

Máquinas de Vector Soporte

Dados $\vec{w} \in \mathbb{R}^p$ y $b \in \mathbb{R}$, fijos, se define una aplicación lineal $f : \mathbb{R}^p \rightarrow \mathbb{R}$ como:

$$f(\vec{x}) = \langle \vec{w}, \vec{x} \rangle + b = \vec{w}' \vec{x} + b = \sum_{j=1}^p w_j x_j + b$$

Se llama hiperplano, π , al subconjunto de puntos $\mathbb{R}^p$ que verifican:

$$\pi = \{ \vec{x} | f(\vec{x}) = 0 \}$$

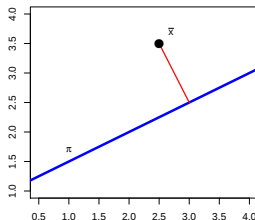
Cualquier hiperplano de $\mathbb{R}^2$ define una recta

Máquinas de Vector Soporte

El vector $\vec{w}$ es ortogonal a todos los puntos del hiperplano

La distancia entre un hiperplano π y un punto $\vec{x} \in \mathbb{R}^p$ viene dada por la expresión:

$$\text{dist}(\pi, \vec{x}) = \frac{f(\vec{x})}{\|\vec{w}\|} \quad \text{con} \quad \|\vec{w}\| = \sqrt{\sum_{j=1}^p w_j^2}$$



Conjunto linealmente separable

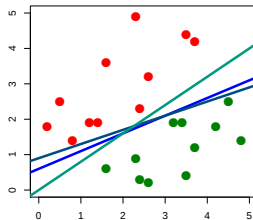
Un conjunto de puntos $(\vec{x}_i, y_i)_{i=1}^n$ con $\vec{x}_i \in \mathbb{R}^p$, $y_i \in \{-1, +1\}$, es linealmente separable si existe un hiperplano π tal que:

- ▶ $f(\vec{x}_i) = (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 0$ si $y_i = +1$
- ▶ $f(\vec{x}_i) = (\langle \vec{w}, \vec{x}_i \rangle + b) \leq 0$ si $y_i = -1$

Esas dos condiciones son equivalentes a:

$$y_i f(\vec{x}_i) = y_i (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 0 \quad i = 1, \dots, n$$

Máquinas de Vector Soporte



Cuando existe un hiperplano de separación un criterio razonable para clasificar un punto $\vec{x}$ es:

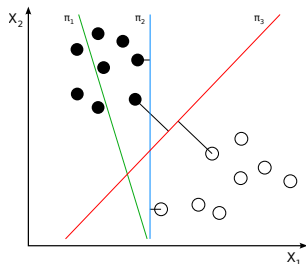
$$G(\vec{x}) = \text{signo}(f(\vec{x}))$$

Margen de un hiperplano de separación

Dado un conjunto de puntos $(\vec{x}_i)_{i=1}^n$ con $\vec{x}_i \in \mathbb{R}^p$, y un hiperplano de separación $\pi = \{\vec{x} \mid \vec{w}' \cdot \vec{x} + b = 0\}$ se define su margen como el valor M tal que

$$M = \min_{i=1, \dots, n} \text{dist}(\vec{x}_i, \pi)$$

- ▶ π_1 no es hiperplano de separación
- ▶ π_2 es hiperplano de separación con margen pequeño
- ▶ π_3 es el hiperplano de separación con margen máximo



Hiperplano óptimo

Es el hiperplano de separación con margen máximo.

La búsqueda de un hiperplano óptimo se puede plantear de la siguiente forma:

Maximizar M

$$y_i \cdot (\langle \vec{w}, \vec{x} \rangle + b) \geq M$$

Imponiendo la condición

$$M \cdot \|\vec{w}\| = 1 \iff M = \frac{1}{\|\vec{w}\|}$$

$$\begin{array}{ccc} \text{Maximizar } \frac{1}{\|\vec{w}\|} & \iff & \text{Minimizar } \|\vec{w}\| \\ y_i [\langle \vec{w}, \vec{x} \rangle + b] \geq 1 & & y_i [\langle \vec{w}, \vec{x} \rangle + b] - 1 \geq 10 \end{array}$$

Máquinas de Vector Soporte

- ▶ Problema de optimización cuadrática con restricciones lineales.
- ▶ La función lagrangiana es

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2} \langle \vec{w}, \vec{w} \rangle - \sum_{i=1}^n \alpha_i [y_i (\langle \vec{w}, \vec{x}_i \rangle + b) - 1]$$

- ▶ Aplicando las condiciones de Karush-Kuhn-Tucker se obtiene

$$\frac{\partial L}{\partial \vec{w}} = 0 \iff \vec{w}^* = \sum_{i=1}^n \alpha_i \cdot y_i \cdot \vec{x}_i \quad i = 1, \dots, n$$

$$\frac{\partial L}{\partial b} = 0 \iff \sum_{i=1}^n \alpha_i \cdot y_i = 0 \quad i = 1, \dots, n$$

$$\alpha_i [1 - y_i (\langle \vec{w}, \vec{x}_i \rangle + b^*)] = 0$$

Se puede plantear el problema inicial en su forma dual:

Maximizar $L(\vec{\alpha})$

$$\sum_{i=1}^n \alpha_i \cdot y_i = 0 \quad \alpha_i \geq 0 \quad i = 1, \dots, n$$

$$\alpha_i - \alpha_i \cdot y_i \cdot \langle \vec{w}^*, \vec{x}_i \rangle + b^* = 0$$

Una vez calculado α^* , se obtiene

$$f(\vec{x}) = \sum_{i=1}^n \alpha_i^* \cdot y_i \cdot \langle \vec{x}, \vec{x}_i \rangle + b^*$$

Máquinas de Vector Soporte

Si $\alpha_i > 0$,

$$1 = y_i \cdot (\langle \vec{w}^*, \vec{x}_i \rangle + b^*)$$

que corresponde con los vectores soporte

$$y_i = \langle \vec{w}^*, \vec{x}_i \rangle + b^* \iff b^* = y - \langle \vec{w}^*, \vec{x}_i \rangle$$