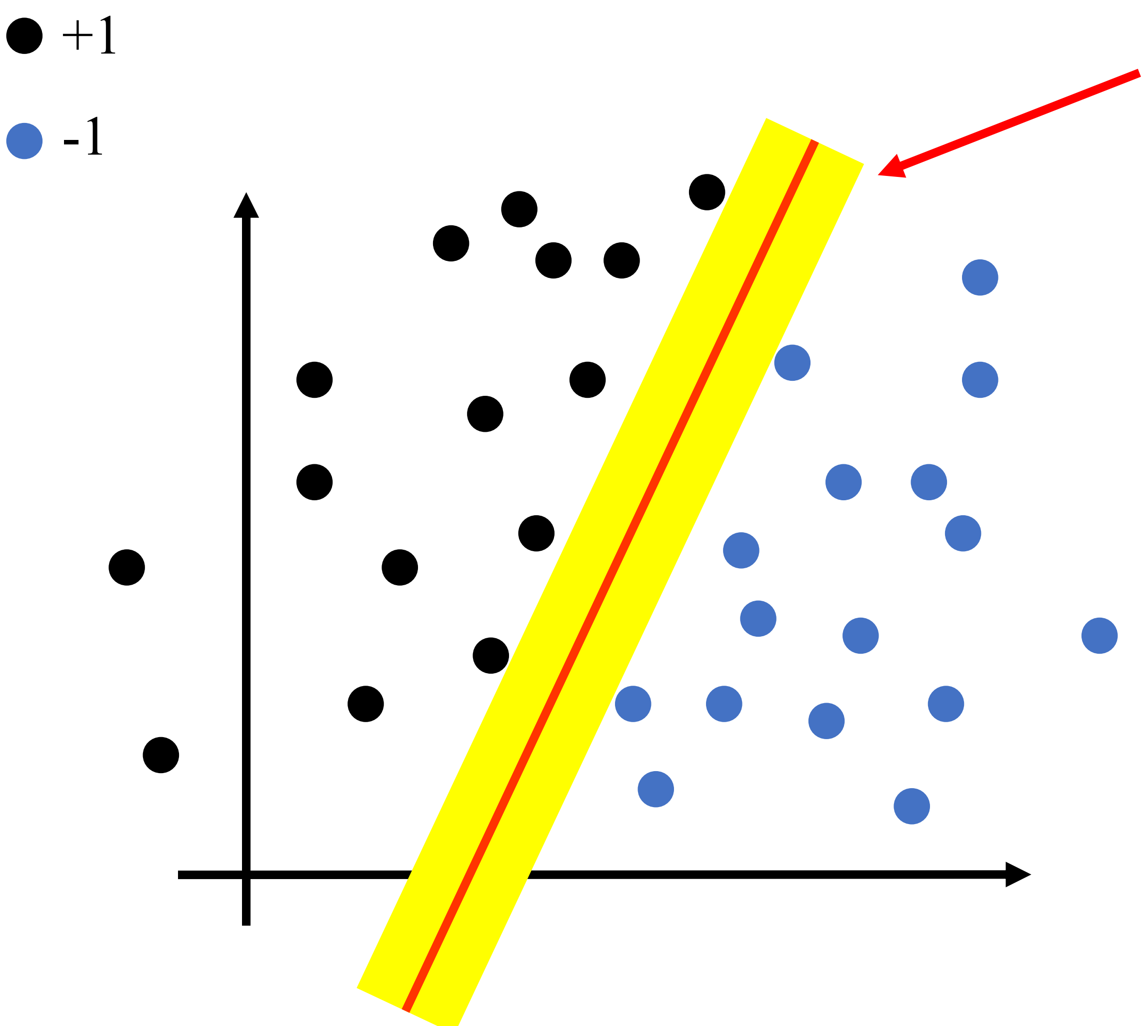
● +1

● -1



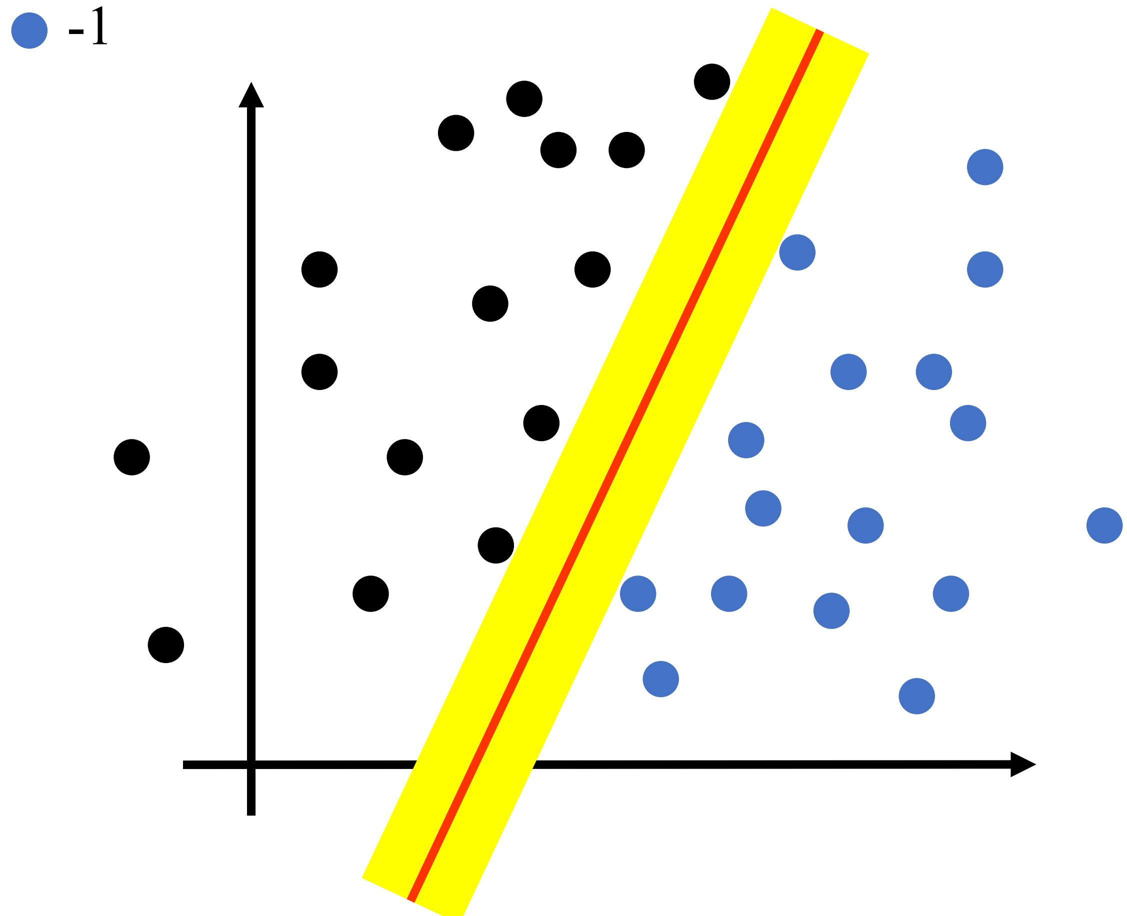
Definimos el margen

Interpretación geométrica



La idea es maximizar el margen.

