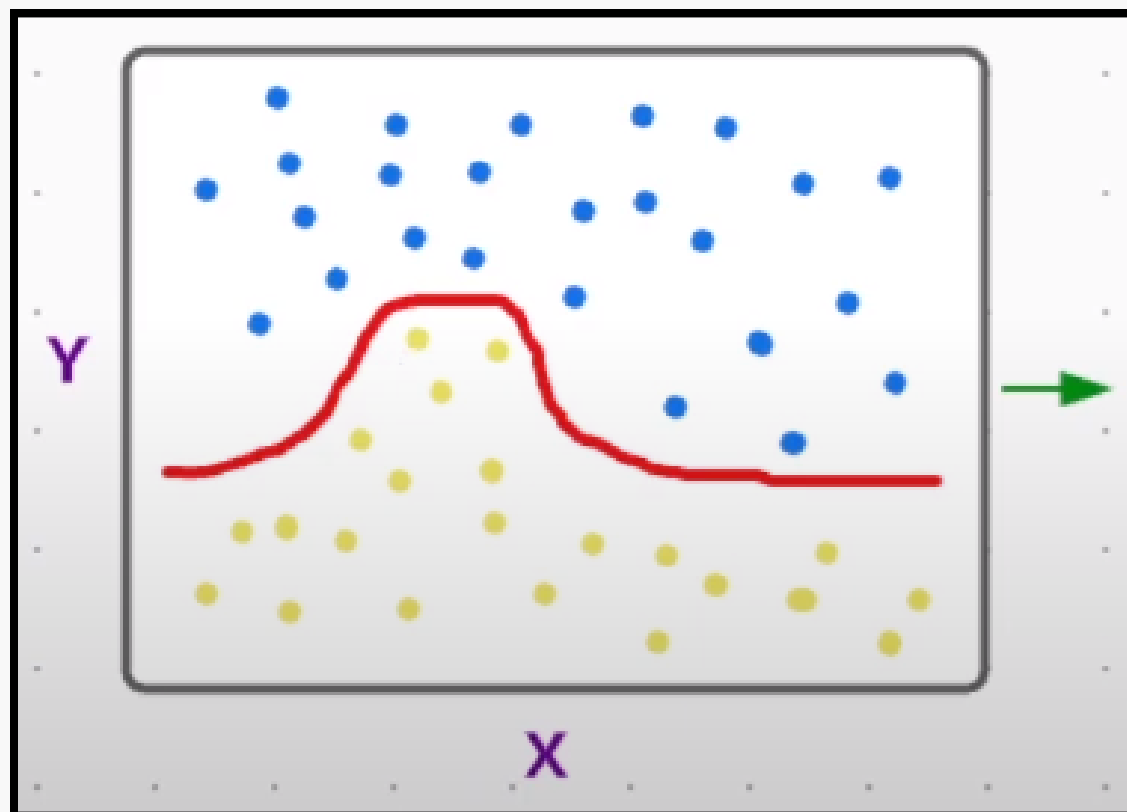


Minería de Datos

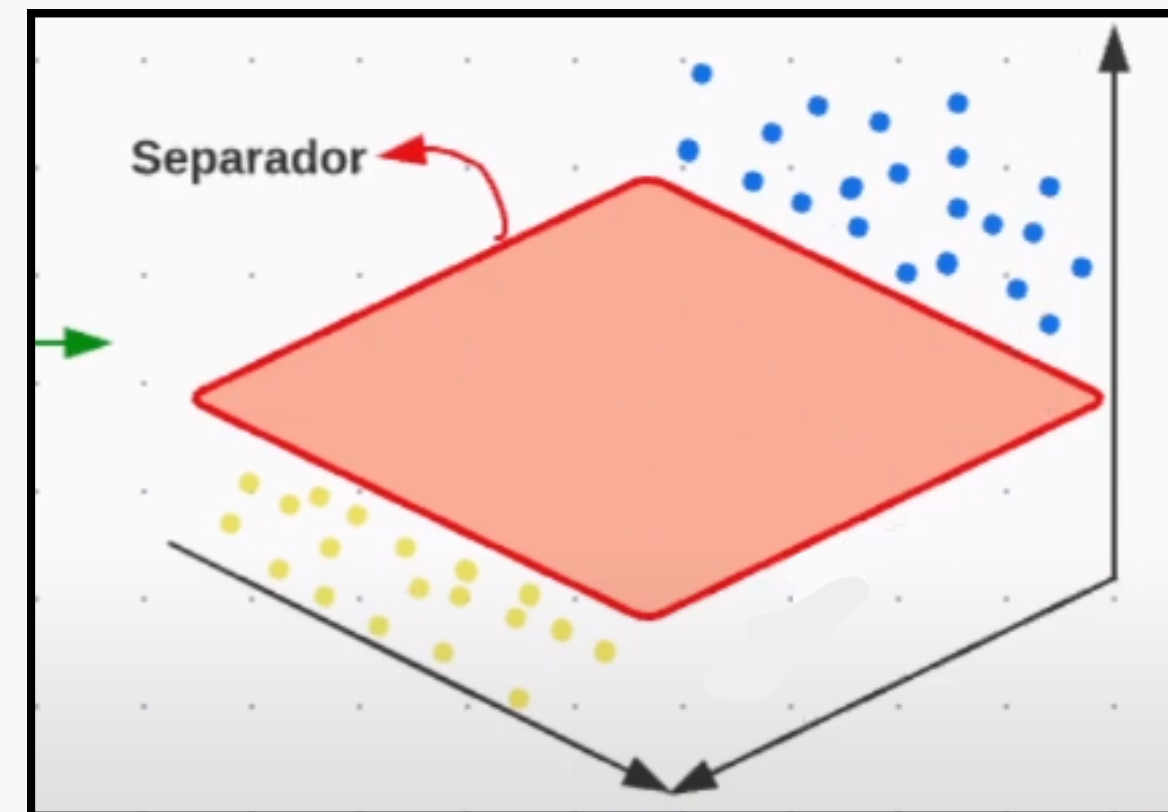
MÁQUINAS DE SOPORTE VERCTORIAL

En términos simples, se trata de un algoritmo supervisado que se encarga de clasificar datos encontrando un separador

¿Cómo funciona?



Debido a que no se pueden clasificar en este plano, **entonces vamos a transformar los datos para llevarlos a un plano tridimensional**

