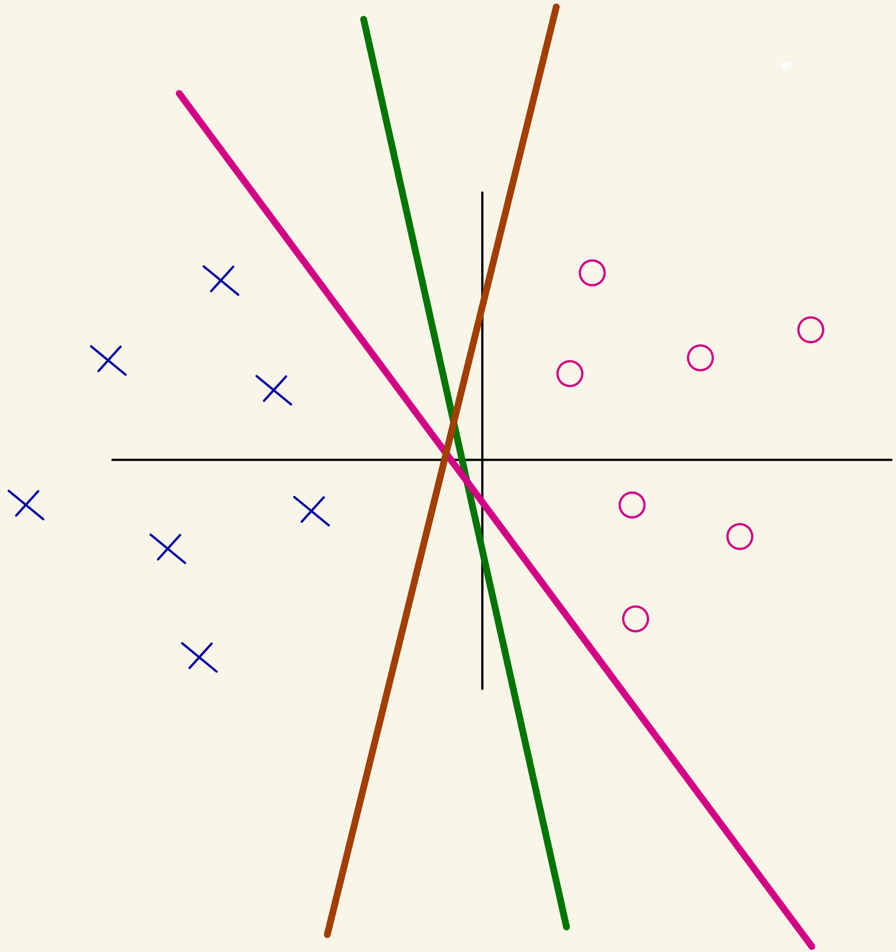
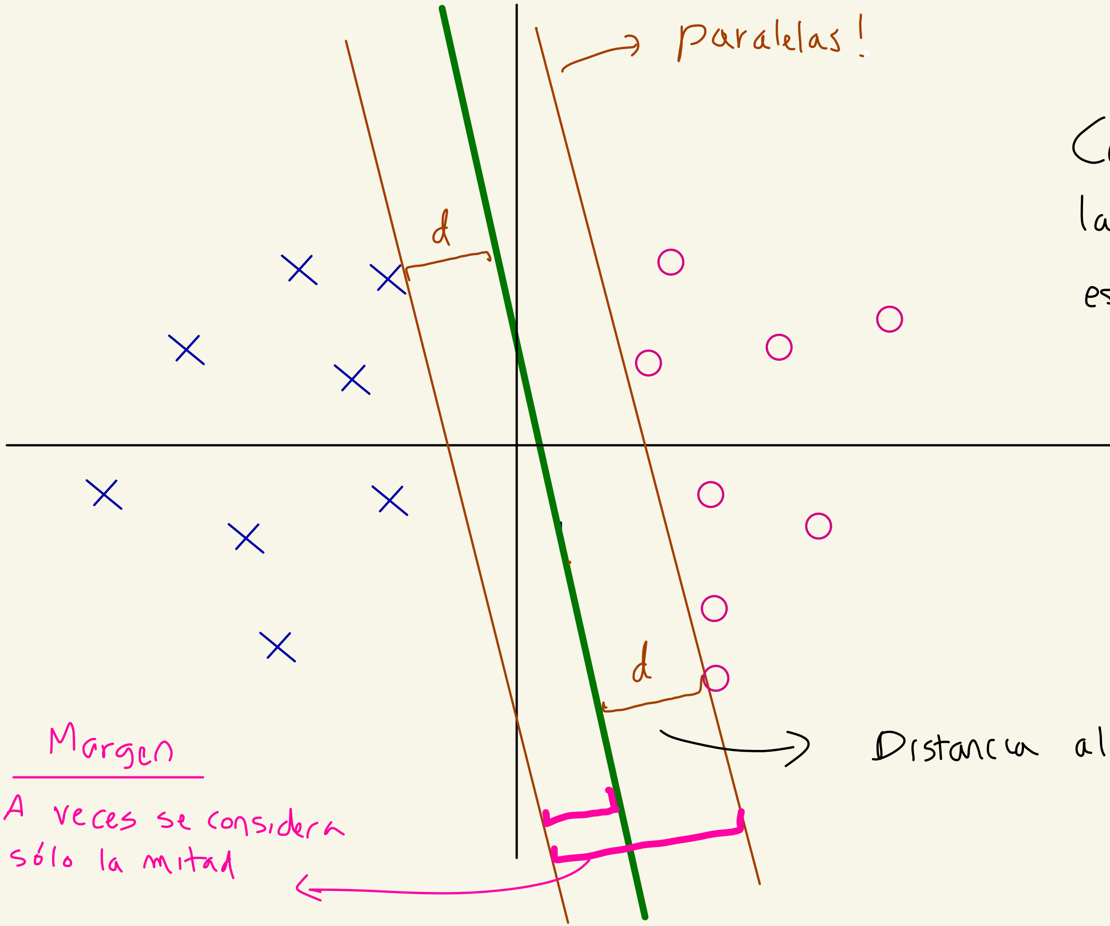
Máquinas de soporte vectorial