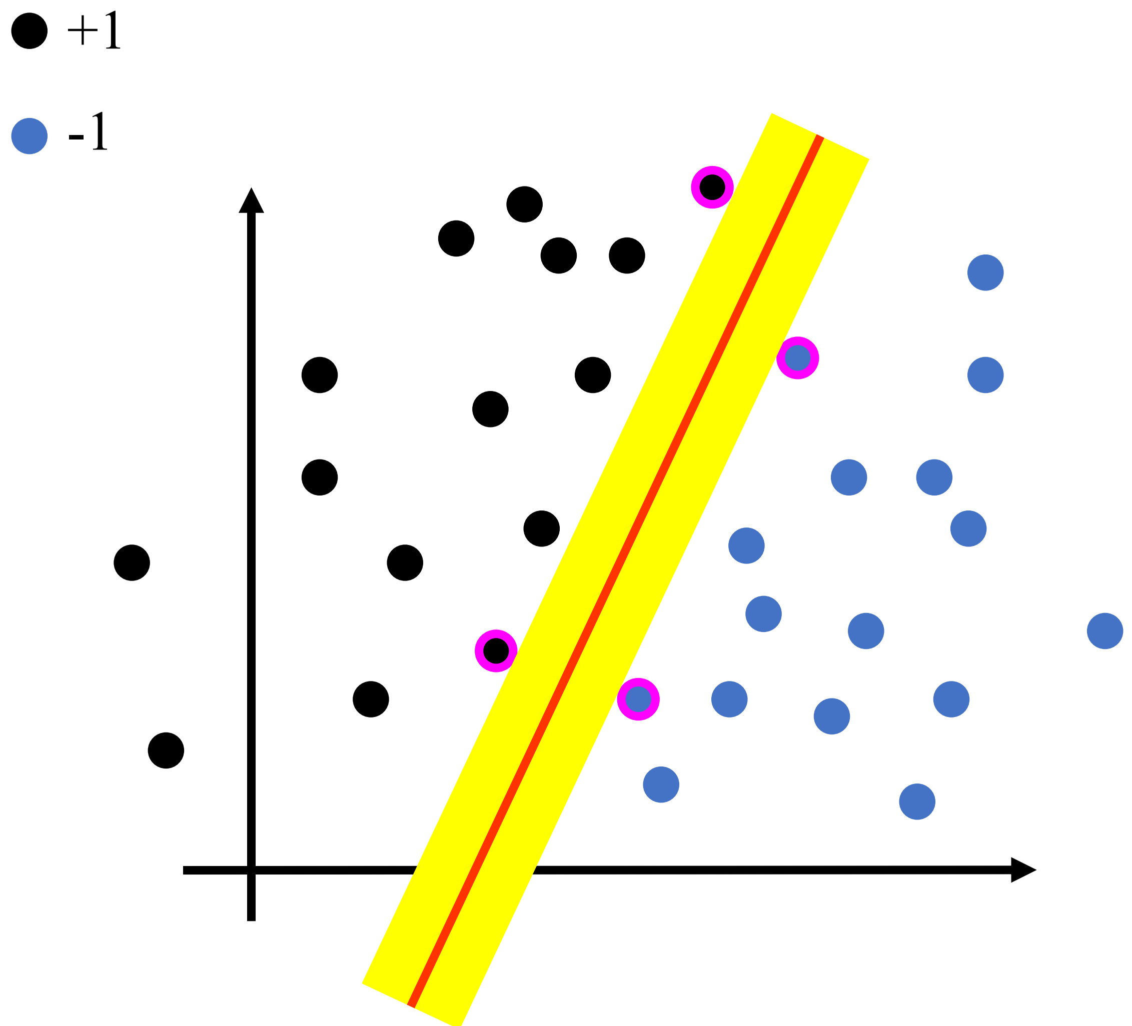
Interpretación geométrica



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