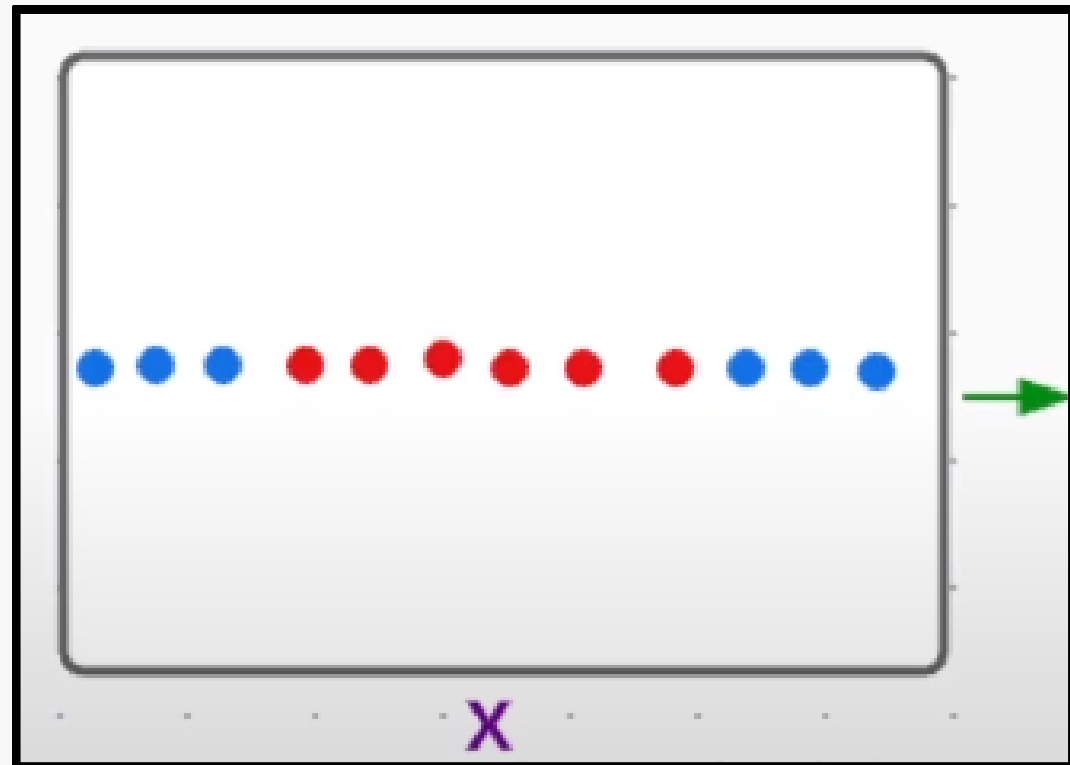


Partimos de un conjunto bidimensional de datos que no son linealmente separables

Estos datos están en un plano bidimensional

Al transformar los datos en una dimensión mayor, ya podemos clasificarlo con un hiperplano o separador