¿Cuál es mejor entre las 3 rectas para clasificar estos datos?

Máquinas de soporte vectorial



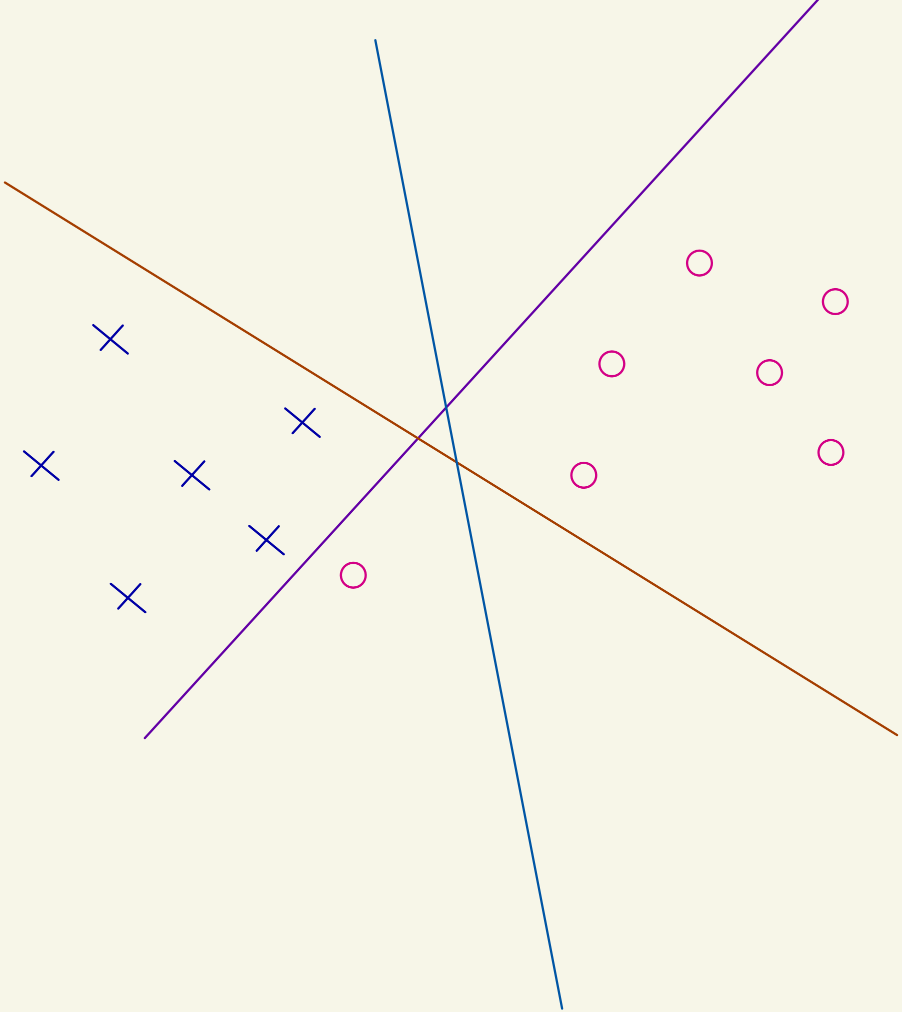
¿Cuál es mejor entre las 3 rectas para clasificar estos datos?

