



Módulo 5 – Aprendizaje de Máquina Supervisado

Máquina de Soporte Vectorial

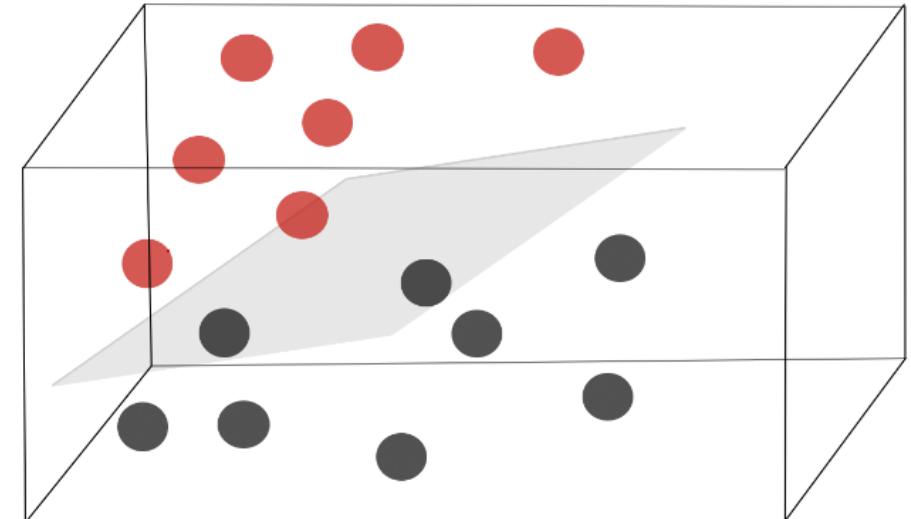
Especialización en Ciencia de Datos

Máquina de Soporte Vectorial



Support Vector Machine

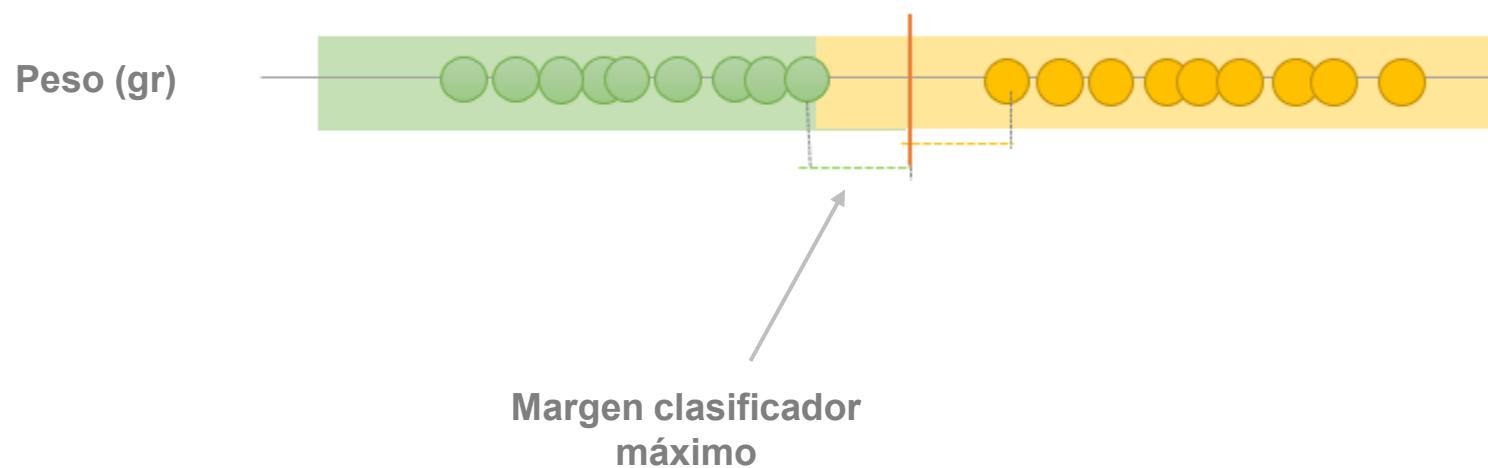
- SVMs son modelos de aprendizaje supervisado con algoritmos que analizan la data y reconocen patrones, usados para problemas de **clasificación y regresión**.
- Dado un set de entrenamiento etiquetado para una o dos categorías, un algoritmo de SVM construye un modelo que asigna nuevos ejemplos en una categoría o en la otra, haciendo una clasificación binaria lineal no probabilística.
- El algoritmo SVM es popular debido a su capacidad para **manejar eficientemente grandes conjuntos de datos** y su capacidad para manejar tanto datos **linealmente separables como no linealmente separables**. También, es **robusto en presencia de datos atípicos** y es capaz de generalizar bien a datos no vistos.



SVM

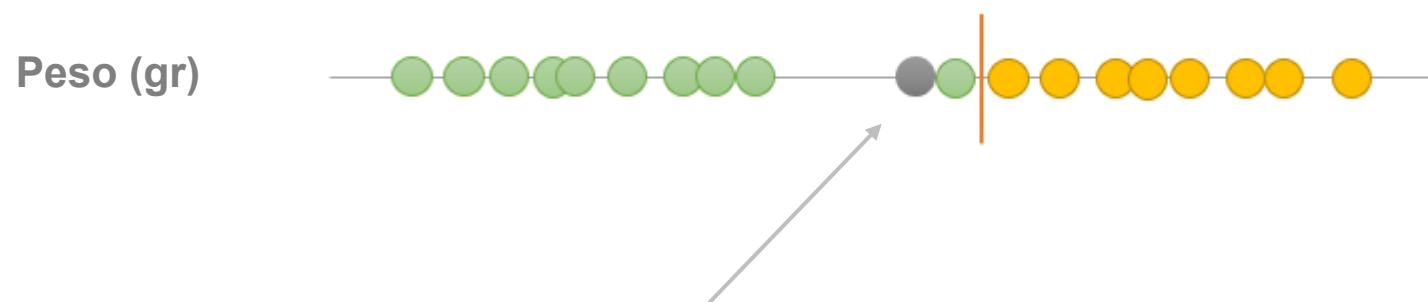
Margen Clasificador

Suponga que tenemos dos cítricos (limones y naranjas), y necesitamos hacer un clasificador que permita distinguir entre cada una de ellas a partir de su peso.



