Aprendizaje automatizado

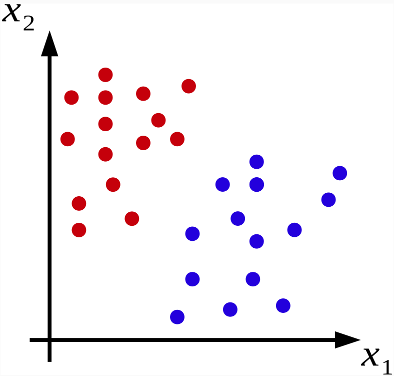
Máquinas de vectores de soporte

Gibran Fuentes-Pineda

Mayo 2023

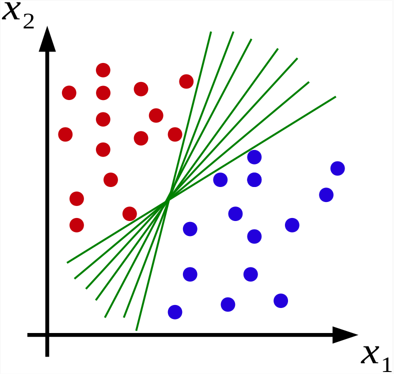
Caso 1: Clasificación binaria linealmente separable

- ¿Cómo separamos las clases?



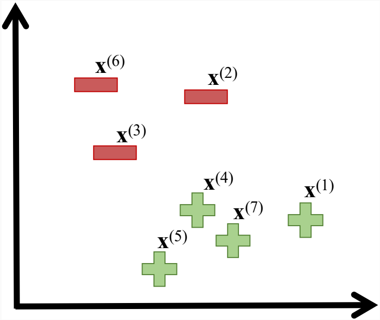
Caso 1: Clasificación binaria linealmente separable

- ¿Qué hiperplano elegimos?

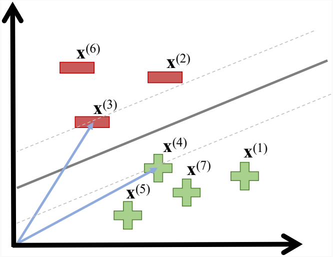


Clasificadores de margen máximo (1)

- El del margen más grande: hiperplanos paralelos a región de decisión que pasan por datos se llaman *vectores de soporte*



Clasificadores de margen máximo (2)



- Consideremos la frontera de decisión generada por $\mathbf{w}$ y una constante c. Dado un punto $\mathbf{x}^{(i)}$, la regla de decisión está definida por

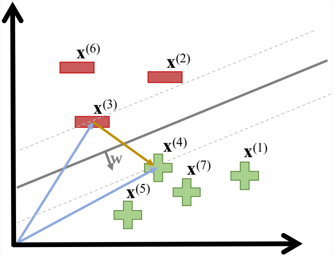
$$\mathbf{w}^\top \mathbf{x}^{(i)} \geq c$$

- La cual podemos reescribir como ($b = -c$)

$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \geq 0$$

- $\mathbf{w}$ es perpendicular a la frontera de decisión

Clasificadores de margen máximo (3)



- Restricciones

$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \geq 1, \text{ si } y^{(i)} = 1$$
$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \geq -1, \text{ si } y^{(i)} = -1$$

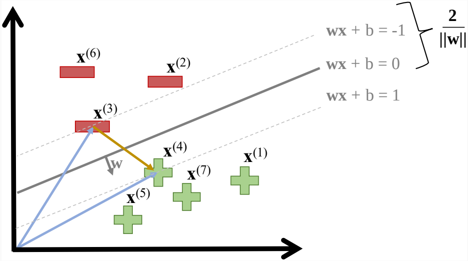
- Sea $y^{(i)} = 1$ para positivos y $y^{(i)} = -1$ para negativos, podemos reescribir

$$y^{(i)} \cdot (\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1$$
$$y^{(i)} \cdot (\mathbf{w}^\top \mathbf{x}^{(i)} + b) - 1 \geq 0$$

- Si $\mathbf{x}^{(i)}$ está exactamente en los hiperplanos de soporte

$$y^{(i)} \cdot (\mathbf{w}^\top \mathbf{x}^{(i)} + b) - 1 = 0$$

Clasificadores de margen máximo (4)



- Dados $\mathbf{x}_{pos}^{(i)}$ (ejemplo positivo) y $\mathbf{x}_{neg}^{(j)}$ (ejemplo negativo), el margen se puede calcular como

$$\frac{\mathbf{w}^\top}{||\mathbf{w}||} \cdot (\mathbf{x}_{pos}^{(i)} - \mathbf{x}_{neg}^{(j)}) = \frac{2}{||\mathbf{w}||}$$

- Queremos encontrar la $\mathbf{w}$ que maximice el ancho o de forma equivalente minimizar

$$\frac{||\mathbf{w}||}{2}$$

Optimización con restricciones

- Podemos convertir el problema a una optimización con restricciones

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{sujeto a } y^{(i)}(\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1$$

donde $y^{(i)} \in \{-1, +1\}$

Optimización con restricciones

- Podemos convertir el problema a una optimización con restricciones

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{sujeto a } y^{(i)}(\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1$$

  donde $y^{(i)} \in \{-1, +1\}$

- Optimización cuadrática con restricciones lineales y estrictamente convexa con solución única para problemas linealmente separables

Caso 2: No linealmente separables

- Penalizando suavemente clasificacionnes erróneas a través de *variables flojas*, $\xi^{(i)} \geq 0, i = 1, \ldots, n$

$$\min_{\mathbf{w},b} \left[C \sum_{i=1}^{n} \xi^{(i)} + \frac{1}{2} \|\mathbf{w}\|^2 \right]$$

$$\text{sujeto a } y^{(i)}(\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}$$

