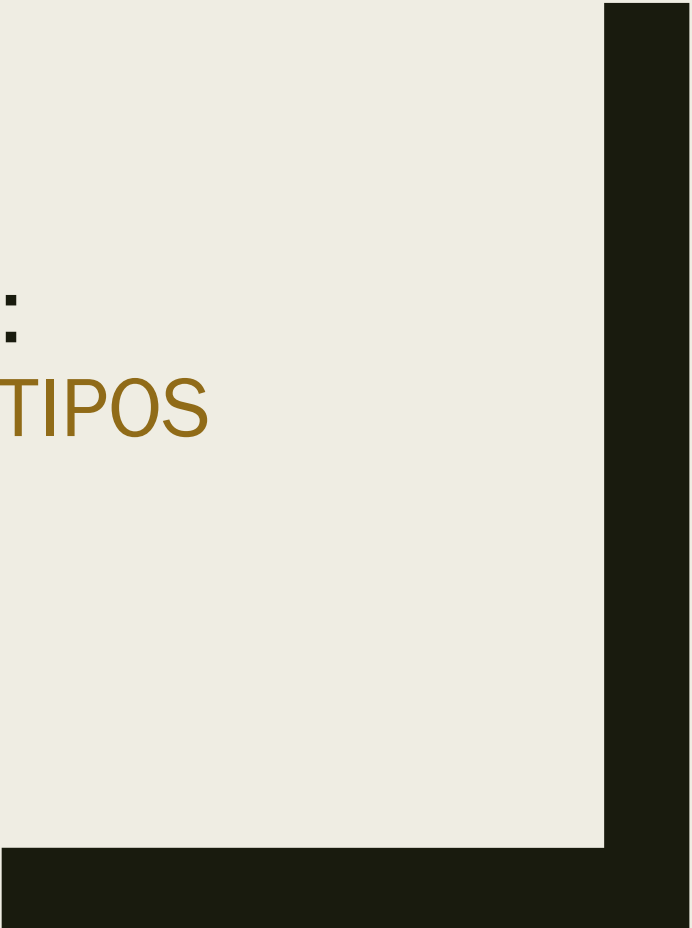




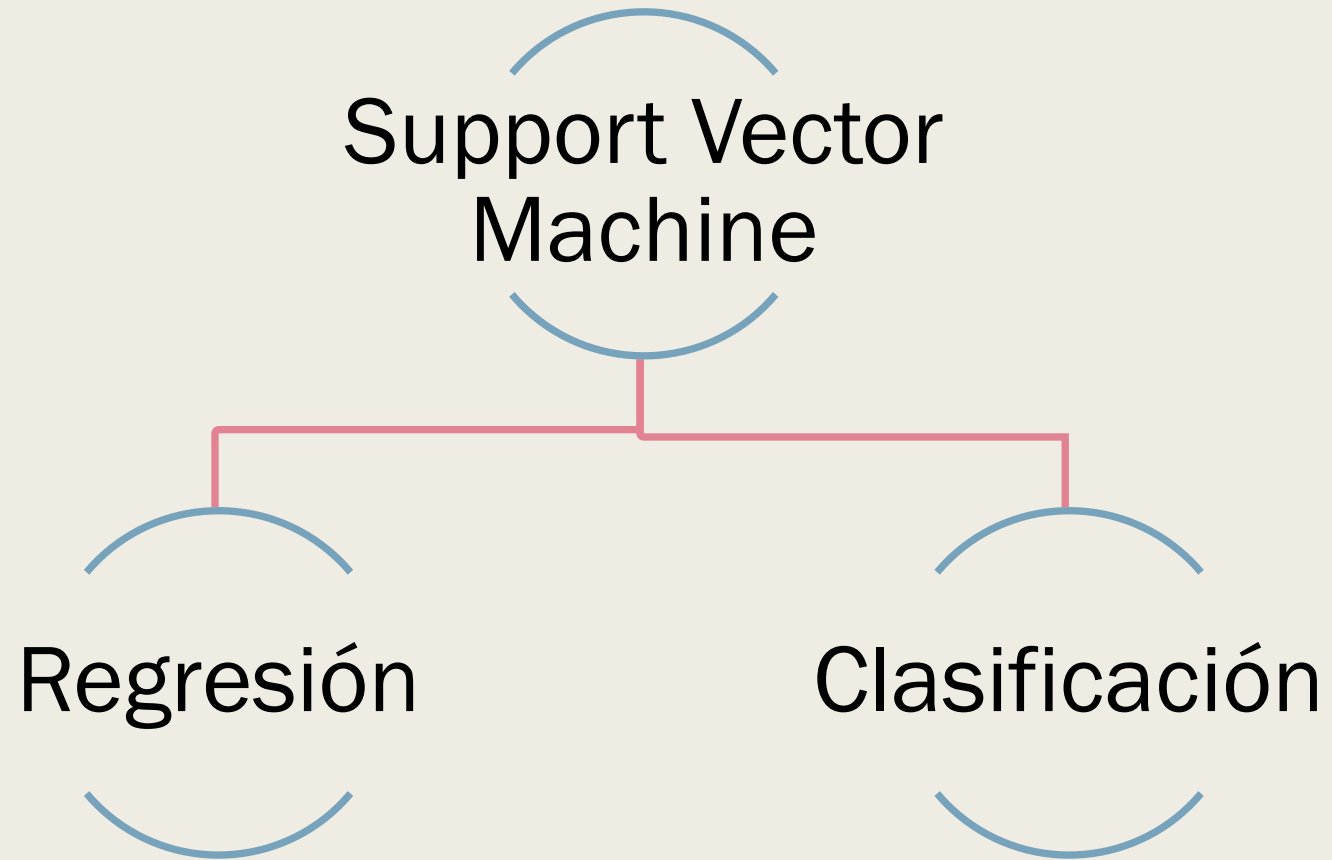
ANALÍTICA AVANZADA DE DATOS: MÁQUINAS DE SOPORTE VECTORIAL

A. Alejandra Sánchez Manilla
asanchezm.q@gmail.com



Máquinas de soporte vectorial

Puede ser aplicado para:





CLASIFICACIÓN

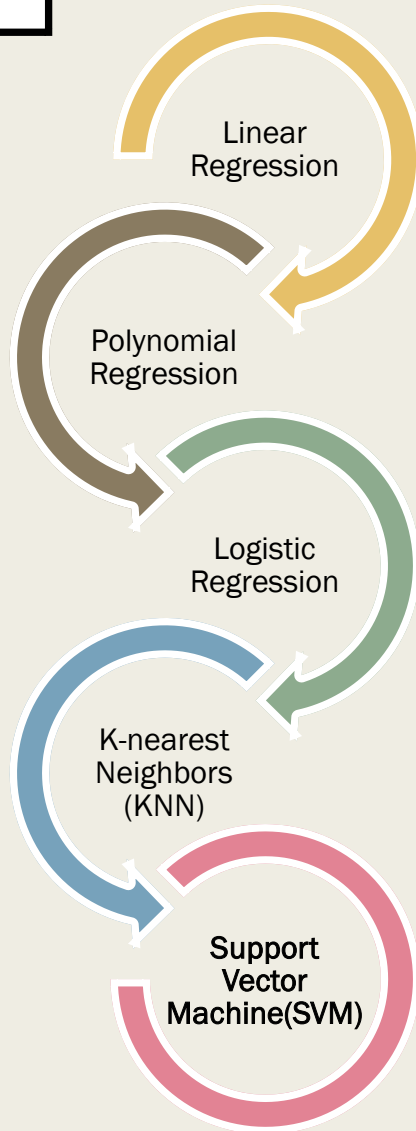


Máquinas de soporte vectorial

Las Máquinas de Soporte Vectorial (SVM de sus siglas en inglés), son discriminadores lineales propuesto para problemas de clasificación binaria (2 clases).

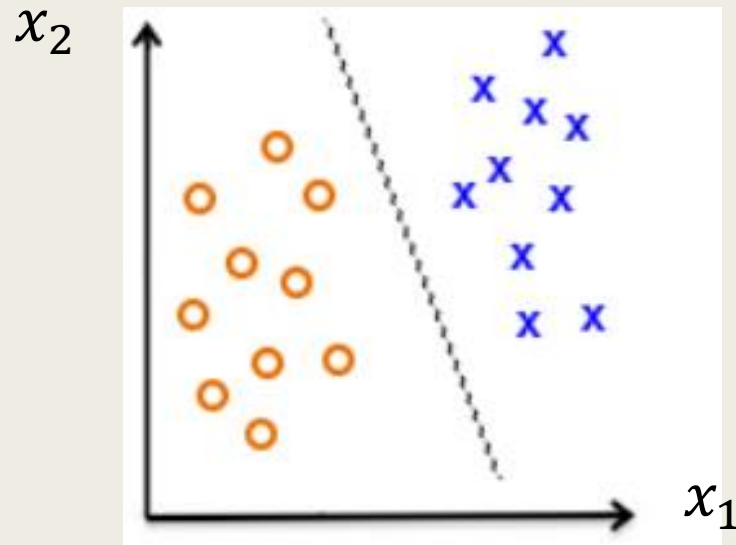
Las SVM se basan en 2 ideas principales:

- Pre-procesamiento de los datos originales para representarlos en un espacio dimensional superior.
- Identificar una frontera de decisión lineal en el nuevo espacio dimensional.



Máquinas de soporte vectorial

Es un clasificador discriminatorio definido formalmente por un **hiperplano de separación**



Dados los datos de entrenamiento etiquetados el algoritmo genera un hiperplano óptimo que clasifica los nuevos ejemplos en dos espacios dimensionales, este hiperplano es una **línea** que divide un plano en dos partes donde en cada clase se encuentra en cada lado

Máquinas de soporte vectorial

Con una función de mapeo no lineal a un espacio dimensional lo suficientemente alto, los datos provenientes de 2 clases siempre serán separables por un hiperplano.

SVM abstraen una frontera de decisión en un espacio multidimensional utilizando un subconjunto apropiado de patrones del conjunto de entrenamiento. Este subconjunto de patrones son los vectores de soporte.

Geométricamente los vectores de soporte son los patrones de entrenamiento que se encuentran más cerca de la frontera de decisión.