Separarlos mientras se maximiza d

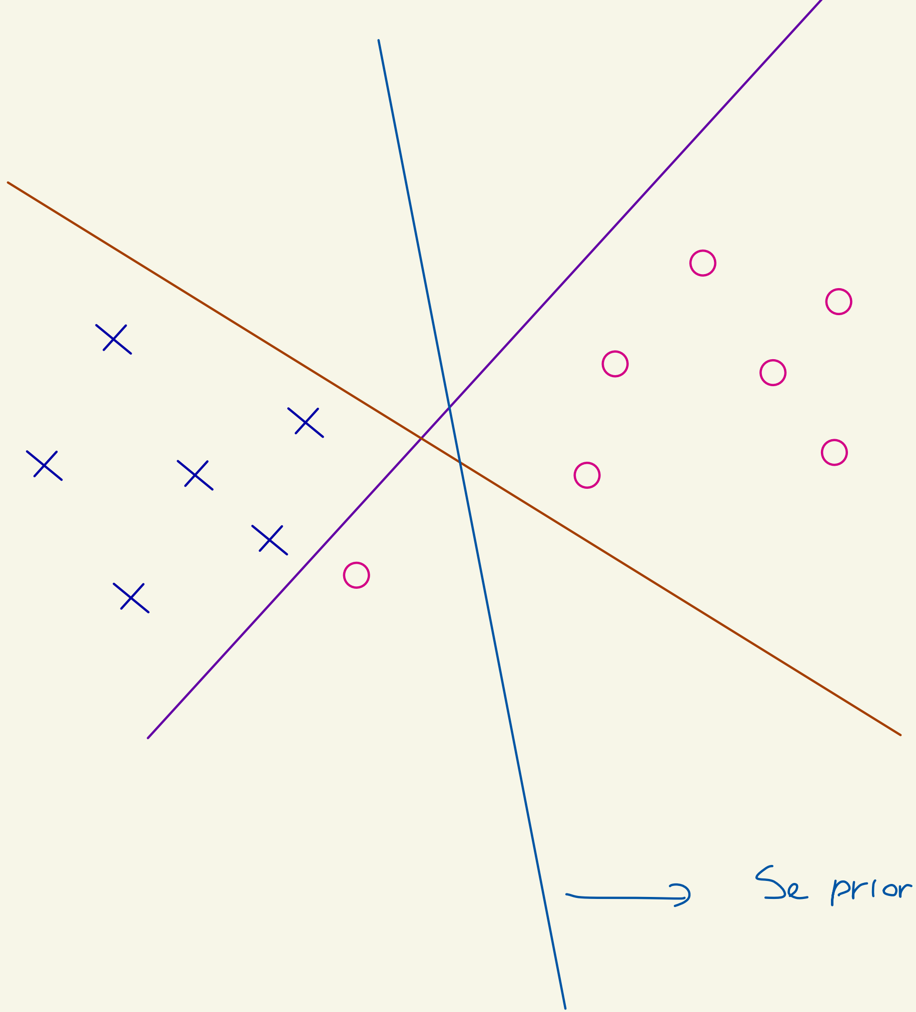
Distancia al punto más cercano

Margen

A veces se considera sólo la mitad



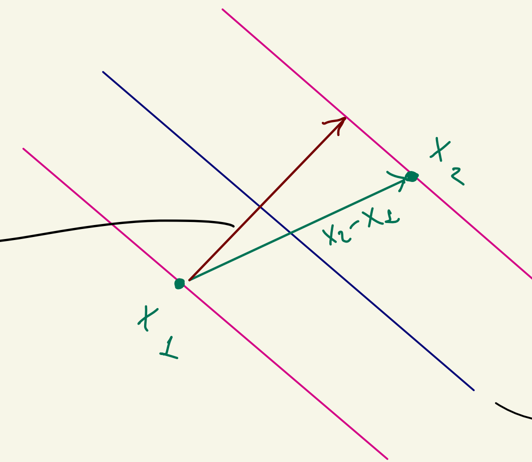
¿Cuál es mejor según un clasificador de soporte vectorial?



