Máquinas de Vectores de Soporte

Casos No Lineales

Luis Norberto Zúñiga Morales

24 de agosto de 2022

Contenido

Funciones de Decisión No Lineales

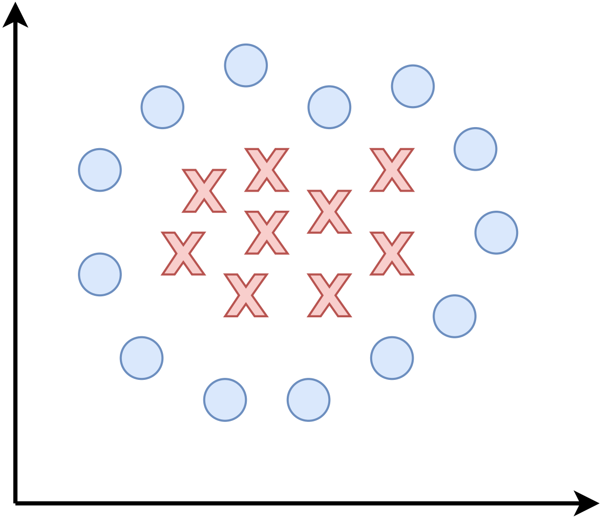


Figura: ¿Cómo separarían los datos?

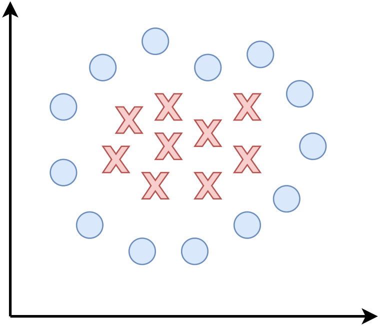
Funciones de Decisión No Lineales

- Una opción sería utilizar funciones complejas, como polinomios de alto grado. Por ejemplo, $y = 1$ si

  $$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \ldots \geq 0$$

  y 0 en cualquier otro caso.

- No es buena idea:
  - Son computacionalmente caros.
  - ¿Cómo elegir la mejor función?



MVS Caso No Lineal

Durante la construcción del clasificador se llegó a la ecuación

$$\sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i H_{ij} \alpha_j$$

donde se construyó la matriz

$$H_{ij} = y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$$

El producto punto de los vectores de entradas en la matriz anterior se puede representar como una función:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j = \boldsymbol{x}_i^T \boldsymbol{x}_j \tag{1}$$

MVS Caso No Lineal

Kernel

La función $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ es un ejemplo de una familia de funciones llamadas **kernels**, las cuales se basan en el calculo de productos puntos de los vectores de entrada del conjunto de datos.

Los kernels son funciones $x \mapsto \phi(x)$ que permiten mapear los datos a diferentes dimensiones **sin la necesidad de determinar la función** ϕ que realiza el mapeo.

MVS Caso No Lineal

Truco del Kernel

El **truco del kernel** permite atacar problemas que no son linealmente separables en el espacio en turno, y al realizar el mapeo por medio del kernel, es posible que en otro espacio sí sea separable.

MVS Caso No Lineal

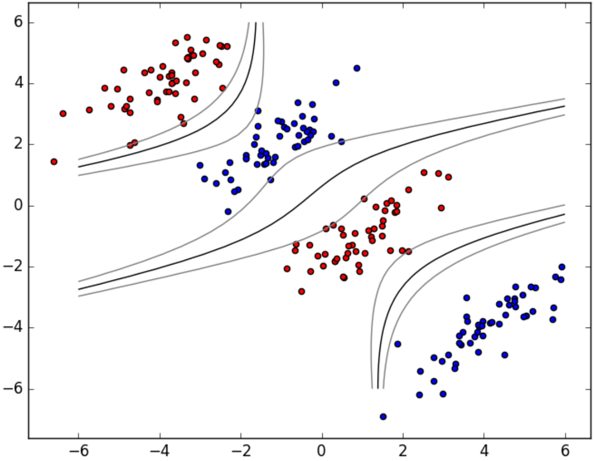


Figura: Ejemplo de un caso no linealmente separable y como el truco del kernel puede ser de utilidad.