Máquinas de soporte vectorial

SVM es especialmente útil cuando se trata de datos que no son linealmente separables, lo que significa que no se puede separar simplemente con una línea recta o un plano

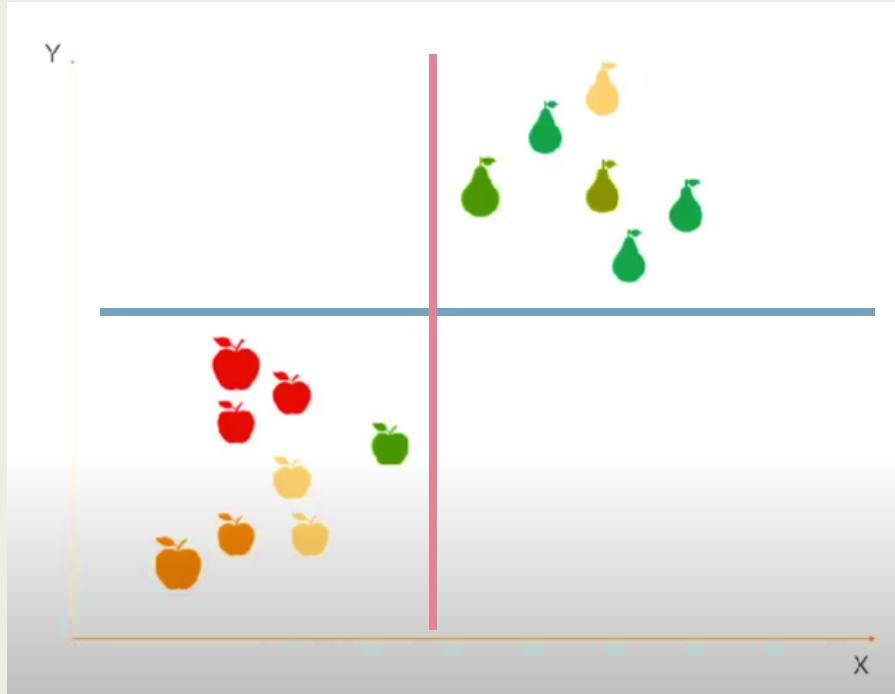
Puede ser adecuado en casos como la detección de fraudes, detección de Spam, análisis de sentimientos y diagnóstico médico

Dos objetivos principales:

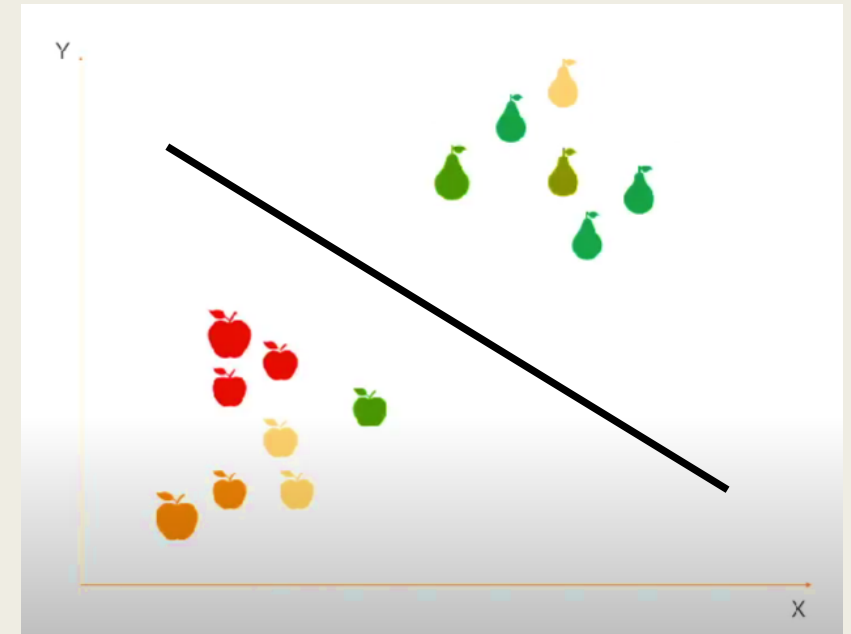
- Encontrar el hiperplano que separa los datos
- Maximizar el margen entre las clases. Al hacerlo intenta clasificar datos desconocidos en su posición en relación al hiperplano

Máquinas de soporte vectorial

Ejemplo



1. Establecer el hiperplano (separar en dos el plano)
2. En ambos casos el hiperplano separa perfectamente en dos los datos



Máquinas de soporte vectorial

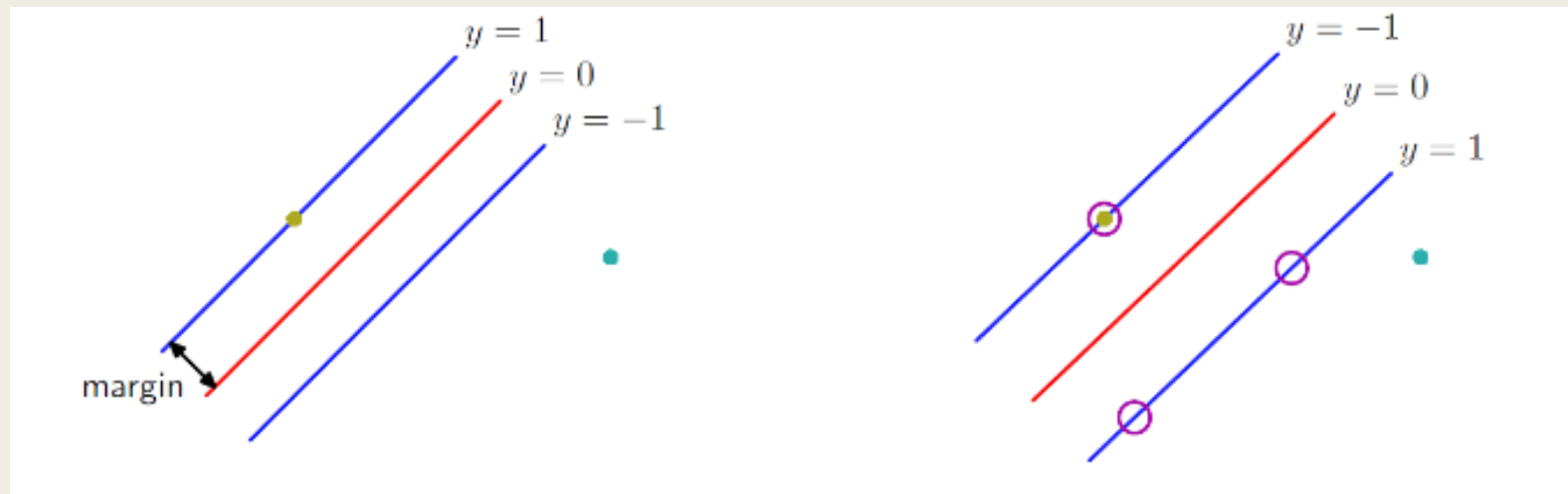
Es importante saber qué línea separa mejor los datos porque a pesar de que en el ejemplo solo tenemos dos clases (peras y manzanas) en el mundo real, esta línea nos puede decir:

- *Apto para un préstamo*
- *Alguien que cumple los requisitos para un medicamento*
- *Cómo separar spams en tu correo*

Máquinas de soporte vectorial

Si existen varias soluciones posibles que permiten dividir correctamente las clases, se debería escoger aquella con el mínimo error

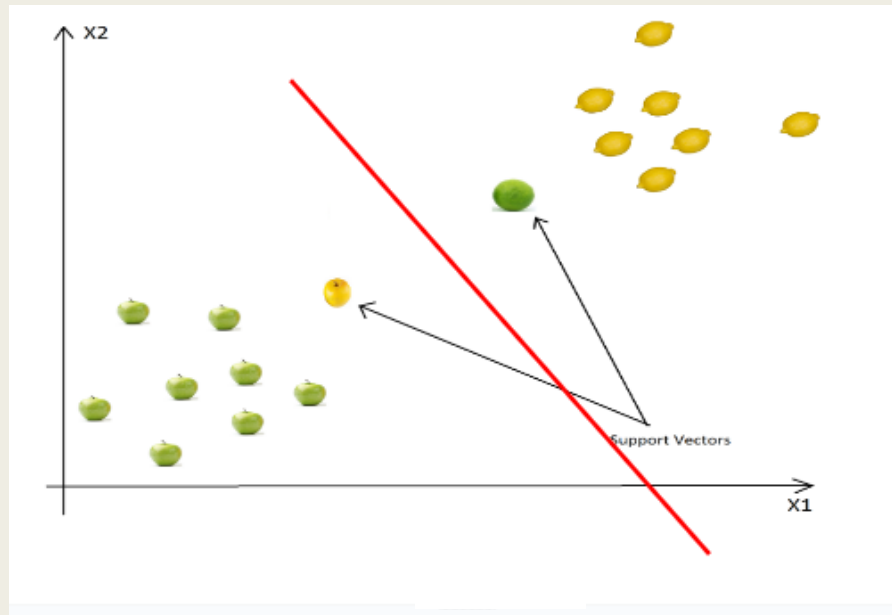
Las SVM atacan este problema a través del concepto de márgenes, el cual es definido como la distancia mínima entre la frontera de decisión y los patrones de soporte



Máquinas de soporte vectorial

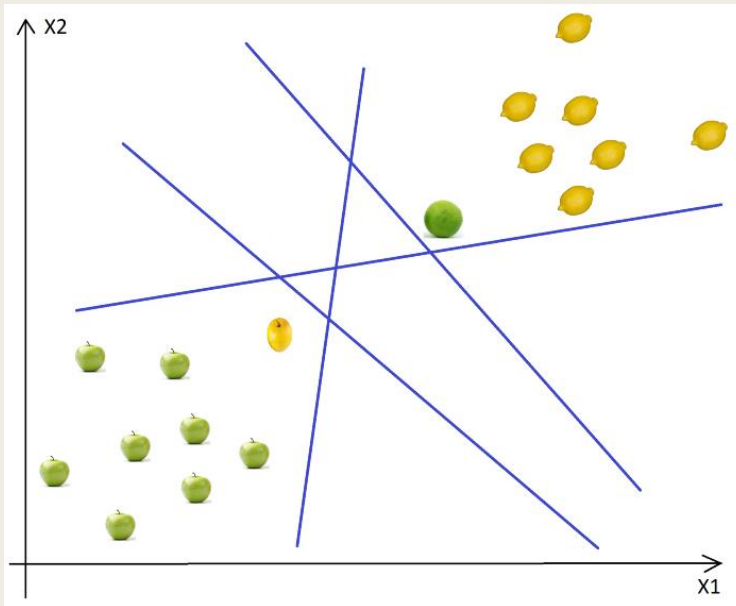
Por lo general, un algoritmo de aprendizaje intenta aprender las características más comunes de una clase y la clasificación se basa en esas características representativas

La SVM funciona al revés. Encuentra los ejemplos más similares entre clases
Esos serán los vectores de soporte

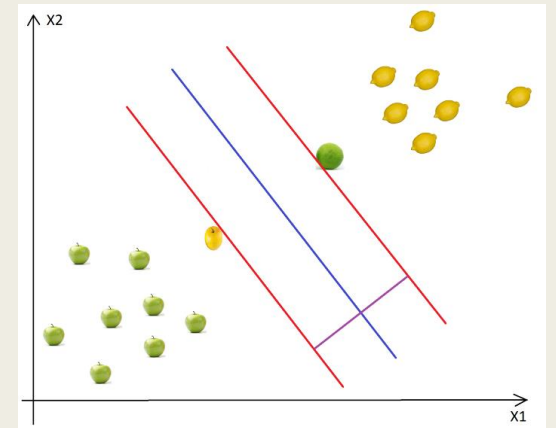
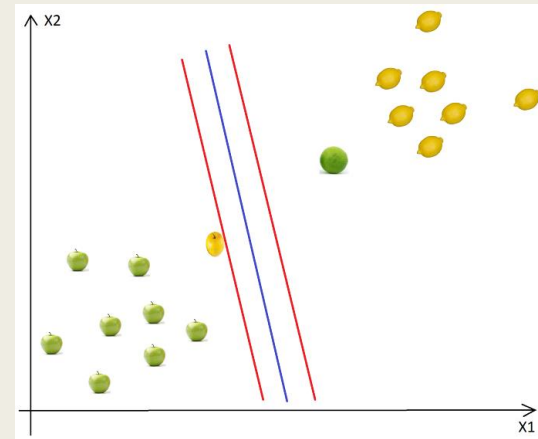