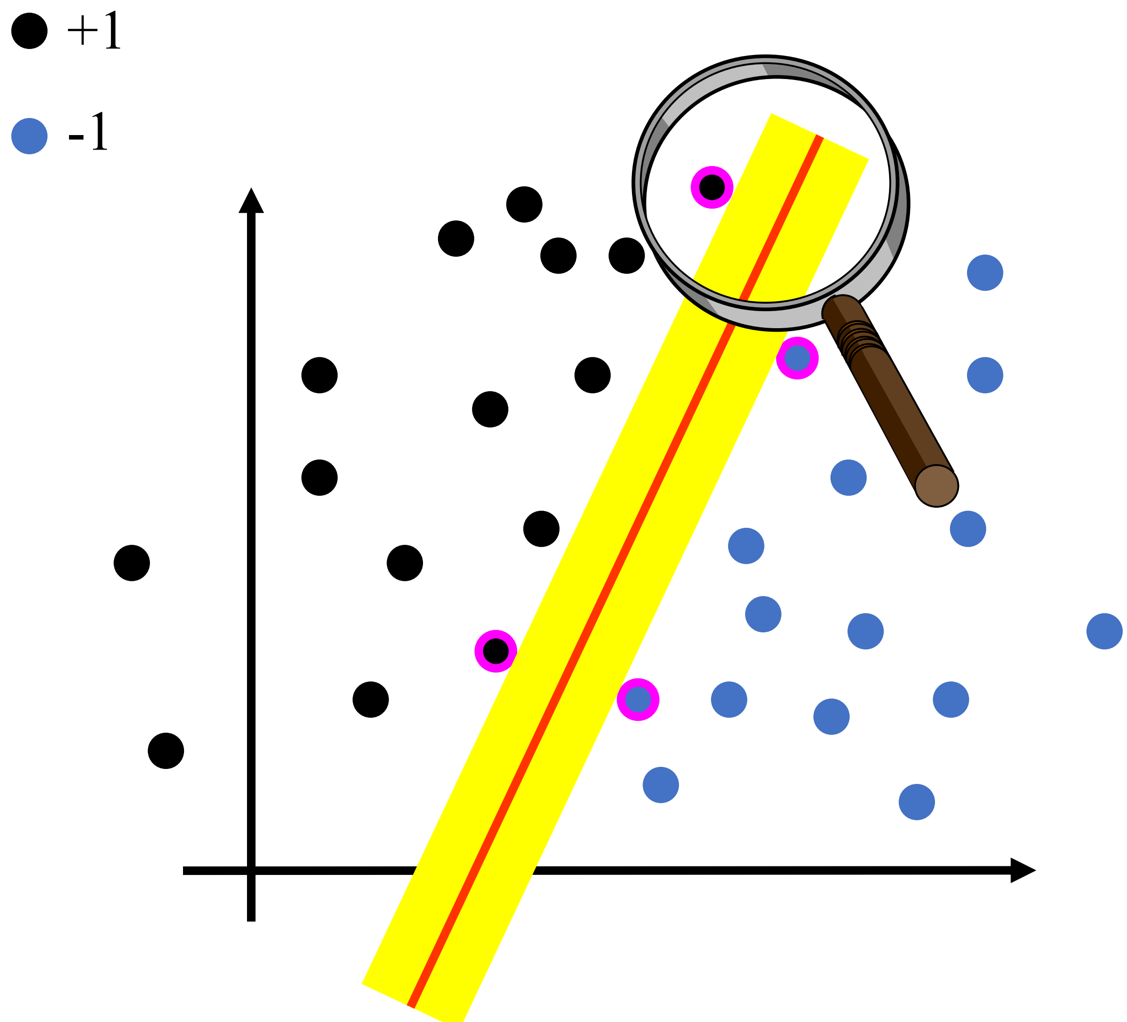
Esta es la clase más simple de SVM, la LSVM.

