

Máquinas de vectores de soporte y selección de modelos

Deep Learning y Series de tiempo



Marco Teran
Universidad Sergio Arboleda

2023

Contenido

- 1 Introducción
- 2 Definición de SVM
 - Características importantes de SVM
- 3 Matemáticas de las SVM
- 4 Recomendaciones para utilizar SVM

Introducción

Introducción a SVM

- SVM es una técnica de aprendizaje automático utilizada para resolver problemas de clasificación y regresión.
- SVM busca encontrar la mejor separación posible entre las clases de datos mediante la búsqueda de un hiperplano que maximice la distancia entre las muestras de cada clase.

Introducción a SVM

- SVM puede manejar datos linealmente separables y no linealmente separables mediante el uso de trucos de kernel.
- SVM es ampliamente utilizado en diversas aplicaciones, incluyendo la detección de spam, reconocimiento de imágenes y análisis de texto.

Definición de SVM

Definición de SVM

¿Qué es SVM?

Una máquina de vectores soporte (SVM) es un modelo de aprendizaje automático supervisado que utiliza algoritmos de clasificación para problemas de clasificación de dos grupos. Tras proporcionar a un modelo SVM conjuntos de datos de entrenamiento etiquetados para cada categoría, es capaz de clasificar texto nuevo.

Definición de SVM

SVM encuentra el mejor hiperplano que separa las muestras de cada clase en un espacio de alta dimensión.

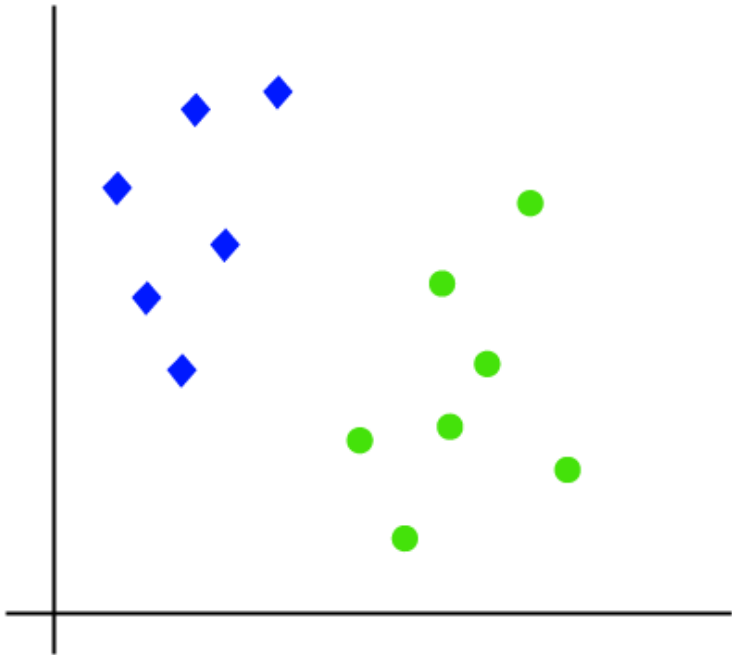
¿Cómo funciona SVM?