Máquinas de soporte vectorial

Pero esa línea podría dibujarse de muchas maneras



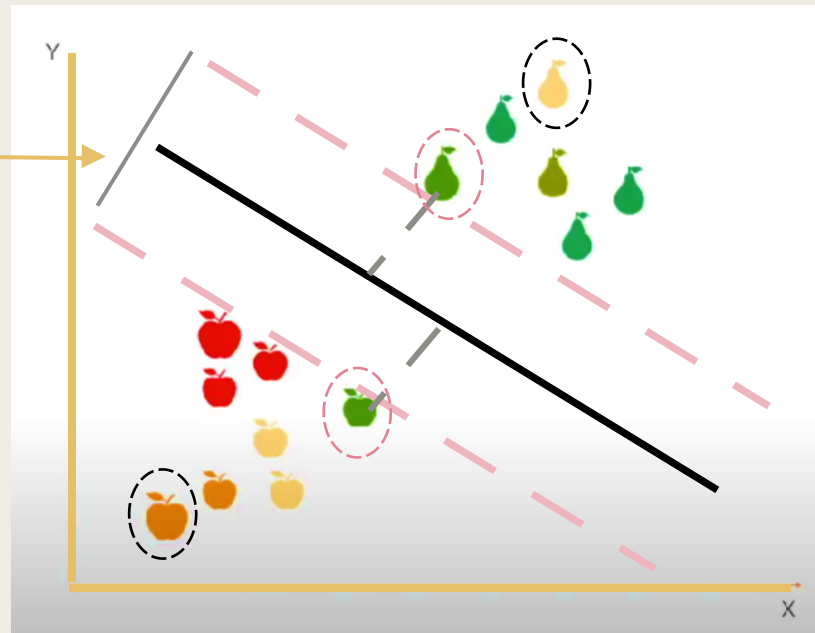
¿Cómo encontrar la óptima?



Máquinas de soporte vectorial

3. Una vez definido el hiperplano, creamos dos líneas paralelas tanto en el lado positivo como en el lado negativo.
4. Estas líneas se crean utilizando los vectores de soportes, es decir, los datos de ambos planos más cercanos al hiperplano se usarán como soporte y como guías para estas nuevas líneas, que crearán una banda o rango entre los datos

Máximo Margen

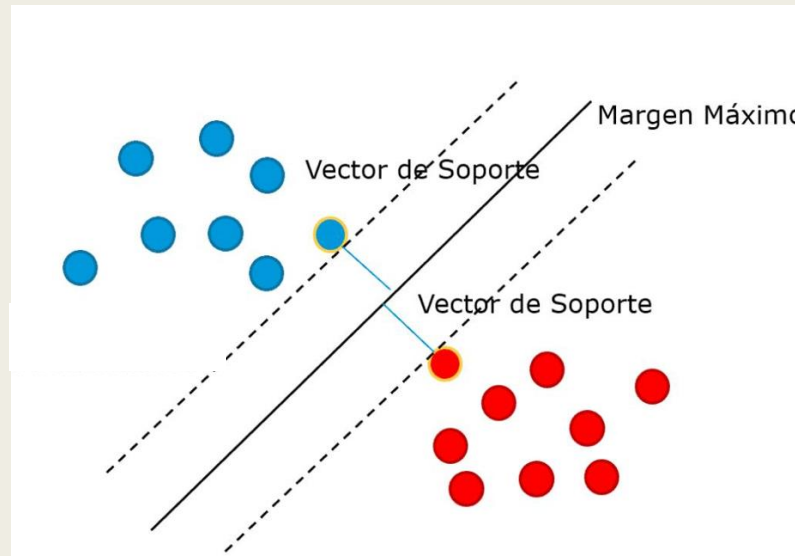


Vectores de soporte

Máquinas de soporte vectorial

SVM encuentra una línea o un plano que separa los datos en diferentes clases. Para hacer esto, el algoritmo busca la línea o plano que tiene la máxima separación entre las dos clases.

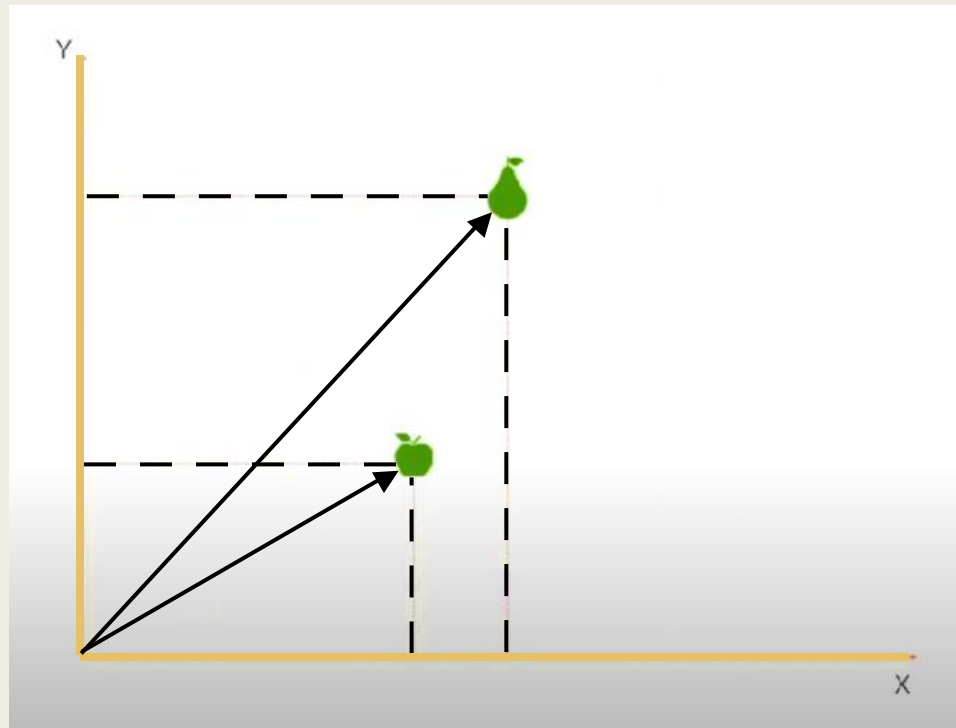
Esta línea se llama “**margen**” y se encuentra buscando los puntos de datos más cercanos de ambas clases, llamados “**vectores de soporte**”



Máquinas de soporte vectorial

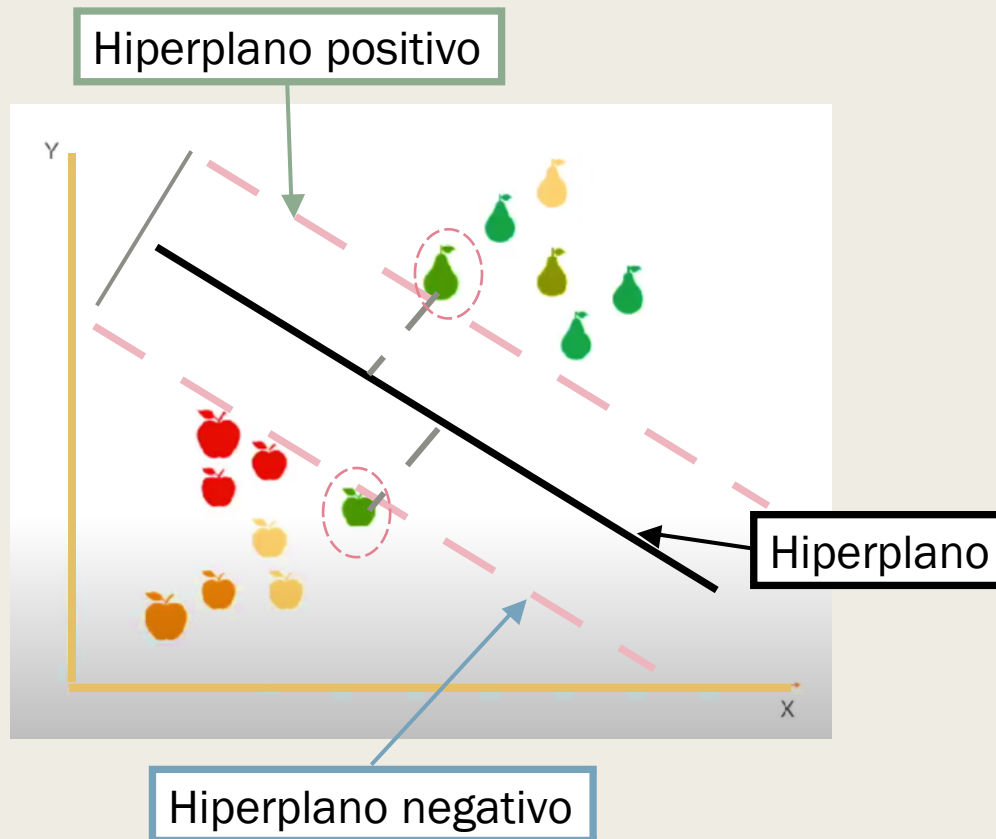
¿Por qué se llaman vectores de soporte?

Porque se vean representados como un punto o una figura, para este caso, en el plano, al final son vectores:



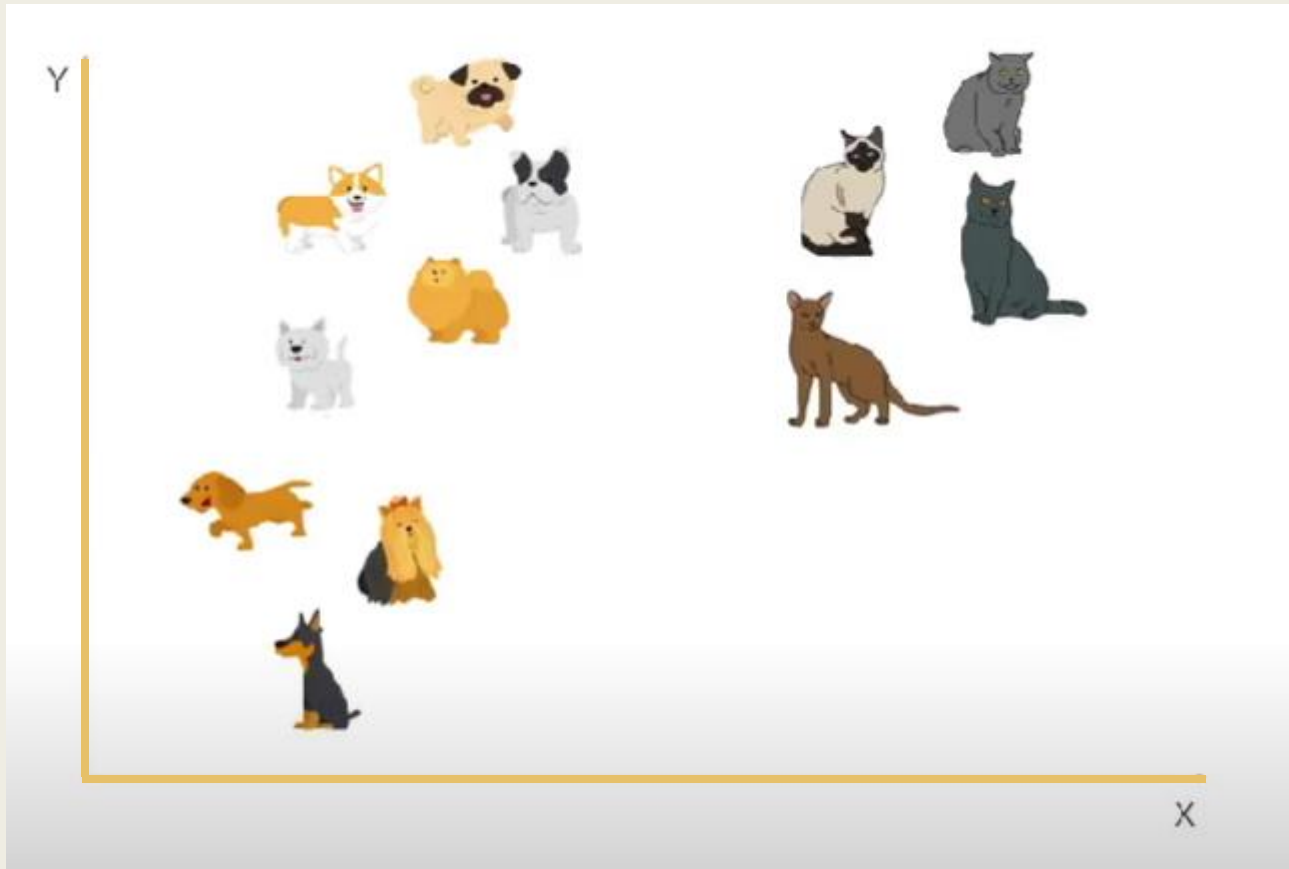
Estos vectores los utilizaremos para crear nuestras bandas