¿Cómo se transforman los datos?



Supongamos este conjunto de datos en un plano unidimensional

Como se puede observar no son linealmente separables

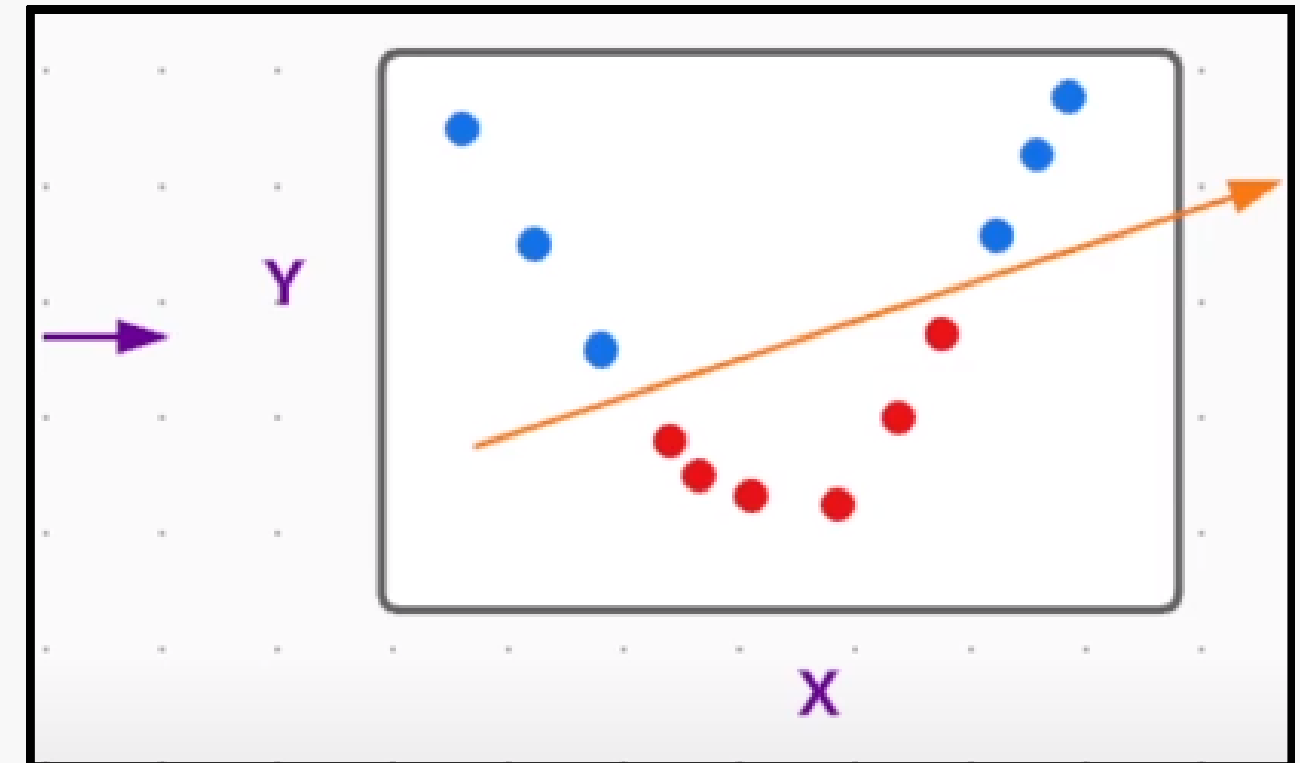
Para hacer la transformación necesitamos una función de núcleo, que puede ser:

Polinomial

Lineal

RBF

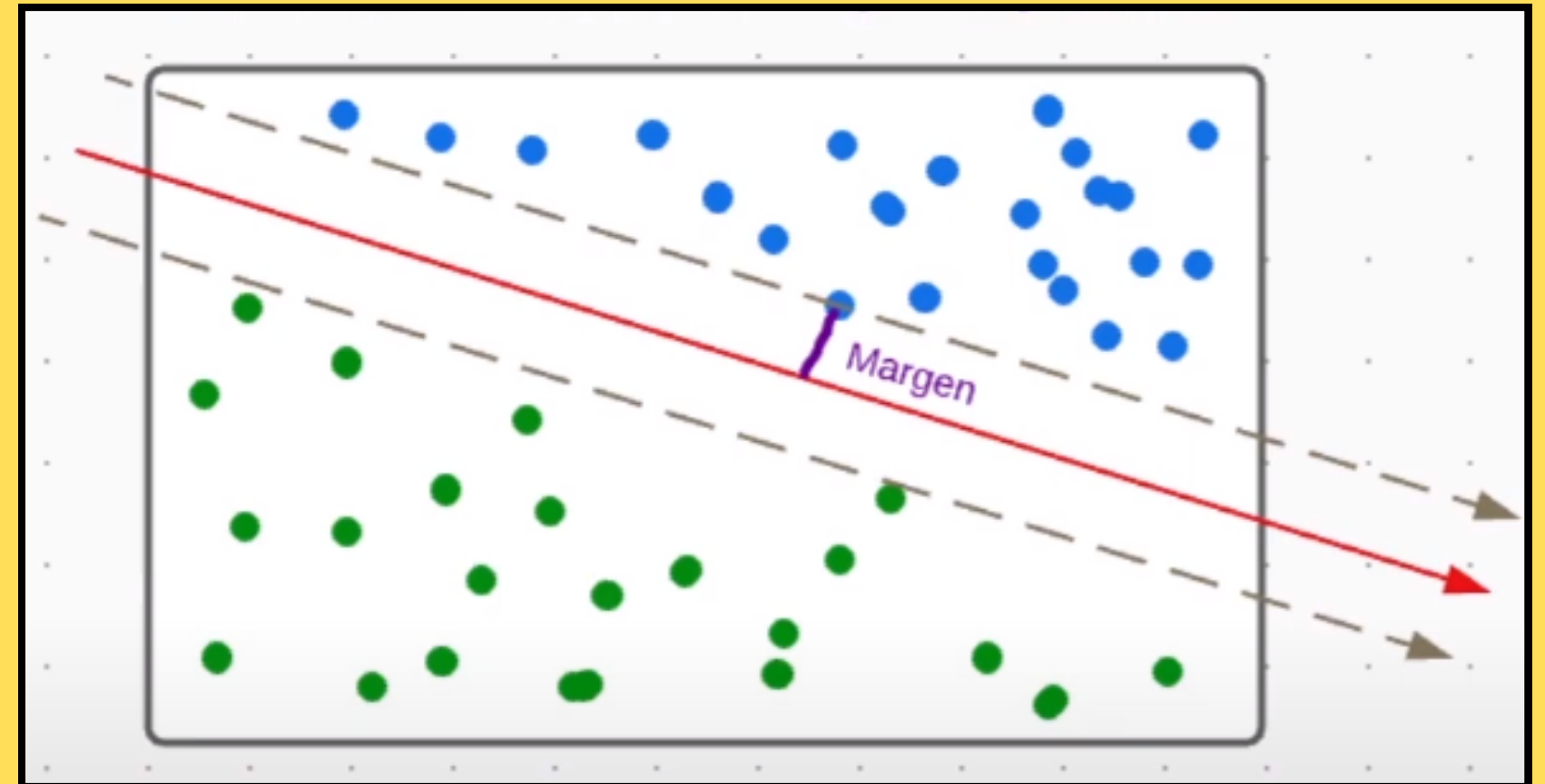
Sigmoide



Cada función es diferente, y con ella vienen sus PROS y CONTRAS, así que lo recomendable es entrenar nuestro modelo con diferentes funciones de núcleo y ver cual es la que brinda mejores resultados a nuestra problemática

¿Cómo encontramos el mejor separador?

El mejor separador es aquel que represente mejor la separación o margen entre las clases, ya que los puntos más cercanos a nuestro separador serán nuestros datos de entrenamiento para el soporte vectorial



Kernels

Kernel Lineal:

La función de kernel lineal se utiliza para problemas de clasificación lineal. La función del kernel lineal es simplemente el producto escalar entre dos vectores de características. Para dos vectores x y y , el kernel lineal se define como

$$K(x, y) = (x \cdot y + c)^d,$$

Kernel Polinomial:

El kernel polinomial permite que las SVM realicen clasificaciones polinomiales. La función del kernel polinomial de grado d se define como

$$K(x, y) = x \cdot y.$$

donde d es el grado del polinomio y c es una constante.

Kernels

Kernel RBF:

El kernel RBF es uno de los kernels más utilizados y es adecuado para problemas de clasificación no lineal. La función del kernel RBF se define como

$$K(x, y) = \exp(-\gamma \cdot ||x - y||^2)$$

Donde:

- Gamma es un parámetro positivo que controla la forma de la función de base radial. Valores más grandes de gamma hacen que el efecto de $||x - y||^2$ sea más pronunciado y, por lo tanto, el límite de decisión de la SVM se ajusta más estrechamente a los puntos de datos en el espacio de características. Valores demasiado grandes de gamma pueden conducir a un sobreajuste.

Kernel Sigmoid:

- El kernel sigmoide realiza transformaciones no lineales utilizando funciones sigmoideas. La función del kernel sigmoide se define como