→ Se prioriza la margen máxima

Matemáticamante hablando, cómo se determina la recta?

Vector a maximizar

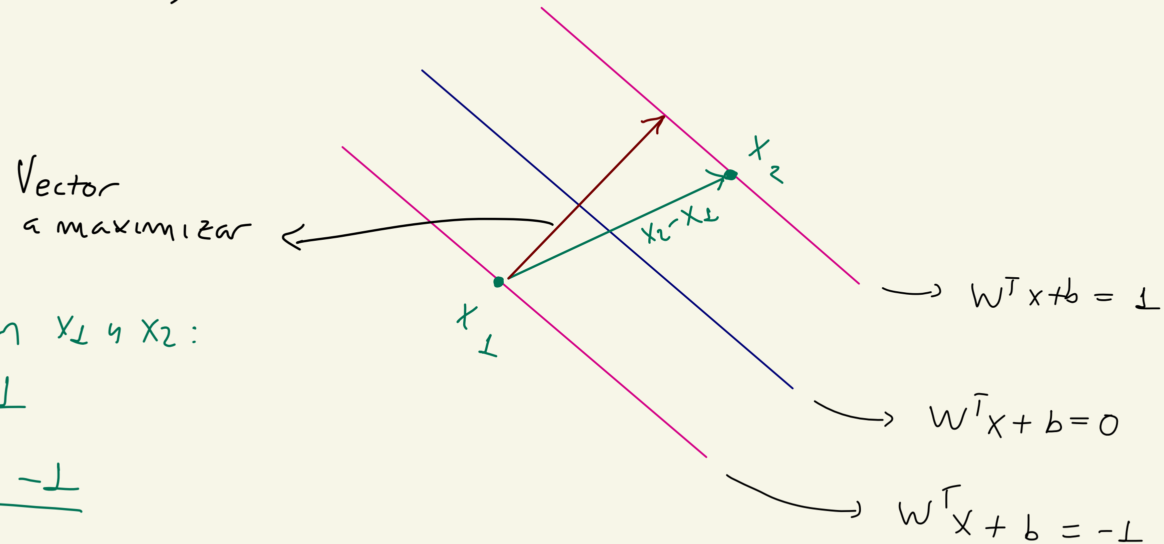


$w^T x + b = 1$

$w^T x + b = 0$

$w^T x + b = -1$

Matemáticamante hablando, cómo se determina la recta ?



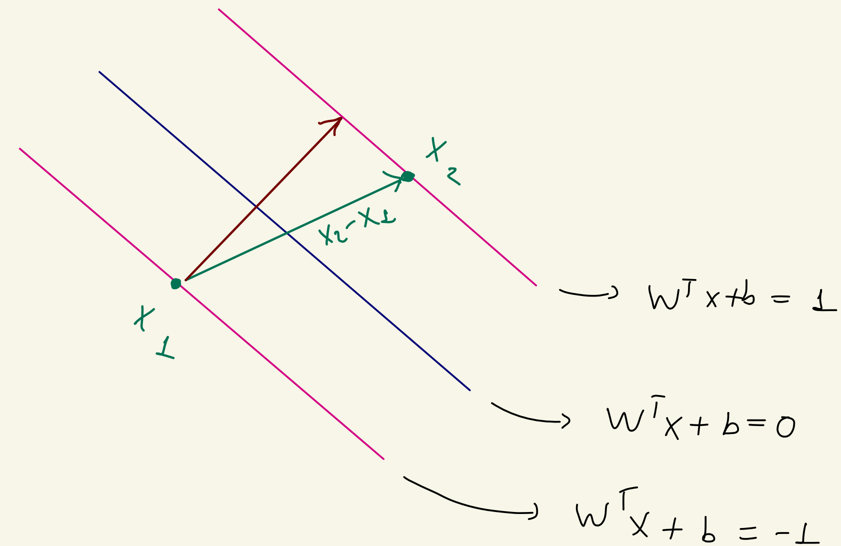
si restamos las dos ecuaciones que satisfacen x_1 y x_2 :

$$
\begin{array}{rl}
 & w^T x_1 + b = 1 \\
- & w^T x_2 + b = -1 \\
\hline
 & w^T (x_2 - x_1) = 2
\end{array}
$$

dividimos entre $\|w\|$

$$\frac{w^T (x_2 - x_1)}{\|w\|} = \frac{2}{\|w\|}$$

Matemáticamante hablando, cómo se determina la recta?



si restamos las dos ecuaciones que satisfacen x_1 y x_2:

$$
\begin{array}{rl}
 & w^T x_1 + b = 1 \\
- & w^T x_2 + b = -1 \\
\hline
 & w^T (x_2 - x_1) = 2
\end{array}
$$

dividimos entre $\|w\|$

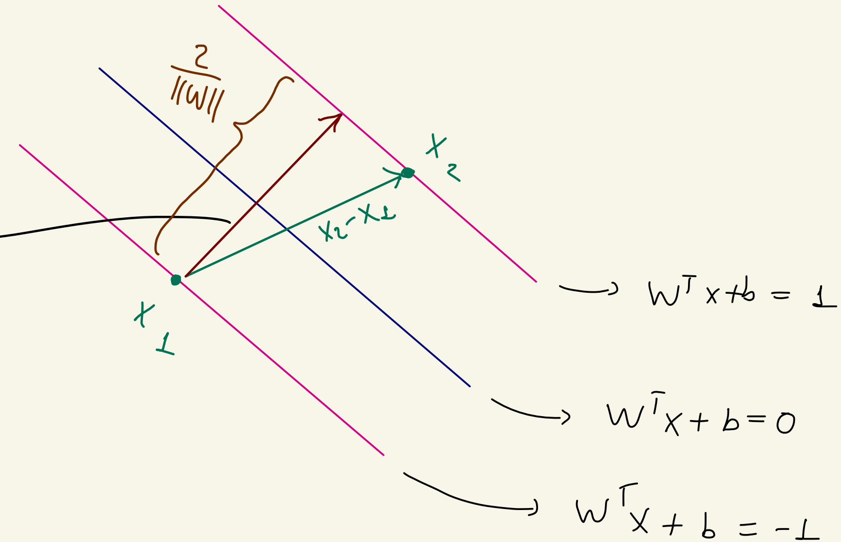
$$\boxed{\frac{w^T (x_2 - x_1)}{\|w\|}} = \boxed{\frac{2}{\|w\|}}$$

$$\text{proy}_{\vec{w}} \vec{v} = \frac{v \cdot w}{\|w\|} w$$

ojo! w es perpendicular a la recta/plano/hiperplano

proyección de $x_2 - x_1$ sobre w. Luego esta es la medida a maximizar, la cual es igual a $\frac{2}{\|w\|}$

Matemáticamente hablando, cómo se determina la recta ?



Magnitud a maximizar

si restamos las dos ecuaciones que satisfacen x_1 y x_2:

$$
\begin{array}{r}
w^T x_1 + b = 1 \\
- \quad w^T x_2 + b = -1 \\
\hline
w^T (x_2 - x_1) = 2
\end{array}
$$

dividimos entre $\|w\|$

$$\boxed{\frac{w^T (x_2 - x_1)}{\|w\|}} = \boxed{\frac{2}{\|w\|}}$$

$$\text{proy}_{\vec{w}} \vec{v} = \frac{v \cdot w}{\|w\|} w$$

ojo! w es perpendicular al (hiper)plano/recta

proyección de $x_2 - x_1$ sobre w. Luego esta es la medida a maximizar

Maximizar $\frac{2}{\|W\|}$, equivalente a minimizar $\boxed{\|W\|^2}$ (*)

Obs A la vez que maximizamos el margen, quisiéramos que los puntos queden bien clasificados

¿Qué quiere decir bien clasificados (por el modelo lineal)?

OjO
Aquí usamos las etiquetas $\{-1, 1\}$ en lugar de $\{0, 1\}$

$$\underbrace{y_i}_{1} \, (\underbrace{W^T x_i - b}_{\geq 1}) \geq 1 \quad (**)$$

$$\underbrace{-1 \quad \leq -1}_{\text{Casos en que el punto está bien clasificado}}$$

Finalmente, para encontrar los parámetros de la mejor recta/plano/hiperplano, se minimiza la siguiente expresión

$$\min \quad \underbrace{C\|w\|^2}_{(*)} + \left[\frac{1}{n} \sum_{i=1}^{n} \underbrace{\max\left\{0,\ 1 - y_i(w^T x_i - b)\right\}}_{(**)} \right]$$

