Máquinas de Vectores de soporte/ Support Vector Machines

A viernes 23 de mayo del 2025

En la División de Ingenierías Irapuato - Salamanca

Autores: Jashiel Robles Lecona, Diego Perez Perez, Mario Díaz Ramos

# Resumen

En este reporte se describirán, definirán y mostrarán que son las Support Vector Machines y/o máquinas de vectores de soporte.

Mostrando como sería una forma gráfica de representarlas, sus componentes, una breve comparación con la inteligencia artificial, algunos casos de uso y sus ventajas y desventajas frente a otro tipo de clasificadores computacionales.

1. **Introducción**

En este reporte se plantea indagar en las máquinas de vectores de soporte / Support Vector machines (SVM), las cuáles son de breve manera clasificadores binarios, es decir, solo dan dos respuestas, si o no, esto nos permite de alguna manera poder clasificar ciertos datos en grupo e incluso lograr discernir entre distintos tipos de datos pudiendo hacer una correcta diferenciación entre estos datos.

Se explicará cómo es su funcionamiento, las partes que lo conforman en una representación gráfica, como determina que un resultado es correcto o no. Sus distintas ventajas y desventajas.

1. **Desarrollo**

¿Qué son las support vector machines?

Las máquinas de vectores de soporte (SVM) son un algoritmo de aprendizaje automático supervisado que clasifica los datos al encontrar una línea o hiperplano óptimo que maximice la distancia entre cada clase en un espacio N-dimensional.

Las SVM fueron desarrolladas en la década de 1990 por Vladimir n. Vapnik y sus colegas, publicaron este trabajo en un artículo titulado “Support Vector Method For Function Approximation, Regression Estimation, and Signal Processing” en 1995.

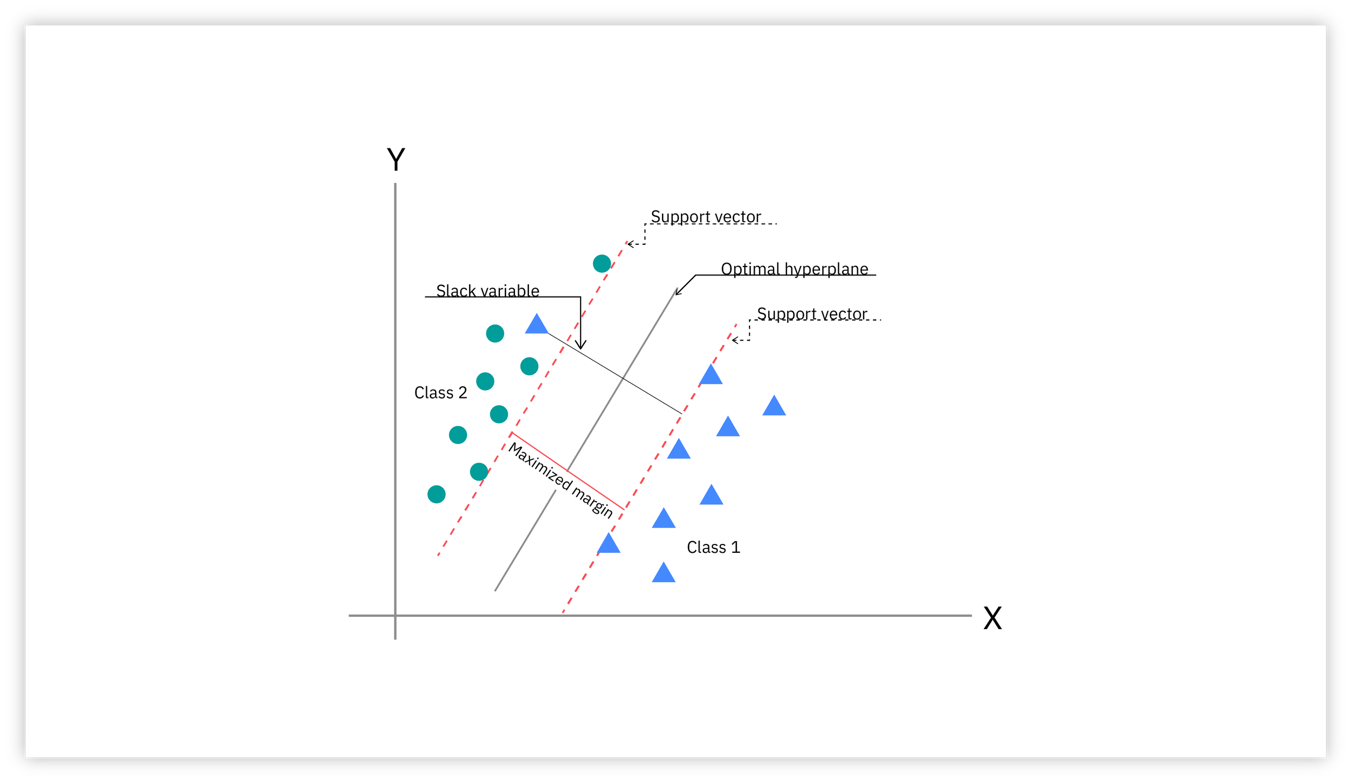
Aunque inicialmente se desarrolló como un método de clasificación binaria, su aplicación se ha extendido a problemas de clasificación múltiple y regresión. SVMs ha resultado ser uno de los mejores clasificadores para un amplio abanico de situaciones, por lo que se considera uno de los referentes dentro del ámbito de aprendizaje estadístico y machine learning.