Máquinas de soporte vectorial



Todo lo que cae en cada uno de los hiperplanos será clasificado igual a los datos correspondientes a ese hiperplano, en el ejemplo será: manzanas o peras

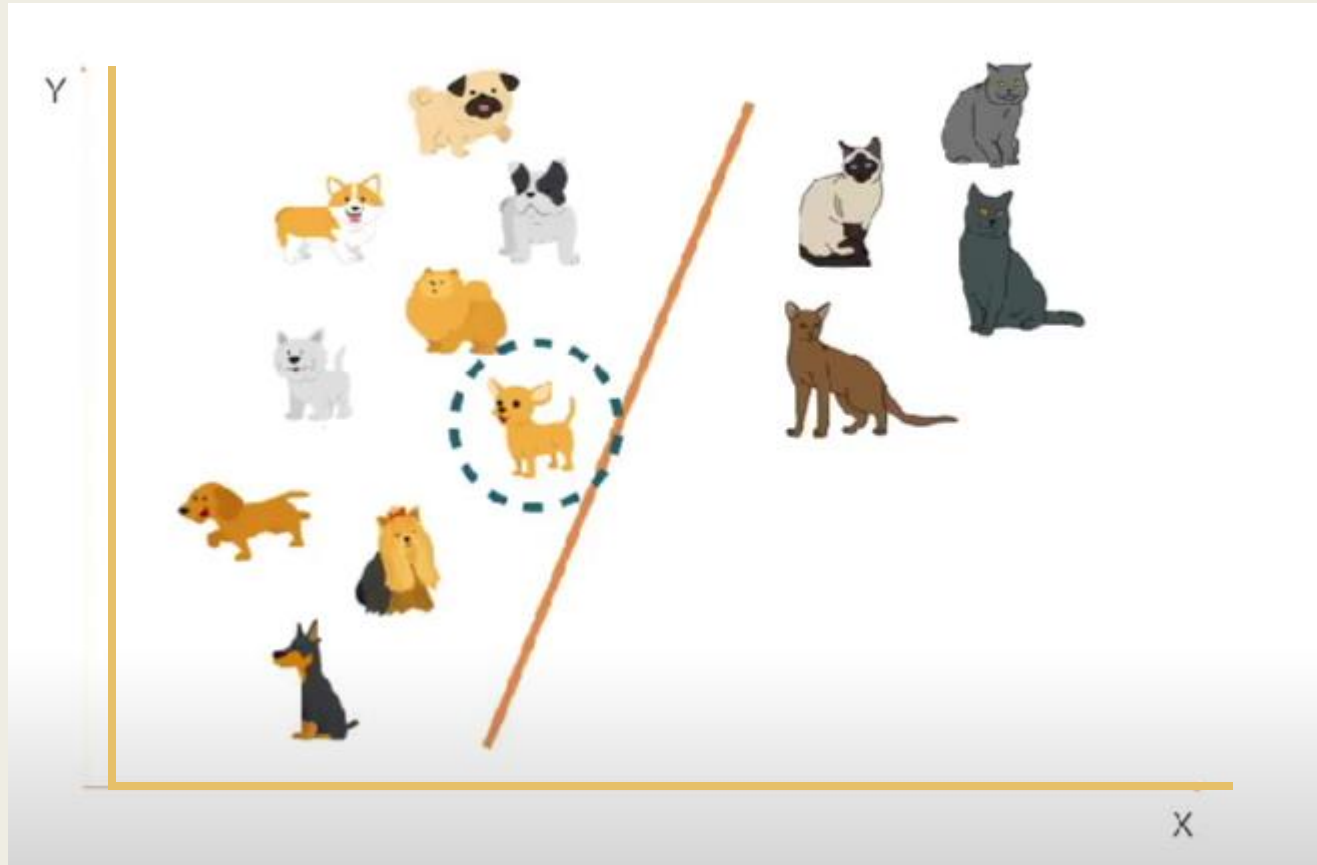
Máquinas de soporte vectorial



Clasificar un nuevo animal??



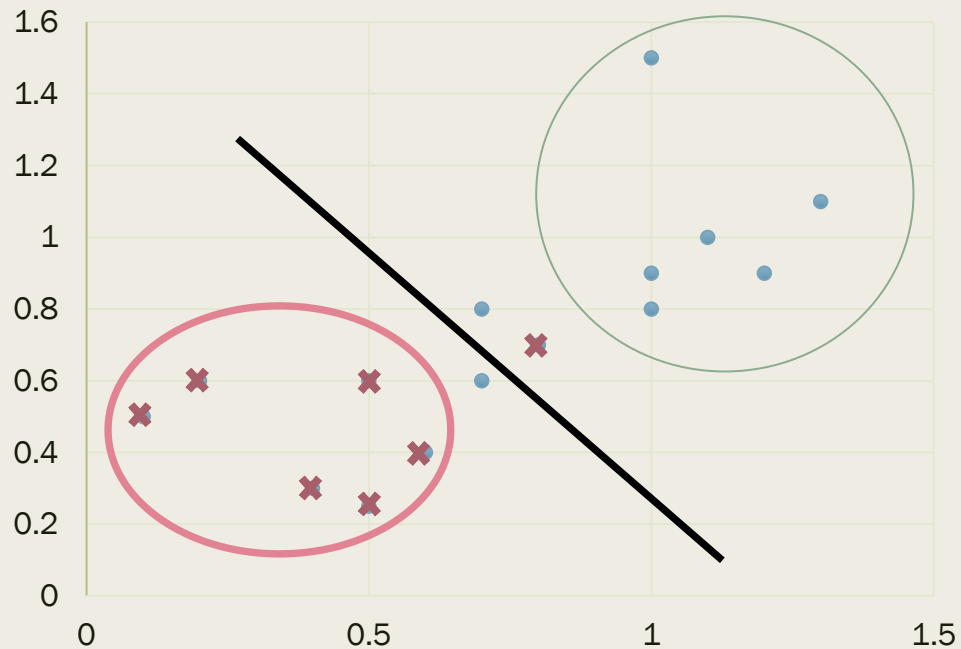
Máquinas de soporte vectorial



- Hiperplano negativo (Perros)
- Clasificación: Perro

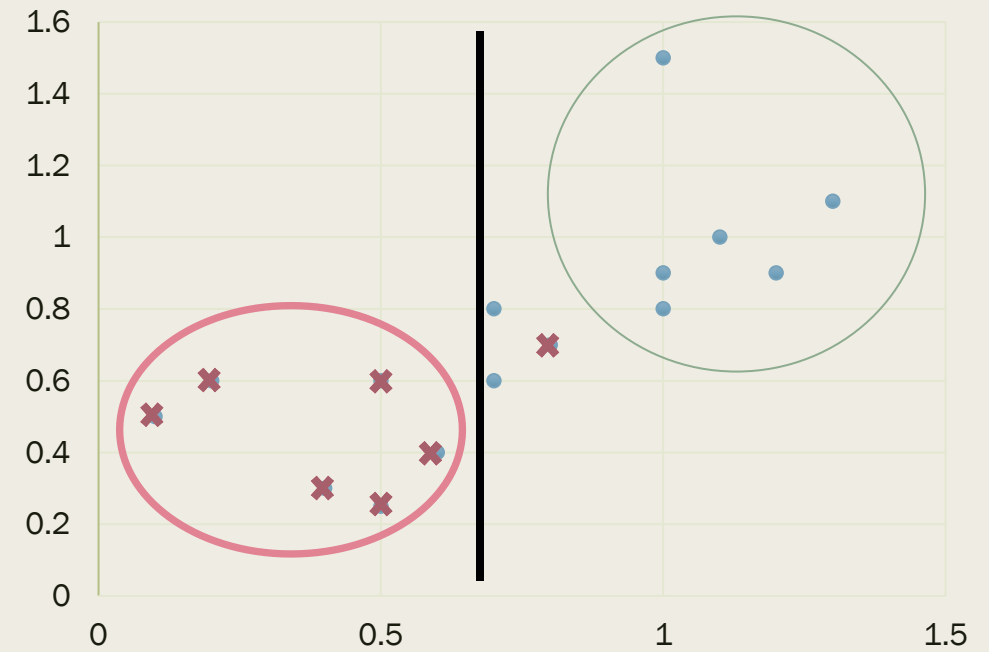
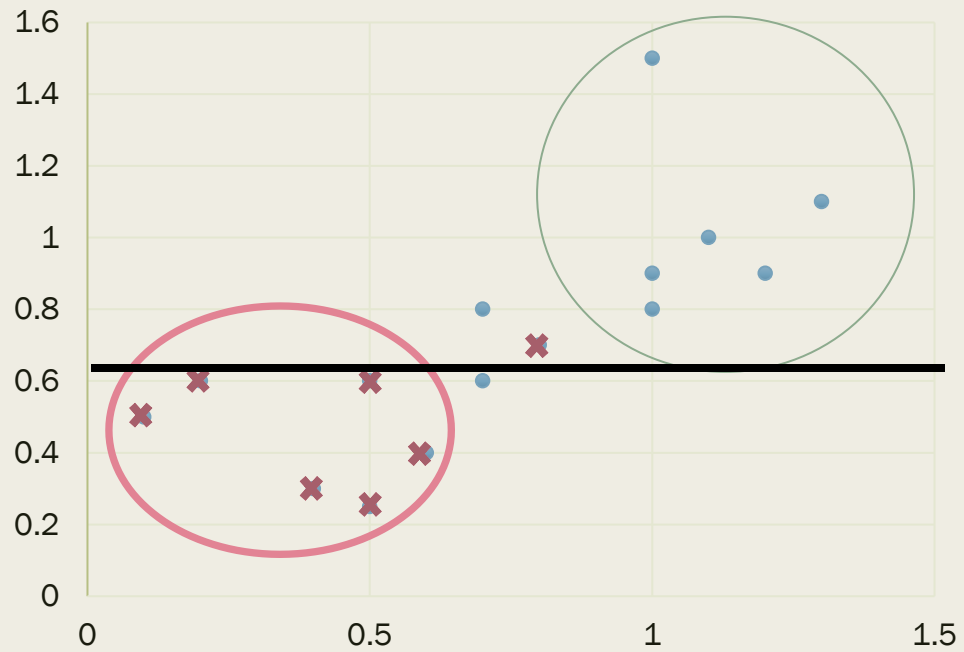
Máquinas de soporte vectorial

