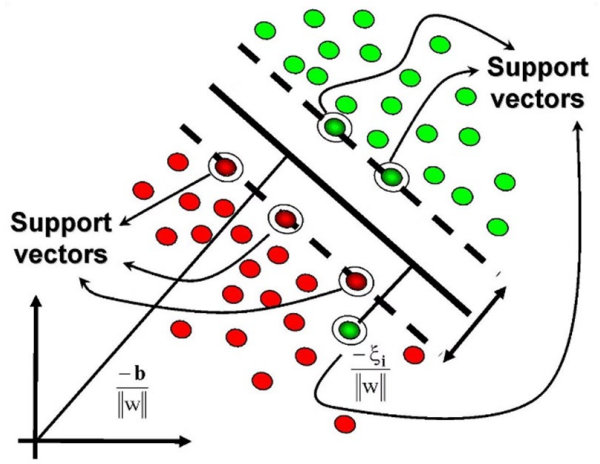


SISTEMAS INTELIGENTES

Lecture 08:Maquinas de Vectores de Soporte - SVM

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024

Introduccion

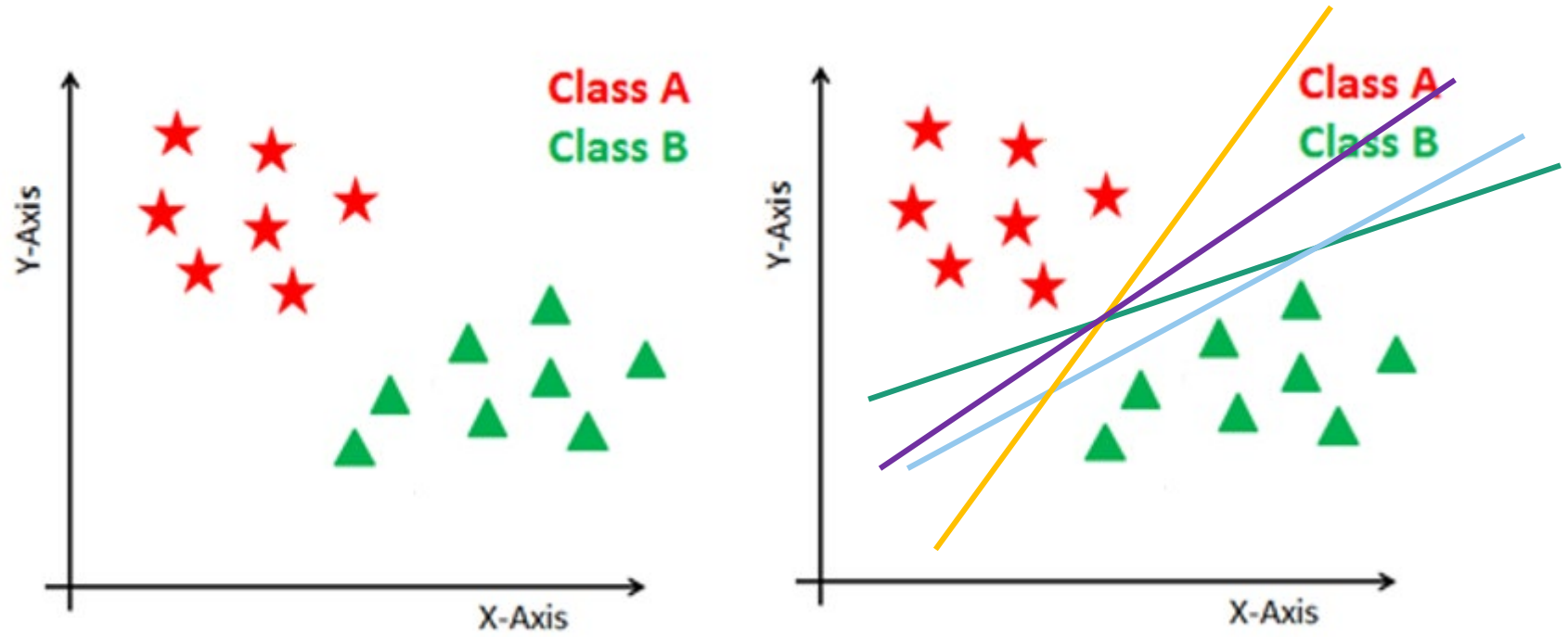
Las máquinas de vectores de soporte (SVM, por sus siglas en inglés) son algoritmos de aprendizaje automático que se utilizan para fines de clasificación y regresión. Las SVM son uno de los algoritmos de aprendizaje automático más potentes para fines de clasificación, regresión y detección de valores atípicos. Un clasificador SVM crea un modelo que asigna nuevos puntos de datos a una de las categorías dadas. Por lo tanto, puede verse como un clasificador lineal binario no probabilístico.

El algoritmo SVM original fue desarrollado por Vladimir N Vapnik y Alexey Ya. Chervonenkis en 1963. En ese momento, el algoritmo estaba en sus primeras etapas. La única posibilidad era dibujar hiperplanos para el clasificador lineal. En 1992, Bernhard E. Boser, Isabelle M Guyon y Vladimir N Vapnik sugirieron una forma de crear clasificadores no lineales aplicando el truco del núcleo (kernel trick) a los hiperplanos de margen máximo. El estándar actual fue propuesto por Corinna Cortes y Vapnik en 1993 y publicado en 1995.

Las SVM se pueden utilizar para fines de clasificación lineal. Además de realizar una clasificación lineal, las SVM pueden realizar de manera eficiente una clasificación no lineal utilizando el kernel trick. Esto nos permite mapear implícitamente las entradas en espacios de características de alta dimensión.

Como funciona?

- ✓Supongamos que nos dan estos puntos de datos de dos clases diferentes y queremos encontrar un clasificador lineal que los separe.



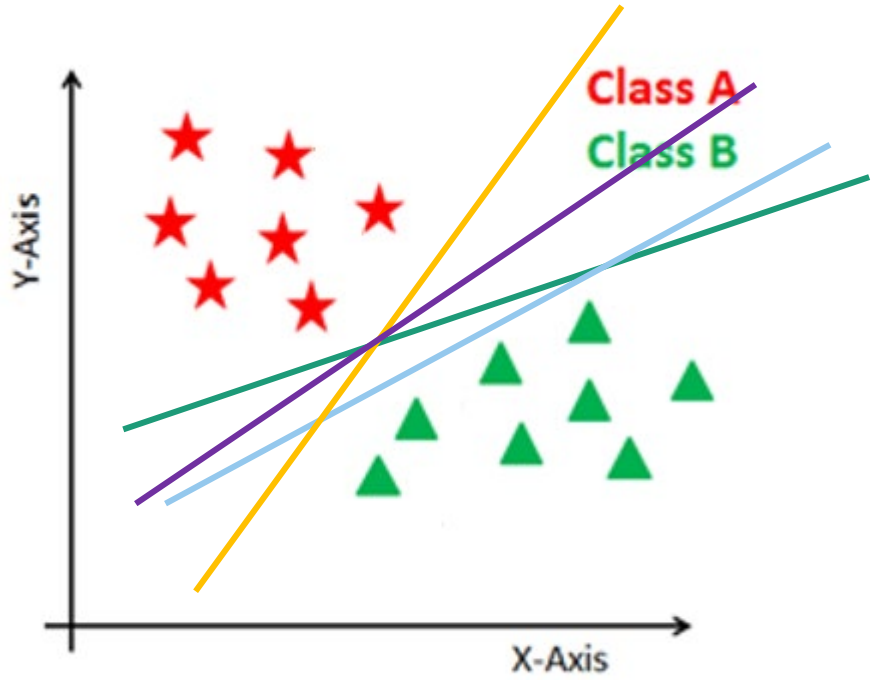
Como funciona?