Interpretación geométrica



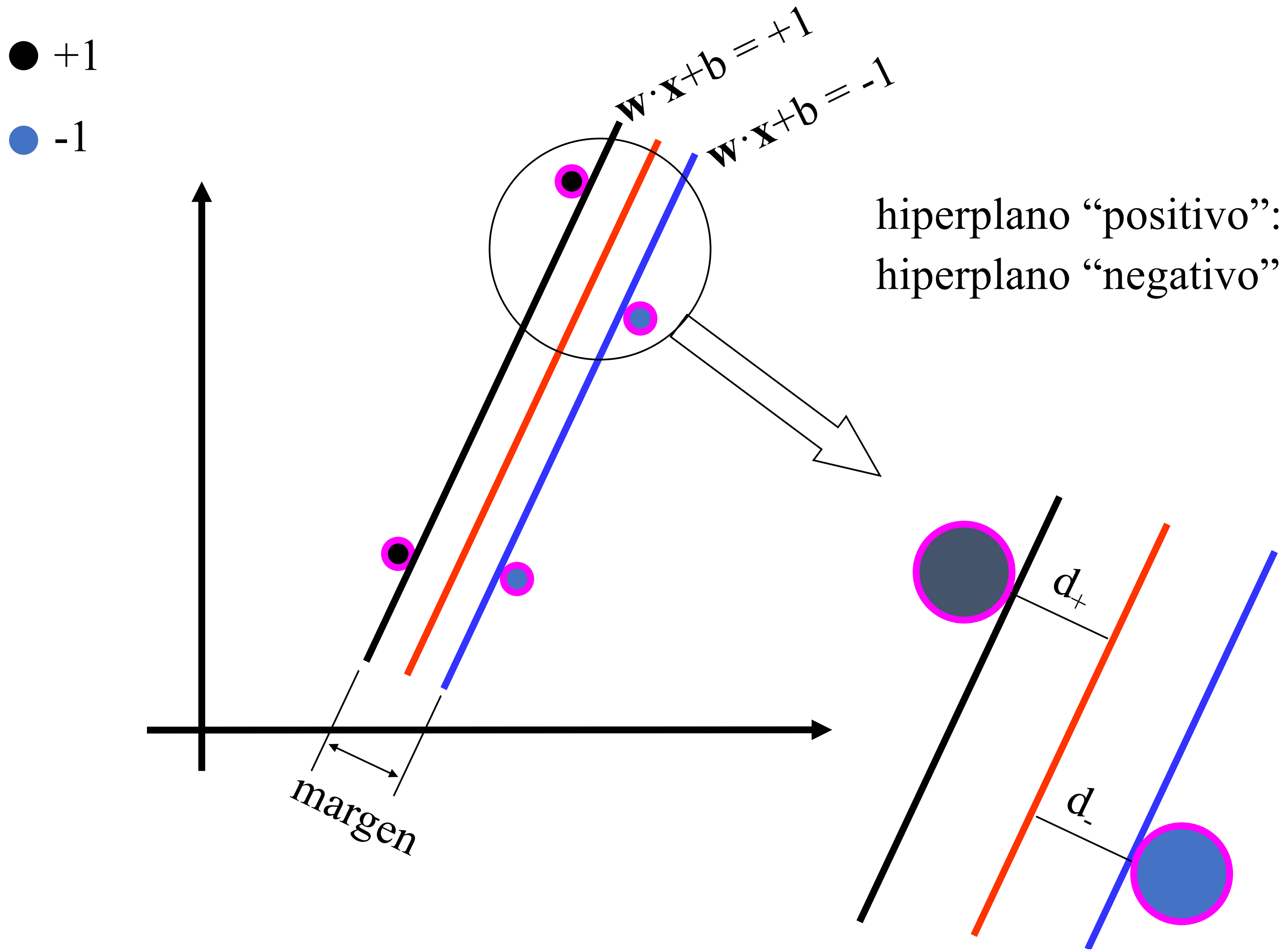
Los vectores de soporte son los puntos que tocan el límite del margen.

Interpretación geométrica



Veamos los hiperplanos "positivo" y "negativo"