Existe un parámetro en SVM que puedes utilizar cuando la información que tienes esta mezclada



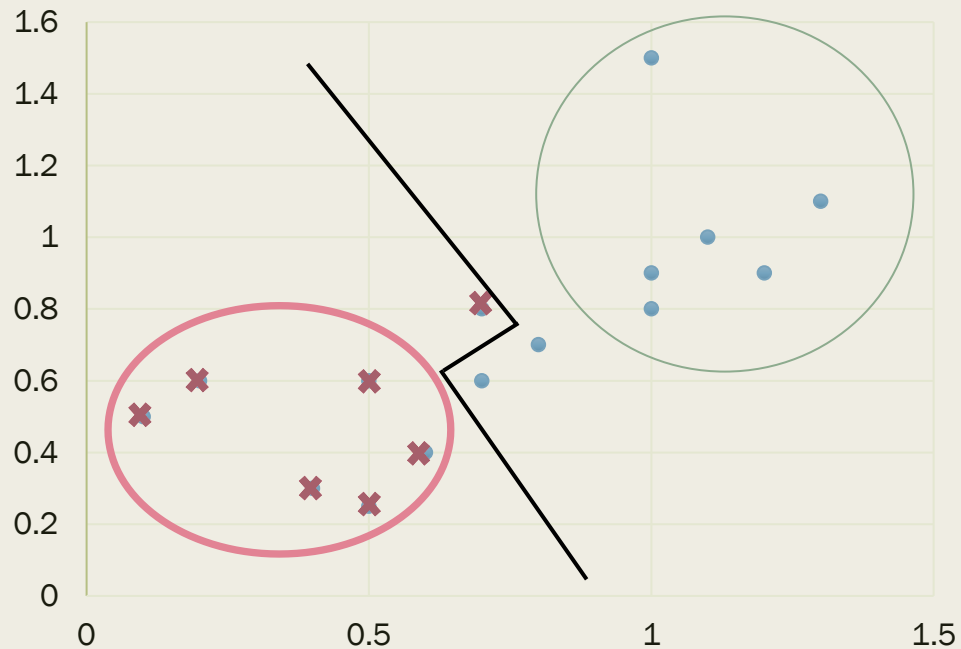
Si trazamos una línea, clasificaríamos mal y por más hiperplanos que tratemos no encontraremos uno perfecto

Máquinas de soporte vectorial



Máquinas de soporte vectorial

Utilizamos un **parámetro C** el cuál reajusta el hiperplano para separar los datos



Este parámetro es ajustable y debe ajustarse al mejor valor para evitar clasificar mal los datos

¿Y si los datos no son lineales?

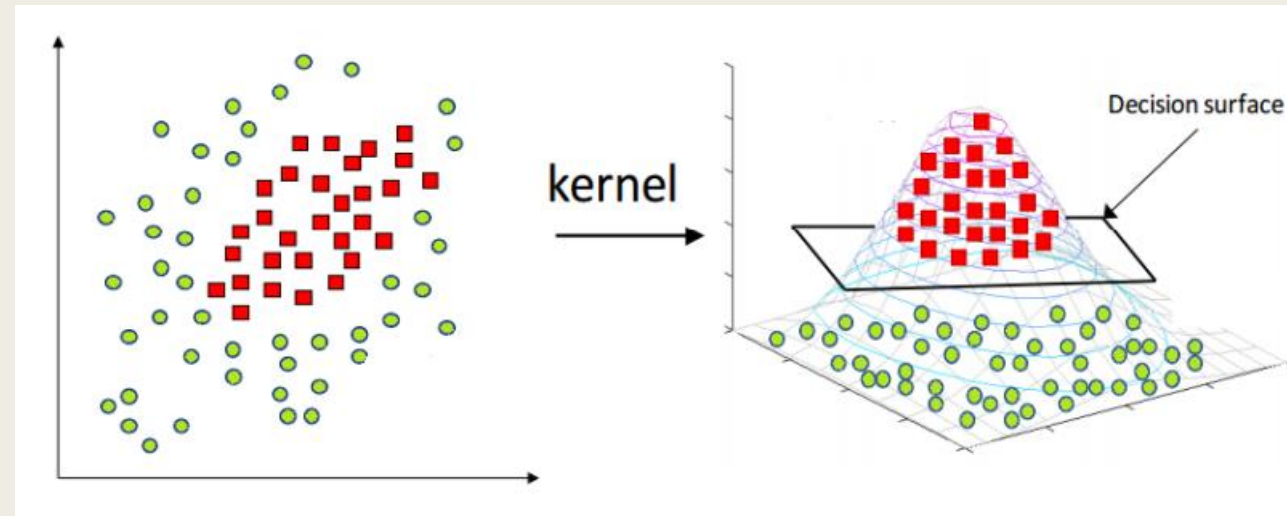
Máquinas de soporte vectorial

Kernel

Deben su nombre a al uso de las funciones de núcleo, que les permiten operar en un espacio de características implícito y de alta dimensión sin tener que calcular nunca las coordenadas de los datos en ese espacio, sino que simplemente calculando los **productos internos entre las imágenes** de todos los pares de datos en el espacio de características

Máquinas de soporte vectorial

Si encontramos una manera de mapear los datos desde el espacio bidimensional al espacio tridimensional, podremos encontrar una superficie de decisión que se divide entre las diferentes clases



Máquinas de soporte vectorial

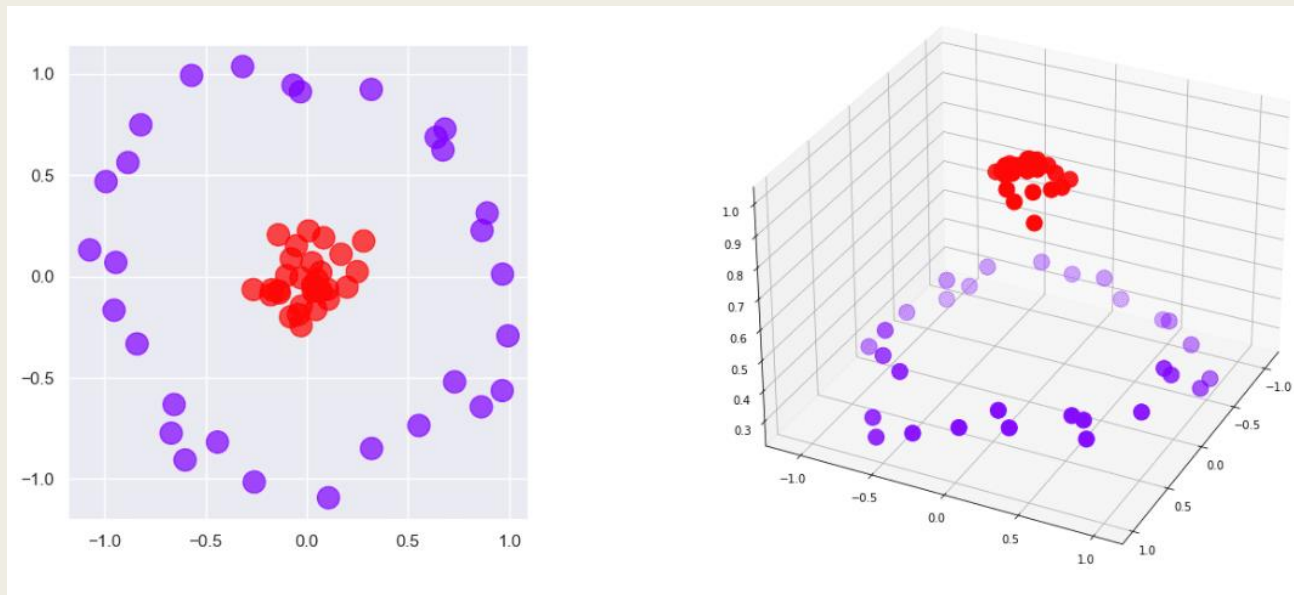
La idea principal es:

- *Mapear todos los datos a una dimensión superior, en este caso 3 dimensiones*
- *Encontrar el límite*
- *Hacer la clasificación*

Sin embargo, cuando hay más dimensiones, los cálculos dentro de este espacio se vuelven complicados. Aquí es donde utilizamos el kernel

Máquinas de soporte vectorial

Kernel



Nos permite operar en el espacio de la característica original sin calcular las coordenadas de los datos en un espacio dimensional superior

Truco del kernel:

Ofrecer una manera más eficiente y menos costosa de transformar los datos en dimensiones superiores

Aplicación del truco del kernel:

No solo se limita al algoritmo de SVM, cualquier calculo que involucre los productos de puntos (x, y) puede utilizar este método

Máquinas de soporte vectorial

Kernel Lineal

- $K(x, xi) = \text{sum}(x * xi)$
- El producto entre dos vectores es la suma de la multiplicación de cada par de valores
- Puede ser utilizado como el producto normal de dos observaciones dadas

Kernel Polinomial

- $K(x, xi) = 1 + \text{sum}(x * xi)^d$
- Forma más generalizada del núcleo lineal
- No solo observa las características dadas por las muestras de entrada para determinar la similitud, sino también las combinaciones de éstas con n características originales y d grados de polinomio produce n^d características expandidas

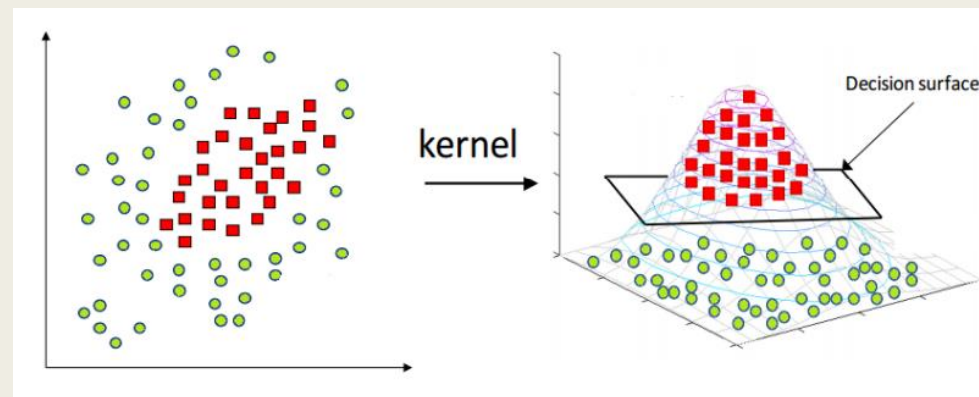
Kernel RBF (gaussiano)