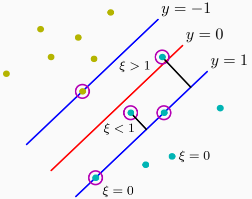


Imagen tomada de Bishop, PRML 2007

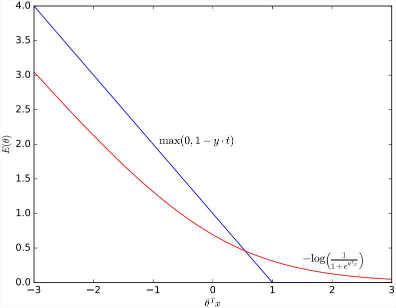
- $\xi^{(i)} = 0$, si están del lado correcto
- $\xi^{(i)} = \left| y^{(i)} - (\mathbf{w}^\top \mathbf{x}^{(i)} + b) \right|$ para otros puntos

Función de pérdida bisagra

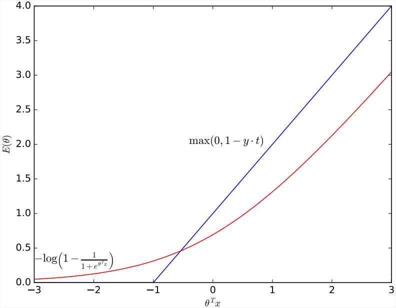
- Error respecto a parámetros está dado por función bisagra

$$B(\hat{y}, y) = \text{máx}(0, 1 - \hat{y} \cdot y)$$

$y = 1$



$y = -1$



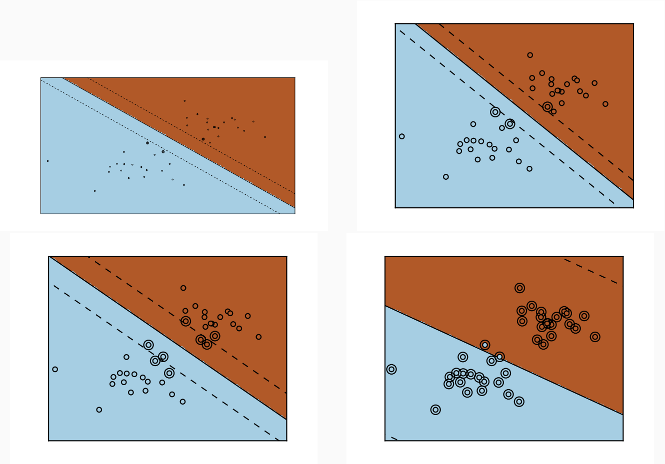
Encontrando el clasificador margen máximo

- El problema de optimización

$$\min_{\mathbf{w},b}\left[C\cdot\sum_{i=1}^{n}B(\hat{y}^{(i)},y^{(i)})+\frac{1}{2}\|\mathbf{w}\|^2\right]$$

Caso 2: No linealmente separables

- Clasificación con diferentes valores de C



Máquinas de vectores de soporte para regresión

- Extensión que preserva dispersidad en datos para regresión

Máquinas de vectores de soporte para regresión

- Extensión que preserva dispersidad en datos para regresión

- Usa función de pérdida ϵ-*sensible*

$$E(\hat{y}^{(i)}, y^{(i)}) = \begin{cases} 0 & \text{si } |\hat{y}^{(i)} - y^{(i)}| < \epsilon \\ |\hat{y}^{(i)} - y^{(i)}| - \epsilon & \text{en caso contrario} \end{cases}$$

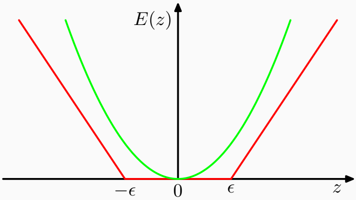


Imagen tomada de Bishop, PRML 2006

Problema de optimización para regresión

- Se busca resolver

$$\min_{\mathbf{w},b}\left[C\sum_{i=1}^{n}E(\hat{y}^{(i)},y^{(i)})+\frac{1}{2}\|\mathbf{w}\|^2\right]$$

- Expresado con variables flojas ξ

$$\min_{\mathbf{w},b}\left[C\sum_{i=1}^{n}(\xi^{(i)}+\hat{\xi}^{(i)})+\frac{1}{2}\|\mathbf{w}\|^2\right]$$

$$\text{sujeto a } \hat{y}^{(i)}+\epsilon+\xi^{(i)}\geq y^{(i)}$$

$$\hat{y}^{(i)}-\epsilon-\hat{\xi}^{(i)}\leq y^{(i)}$$

Tubo ϵ

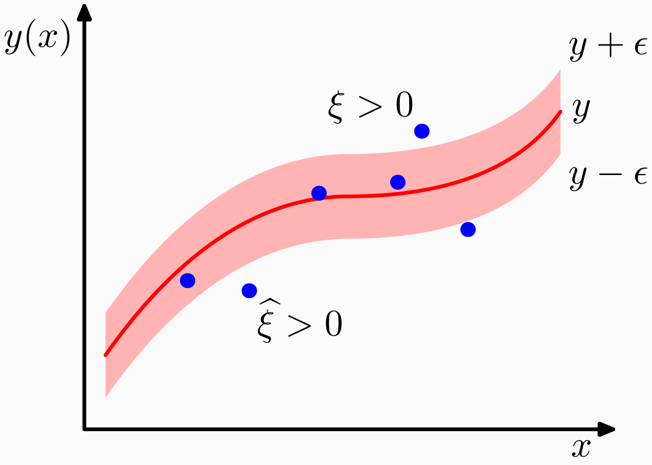


Imagen tomada de Bishop, PRML 2006