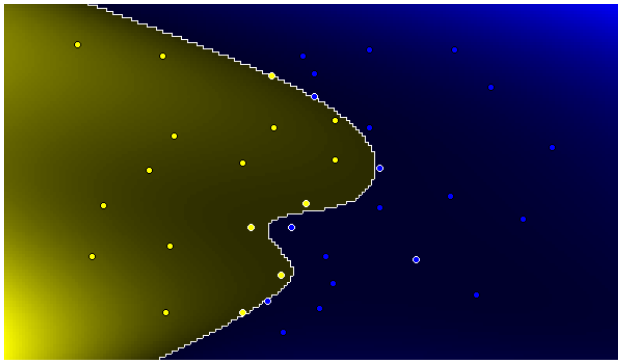
Interpretación geométrica

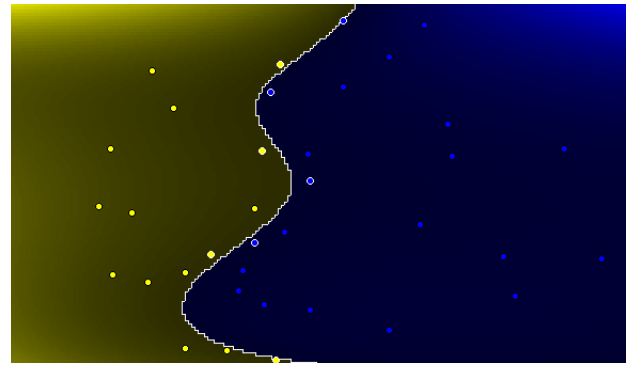


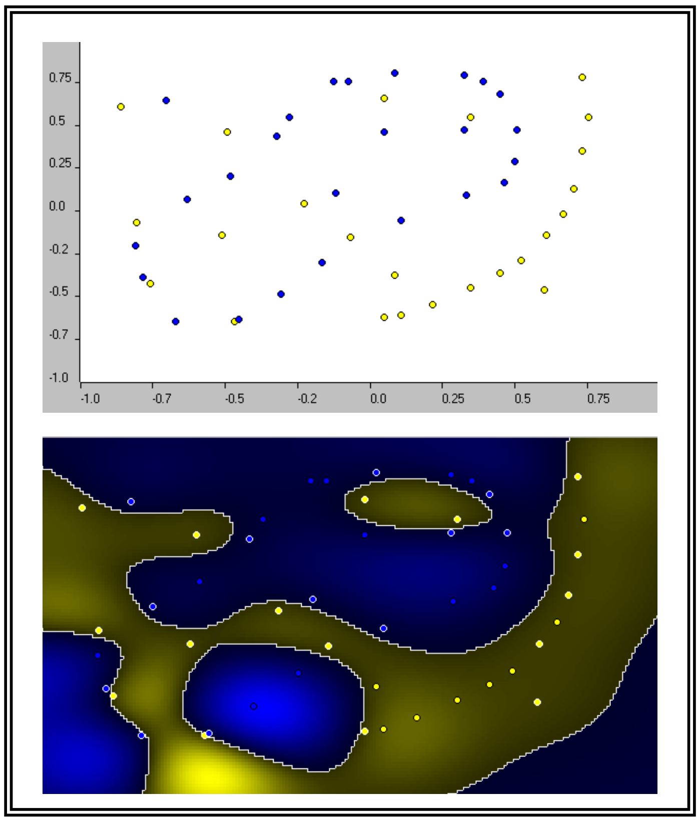
hiperplano "positivo": $\mathbf{w}\cdot\mathbf{x}+b = +1$

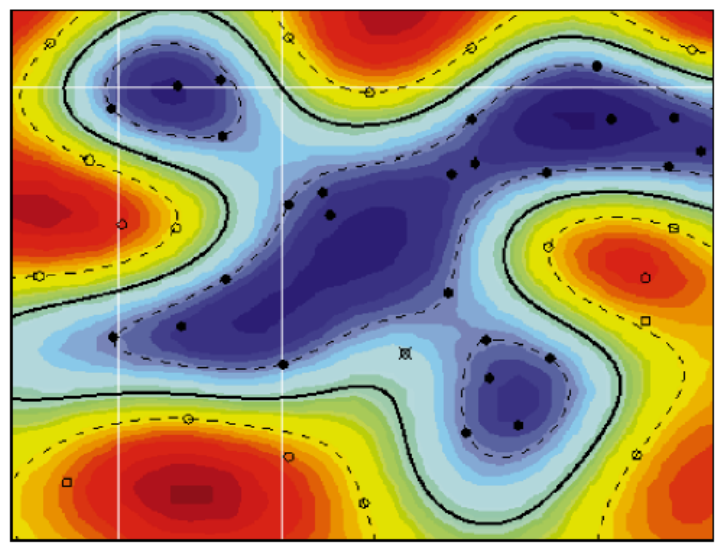
hiperplano "negativo": $\mathbf{w}\cdot\mathbf{x}+b = -1$

Separación polinómica y con RBF









Gracias!!!