- $K(x, xi) = \exp(-\text{gamma} * \text{sum}(x * xi)^2)$
- Puede mapear un espacio de entrada en un espacio dimensional infinito
- Gamma es un parámetro que va desde 0 a 1. Un valor más alto de gamma encajará perfectamente en el conjunto de datos de entrenamiento, lo que provoca un sobreajuste
- Gamma igual a 0.1 se considera un buen valor por defecto

Máquinas de soporte vectorial

Una cosa a tomar en cuenta es cuando mapeamos los datos a una dimensión superior, hay posibilidades de que podamos adaptar el modelo

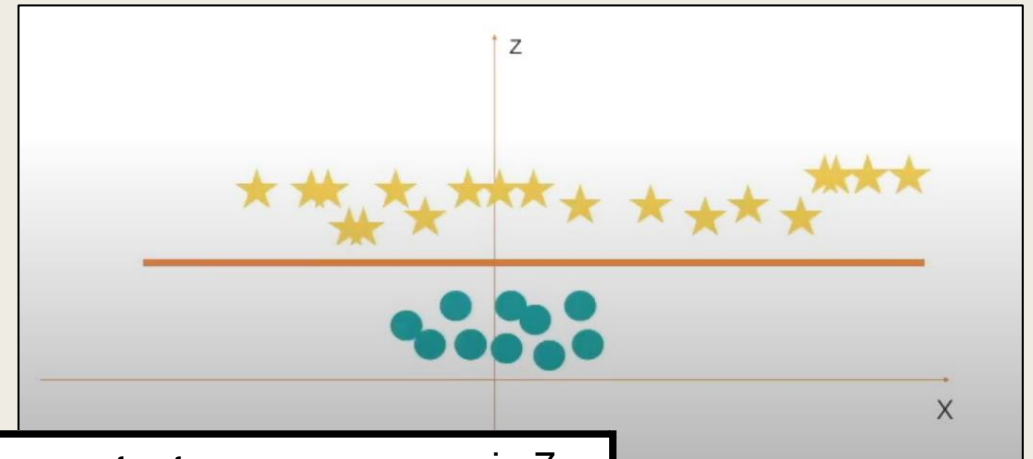
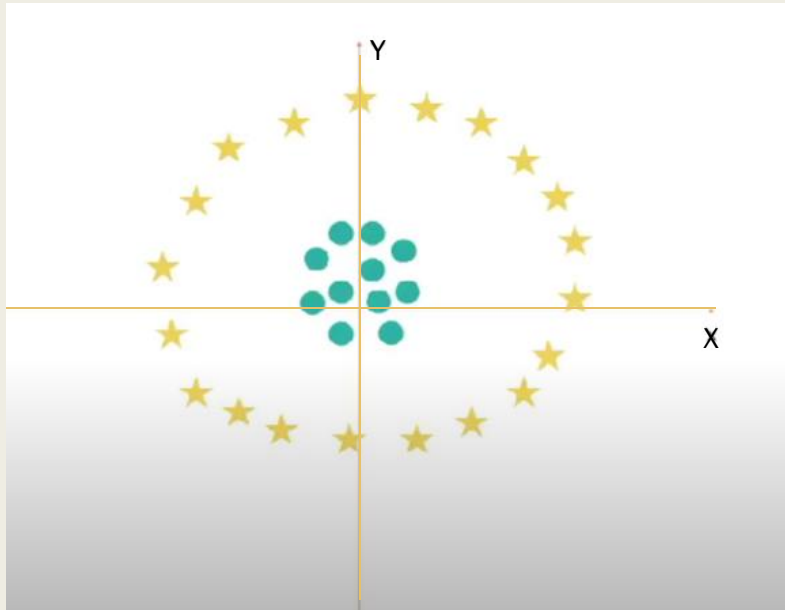
Por lo tanto, la elección de la función correcta del Kernel, incluyendo los parámetros correctos son de gran importancia



Máquinas de soporte vectorial

¿Y si el problema no es lineal?

Este algoritmo permite utilizar funciones llamadas **Kernel**, en donde se puede llevar los datos al espacio donde convierte el hiperplano en una solución no lineal lo que hace más sencillo utilizar el modelo



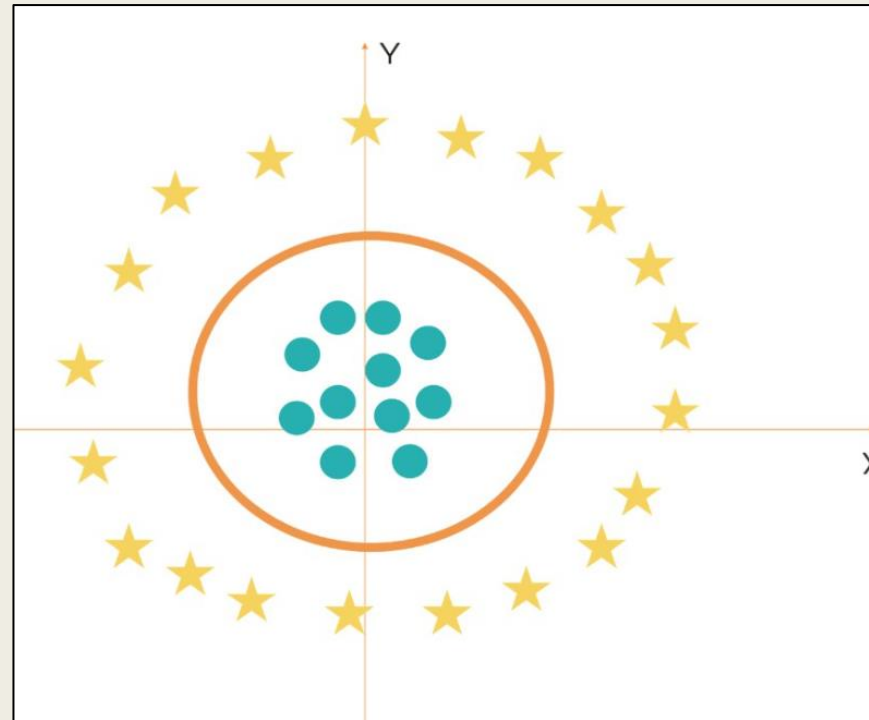
Como agregamos un nuevo eje Z, los datos se convirtieron en lineales, por lo que crear el hiperplano es más fácil

Máquinas de soporte vectorial

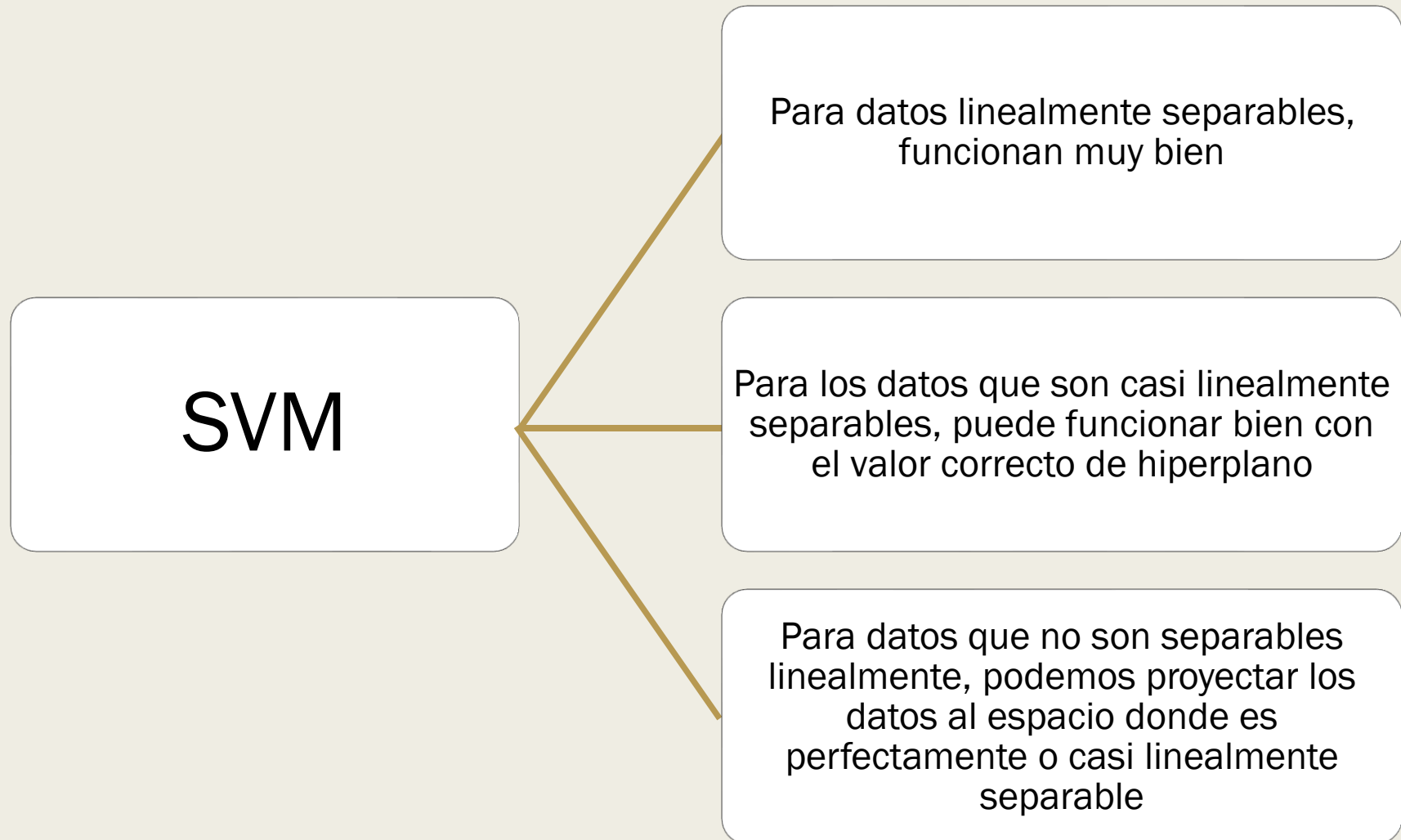
¿Y si el problema no es lineal?

Una vez es conseguida la solución, se transforma de nuevo al espacio original:

- Los datos originales, pero a su vez el hiperplano



Máquinas de soporte vectorial



Máquinas de soporte vectorial

Las SVM son importantes debido a las siguientes razones teóricas:

1. Robustas a una gran cantidad de variables y a una cantidad pequeña de muestras.
2. Capaces de aprender modelos de clasificadores simples y muy complejos.
3. Emplean modelos matemáticos sofisticados para evitar sobreentrenamiento.