El objetivo del algoritmo SVM es crear un límite o línea de decisión que pueda separar eficazmente un conjunto de datos determinado en diferentes clases.

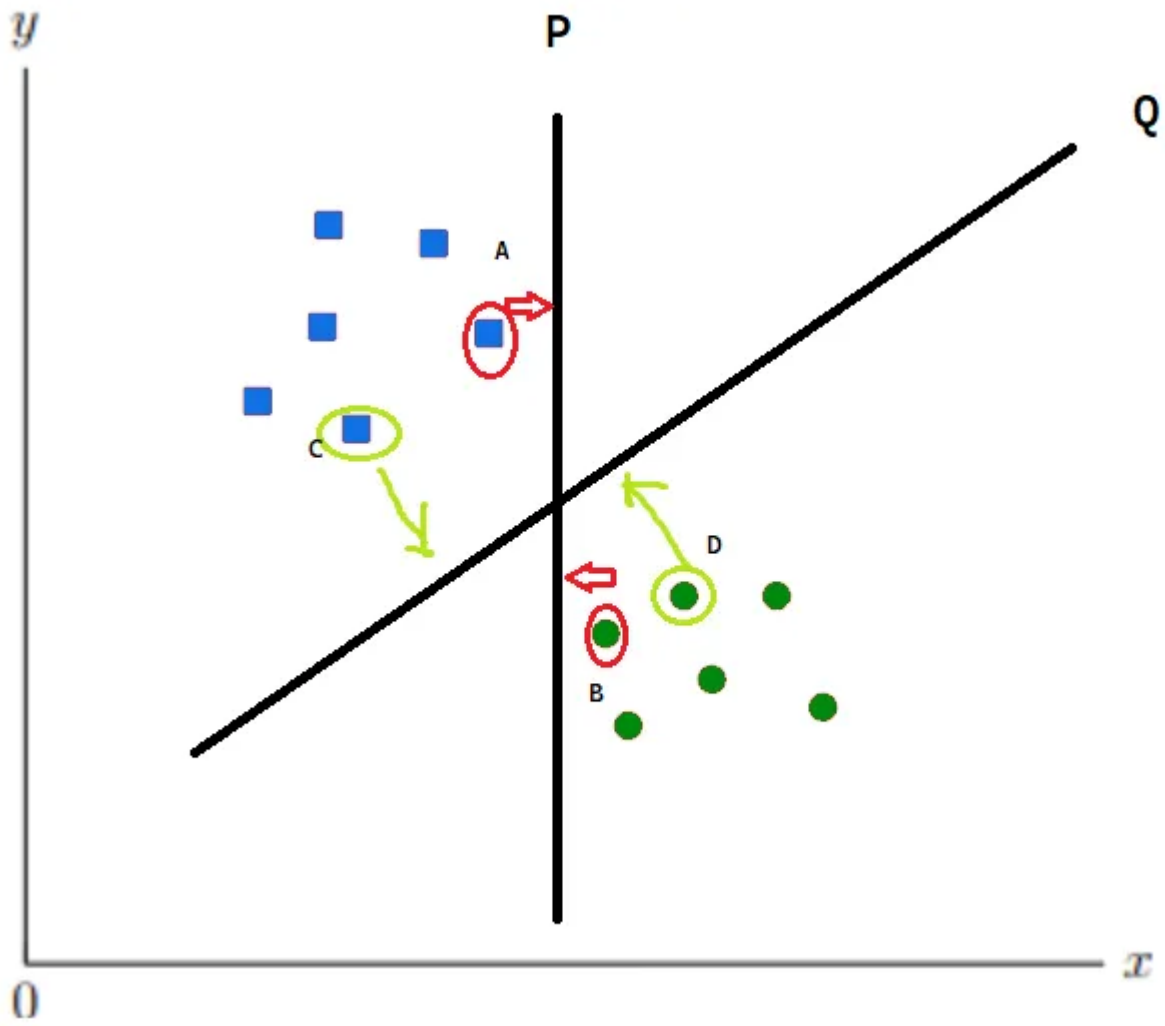
Una vez que se establece el límite de decisión, se pueden clasificar nuevos ejemplos en las clases adecuadas con relativa facilidad.

En el algoritmo SVM, este límite de decisión se conoce como hiperplano .

El desafío entonces es dibujar este hiperplano con precisión.



Como funciona?



- ✓En este gráfico, los puntos A y B son los más cercanos para este hiperplano P. Por lo tanto , en este caso, A y B se consideran vectores de soporte del hiperplano P.
- ✓y C y D son los otros puntos más cercanos al hiperplano Q.

SVM – Definiciones importantes

Vectores de soporte

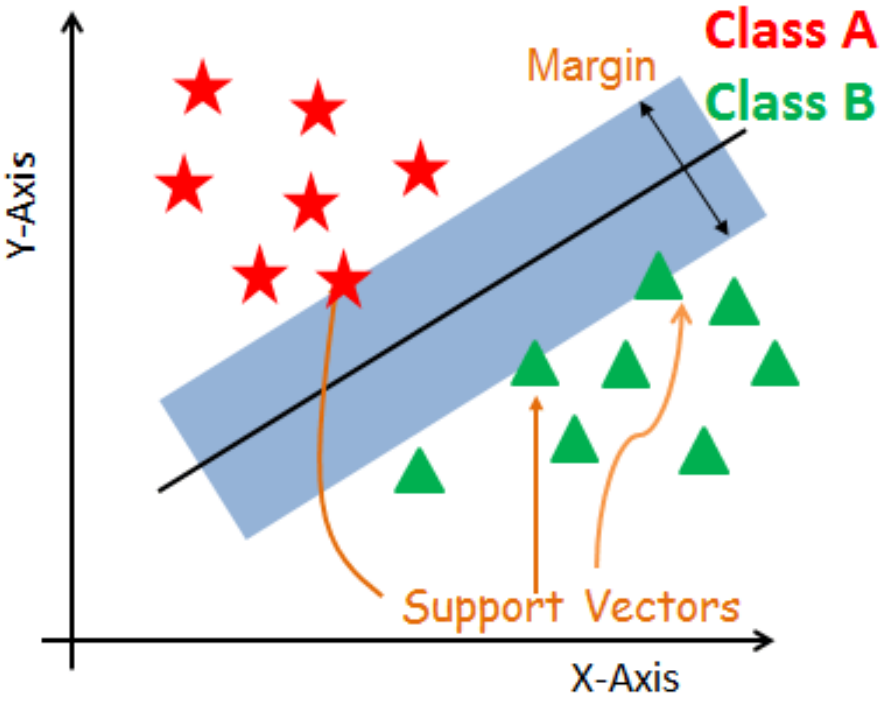
- Los vectores de soporte son los puntos de datos más próximos al hiperplano. Estos puntos definirán mejor la línea de separación calculando los márgenes. Estos puntos son más relevantes para la creación del clasificador.