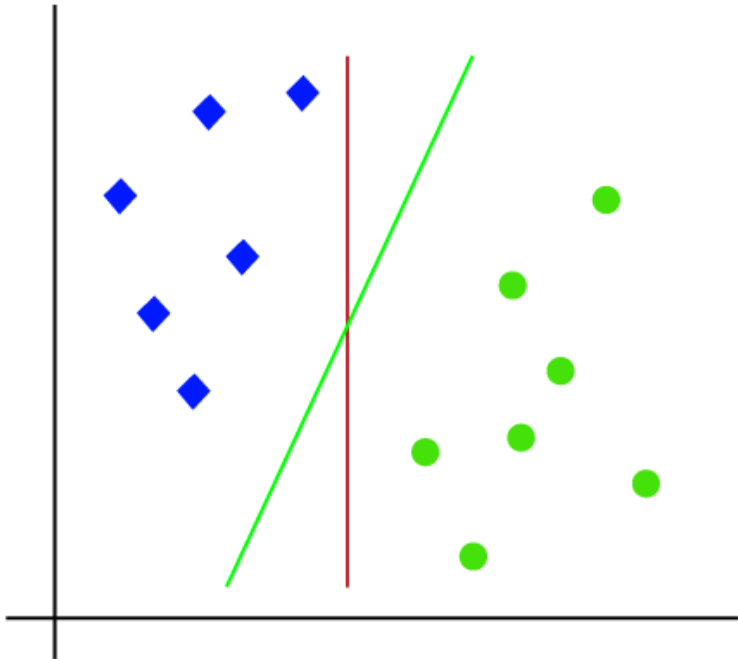
SVM busca el hiperplano que maximiza la distancia entre las muestras de cada clase, lo que se conoce como margen. Las muestras más cercanas al hiperplano son conocidas como vectores de soporte. SVM utiliza estos vectores de soporte para encontrar la mejor solución posible.

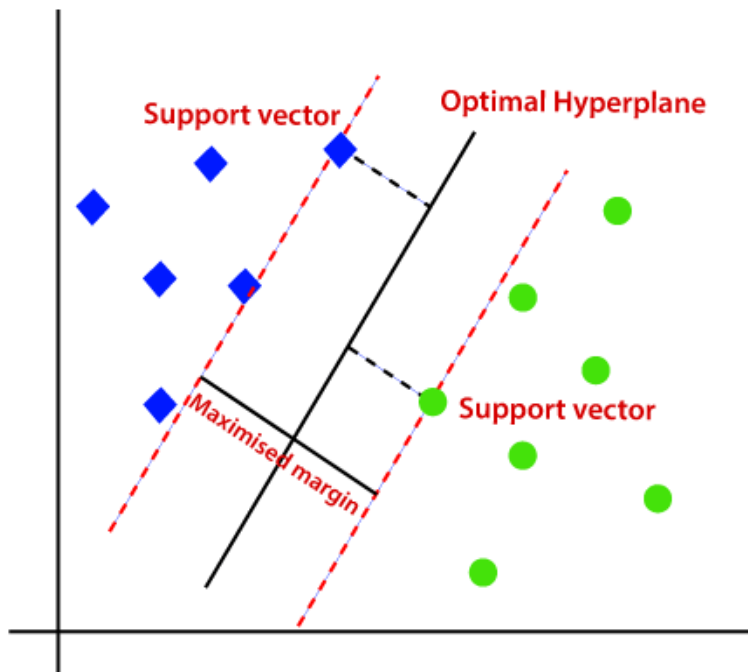
Definición de SVM

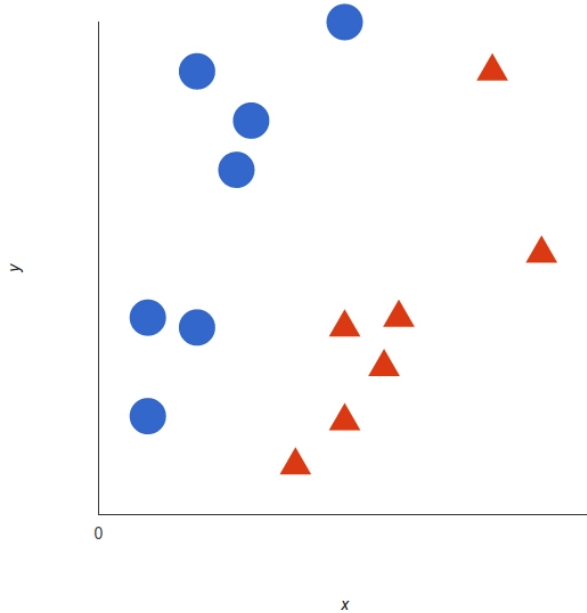
¿Cómo maneja SVM datos no linealmente separables?

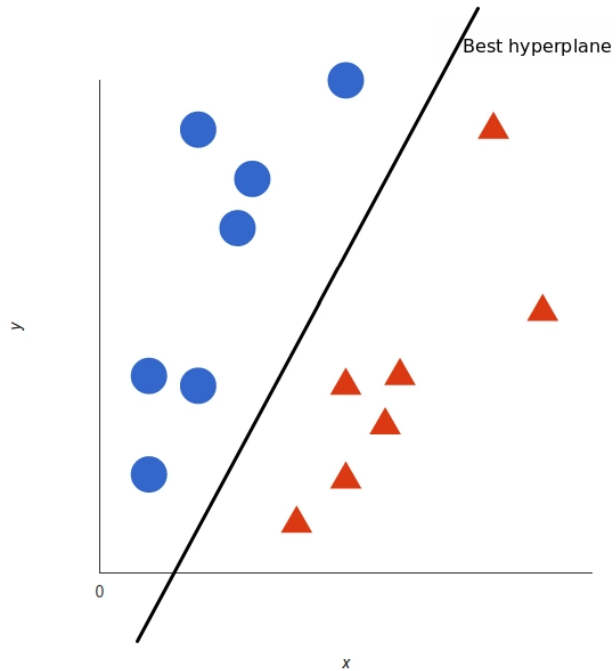
SVM utiliza trucos de kernel para transformar los datos a un espacio de características de mayor dimensión en el que los datos son más fáciles de separar linealmente. Los trucos de kernel más comunes son el kernel lineal, el kernel polinómico y el kernel RBF.

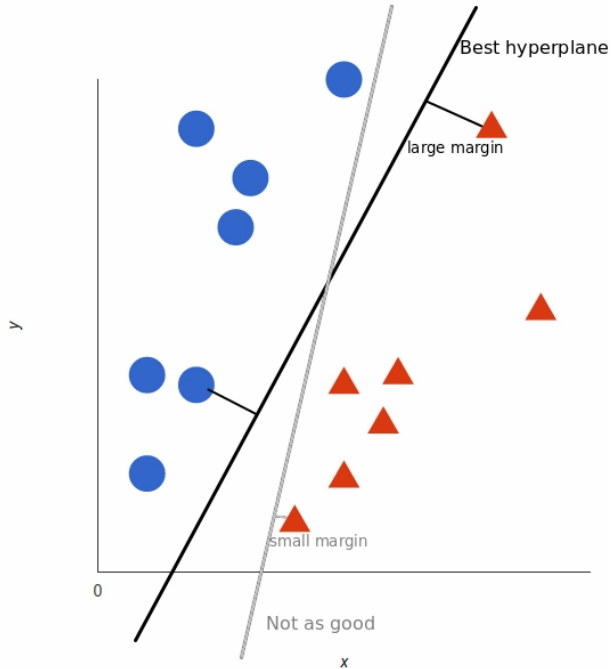


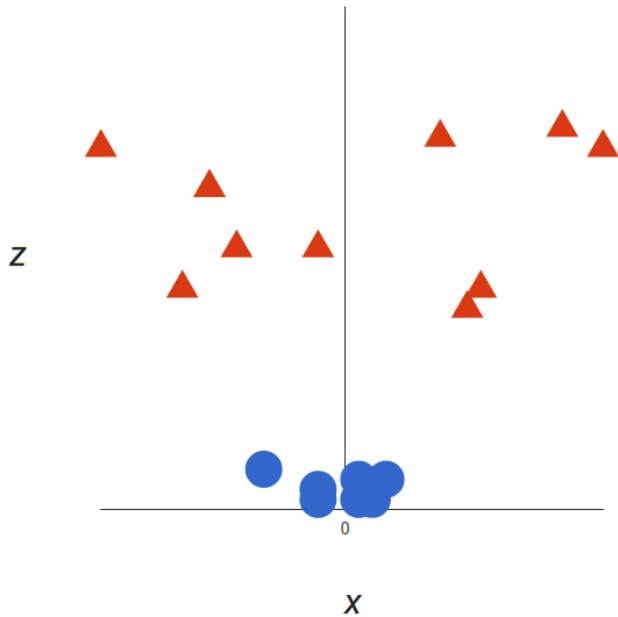


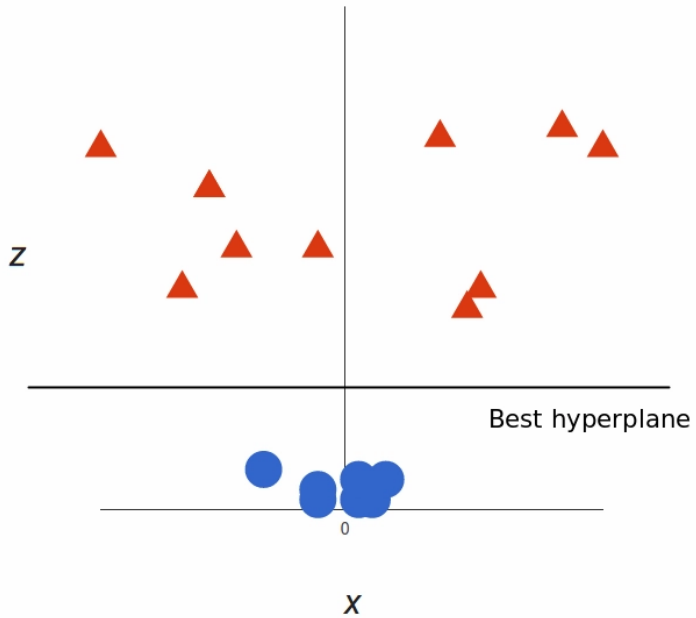


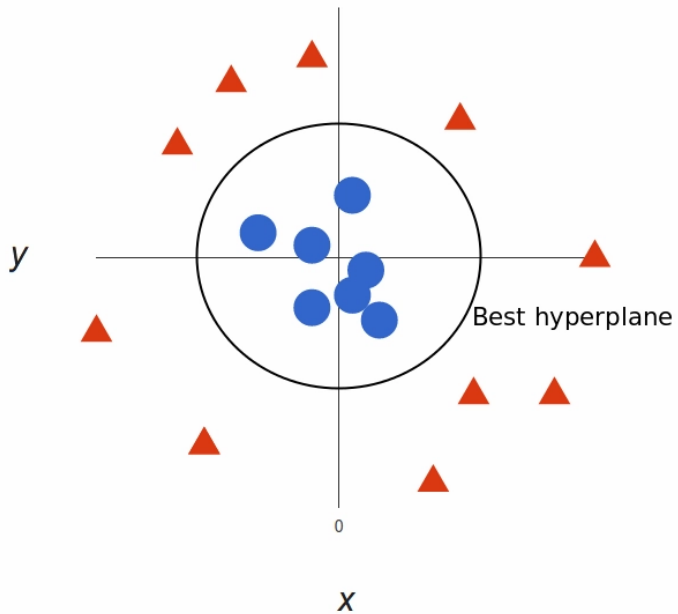












Hiperplano y Vectores de Soporte en el algoritmo SVM

Hiperplano

- La mejor línea o límite de decisión que ayuda a clasificar los puntos de datos en el espacio n -dimensional
- La dimensión del hiperplano depende de las características presentes en el conjunto de datos.

Vectores de Soporte:

- Son los puntos o vectores de datos más cercanos al hiperplano y que afectan la posición del mismo
- Estos vectores soportan al hiperplano, de ahí su nombre "Vectores de Soporte".

Funciones Kernel en SVM

- La función kernel permite transformar el conjunto de datos original en otro espacio de mayor dimensión, permitiendo separar las clases de forma más efectiva.
- Los kernels más comunes son el lineal, polinómico, RBF y sigmoideal, cada uno con sus propias ventajas y desventajas dependiendo del conjunto de datos.

Características importantes de SVM

Características importantes de SVM

- SVM es un algoritmo de aprendizaje supervisado utilizado para clasificación y regresión.
- SVM busca el mejor hiperplano que separa las muestras de cada clase mediante la maximización del margen entre las muestras de cada clase.
- SVM puede manejar datos no linealmente separables mediante el uso de trucos de kernel.
- SVM es sensible a la selección de parámetros, como el parámetro de regularización C y el tipo de kernel utilizado.
- SVM es utilizado en una variedad de aplicaciones, incluyendo la detección de spam, la clasificación de imágenes y el análisis de texto.

Ventajas del algoritmo SVM

- En comparación con algoritmos más nuevos como las redes neuronales, tiene dos ventajas principales: mayor velocidad y mejor rendimiento con un número limitado de muestras (en el rango de miles).
- Esto hace que el algoritmo sea muy adecuado para problemas de clasificación de texto, donde es común tener acceso a un conjunto de datos de solo un par de miles de muestras etiquetadas.

Matemáticas de las SVM

Matemáticas de las SVM

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

Sujeto a:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, 2, \dots, n$$

Donde w es el vector de pesos y b es el término de sesgo, $|w|$ es la norma euclidiana de w y y_i es la etiqueta de clase ($y_i \in -1, 1$) para cada ejemplo de entrenamiento x_i .

Matemáticas de las SVM

El problema de optimización anterior se puede reformular utilizando la teoría de Lagrange para obtener el problema dual:

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

Sujeto a:

$$0 \leq \alpha_i \leq C \quad \forall i = 1, 2, \dots, n$$

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

Donde α_i son las variables de holgura Lagrange y C es el parámetro de regularización.

Matemáticas de las SVM

La solución óptima del problema dual se puede utilizar para calcular los pesos y el término de sesgo del modelo de SVM:

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

$$b = \frac{1}{n_{SV}} \sum_{i \in SV} (y_i - w^T x_i)$$

Donde n_{SV} es el número de vectores de soporte y SV es el conjunto de índices de los vectores de soporte. Los vectores de soporte son los ejemplos de entrenamiento más cercanos al hiperplano de separación.

Recomendaciones para utilizar SVM

Recomendaciones para utilizar SVM

- Escala tus datos antes de entrenar un modelo SVM, para evitar que las variables con rangos más grandes dominen el entrenamiento del modelo.
- Asegúrate de que tus datos sean linealmente separables antes de utilizar un kernel no lineal. Si tus datos son linealmente separables, utilizar un kernel no lineal podría aumentar la complejidad innecesariamente.
- Experimenta con diferentes kernels no lineales (como el kernel polinómico y el kernel RBF) para encontrar el mejor ajuste para tus datos.

Recomendaciones para utilizar SVM

- Ajusta el parámetro de regularización C para evitar el sobreajuste o subajuste del modelo. Si C es demasiado grande, el modelo puede sobreajustar los datos de entrenamiento, mientras que si C es demasiado pequeño, el modelo puede subajustar los datos de entrenamiento.
- Utiliza técnicas de selección de características para reducir la dimensionalidad de tus datos antes de entrenar un modelo SVM. Esto puede mejorar el rendimiento del modelo y reducir el tiempo de entrenamiento.
- Si tienes un conjunto de datos desequilibrado, considera utilizar técnicas de muestreo para equilibrar las clases antes de entrenar un modelo SVM.
- Si tienes un conjunto de datos grande, considera utilizar métodos de descenso de gradiente estocástico (SGD) para entrenar el modelo SVM. Esto puede reducir significativamente el tiempo de entrenamiento.

¡Muchas gracias por su atención!

¿Preguntas?



Contacto: Marco Teran
webpage: marcoteran.github.io/
e-mail: marco.teran@usa.edu.co