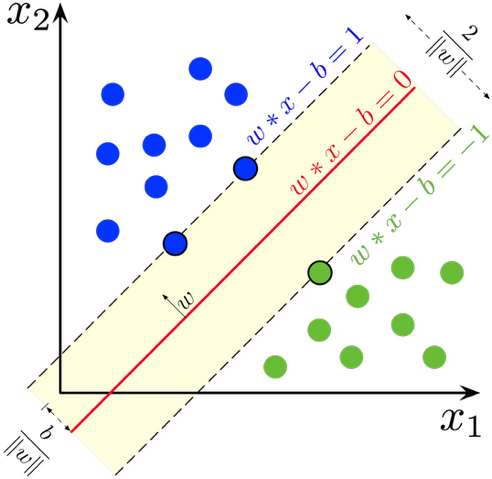
Figure 1 Representación del Hiperplano

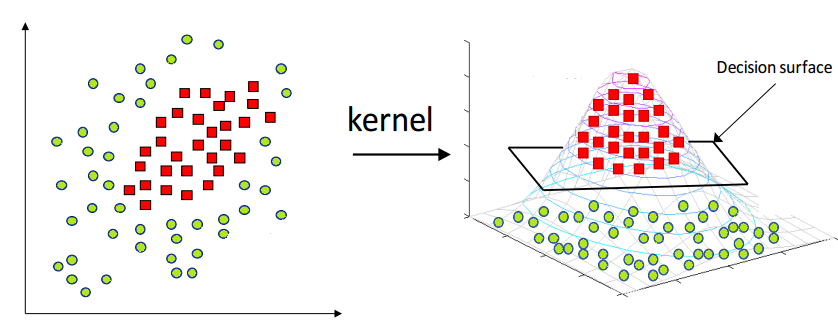
Los SVM se emplean comúnmente en problemas de clasificación. Distinguen entre dos clases encontrando el hiperplano óptimo que maximiza el margen entre los puntos de datos más cercanos de clases opuestas. El número de características en los datos de entrada determina si el hiperplano es una línea en un espacio bidimensional o un plano en un espacio n-dimensional. Dado que se pueden encontrar múltiples hiperplanos para diferenciar clases, la maximización del margen entre puntos permite al algoritmo encontrar la mejor frontera de decisión entre clases. Esto a su vez le permite generalizar bien los nuevos datos y hacer predicciones de clasificación precisas.