Máquinas de soporte vectorial

- Es resistente al sobreajuste, es decir, puede generalizar bien datos desconocidos
- No funciona bien en grandes conjuntos de datos, debido a que el tiempo de entrenamiento y predicción puede ser muy largo
- Es sensible a la selección del kernel y a los parámetros de ajuste, por lo que se requiere una selección correcta
- Puede ser sensible a los datos atípicos



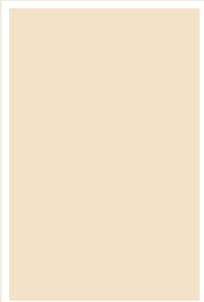
REGRESIÓN



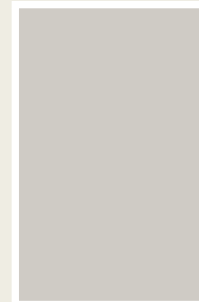
Máquinas de soporte vectorial

- Se basa en buscar la curva o hiperplano que modele la tendencia de los datos de entrenamiento y según ella predecir cualquier dato en el futuro
- Se basa en predecir valores numéricos, dado que la salida es un número real, se vuelve muy difícil predecir la información disponible, que tiene infinitas posibilidades, sin embargo, la idea principal es siempre la misma: minimizar el error, individualizar el hiperplano que maximiza el margen, teniendo en cuenta que se tolera parte del error

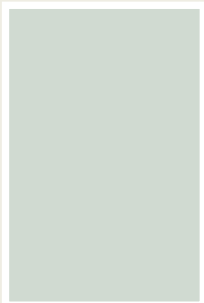
Máquinas de soporte vectorial



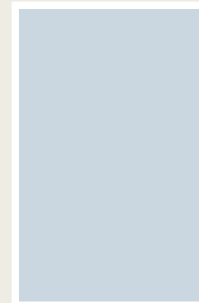
Dado que la salida es un número o real, se vuelve muy difícil predecir la información disponible, que tiene infinitas posibilidades



Minimizar el error

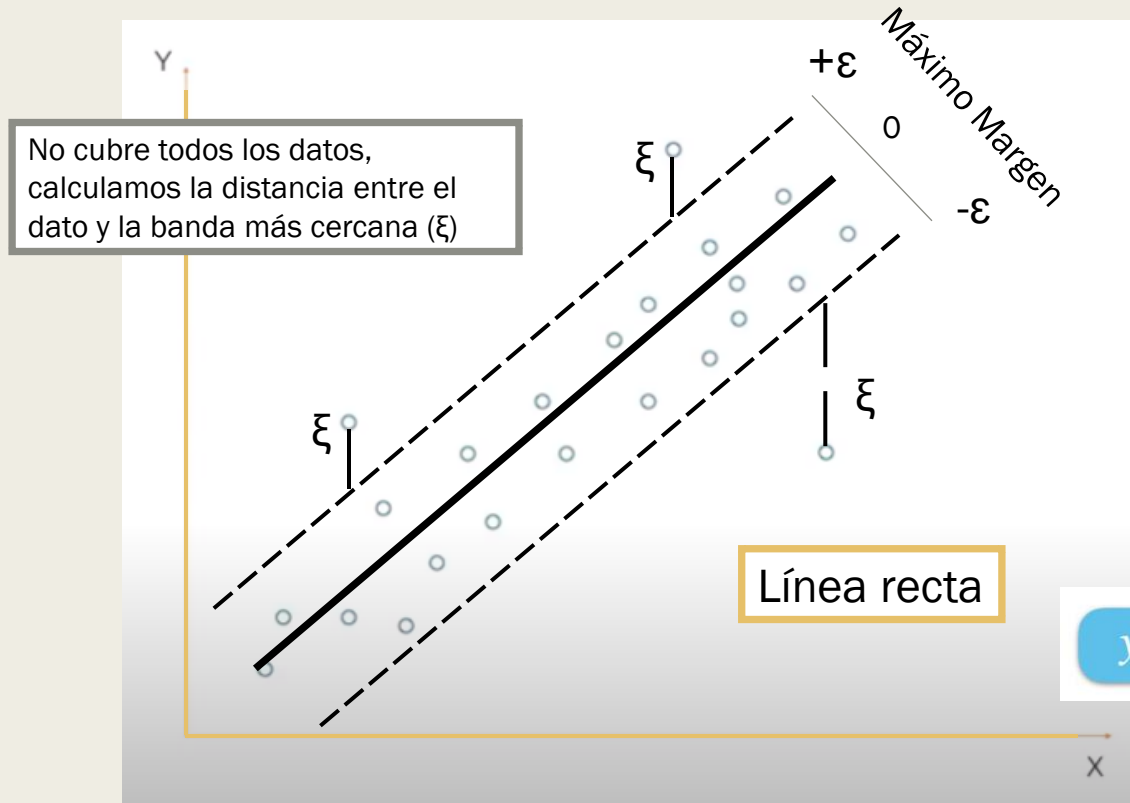


Individualizar el hiperplano que maximiza el margen



Se tolera parte del error

Máquinas de soporte vectorial



1. Construir una curva o hiperplano que modele la tendencia de los datos
2. A partir de este hiperplano, se crean dos bandas, una en el lado positivo y otra en el lado negativo (vectores de apoyo o soporte)

Toma en cuenta que la distancia desde el hiperplano será la misma en ambas

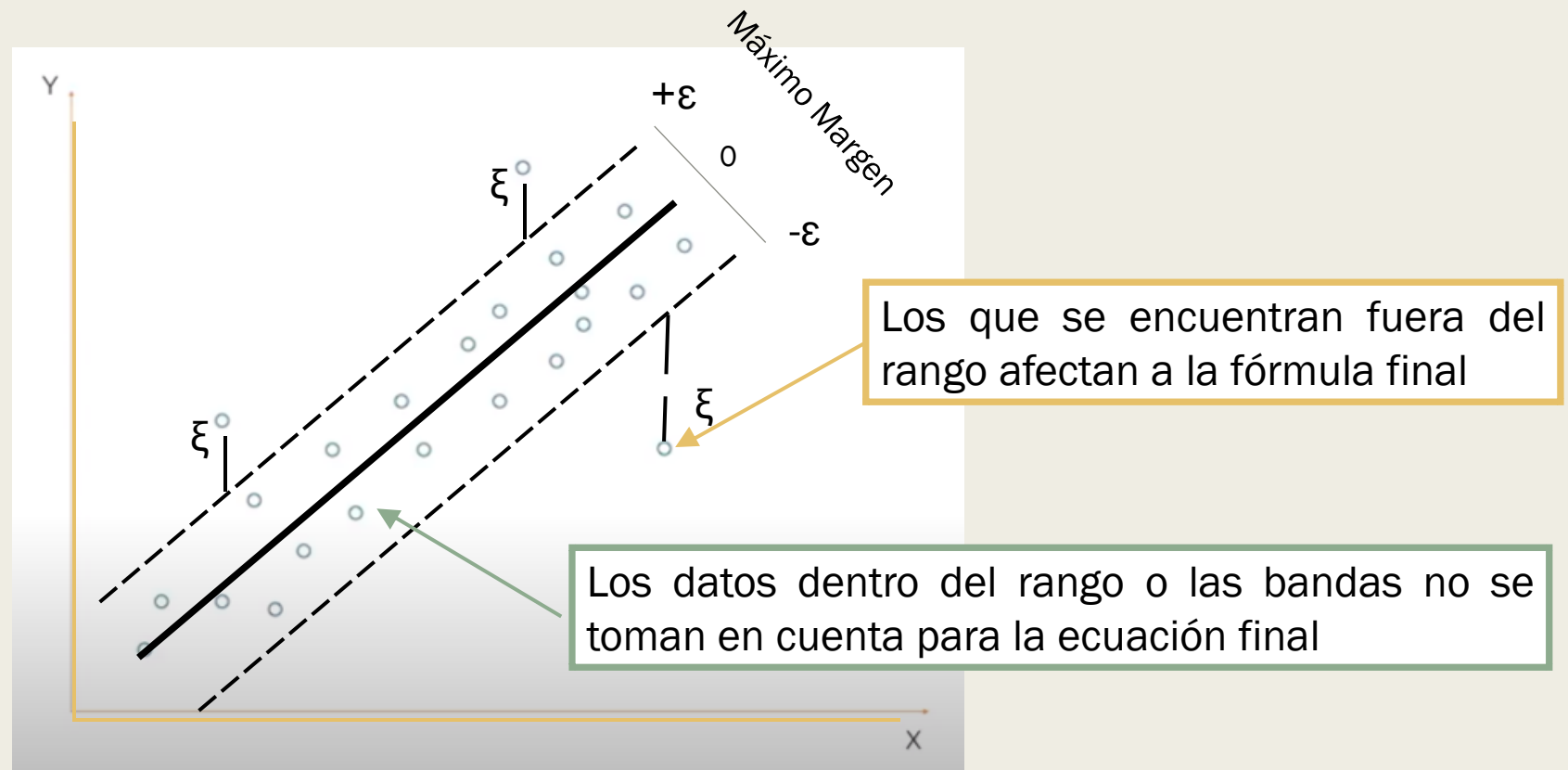
La distancia total se le llama **Máximo Margen**

Idea principal:

Que las bandas cubran la mayor cantidad posible de datos que queremos modelar en el algoritmo

Máquinas de soporte vectorial

- Justo épsilon (ξ) es lo que vamos a utilizar para nuestra ecuación del modelo



Máquinas de soporte vectorial

Vectores de soporte

Al final la formula completa para el cálculo de este algoritmo, utilizando datos lineales:

$$\min \underbrace{\frac{1}{2} \| w \|^2}_{\text{maximizar margen}} + C \underbrace{\sum_{i=1}^N (\xi + \xi^*)}_{\text{minimizar error entrenamiento}}$$

Donde:

- w es la magnitud del vector o hiperplano
- C es una constante y debe ser mayor a 0, determina el equilibrio entre la regularidad de la función y la cuantía hasta la cual toleramos desviaciones mayores que las bandas de soporte
- ξ y ξ^* son las variables que controlan el error cometido por la función de regresión al aproximar a las bandas

Máquinas de soporte vectorial

Vectores de soporte

$$\min \underbrace{\frac{1}{2} \| w \|^2}_{\text{maximizar margen}} + C \underbrace{\sum_{i=1}^N (\xi + \xi^*)}_{\text{minimizar error entrenamiento}}$$

- Si el valor de la constante C es muy grande, en el caso del límite C tendiendo a infinito estaríamos considerando que el conjunto está perfectamente representado por nuestro hiperplano predictor, es decir, ϵ tiende a 0
- Al contrario, un número demasiado pequeño para C permitirá valores de ϵ elevados, es decir, estaríamos admitiendo un número elevado de ejemplos mal representados

Máquinas de soporte vectorial

- Recordamos que en Regresión Lineal el modelo era simplemente una línea, y con ella íbamos a tratar de cubrir la mayor parte de nuestros datos para de esta forma reducir el error
- En este caso, no es una línea sino un rango, la cual cubre la mayor cantidad de datos y todos los datos que se encuentren fuera de este rango o banda, son los errores que puede haber
- Por lo que tenemos que considerarlos en la fórmula final