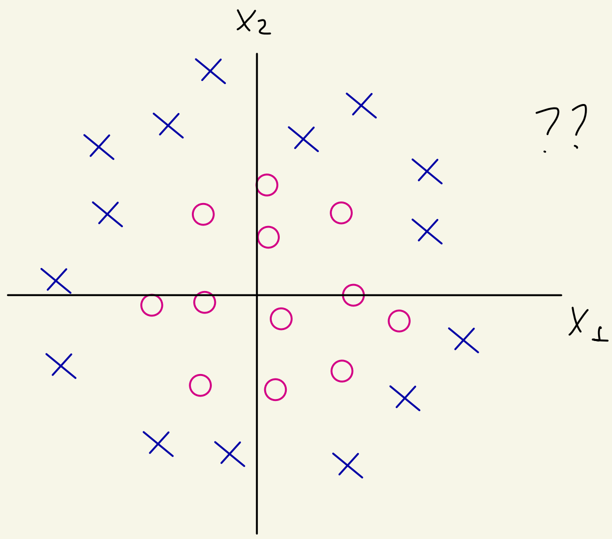
Finalmente, para encontrar los parámetros de la mejor recta/plano/hiperplano, se minimiza la siguiente expresión

$$\min \quad C \underbrace{\|w\|^2}_{\text{maximizando el margen}} + \left[\underbrace{\frac{1}{n} \sum_{i=1}^{n} \max \{0, 1 - y_i(w^T x_i - b)\}}_{\text{maximizando la cantidad de puntos bien clasificados}} \right]$$

hiperparámetro

establece equilibrio entre mayor margen y puntos bien clasificados

Datos no separables linealmente



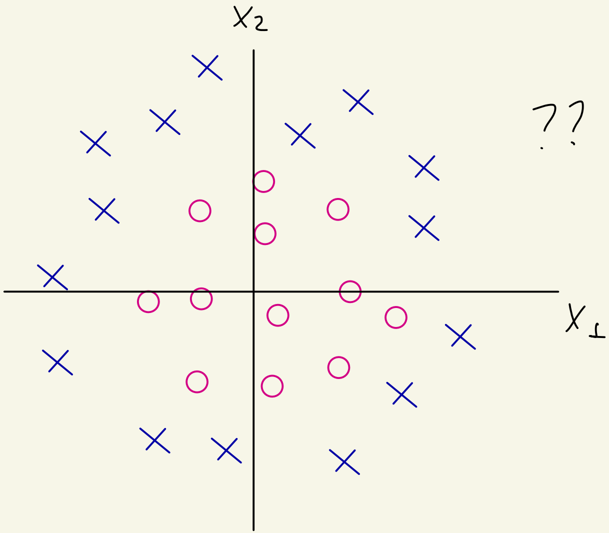
Datos no separables linealmente

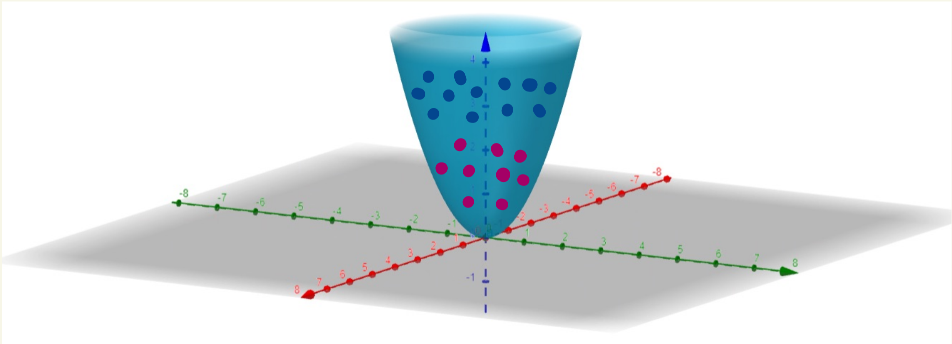
podemos agregar una dimension/atributo

(x_1, x_2, x_3)

$$x_3 = x_1^2 + x_2^2$$

"$z = x^2 + y^2$"