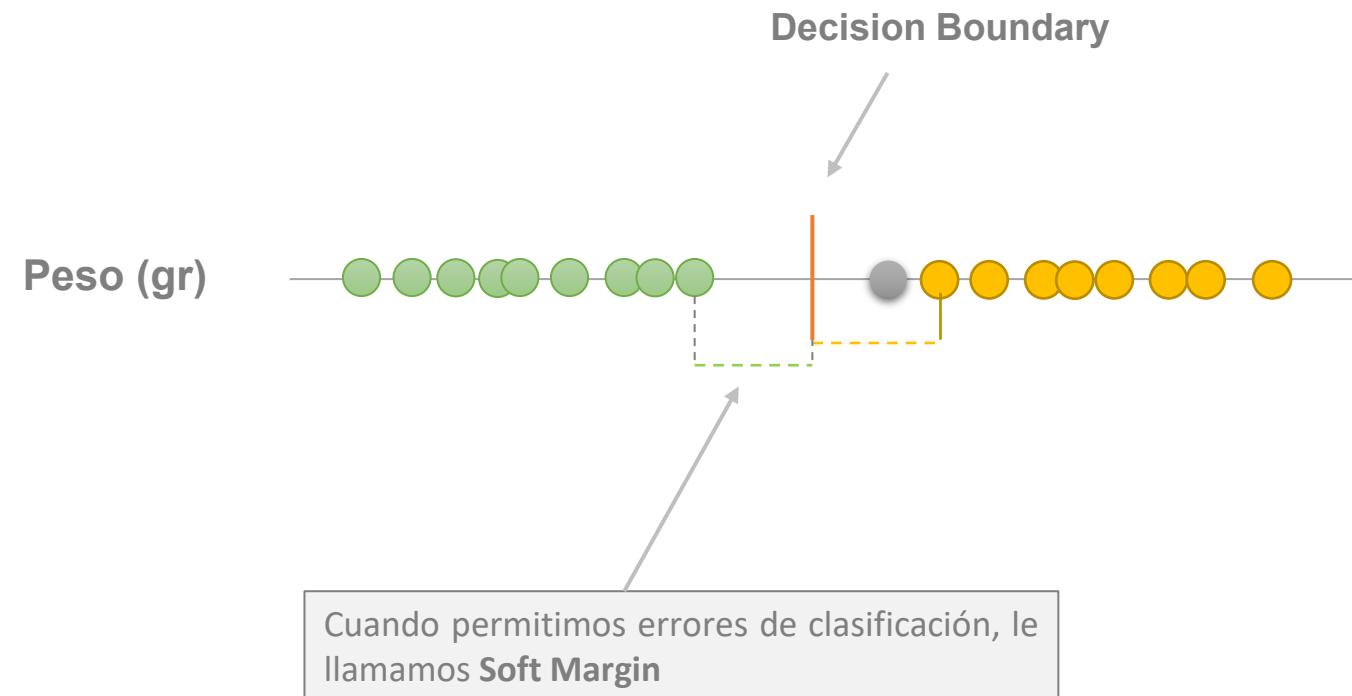
Margen Clasificador

Suponga que una nueva medición de un limón, queda posicionado en la cercanía del grupo de las naranjas, lo cual hace que el margen de clasificación máxima sea más estrecho.