$$K(x, y) = \tanh(\alpha x \cdot y + c)$$

donde α y c son parámetros ajustables.

Comparación de los Kernel

Kernel	Ventajas	Desventajas
Lineal	<ul style="list-style-type: none">-Eficiente en conjuntos de datos lineales.-Menos propenso al sobreajuste.	<ul style="list-style-type: none">-No puede capturar patrones no lineales.-Limitado en la capacidad de clasificación.
Polinomial	<ul style="list-style-type: none">-Puede manejar patrones no lineales.-Flexibilidad para límites de decisión complejos.	<ul style="list-style-type: none">-Sensible al grado del polinomio.-Puede sobreajustarse fácilmente.-Computacionalmente costoso para grados altos.
RBF	<ul style="list-style-type: none">-Efectivo para datos no lineales y de alta dimensionalidad.-Puede capturar patrones complejos.	<ul style="list-style-type: none">-Requiere ajuste cuidadoso del parámetro gamma para evitar sobreajuste.-Costoso computacionalmente en conjuntos de datos grandes.-Difícil de interpretar.
Sigmoidal	<ul style="list-style-type: none">-Puede manejar patrones no lineales.-Ampliamente utilizado en redes neuronales.	<ul style="list-style-type: none">-Sensible a la elección de los parámetros.-Puede requerir ajuste fino para obtener buenos resultados.

Ventajas

Funciona bien en un conjunto de datos que tiene muchas características.

Proporciona un claro margen de separación.

Es muy eficaz para conjuntos de datos donde la cantidad de características es mayor que los puntos de datos.

Puede especificar diferentes funciones del kernel para establecer un límite de decisión adecuado.

Desventajas

Requiere un tiempo de entrenamiento muy elevado, por lo que no se recomienda para conjuntos de datos grandes.

Es muy sensible a los valores atípicos.