Truco del Kernel

Los kernels más comunes en la práctica son:

- Kernel Lineal: $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$
- Kernel Polinomial: $[\gamma(\boldsymbol{x}_i \cdot \boldsymbol{x}_j) + r]^d$
- Kernel Función de Base Radial: $\exp\left(-\gamma \cdot |\boldsymbol{x}_i - \boldsymbol{x}_j|^2\right)$
- Kernel Sigmoide: $\tanh\left(\boldsymbol{x}_i \cdot \boldsymbol{x}_j + r\right)$

donde $\gamma > 0$ y $r, d \in \mathbb{R}$.

Truco del Kernel

Expresado en la formulación del clasificador en la ecuación del primal:

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{L}\xi_i \\
\text{sujeto a} \quad & y_i(\boldsymbol{w}^T \cdot \phi(\boldsymbol{x_i}) + b) \geq 1 - \xi_i \quad \forall i
\end{aligned}
\tag{2}
$$

Y su dual en la ecuación

$$
\begin{aligned}
\max_{\alpha} \quad & \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{H}_k \boldsymbol{\alpha} \\
\text{sujeto a} \quad & 0 \leq \alpha_i \leq C \ \ \forall_i \\
& \sum_{i=1}^{L} \alpha_i y_i = 0
\end{aligned} \tag{3}
$$

donde

$$
\boldsymbol{H}_k = y_i y_j k(\boldsymbol{x_i}, \boldsymbol{x_j}). \tag{4}
$$

Truco del Kernel

1. Elegir de antemano cual es el kernel que se aplicará en la MVS y la función de mapeo $\phi(\boldsymbol{x})$. En práctica, el kernel de función de base radial funciona mejor.

Truco del Kernel

1. Elegir de antemano cual es el kernel que se aplicará en la MVS y la función de mapeo $\phi(\boldsymbol{x})$. En práctica, el kernel de función de base radial funciona mejor.
2. Crear $\boldsymbol{H}_k$, donde $H_{ij} = y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Truco del Kernel

1. Elegir de antemano cual es el kernel que se aplicará en la MVS y la función de mapeo $\phi(\boldsymbol{x})$. En práctica, el kernel de función de base radial funciona mejor.
2. Crear $\boldsymbol{H}_k$, donde $H_{ij} = y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
3. Elegir el valor del parámetro C, el cual permitirá penalizar clasificaciones erróneas.

Truco del Kernel

4. Encontrar las α_i que maximicen

$$\sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}_k \boldsymbol{\alpha}$$

sujeto a las restricciones

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{L} \alpha_i y_i = 0.$$

mediante un programa para resolver problemas de optimización cuadrática.

Truco del Kernel

4. Encontrar las α_i que maximicen

$$\sum_{i=1}^{L} \alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{H}_k \boldsymbol{\alpha}$$

   sujeto a las restricciones

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{L} \alpha_i y_i = 0.$$

   mediante un programa para resolver problemas de optimización cuadrática.

5. Calcular $\boldsymbol{w} = \sum_{i=1}^{L} \alpha_i y_i \phi(\boldsymbol{x_i})$.

Truco del Kernel

6. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.

Truco del Kernel

6. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.
7. Calcular el valor de b mediante la ecuación

$$b = \frac{1}{N_S} \sum_{s \in S} (y_s - \sum_{k \in S} \alpha_k y_k k(\boldsymbol{x}_k, \boldsymbol{x}_s)).$$

Truco del Kernel

6. Determinar el conjunto de vectores de soporte S mediante la identificación de los índices i tales que $0 \leq \alpha_i \leq C$.
7. Calcular el valor de b mediante la ecuación

$$b = \frac{1}{N_S} \sum_{s \in S} (y_s - \sum_{k \in S} \alpha_k y_k k(\boldsymbol{x}_k, \boldsymbol{x}_s)).$$

8. Cada elemento del conjunto de prueba x_t se clasifica evaluando

$$y_t = sgn(\boldsymbol{w}^T \cdot \phi(\boldsymbol{x}_t) + b).$$

Consideraciones sobre el Truco del Kernel

1. Usar el truco del kernel puede generar sobreajuste si se aplica en conjuntos de datos pequeños. Una solución es usar kernels no tan complejos.
2. Un kernel debe seguir ciertas reglas, las cuales no se cubren en este curso.

Tarea

1. Investigar sobre el kernel de la función de Bessel.
2. Investigar sobre el kernel radial ANOVA.
3. Investigar sobre el kernel gaussiano.