Hiperplano

- Un hiperplano es un plano de decisión que separa un conjunto de objetos que pertenecen a diferentes clases.

Margen

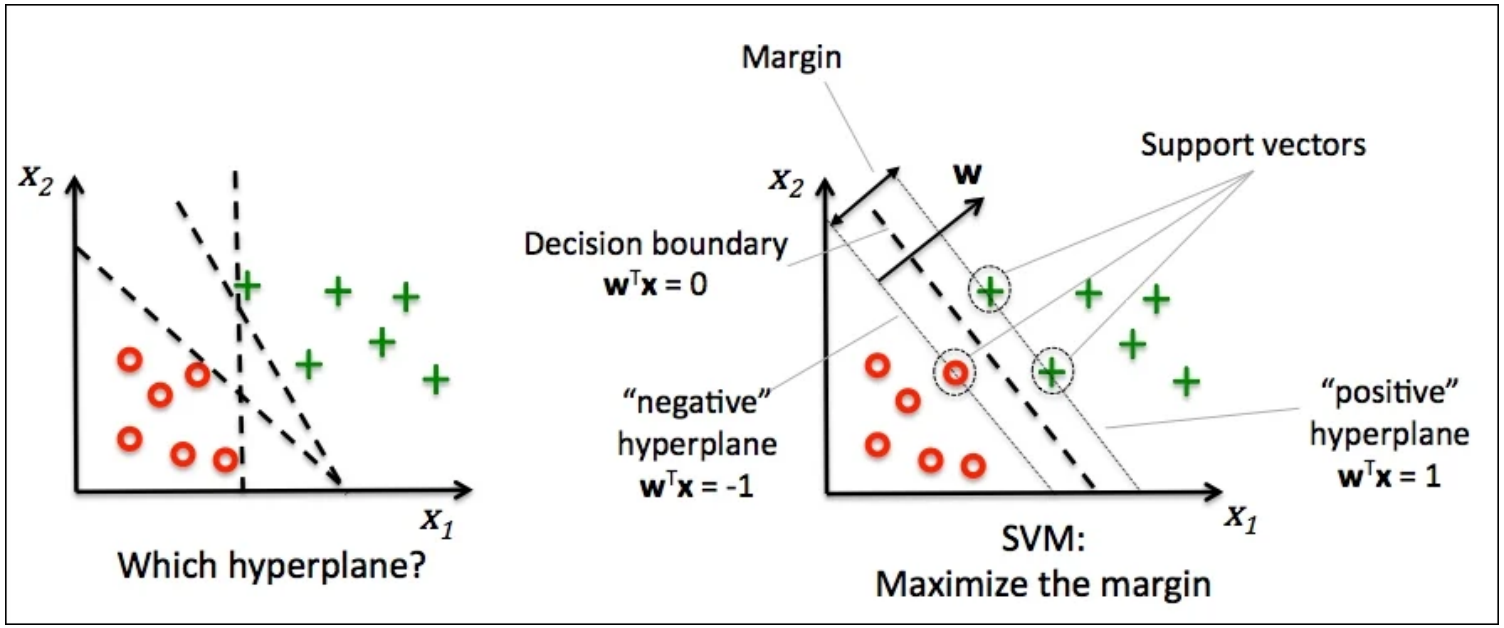
- Un margen es un espacio entre las dos líneas de los puntos de clase más cercanos. Se calcula como la distancia perpendicular de la línea a los vectores de soporte o puntos más cercanos. Si el margen es mayor entre las clases, se considera un buen margen; un margen menor es un mal margen.



SVM al detalle

- ✓ **En SVM, nuestro objetivo principal es seleccionar un hiperplano con el máximo margen posible entre los vectores de soporte en el conjunto de datos dado. SVM busca el hiperplano de máximo margen en el siguiente proceso de 2 pasos:**
  1. **Generar hiperplanos que segreguen las clases de la mejor manera posible. Hay muchos hiperplanos que podrían clasificar los datos. Debemos buscar el mejor hiperplano que represente la mayor separación, o margen, entre las dos clases.**
  2. **Por lo tanto, elegimos el hiperplano de modo que la distancia desde él hasta los vectores de soporte en cada lado sea maximizada. Si existe un hiperplano de este tipo, se lo conoce como *hiperplano de máximo margen* y el clasificador lineal que define se conoce como *clasificador de máximo margen*.**
- ✓ **El siguiente diagrama ilustra el concepto de margen máximo e hiperplano de máximo margen de manera clara.**

Hiperplano de máximo margen



SVM al detalle

- ✓Las máquinas de vector de soporte (SVM) se pueden considerar el primer modelo que utiliza un enfoque completamente distinto y que está basado en un modelado geométrico y que se resuelve mediante un problema de optimización con restricciones.
- ✓El objetivo de las redes de vectores de soporte es buscar un plano que separe las clases en feature space. Por feature space se entiende un nuevo espacio de dimensiones diferente (por lo general con un mayor número de dimensiones) al espacio original.
- ✓Debido a que este objetivo de buscar un plano que separe las clases por lo general no es posible obtener, en las SVM se proponen las siguientes modificaciones:
  1. Relajar la definición de «separar».
  2. Mejorar y enriquecer el feature space para que la separación sea posible.

SVM al detalle: Hiperplanos

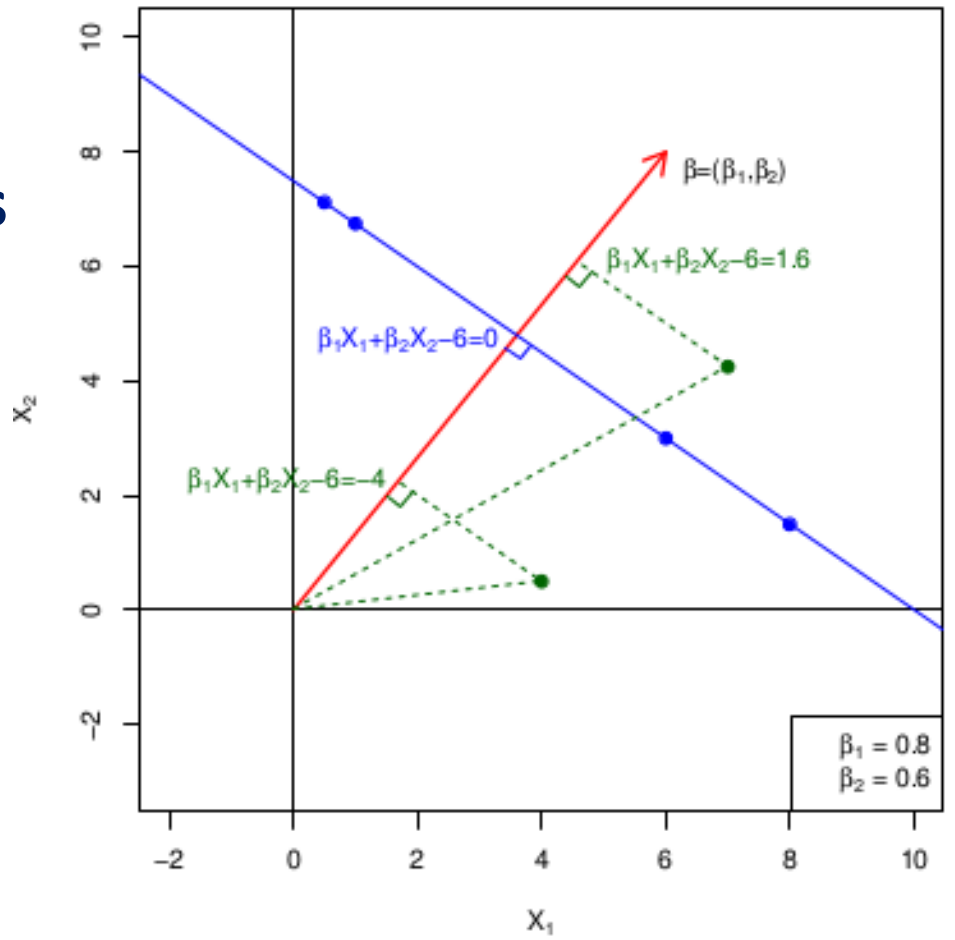
- ✓En un espacio de dos dimensiones el hiperplano es una recta. Pero, un hiperplano se puede definir para un espacio de p dimensiones. La ecuación general del hiperplano para un espacio de p dimensiones es:

$$\beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + B_p X_p = 0$$

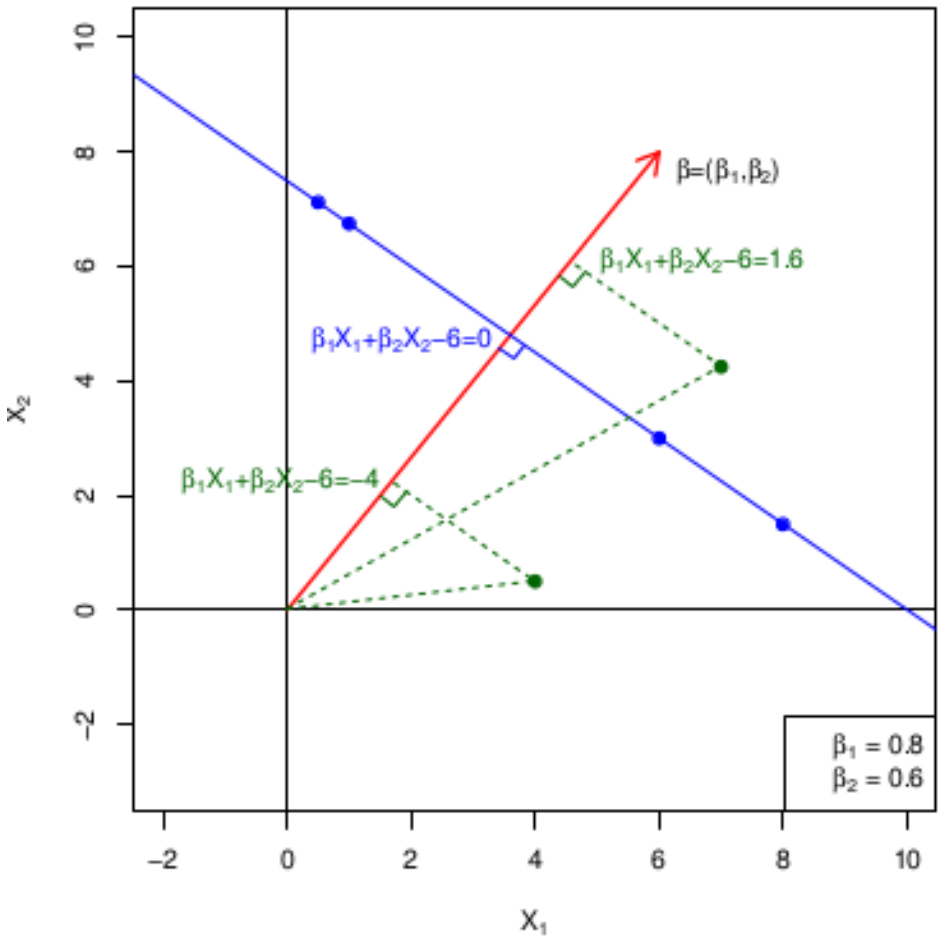
donde el vector:

$$\beta = (\beta_1, B_2, \ldots, B_p)$$

- ✓Se llama vector normal, y β es un vector unitario en el cual la suma de los cuadrados es 1. Este vector apunta a una dirección ortogonal a la superficie del hiperplano (vector de color rojo de la figura 1.).



SVM al detalle: Hiperplanos



- ✓Si proyectamos cada uno de los puntos sobre el vector normal, los puntos que caen sobre el hiperplano tienen valor 0 al proyectarse.
- ✓Los puntos por encima del hiperplano tienen un valor positivo, los puntos por debajo del hiperplano un valor negativo y además estos valores son mayores cuanto más lejanos están del hiperplano. Es decir, el valor que se obtiene al proyectar los puntos sobre el vector normal es proporcional a la distancia de los puntos al hiperplano.
- ✓Esta característica geométrica hace posible el uso de las máquinas de vector de soporte para buscar los patrones necesarios.

SVM al detalle: Hiperplanos

La cuestión es que el hiperplano busca separar dos conjuntos de puntos en dos regiones distintas. Por ejemplo, en un problema de clasificación el hiperplano puede separar aquellos puntos que corresponden a personas con un tumor determinado de aquellas personas sanas.

Sin embargo, para un conjunto de puntos determinados existen múltiples hiperplanos posibles que nos separan los puntos en dos regiones.

El objetivo de las máquinas de vector de soporte es buscar un hiperplano que separe las clases en feature space. Por feature space se entiende un nuevo espacio de dimensiones diferente al espacio original.

SVM al detalle: Maximal-margin Classifier

- ✓ En el caso de un problema de clasificación binaria, de todos los hiperplanos posibles es necesario buscar aquel que nos proporciona la mayor diferencia entre las dos clases, lo cual se traduce en la mayor distancia entre los puntos que pertenecen a una clase y a otra.
- ✓ La hipótesis que hay detrás de esto, es porque suponemos que este hiperplano será el que tendrá una mayor distancia en el conjunto de test y en las predicciones futuras.
- ✓ Esta situación se puede modelar como un problema de optimización con restricciones donde es necesario maximizar el margen y, matemáticamente, se define:

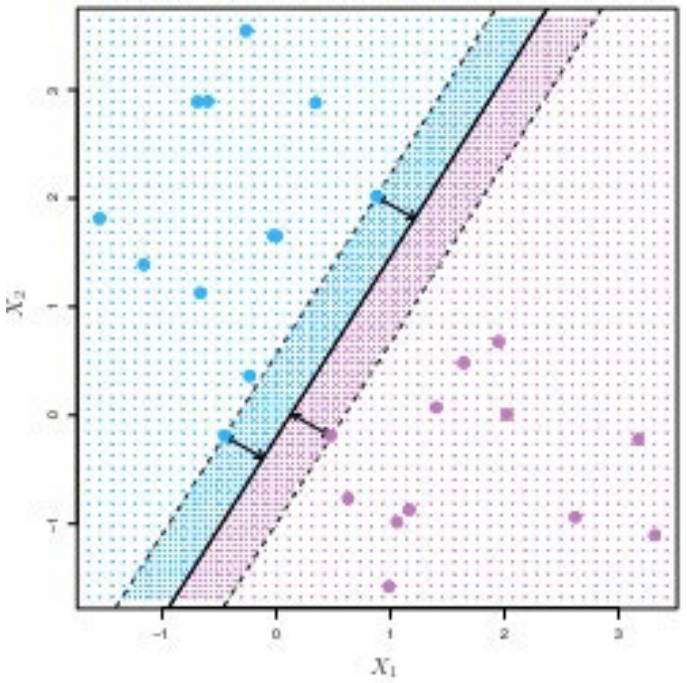
$$maximize\ M$$

$$\beta_0, \beta_1, \dots, \beta_p$$

$$subject\ to \sum_{j=1}^{p} \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p X_{ip}) \geq M$$

$$for\ all\ i = 1, \dots, N$$



- ✓ En la siguiente imagen se muestra de forma gráfica el concepto donde la línea continua resaltada es el maximal-margin classifier, y se muestran dos bandas con líneas discontinúas que contienen la distancia del hiperplano a los primeros puntos de cada una de las clases, el objetivo es maximizar la distancia de estas dos bandas.

Optimización del Margen Máximo

Restricciones

Para encontrar el margen máximo, se deben considerar algunas restricciones. La restricción es que "calcularemos la distancia (d) de tal manera que ningún punto positivo o negativo pueda cruzar la línea de margen".

Función de Optimización

La función de optimización se puede escribir como: Minimizar [||w|| / 2] + c. i = 1 -> n Σ E, donde E es el error de clasificación y c es un hiperparámetro que controla el equilibrio entre el margen y los errores de clasificación.

Margen Suave

En los casos en los que los datos del mundo real rara vez son perfectamente linealmente separables, las máquinas de vectores de soporte se vuelven cruciales. Pueden manejar conjuntos de datos casi lineales, separables y no linealmente separables, proporcionando una solución robusta a los problemas de clasificación.

Hiperparámetro c

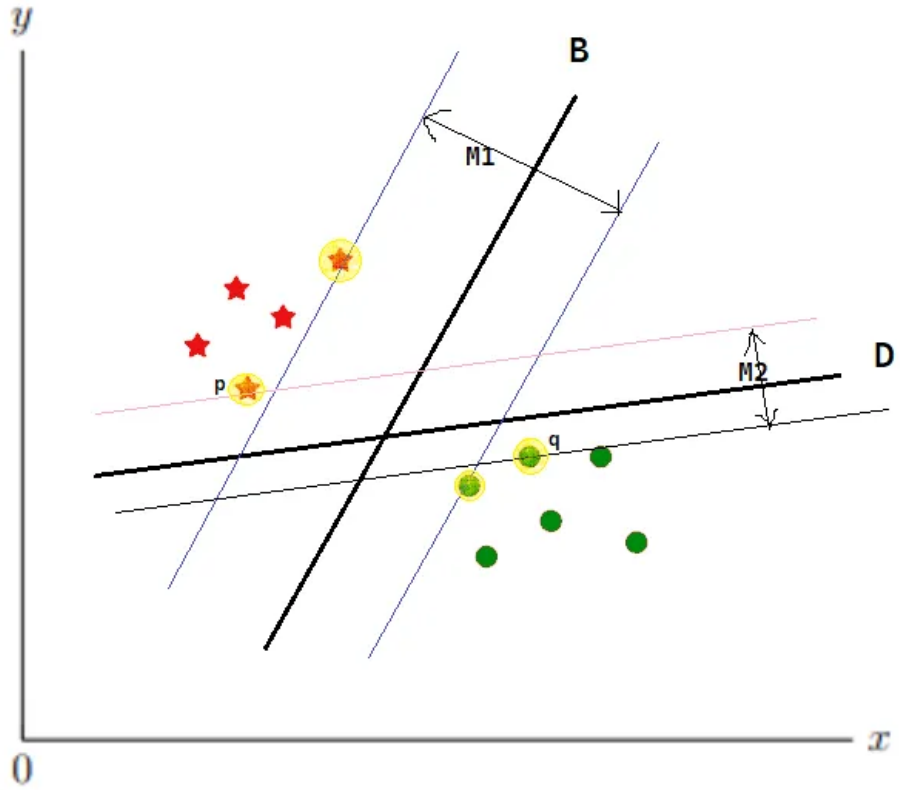
El valor de c es un hiperparámetro importante que se debe ajustar cuidadosamente. Cuanto mayor sea c, menor será el margen, pero habrá menos errores de clasificación. Por lo tanto, se debe encontrar un equilibrio óptimo entre el margen y los errores de clasificación.

TIPOS DE SVM

- ✓**SVM LINEAL: el SVM lineal se utiliza para datos separables linealmente, lo que significa que si un conjunto de datos se puede clasificar en dos clases utilizando una sola línea recta, entonces dichos datos se denominan datos separables linealmente y se utiliza un clasificador llamado clasificador SVM lineal .**

SVM LINEAL

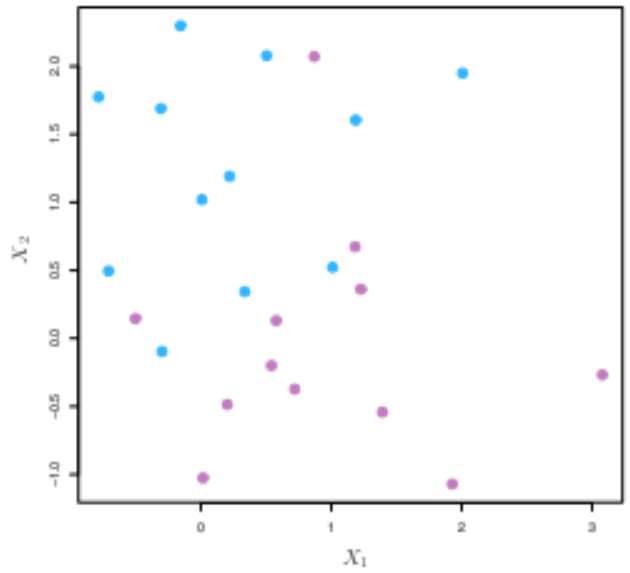
Supongamos que tenemos un conjunto de datos que tiene dos clases (estrellas y puntos) y el conjunto de datos tiene dos características **x e y** . Podemos separar este conjunto de datos utilizando las líneas rectas **B y D**. Cuando dibujamos varias líneas rectas para separar los datos, ¿cuál debe considerarse como el **hiperplano** ?