Las líneas adyacentes al hiperplano óptimo se conocen como vectores de soporte, ya que estos vectores atraviesan los puntos de datos que determinan el margen máximo.

Una breve analogía de cómo se componen las partes del hiperplano y de comprender mejor las matemáticas detrás de este asombroso algoritmo de machine learning.

La ecuación del hiperplano se parece a la ecuación de la recta y esto es simple, la ecuación del hiperplano es la ecuación de la recta. W \* x - b = 0

* Donde **W s**e refiere a al vector de dirección del hiperplano.
* **X** hacen referencia a nuestros datos de entrada.
* **B** es el sesgo hacía donde se dirige el plano.
* El área sombreada corresponde al margen de error que es el margen aceptable para poder generalizar los datos.

El algoritmo SVM se emplea ampliamente en el [aprendizaje automático](https://www.ibm.com/mx-es/think/topics/machine-learning) ya que puede manejar tareas de clasificación tanto lineales como no lineales. Sin embargo, cuando los datos no son separables linealmente, las funciones del kernel se utilizan para transformar el espacio multidimensional de los datos y permitir la separación lineal. A esta aplicación de las funciones del kernel (o núcleo) se le conoce como el “truco del kernel”, y la elección de funciones del kernel, como núcleos lineales, núcleos polinomiales, núcleos de función de base radial (RBF, sigla en inglés de radial basis function) o núcleos sigmoides, depende de las características de los datos y del caso de uso específico.

SVM lineales

Las SVM lineales se emplean con datos separables linealmente. Esto significa que los datos no necesitan someterse a ninguna transformación para separar los datos en diferentes clases. El límite de decisión y los vectores de soporte forman la apariencia de una calle, y el profesor Patrick Winston del MIT emplea la analogía de "[ajustar la calle más ancha posible](https://ocw.mit.edu/courses/6-034-artificial-intelligence-fall-2010/resources/mit6_034f10_svm/)"2 (enlace externo a ibm.com) para describir este problema de optimización cuadrática. Matemáticamente, este hiperplano de separación se puede representar de la siguiente manera:

Figure 2 Partes y ecuación del hiperplano

**wx + b = 0**

donde w es el vector de ponderación, x es el vector de entrada y b es el término de sesgo.

Existen dos enfoques para calcular el margen, o la distancia máxima entre clases, que son la clasificación de margen duro y la clasificación de margen blando. Si empleamos una SVM de margen duro, los puntos de datos estarán perfectamente separados fuera de los vectores de soporte, o "fuera de la calle" para seguir con la analogía del profesor Hinton. Esto se representa con la fórmula,

Figure 3 Kernel

**(wxj + b) yj ≥ a,**

y luego se maximiza el margen, que se representa como: max ɣ= a / ||w||, donde a es el margen proyectado sobre w.

La clasificación de margen suave es más flexible, ya que permite algunos errores de clasificación mediante el uso de variables de holgura (`ξ`). El hiperparámetro C ajusta el margen; un valor C mayor estrecha el margen para una clasificación errónea mínima, mientras que un valor C menor lo amplía, permitiendo más datos clasificados erróneamente 3.

SVM no lineales

Gran parte de los datos del mundo real no son linealmente separables, y ahí es donde entran en juego las SVM no lineales. Para que los datos sean linealmente separables, se aplican métodos de preprocesamiento a los datos de entrenamiento para transformarlos en un espacio de características de mayor dimensión. Dicho esto, los espacios de mayor dimensión pueden crear más complejidad al aumentar el riesgo de sobreajuste de los datos y convertirse en una carga computacional. El "truco del núcleo" ayuda a reducir parte de esa complejidad, haciendo que el cálculo sea más eficiente, y lo hace sustituyendo los cálculos del producto de punto por una función de núcleo equivalente4.

