1. Máquinas de Soporte vectorial

1.1 Hiperplano

Definición. Un hiperplano es un espacio n-dimensional es una subvariedad lineal de dimensión n-1 que divide el espacio en 2 mitades, se usa para separar los datos en diferentes clases

Formula. $w \cdot x + b = 0$

- w es el vector de pesos, determina la orientación
- x es el vector de características de un pto. en el espacio
- b es el término de sesgo, que determina la posición

Explicación. En un espacio bidimensional, un hiperplano es una línea. En un espacio tridimensional un hiperplano es un plano que no estamos interesados en encontrar el hiperplano que no solo separa las 2 clases de datos, sino que lo hace con el mayor margen posible.

1.2. Fronteras de decisión con margen

Definición. Son hiperplanos paralelos al hiperplano de separación que están a la distancia del margen máximo de los puntos más cercanos de cada clase. Ayudan a maximizar la distancia entre clases

Fórmula. $w \cdot x + b = \pm 1$

- w es el vector de pesos
- x es el vector de características de un pto en el espacio
- b es el término de sesgo

Explicación. Estos hiperplanos paralelos son cruciales para la definición del margen. La distancia entre estos hiperplanos se maximiza para mejorar la robustez del clasificador. Si los datos están correctamente clasificados deberían estar fuera o justo sobre estas fronteras de decisión.

1.3 Vectores de soporte

Definición. Los vectores de soporte son los puntos de datos más cercanos al hiperplano de decisión. Estos puntos determinan las fronteras del margen y son cruciales para definir el hiperplano óptimo. Solo estos puntos afectan la posición y orientación del hiperplano.

Fórmula. $y_i(w \cdot x_i + b) = 1$

- y_i es la etiqueta de clase (+1 o −1)
- w es el vector de pesos
- x_i es el vector de características de los vectores de soporte
- b es el término de sesgo

Explicación. Son fundamentales porque cualquier cambio en ellos afectaría el hiperplano óptimo. Los puntos de datos que están más lejos del margen no afectan el hiperplano, y por lo tanto, no son vectores de soporte.

1.4 Longitud del margen

Definición. La longitud del margen es la distancia entre las dos fronteras de decisión. Un margen mayor generalmente mejora la capacidad del modelo para generalizar a datos no visto

Fórmula. $Margen = \frac{2}{||w||}$

- $||w||$ es la norma del vector de pesos w

Explicación. El objetivo del SVM es maximizar este margen. La maximización del margen conduce a un clasificador más robusto, menos propenso a errores de clasificación en nuevos datos.

1.5 Longitud o distancia del margen para los vectores de soporte

Fórmula. $\frac{1}{||w||}$

- $||w||$ es la norma del vector de pesos w

Explicación. Esta fórmula muestra que la distancia desde un punto de datos al hiperplano de decisión es inversamente proporcional a la norma del vector de pesos. Por lo tanto, para maximizar la distancia, debemos minimizar la norma del vector de pesos, sujeto a las restricciones que los datos estén correctamente clasificados.

1.6 Optimización con los multiplicadores de Lagrange.

Definición. Los multiplicadores de Lagrange se utilizan para convertir un problema de optimización con restricciones en un problema más manejable. En SVM se maximizan los restricciones

Explicación.

Se origina con la fórmula

$$\min_{w,b} \frac{1}{2} ||w||^2$$

sujeto a

$$y_i(w \cdot x_i + b) \geq 1$$

Se convierte en el problema dual usando los multiplicadores de Lagrange α:

$$L(w,b,\alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^{n} \alpha_i [y_i(w \cdot x_i + b) - 1]$$

El objetivo es maximizar la función dual:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

sujeto a

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

1.4, ¿Cuándo es necesario aplicar el truco kernel? ¿En que consiste dicho truco?

Necesidad. El truco de kernel se aplica cuado los datos no son linealmente separables en el espacio original de características. Permite transformar los datos a un espacio de mayor dimensión donde puedan ser separables linealmente

Consiste en. Usar una función kernel $K(x_i, x_j)$ que computa el producto punto en un espacio de características de mayor dimension sin necesidad de calculao explícitamente la transformación

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Tipos de kernels comunes

- lineal $K(x_i, x_j) = x_i \cdot x_j$
- polinomial $K(x_i, x_j) = (x_i \cdot x_j + c)^d$
- gaussiano $K(x_i, x_j) = ex(-\gamma \|x_i - x_j\|^2)$
- sigmoide $K(x_i, x_j) = \tanh(k x_i \cdot x_j + \theta)$

1. ¿Cuál es la finalidad y objetivo de este algoritmo de Maquinas de soporte vectorial

Objetivo. La finalidad de las máquinas de soporte vectorial es encontrar el hiperplano que maximiza el margen entre dos clases para mejorar la precision de clasificacion y la capacidad de generalizacion. Al centrarse en los vectores de soporte, SVM ignora los puntos de datos que no influyen en la decision final, lo que lo hace eficiente y robusto frente a lo sobreajuste.