Ahora se pueden separar por un plano

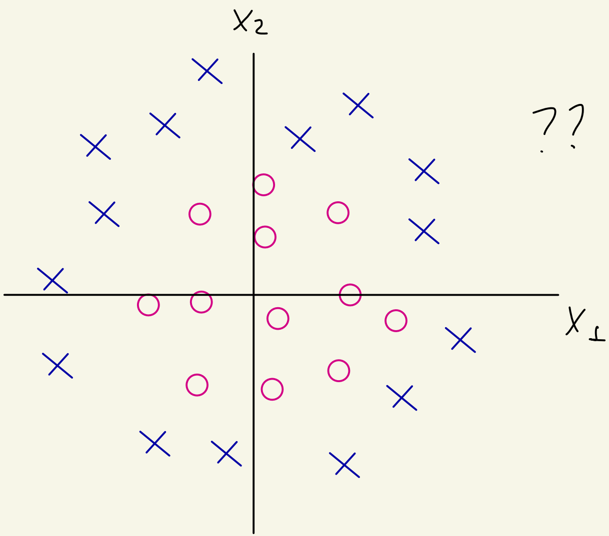
Datos no separables linealmente

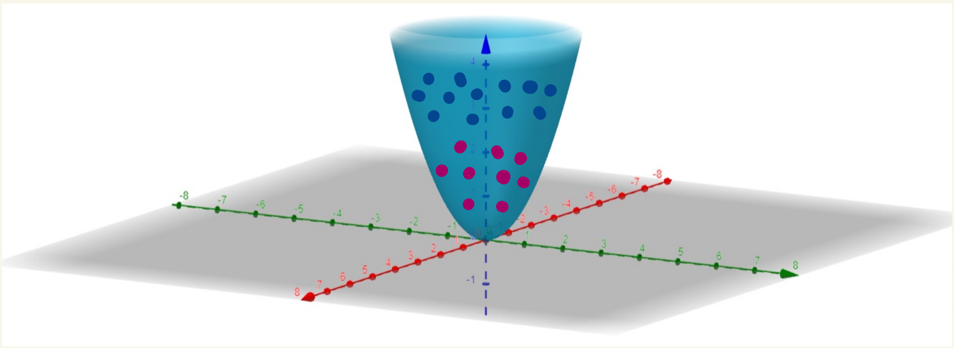
podemos agregar una dimension/atributo

(X_1, X_2, X_3)

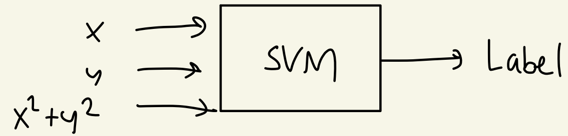
$$X_3 = X_1^2 + X_2^2$$

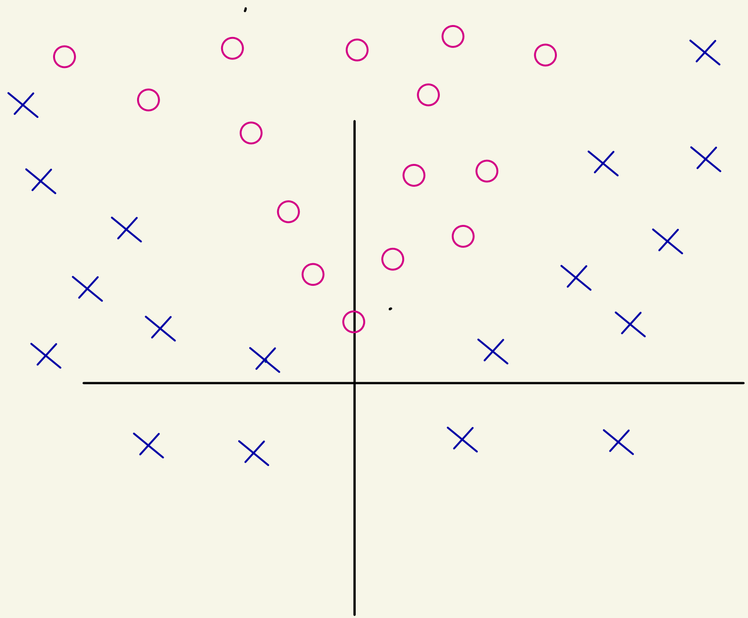
"$z = x^2 + y^2$"



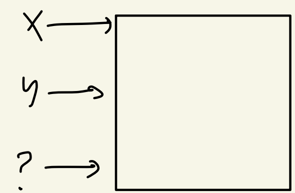


Ahora se pueden separar por un plano





Cuál atributo agregaría?



Kernels

Los Kernel nos dan diferentes formas de involucrar un aumento en la dimensionalidad de los datos en el problema de optimización

$$x, y \xrightarrow{\text{Kernel}} x_1, x_2, x_3, x_4, x_5$$

no separable — separable

Queríamos optimizar el problema siguiente

Minimizar $$C\|w\|^2 + \left[\frac{1}{n}\sum_{i=1}^{n} \max\{0,\ 1 - y_i(w^T x_i - b)\}\right]$$

Este problema equivale al siguiente (programación cuadrática)