Supongamos que **D** es el único hiperplano y que estos **p** y **q** son los vectores de soporte. Por lo tanto, dibujemos una ***línea paralela*** a un hiperplano con el soporte de vectores de soporte. Y la distancia **M2** que necesitamos calcular. Esta distancia la llamamos ***Margen*** . De manera similar, también necesitamos hacer esto con el hiperplano **B.** Y calcular el margen **M1** . Al comparar M1 y M2, el **margen M2 es menor** en comparación con **M1** . Entonces, usando esto, podemos decir que **B es la línea de mejor ajuste (** hiperplano) que divide los datos en dos clases. Deberíamos hacer este cálculo para todos los hiperplanos posibles y el que dé el **margen máximo** será el **hiperplano buscado**.

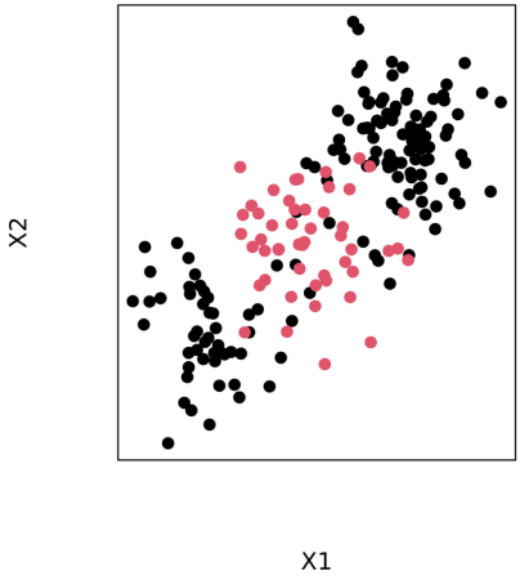
SVM NO LINEAL

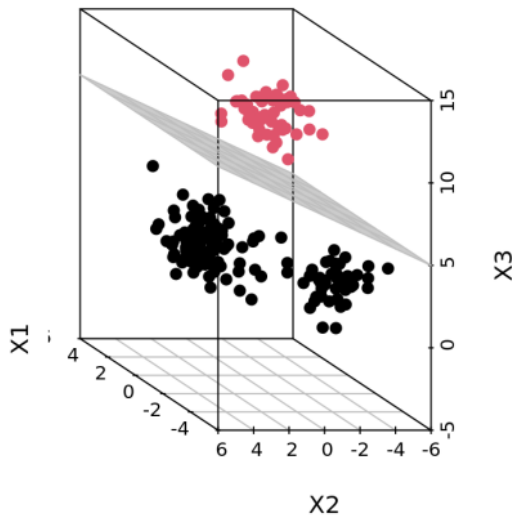
- ✓El principal problema con esta separación óptima del hiperplano es que los datos habitualmente no son linealmente separables con una recta en el caso de un espacio de dos dimensiones, como se observa en la siguiente gráfica, donde es imposible separar los puntos de una clase de los de otra.
- ✓El SVM no lineal se utiliza para datos separados de forma no lineal, lo que significa que si un conjunto de datos no se puede clasificar mediante una línea recta , dichos datos se denominan datos no lineales y el clasificador se denomina clasificador SVM no lineal .

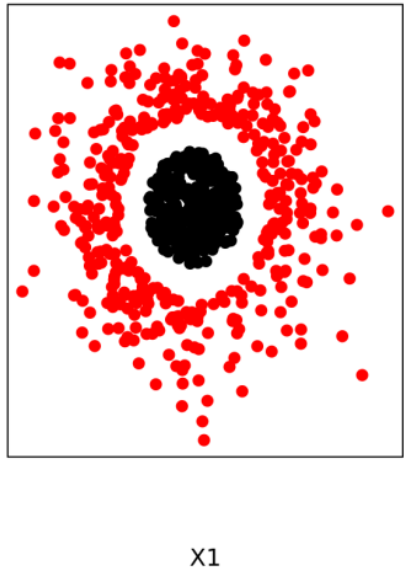


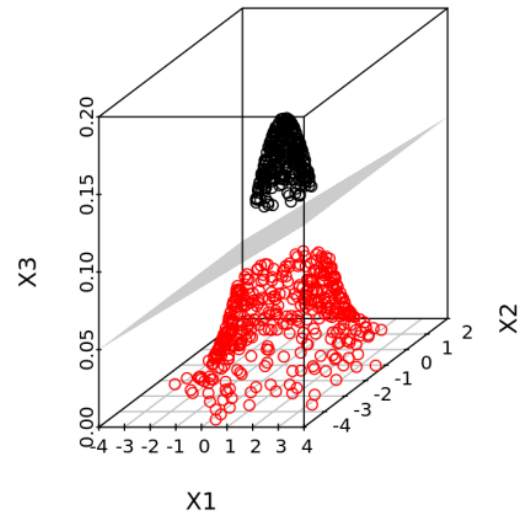
Tratar con planos no lineales e inseparables

- ✓ Algunos problemas no pueden resolverse utilizando el hiperplano lineal, como se muestra en las figuras 1 y 3.
- ✓ En tal situación, SVM utiliza un truco del kernel (kernel trick) para transformar el espacio de entrada en un espacio de mayor dimensión, como se muestra en las figuras 2 y 4. Los puntos de datos se representan en el eje x y el eje z (Z es la suma al cuadrado de x e y: z = x^2 + y^2). Ahora se puede segregar fácilmente estos puntos utilizando la separación lineal.









Kernels SVM

- ✓ El algoritmo SVM se implementa en la práctica utilizando un kernel.
- ✓ Un kernel transforma un espacio de datos de entrada en la forma requerida.
- ✓ SVM utiliza una técnica llamada kernel trick. Aquí, el kernel toma un espacio de entrada de baja dimensión y lo transforma en un espacio de dimensión superior.
- ✓ En otras palabras, se puede decir que convierte el problema no separable en problemas separables añadiéndole más dimensión. Es útil sobre todo en problemas de separación no lineal. El kernel trick ayuda a crear un clasificador más exacto.
- ✓ La función kernel podemos definirlo como:

$$K\left(\overline{x}\right) = \begin{matrix} 1 & \text{if} \left\|\overline{x}\right\| \leq 1 \\ 0 & \text{otherwise} \end{matrix}$$

En el contexto de SVMs, hay 4 kernels populares:

- Linear kernel
- Polynomial kernel
- Radial Basis Funtion (RBF) o Gaussian Kernel
- Sigmoid Kernel

Kernels SVM

- ✓ **El kernel de un operador A denotado por ker es el conjunto de todos los vectores cuya imagen sea el vector nulo:**

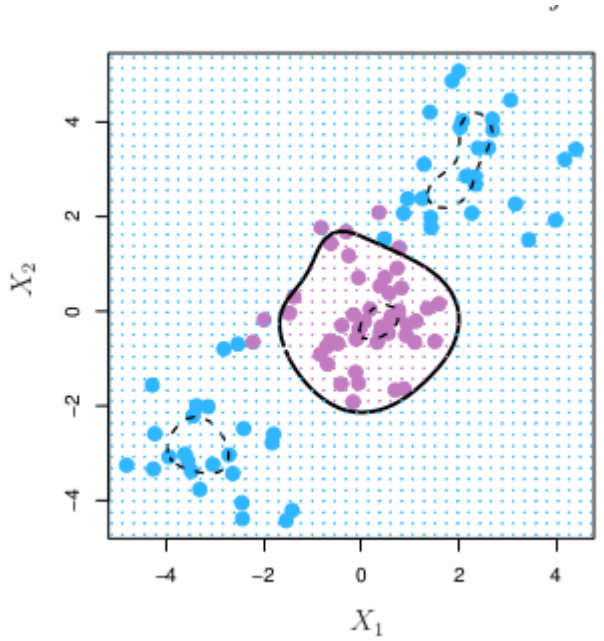
$$\ker A = \{\underset{v}{\rightarrow} \in V : A \underset{v}{\rightarrow} = 0\}$$

- ✓ **Los kernels se apoyan en concepto del producto vectorial de los vectores de soporte. Se trata de funciones que reciben dos vectores como parámetros. Uno de los kernels más utilizados es el de base radial que asume que el feature space es de altas dimensiones y se define con la siguiente función matemática:**

$$K(x_i, x_i') = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p X_{ip}) \geq M$$

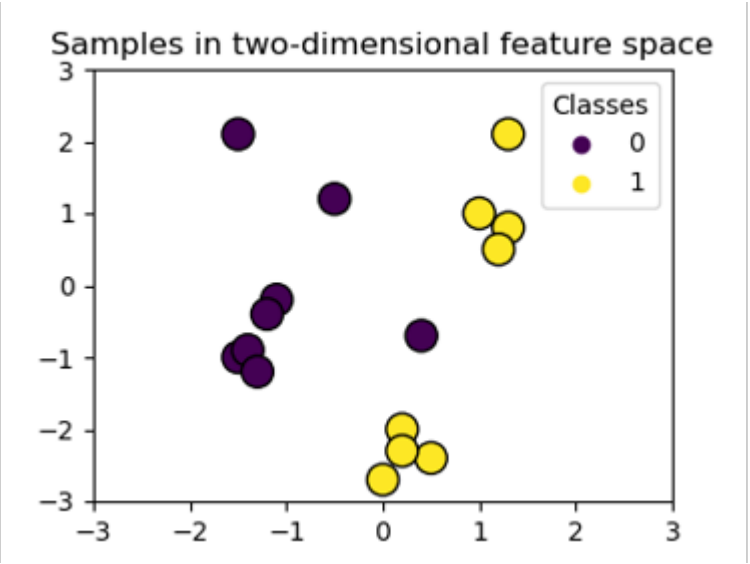
- ✓ **El uso de kernels permite obtener fronteras de decisión no lineales por medio de transformaciones matemáticas sin necesidad de tener que realizar transformaciones con polinomios.**



Ejemplo de las fronteras de decisión no lineales obtenidas por medio del uso de un *kernel* de base radial

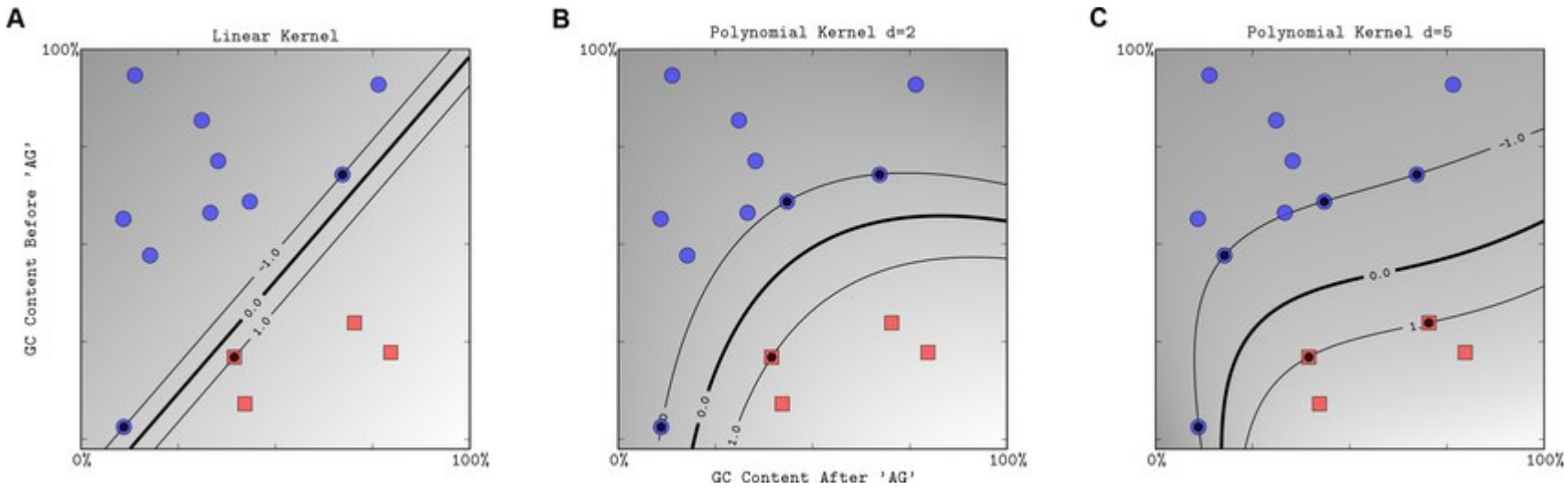
Linear Kernel

- ✓ En el kernel lineal, la función kernel toma la forma de una función lineal de la siguiente manera:
  - ➢ kernel lineal: K(xi, xj) = xiT xj
- ✓ El kernel lineal se utiliza cuando los datos son linealmente separables. Esto significa que los datos se pueden separar utilizando una sola línea. Es uno de los kernels más comunes que se utilizan. Se utiliza principalmente cuando hay una gran cantidad de características en un conjunto de datos. El kernel lineal se utiliza a menudo para fines de clasificación de texto.
- ✓ El entrenamiento con un kernel lineal suele ser más rápido, porque solo necesitamos optimizar el parámetro de regularización C. Cuando se entrena con otros kernels, también necesitamos optimizar el parámetro γ. Por lo tanto, realizar una búsqueda en cuadrícula generalmente llevará más tiempo.
- ✓ El kernel lineal se puede visualizar con la siguiente figura.



Polynomial Kernel

- ✓ Representa la similitud de los vectores (muestras de entrenamiento) en un espacio de características sobre polinomios de las variables originales. El kernel polinomial no solo analiza las características dadas de las muestras de entrada para determinar su similitud, sino también las combinaciones de las muestras de entrada.
- ✓ Para polinomios de grado d, el kernel polinomial se define de la siguiente manera:
  - ➢ Kernel polinomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- ✓ El kernel polinomial es muy popular en el procesamiento del lenguaje natural. El grado más común es d = 2 (cuadrático), ya que los grados mayores tienden a sobreajustarse en los problemas de procesamiento del lenguaje natural. Se puede visualizar con el siguiente diagrama.

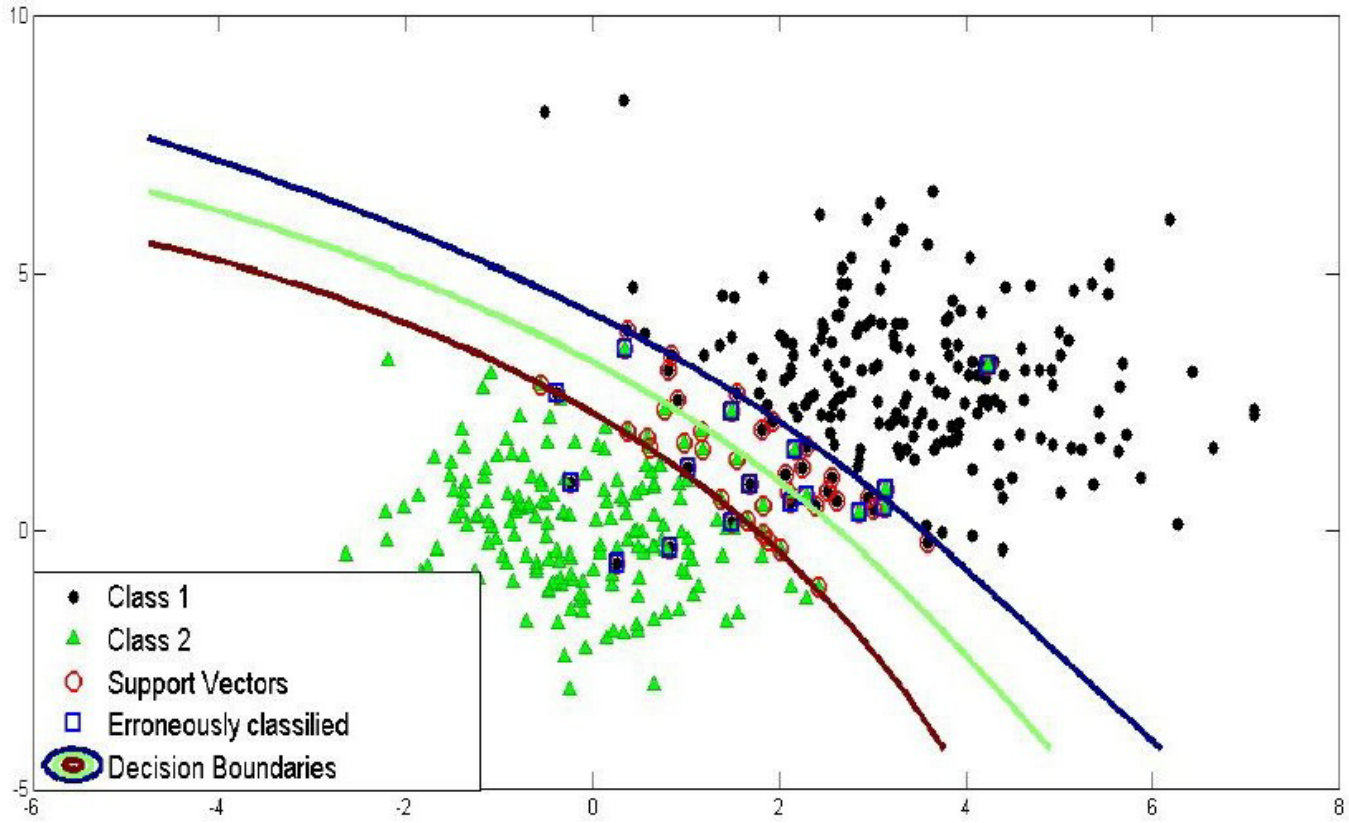


Radial Basis Function Kernel

- ✓El kernel de función de base radial es un kernel de propósito general. Se utiliza cuando no tenemos conocimiento previo sobre los datos. El kernel de función de base radial en dos muestras x e y se define mediante la siguiente ecuación:

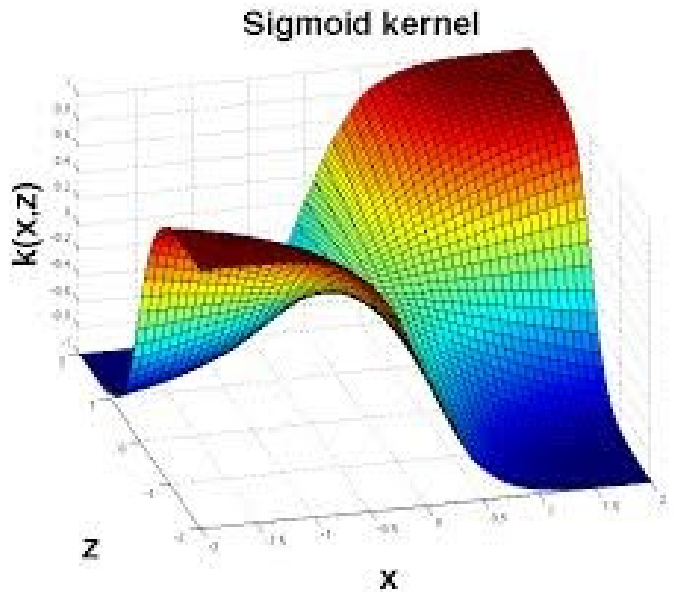
$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$

El siguiente diagrama demuestra la clasificación SVM con kernel rbf.



Sigmoid Kernel

- ✓El núcleo sigmoide tiene su origen en las redes neuronales. Podemos usarlo como proxy para las redes neuronales. El núcleo sigmoide se obtiene mediante la siguiente ecuación:
  - ➢sigmoid kernel : k (x, y) = tanh(αxTy + c)
- ✓El sigmoid kernel se puede visualizar en el siguiente diagrama:



Aplicaciones de SVM

Las SVM se utilizan principalmente para problemas de clasificación, donde el objetivo tiene un número finito de posibilidades, como determinar si un correo electrónico es spam o no.

Aunque las SVM se utilizan principalmente para clasificación, también se pueden utilizar para problemas de regresión, donde el objetivo es continuo, como predecir el aumento de salario de un empleado.

Las SVM se han utilizado con éxito en tareas de reconocimiento de imágenes, como la clasificación de imágenes y la detección de objetos.

Las SVM también se han aplicado a problemas de clasificación de texto, como la clasificación de documentos y el análisis de sentimientos.

Ventajas de SVM

Alta precisión: Las SVM suelen tener un alto rendimiento en términos de precisión de clasificación, especialmente en problemas de clasificación binaria.

Eficiente en Altas Dimensiones: Las SVM funcionan bien incluso cuando el número de características (dimensiones) es mucho mayor que el número de muestras, lo que las hace adecuadas para problemas de alta dimensionalidad

Versatilidad: Las SVM pueden manejar tanto datos linealmente separables como no linealmente separables mediante el uso de funciones kernel, lo que las convierte en un algoritmo muy versátil.

Regularización incorporada: Las SVM tienen una regularización incorporada que ayuda a evitar el sobreajuste, lo que las hace robustas frente al ruido y a los datos atípicos.

Desventajas de SVM

Selección de Kernel: La elección de la función kernel adecuada puede ser un desafío y requiere un conocimiento profundo del problema y los datos.

Sensibilidad a Parámetros: El rendimiento de las SVM depende en gran medida de la selección de los parámetros correctos, como el parámetro de regularización C y los parámetros del kernel.

Complejidad Computacional: El entrenamiento de las SVM puede ser computacionalmente costoso, especialmente cuando se trabaja con grandes conjuntos de datos o un gran número de características.

PREGUNTAS??

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024