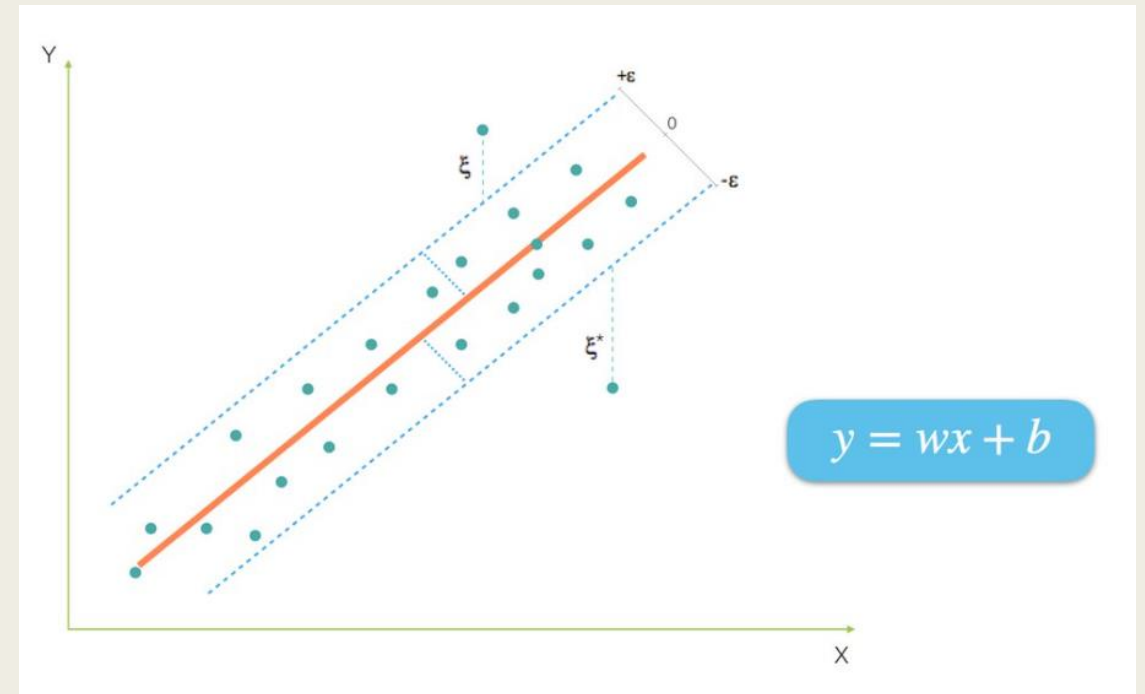
Máquinas de soporte vectorial

Se basa en buscar la curva, con su respectivo máximo margen, que modele la tendencia de los datos y, según ella, predecir cualquier otro dato en el futuro

Esta curva siempre viene acompañada con un margen, el cual tendrá el mismo comportamiento o forma de la curva

Una vez que se tenga la curva y los márgenes, se podrá visualizar los datos que quedaron por fuera

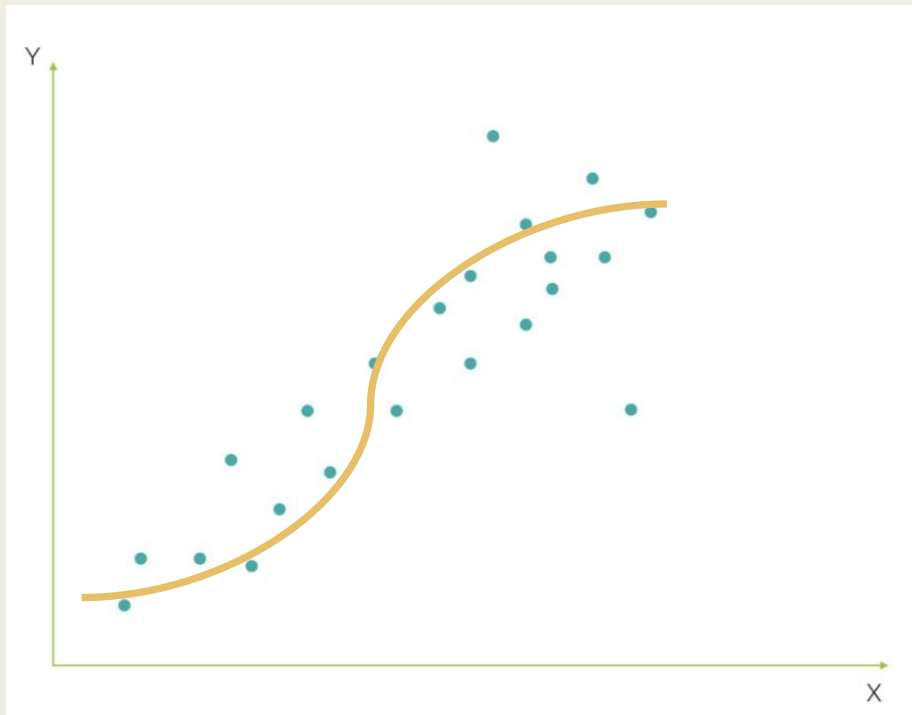
Todo lo que se encuentra fuera de este rango es considerado error



Máquinas de soporte vectorial

¿Y si el problema no es lineal?

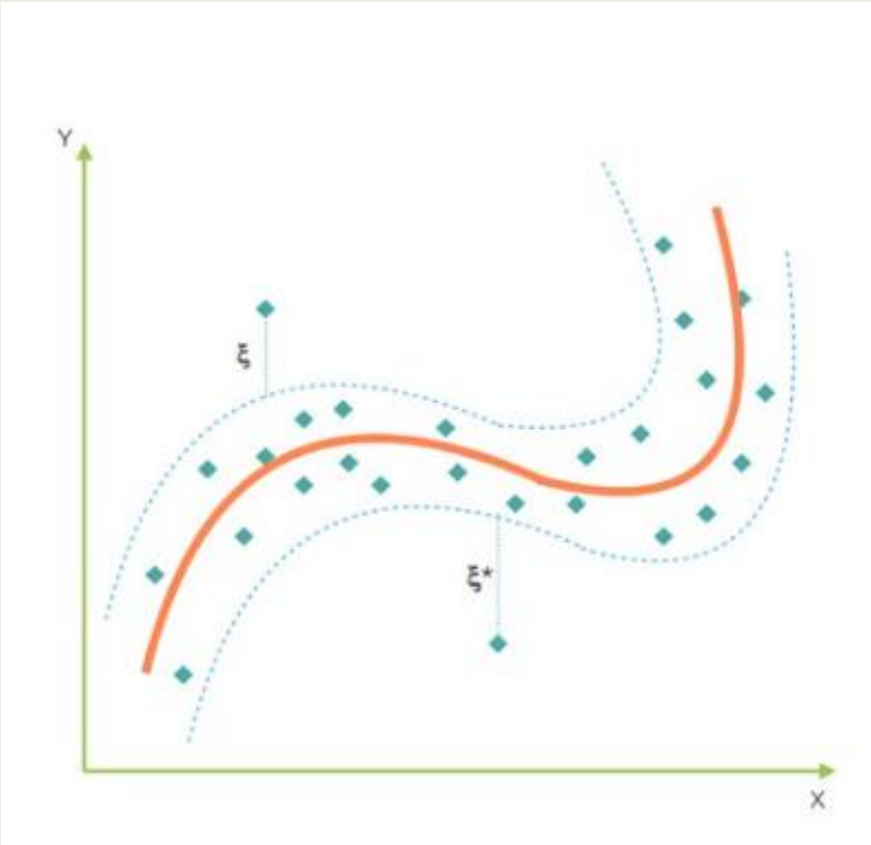
El algoritmo se basa en buscar la curva que mejor se adapte a los datos, en este caso el hiperplano será una curva que siempre buscará adaptarse a los datos



No es necesario buscar otro algoritmo o ecuación, como en la Regresión Lineal o no Lineal, simplemente el algoritmo se adapta a los datos

Máquinas de soporte vectorial

¿Y si el problema no es lineal?

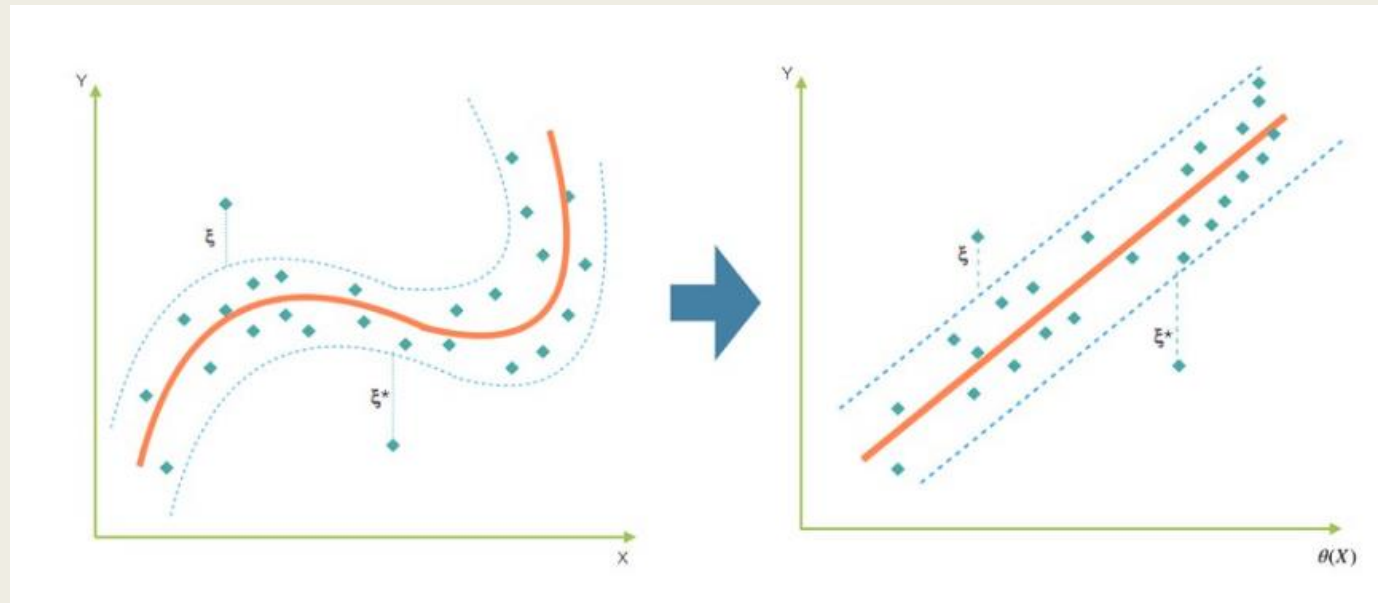


Las bandas o rangos tendrán el mismo comportamiento de la curva, obviamente respetando las distancias para abarcar el mayor número de datos

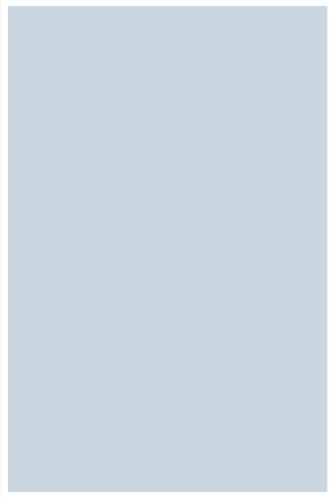
Todos aquellos datos que se encuentren fuera del rango o banda son considerados errores y se les busca el valor de ϵ para incluirlos en la formula

Máquinas de soporte vectorial

- Por su parte, para datos no lineales el procedimiento es exactamente igual, la diferencia es que se implementa un Kernel para convertir los datos lineales



Máquinas de soporte vectorial



El rendimiento del algoritmo SVM para regresión depende de una buena configuración de los parámetros C y de los del Kernel

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi + \xi^*)$$

Se complica por el hecho de que la complejidad del modelo SVM depende de estos parámetros. Podemos utilizar **SCIKIT LEARN**

Máquinas de soporte vectorial

- Este algoritmo tiene algunas premisas que se deben cumplir:

