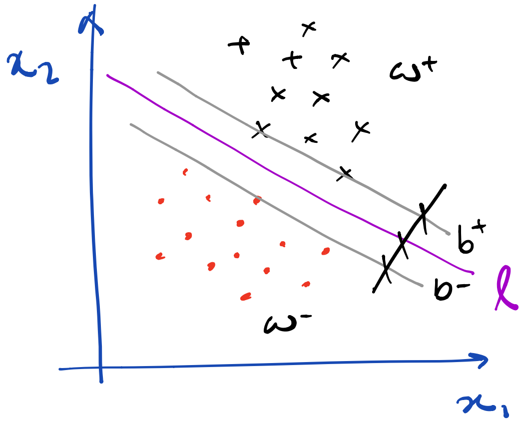
SVM: Support Vector Machine



En SVM se busca una línea de separación ℓ tal que se maximice el margen.

El margen $b = b^+ = b^-$ definido como la distancia perpendicular de la línea ℓ al punto más cercano de cada clase

Definición de ℓ:

$\ell: \ g(\underline{x}) = 0$

donde

$$\boxed{g(\underline{x}) = a_1 x_1 + a_2 x_2 + a_0}$$

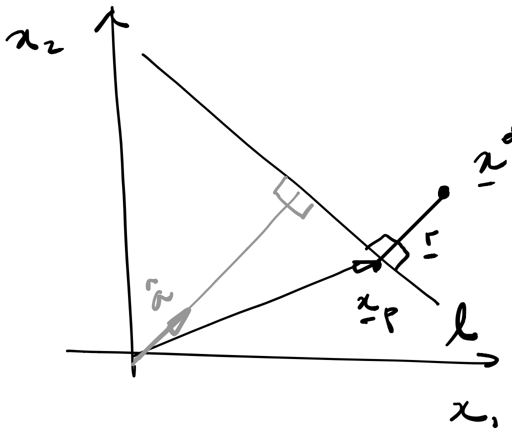
La clasificación de un punto $\underline{x}$ es:

$$g(\underline{x}) \begin{cases} > 0 \longrightarrow \text{clase } +1 \\ \leq 0 \longrightarrow \text{clase } -1 \end{cases}$$

¿Cómo estimar la recta ℓ?

1 $\longrightarrow$ cálculo de distancia de un punto $\underline{x}^*$ a una recta
2 $\longrightarrow$ planteamiento del problema
3 $\longrightarrow$ planteamiento de cómo resolverlo
4 $\longrightarrow$ solución

① Distancia de un punto $\underline{x}^*$ a la recta:



$$\underline{x}^* = \underline{x}_p + \underline{r}$$

$$\underline{r} = r\,\hat{\underline{a}}$$

↳ vector unitario ⊥ l

↳ distancia a l

Se puede demostrar que

$$r = \frac{|g(\underline{x}^*)|}{\|\underline{a}\|}$$

donde $\underline{a} = [a_1\ a_2]^T$

VER DEMOSTRACIÓN AL FINAL

② Planteamiento del problema

Para encontrar l, se debe cumplir que las distancias de los puntos de entrenamiento $\underline{x}_i$ a la recta l deben ser $\geq b$:

$$\frac{|g(\underline{x}_i)|}{\|\underline{a}\|} \geq b \quad \text{para } i = 1 \dots N \quad (1)$$

Se debe encontrar $a_1, a_2, \dots a_0$ tal que se cumpla la condición anterior.

Considerando que $|g(\underline{x}_i)| = z_i\, g(x_i)$

donde $z_i = \pm 1$ si $\underline{x}_i$ pertenece a la clase ± 1.

(1) se puede escribir como:

$$\frac{z_i\, g(\underline{x}_i)}{\|\underline{a}\|} \geq b \qquad (2)$$

Es necesario imponer una restricción adicional a $\underline{a}$ y a_0 para que no escale al infinito:

sujeto a $\quad b\,\|\underline{a}\| = 1$

Entonces (2) es:

$$z_i\, g(\underline{x}_i) \geq 1$$

o bien $\quad \boxed{z_i\, g(\underline{x}_i) - 1 \geq 0} \qquad (3)$

Se persigue entonces maximizar b, esto equivale a minimizar $\|\underline{a}\|$ tal que se cumpla (3).

PROBLEMA:

Minimizar $\quad \|\underline{a}\|$

Sujeto a:

$$z_i\,(\underline{a}\,\underline{x}_i^T + a_0) - 1 \geq 0 \quad \text{para } i \in [1, N]$$

3 ¿Cómo resolver el problema de optimización?

El problema se traduce a:

$$J(\underline{a}, a_0, \underline{\lambda}) = \underbrace{\frac{1}{2}\|\underline{a}\|^2}_{A} \underbrace{-}_{B} \underbrace{\sum_{i=1}^{N} \lambda_i\left[z_i(\underline{a}^T\underline{x}_i + a_0) - 1\right]}_{C} \rightarrow \min \qquad (4)$$

A: Se persigue minimizar $\|\underline{a}\|$

B: Se persigue maximizar segundo término C

C: Con $\lambda_i \geq 0$ se fuerza a encontrar una solución que clasifique correctamente $(z_i(\underline{a}\underline{x}_i^T + a_0) - 1 \geq 0)$

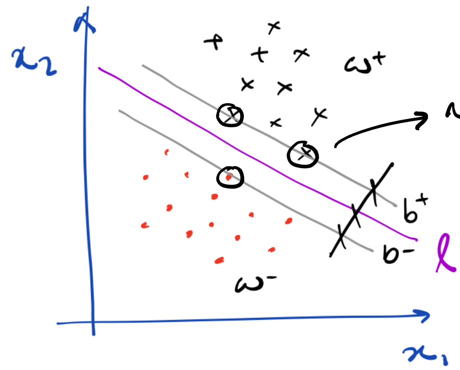
Según Kuhn-Tucker, (4) se puede transformar en:

$$J(\underline{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j z_i z_j \underbrace{\underline{x}_i^T \underline{x}_j}_{\substack{\langle \underline{x}_i^T, \underline{x}_j \rangle \\ \text{producto punto}}} \rightarrow \max \qquad (5)$$

Sujeto a $\lambda_i > 0$

4 Solución

Al resolver (5), sólo los vectores de soporte, i.e, los puntos $\underline{x}_i$ que tienen contacto con los bordes del margen tienen $\lambda_i \neq 0$. El resto son muestras irrelevantes:



→ vectores de soporte con $\lambda_i > 0$

En este ejemplo hay 3 vectores de soporte: $N_{SV} = 3$.

Al resolver (5) la solución es λ_i.

Derivando (4), igualando a cero, y conociendo λ_i se obtiene la solución para $\underline{a}$:

$$\boxed{\underline{a} = \sum_{i=1}^{N} \lambda_i z_i \underline{x}_i} \rightarrow \underline{a} = \sum_{k=1}^{N_{SV}} \lambda_k z_k \underline{x}_k$$

vectores de soporte

Usando los vectores de soporte se obtiene a_0 ya que para ellos se cumple

$$z_i (\underline{a}^T \underline{x}_i + a_0) - 1 = 0 \qquad ({}^{\wedge}/z_i = z_i)$$

$$\underline{a}^T \underline{x}_i + a_0 = z_i$$

$$\boxed{a_0 = z_i - \underline{a}^T \underline{x}_i}$$

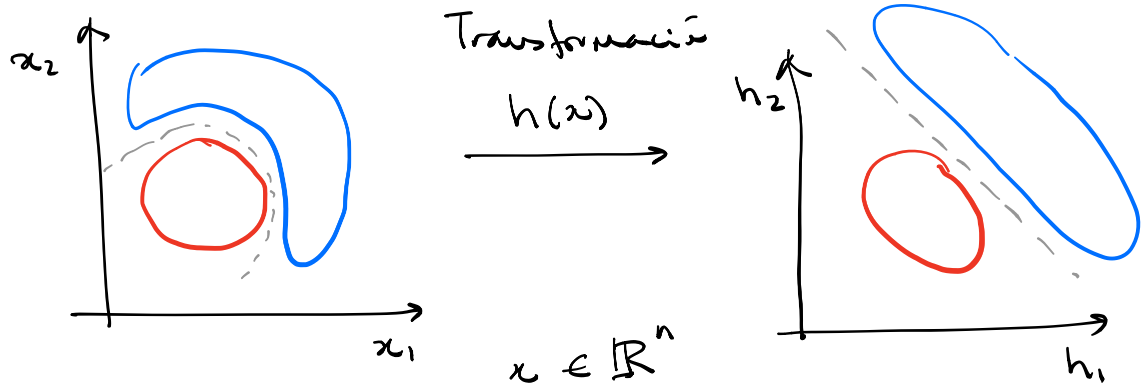
x_i es vector de soporte

El hiperplano queda definido como:

$$g(\underline{x}) = \underline{a}^T \underline{x} + a_0$$

$$\boxed{g(\underline{x}) = \sum_{i=1}^{N} \lambda_i z_i \underbrace{\underline{x}_i^T \underline{x}}_{\langle \underline{x}_i, \underline{x} \rangle} + a_0}$$

⑤ Bonus: líneas de decisión no-lineales.



$\underline{x} \in \mathbb{R}^n$

$\underline{h} \in \mathbb{R}^m$ $\quad m \geq n$

La línea de separación sería

$$g(x) = g(h(\underline{x})) = \underline{a}^T h(\underline{x}) + a_0$$

$$g(\underline{x}) = \sum_{i=1}^{N} \lambda_i z_i \langle h(x_i), h(x) \rangle + a_0$$

<u>IMPORTANTE</u>: no interesa $h(x)$, sino su producto punto

$$\langle h(x_i), h(x) \rangle = \underbrace{K(x_i, x)}_{\text{Función de Kernel}}$$

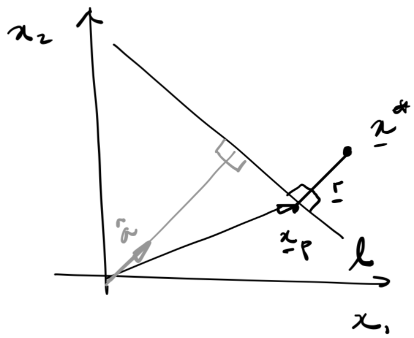
Muy utilizados:

Polinomio grado n : $(1 + \langle x_i, x \rangle)^n$

Radial Basis : $\exp(-\|x_i - x\|^2 / c)$

Neural Network : $\tanh(k_1 \langle x_i, x \rangle + k_2)$

Demo de 7



I Sean dos puntos $\underline{x}_A$ y $\underline{x}_B$ que pertenecen a ℓ, entonces:

$$
\begin{array}{rl}
 & a_1 x_{A1} + a_2 x_{A2} + a_0 = 0 \\
- & a_1 x_{B1} + a_2 x_{B2} + a_0 = 0 \\
\hline
 & [a_1\ a_2](\underline{x}_A - \underline{x}_B) = 0
\end{array}
$$

$\Rightarrow$ $\underline{a} = [a_1\ a_2]^T$ es ortogonal a ℓ.

II Se define:

$\underline{x}_p$: vector desde el origen a punto de ℓ de donde nace la línea perpendicular a $\underline{x}^*$

$\underline{r}$: vector de $\underline{x}_p$ a $\underline{x}^*$

$$\underline{x}^* = \underline{x}_p + \underline{r}$$

$$\underline{r} = \underbrace{r}_{\text{módulo de } \underline{r}}\ \underbrace{\hat{\underline{a}}}_{\text{vector unitario}}$$

$$\hat{\underline{a}} = \underline{a} / \|\underline{a}\|$$

III

$$g(\underline{x}^*) = a_1(x_{p1} + r_1) + a_2(x_{p2} + r_2) + a_0$$

$$= \underbrace{a_1 x_{p1} + a_2 x_{p2} + a_0}_{\cancel{g(\underline{x}_p)}\ 0} + \underbrace{a_1 r_1 + a_2 r_2}_{\underline{a}^T \underline{r}}$$

$$g(\underline{x}^*) = \underline{a}^T \underline{r} = r \|\underline{a}\| \cancel{\cos\theta}\ 1$$

porque son colineales (puede ser −1 de ahí que $|g(\underline{x}^*)| = r\|\underline{a}\|$)

$$\Rightarrow \boxed{r = |g(\underline{x}^*)| / \|\underline{a}\|}$$