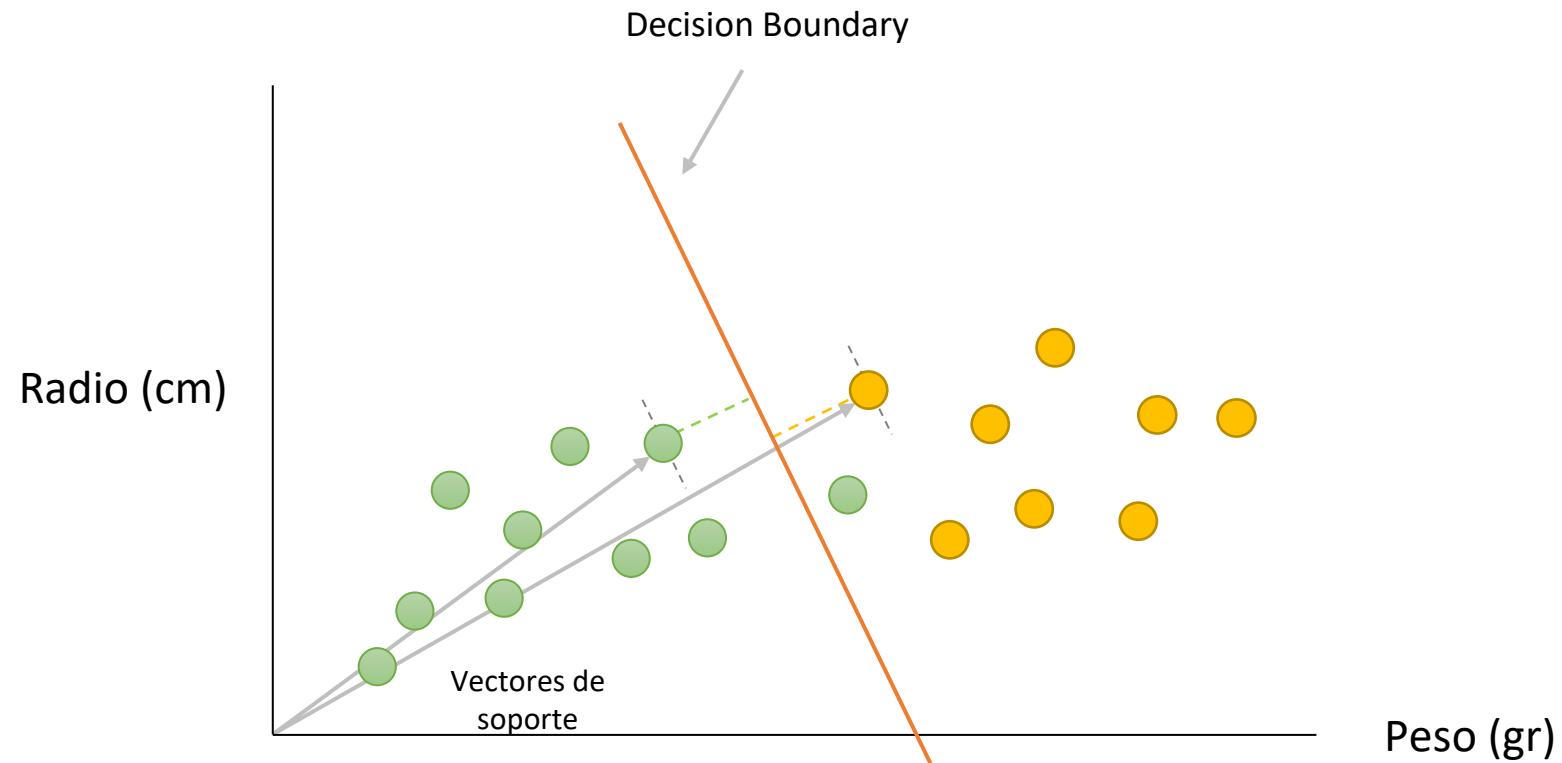
Vector de Soporte

Si tomamos un umbral de error de clasificación (desvío) que lo haga menos sensible a los datos de entrenamiento, el algoritmo preformará mejor cuando se enfrente a nuevos datos (menor varianza). En este caso, el límite de decisión es un punto.



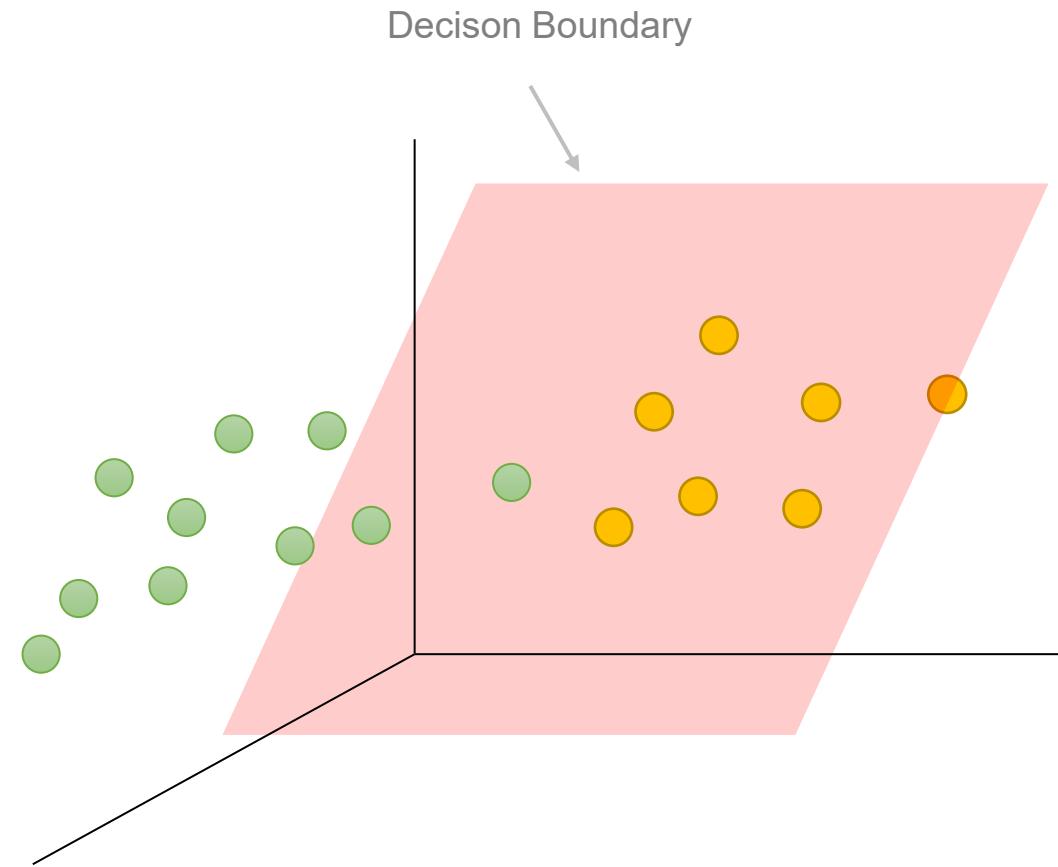
Vectores de Soporte

Llevemos el problema a dos dimensiones. En este caso, el límite de decisión es una línea recta.



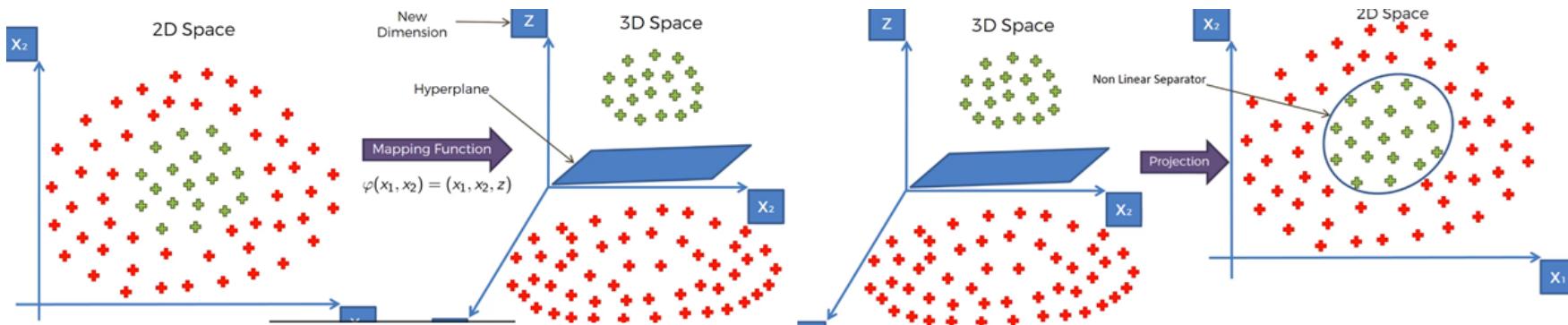
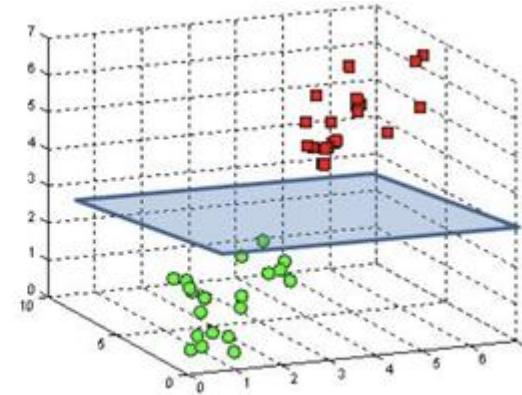
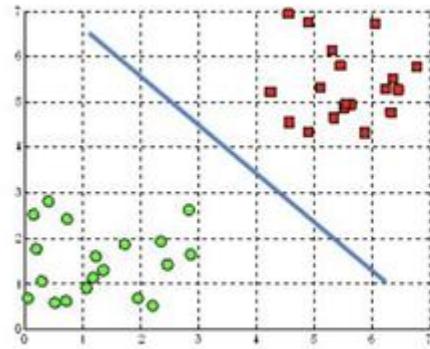
Vectores de Soporte

Llevemos el problema a Tres dimensiones. En este caso, el límite de decisión es un plano.

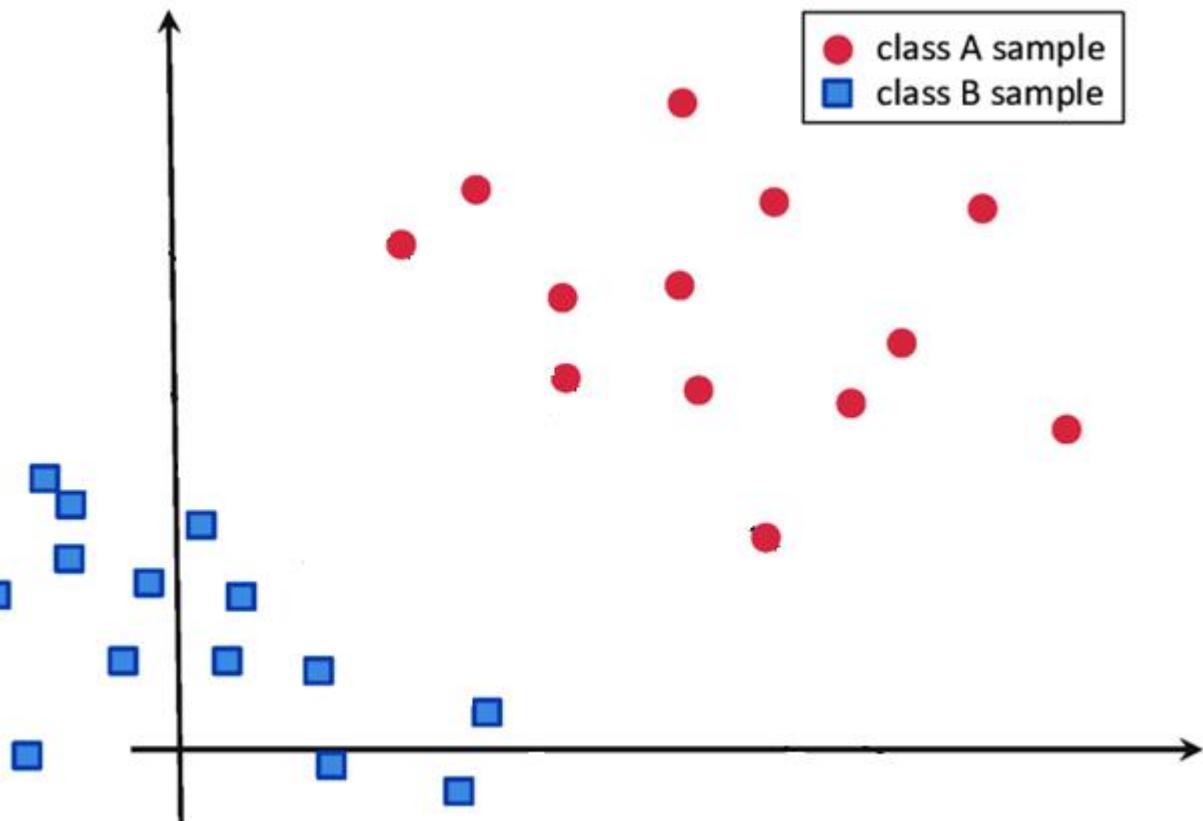


El Hiperplano

En geometría, un hiperplano es un subespacio cuyas dimensiones son una menos que el espacio al cual pertenece. Si el espacio es de 3 dimensiones, entonces su hiperplano es un plano de 2 dimensiones, mientras que si el espacio es 2-dimensional, su hiperplano es una línea de 1 dimensión.

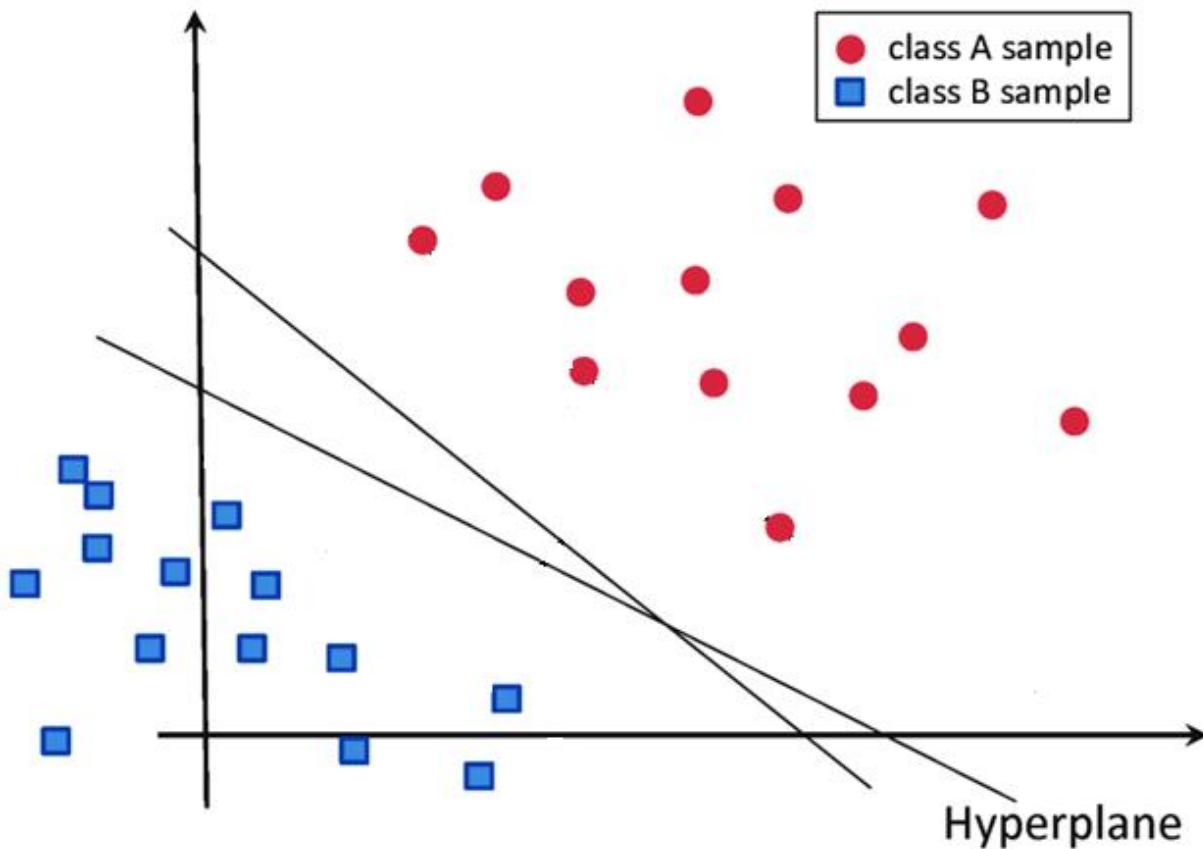


Máquina de Soporte Vectorial



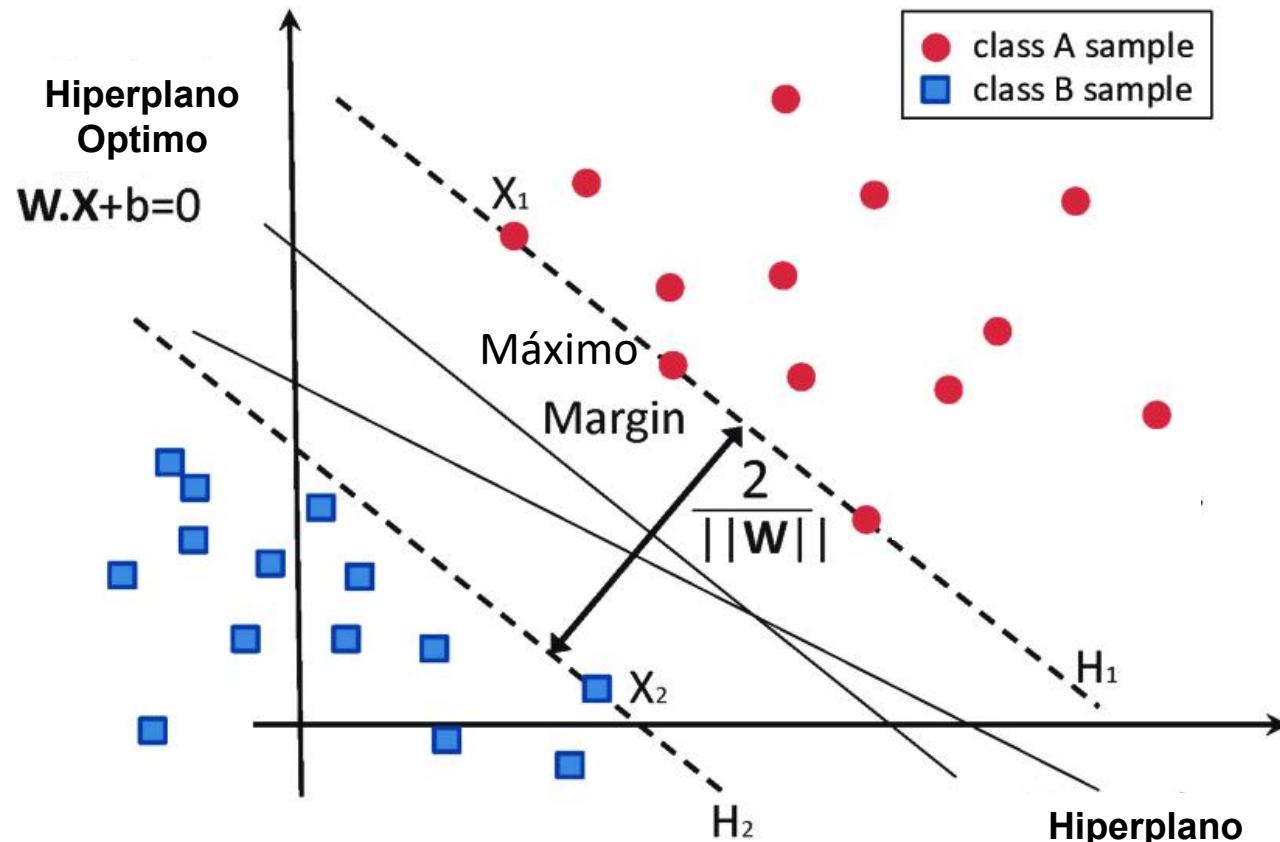
La clasificación se logra encontrando el hiperplano que maximiza los márgenes de clasificación para ambas clases.

Máquina de Soporte Vectorial



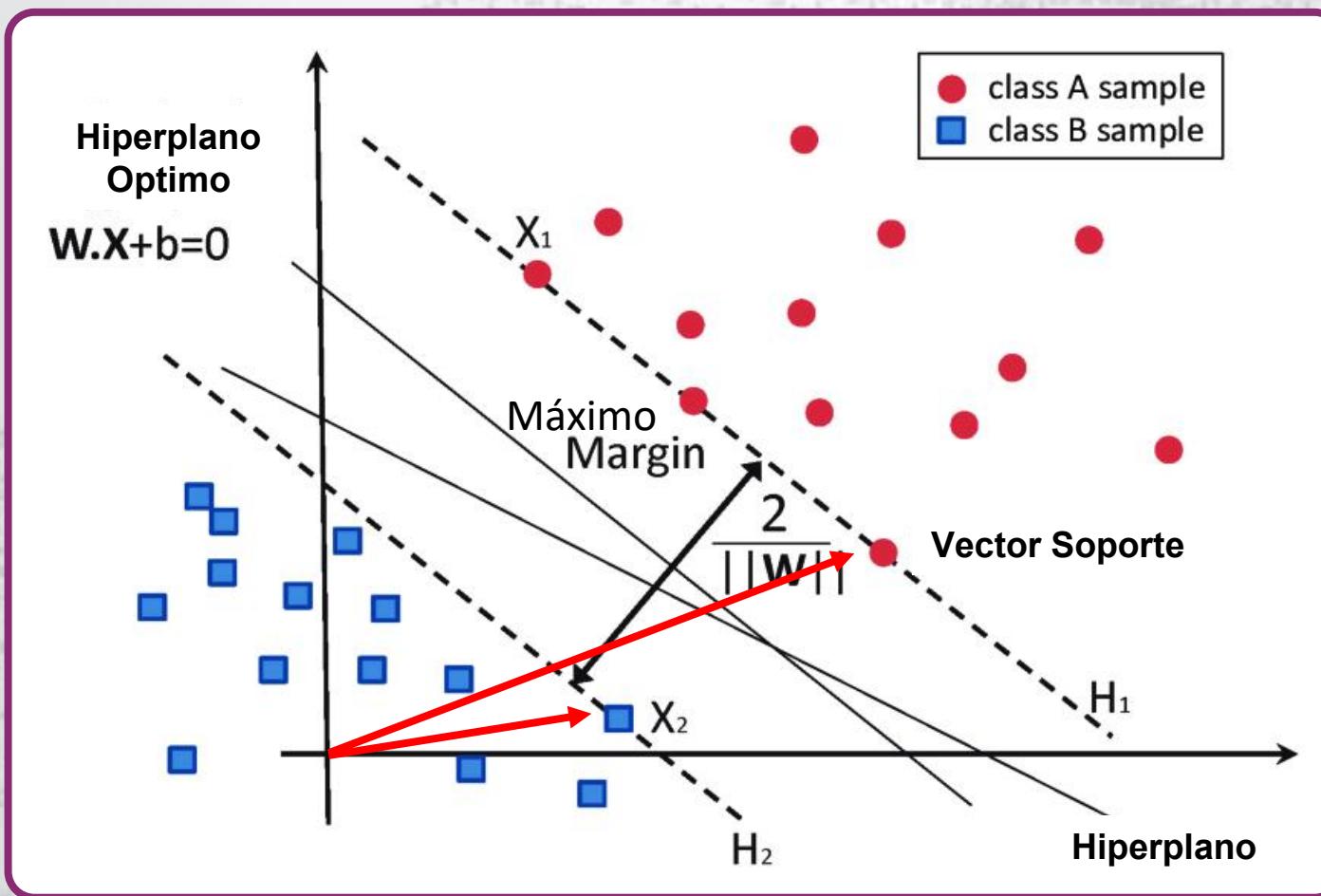
El hiperplano encontrado maximiza la distancia de los puntos más cercanos en todas las clases.

Máquina de Soporte Vectorial



El hiperplano óptimo es aquel que proporciona el mayor margen.

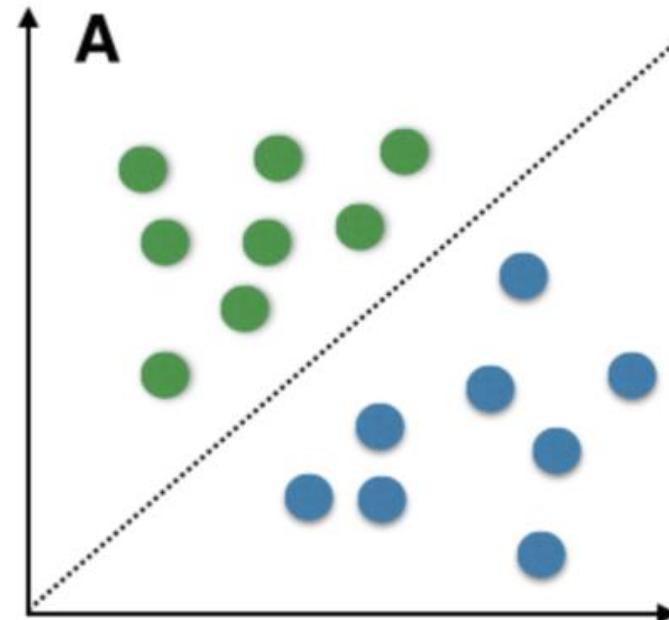
Máquina de Soporte Vectorial



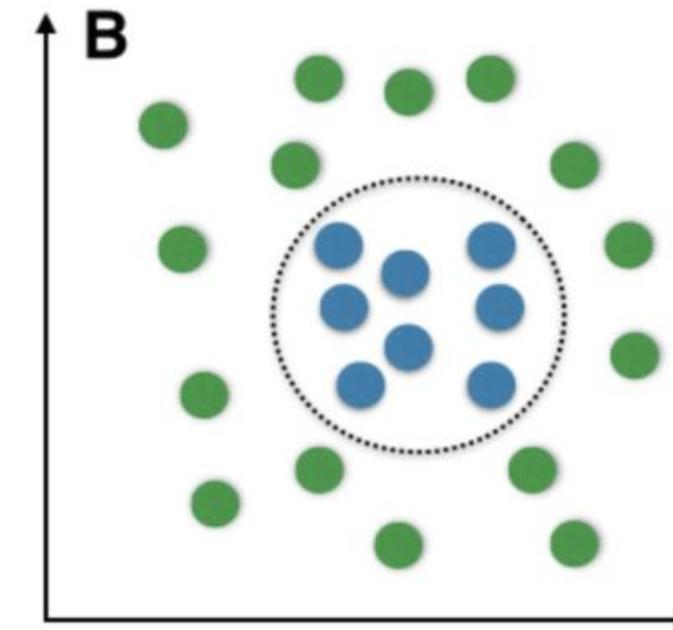
En color rojo se pueden apreciar los vectores que indican los puntos que soportan el hiperplano óptimo.

Linealmente Separable

Hay situaciones en la cual es posible separar los datos con un hiperplano. En ese caso, se dice que el problema es linealmente separable.



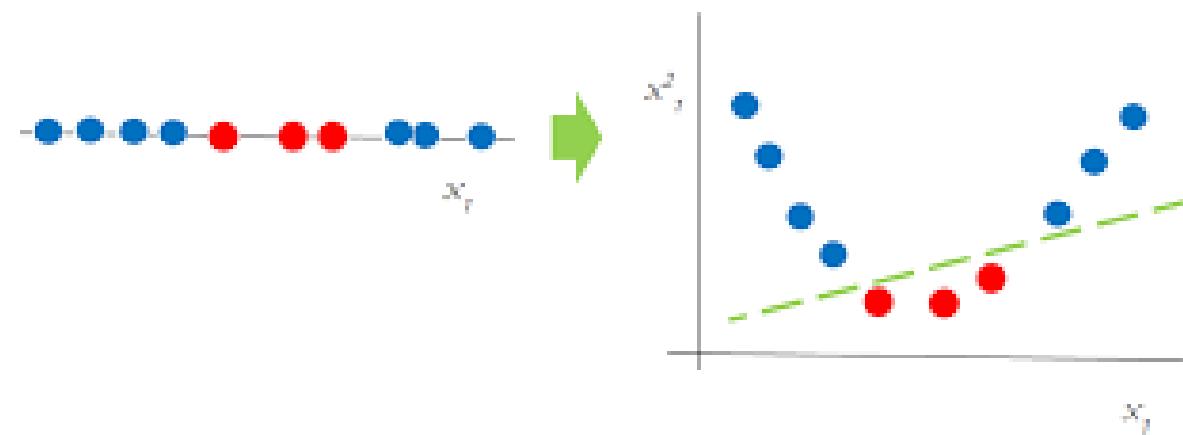
Linealmente separable



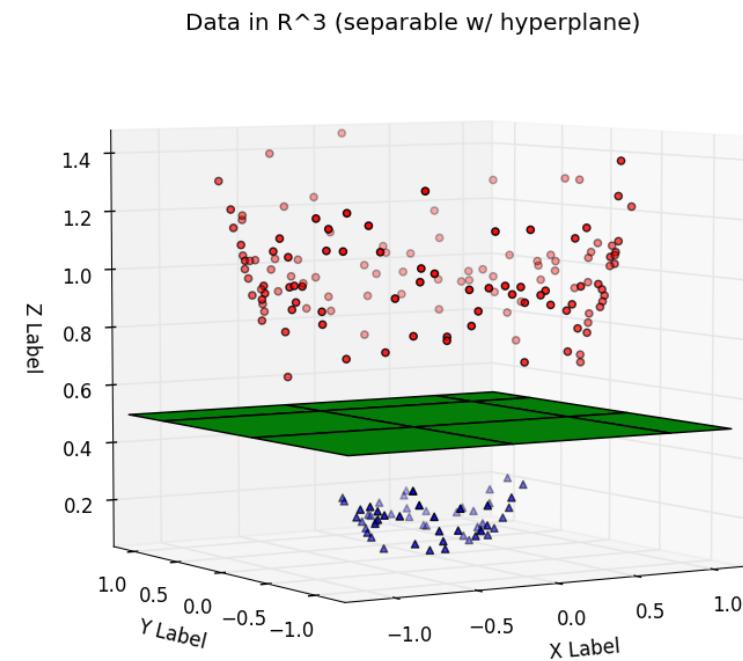
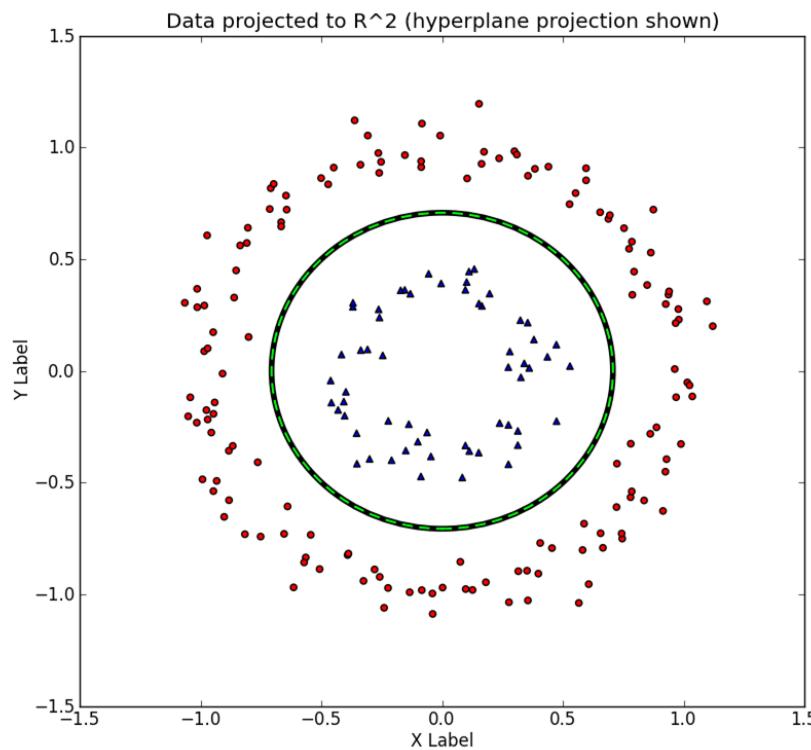
Linealmente no separable