Existen varios tipos de kernel diferentes que se pueden aplicar para clasificar datos. Algunas funciones del kernel populares incluyen:

* Kernel polinómico
* Kernel de función de base radial (también conocido como kernel gaussiano o RBF)
* Kernel sigmoide

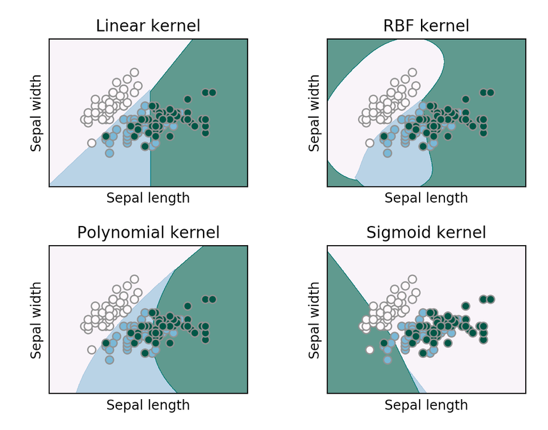


Figure 4 Tipos de Kernel

|  |  |
| --- | --- |
| Kernel | ¿Cuándo Usar? |
| Linear | Datos separados linealmente |
| Polinomial | Relaciones Cuadráticas / cúbicas |
| RFB (Gaussiano) | Patrones complejos, no lineales |
| Sigmoide | Comportamiento similar a redes neuronales |

1. ***Algunas de las ventajas de los SVM***

* **Eficientes:** Son eficientes en el entrenamiento con pocos datos.
* **Versátiles:** Tienen múltiples kernels lo que los hace multiherramientas para resolver distintos problemas.
* **Robustez:** son resistentes a los datos atípicos.
* **Buenos Generalizando**: evitan el sobreajuste por el margen máximo.

Derivado de las ventajas que se tienen, las recomendamos para poderlos usar para casos de usos específicos, en los que estes seguro de describir de manera adecuada las características que buscas discernir de las muestras de datos.

1. ***Algunas de las desventajas de los SVM***

* **Restricción a datos muy grandes:** tiene ciertas limitaciones con la escalabilidad a millones de datos.
* **Muchas características indefinidas**: Son sensibles a definir estrictamente los tipos de características.
* **Interpretabilidad:** Se tiene cierta complejidad al intentar explicar las predicciones.
* **Selección de Kernel:** debido a que se tienen distintos tipos de Kernels por lo que se puede tornar complicado elegir el kernel siendo que puede ser por experimentación la selección de este.
* **Desvalance de clases**: Se puede sesgar hacia la clase mayoritaria.

Con las desventajas mencionadas recomendamos hacer una breve tabla de comparación con los distintos modelos de algoritmos de clasificación de datos, para poder identificar cuál es el mejor para tu caso de uso específico o inclusive el tipo de kernel que puedes usar para lograr tu objetivo.

1. **Resultados o discusión.**

Derivado del contenido en el presente reporte, esta investigacipon puede resultar bastante ilustrativa para las personas que busquen enriquecer su conocimiento en temas derivados a la introducción a las SVM. Comprendiendo desde su funcionamiento, las formulas matematicas detrás de las SVM, como la del Hiperplano.

Pudiendo discernir de cuando conviene usar este algoritmo frente a otros algoritmos.

1. **Conclusiones.**

Por último, podemos concluir que las maquinas de vectores de sporte son un algoritmo simple, eficiente, elegante. Que nos permiten poder lograr una buena clasificación de los datos, minerías de datos, aplicarlos en casos de uso real como filtrar correos de spam, diagnósticos médicos, entre otros casos de uso.

Al ser algoritmos de un nivel de programación “simple” los hace ideales para usarlos en muchos problemas que necesiten una solución práctica, pudiendo resolver desde problemas simples hasta problemas complejos.