Nuevo vector de parámetros

Expresa "similaridad" entre pares de vectores

Maximizar $$W(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_{ij} \alpha_i \alpha_j\, y_i y_j \underbrace{x_i^T x_j}_{K(x_i, x_j)}$$

t.q. $\alpha_i \geq 0$ $\quad \sum \alpha_i y_i$

Esta similaridad o relación entre vectores se puede expresar de diferentes formas

Algunos tipos de kernel

Polinomial: $(x^T y + r)^d$ si $r = 0$, $d = 1$, corresponde al "Kernel lineal"

Radial: $e^{-\gamma \|\vec{x} - \vec{y}\|^2}$ RBF (Radial basis function)

Tang. hiperbólica: $\tanh(\alpha x^T y + \theta)$

Algunos tipos de kernel

$a, b \in \mathbb{R}$

Ejemplo con el polinomial: $(a \cdot b + r)^d$

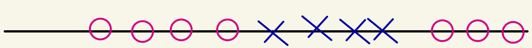
Ejemplo $r = \frac{1}{2}$

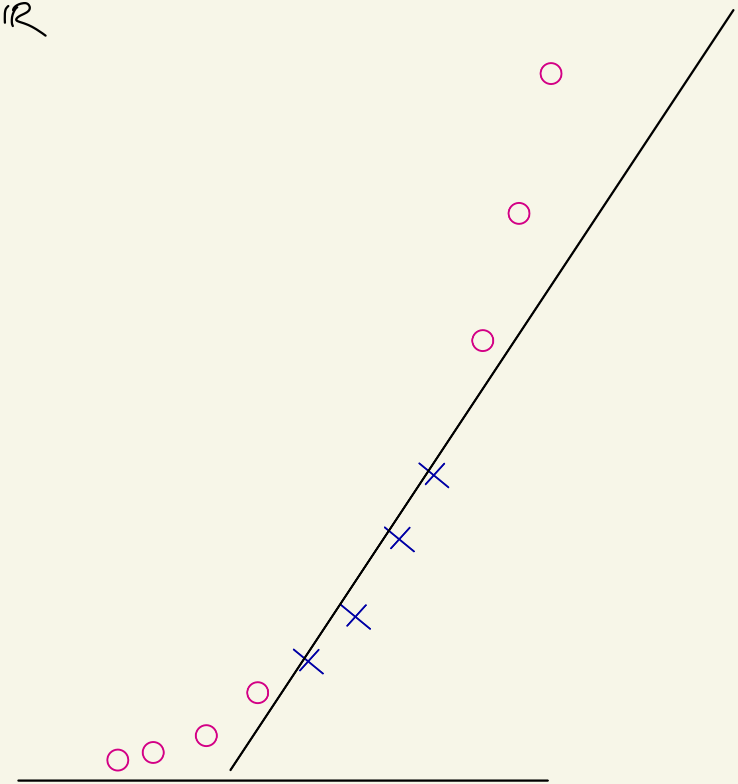
$d = 2$

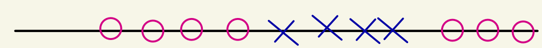
$$\left(a \cdot b + \frac{1}{2}\right)^2 = \left(ab + \frac{1}{2}\right)\left(ab + \frac{1}{2}\right)$$

$$= ab + a^2b^2 + \frac{1}{4}$$

$$= \left(a, a^2, \frac{1}{2}\right)\left(b, b^2, \frac{1}{2}\right)$$







Radial : $e^{-\gamma \|\vec{x}-\vec{y}\|^2}$

2

El kernel radial permite un incremento hacia una dimensionalidad infinita, por medio de la expresión en serie de Taylor de la función e^x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots\ldots$$

para detalles recomiendo el video de Josh Starmer (statquest) "Support Vector machines Part 3: The radial (RBF) Kernel"

En dimensiones finitas, el kernel radial tiene en cuenta la distancia entre vectores; $\|\vec{x}-\vec{y}\|^2$

por lo que puede clasificar de forma similar a KNN (weighted)

$$e^{-\gamma \|\vec{x}-\vec{y}\|^2}$$

Algunos parámetros de SVM en sklearn

Kernel: Linear, $\underbrace{\text{rbf}}_{\text{radial}}$, etc

C : parámetro en la ecuación: $C\|w\|^2 + \left[\frac{1}{n}\sum_{i=1}^{n} \max\{0, 1 - y_i(w^T x_i - b)\}\right]$

establece un equilibrio entre maximizar el margen y clasificar bien los puntos

γ: Determina qué tan lejana es la influencia de un punto de entrenamiento mediante peso que se le da a las distancias

valores bajos - alcance lejano
valores altos - alcance cercano