Mapeo a una Dimensión Superior

Cuando un problema es linealmente no separable, es posible llevar los datos a una dimensión superior, con una adecuada función de transformación, para después encontrar el hiperplano que maximiza el margen de distancia de clasificación.



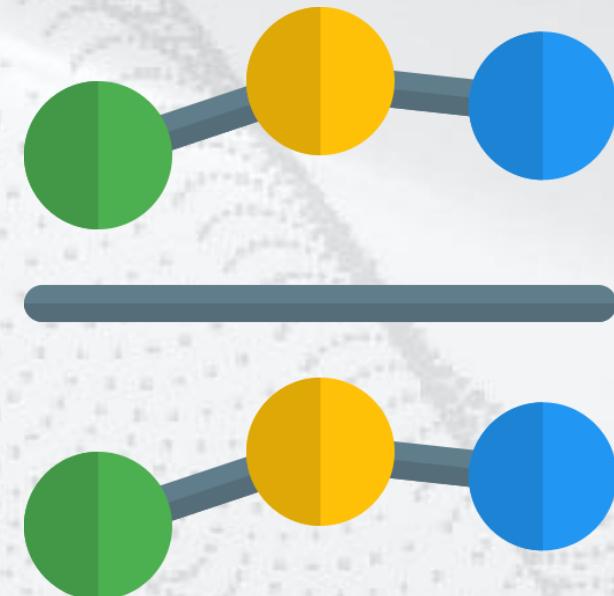
Mapeo a una Dimensión Superior



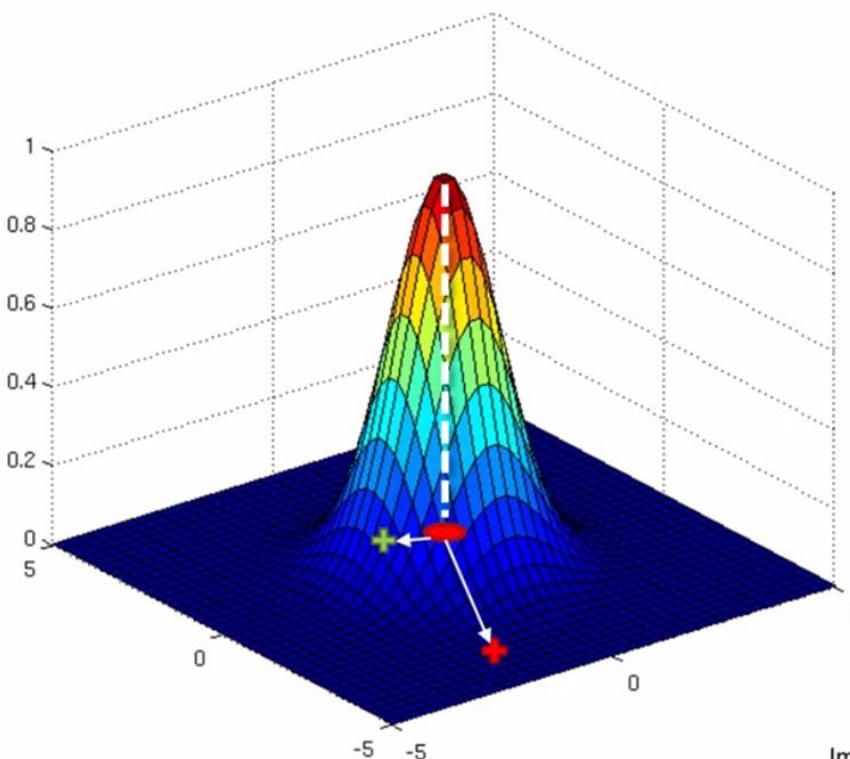
El Truco del Kernel

Si el set de datos que se está analizando es muy grande, es decir, tiene muchas mediciones y dimensiones, entonces realizar el mapeo a una dimensión superior puede ser costoso en procesamiento computacional.

Sin embargo, se puede tener una aproximación realizando lo que se denomina el “truco del kernel”.



Kernel RBF Gausiano

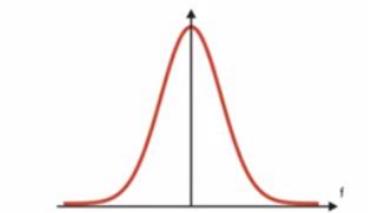


$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$

RBF: Radial Basis Function
(Función Base Radial)

Image source: <http://www.cs.toronto.edu/~duvenaud/cookbook/index.html>

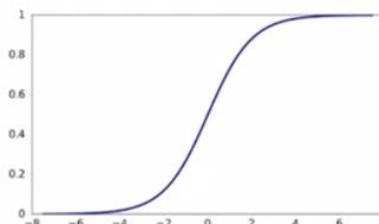
Tipos de Kernel



Gaussian RBF Kernel

(Kernel Gausiano)

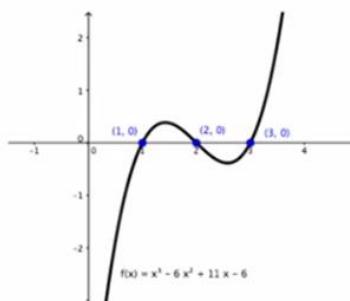
$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$



Sigmoid Kernel

(Kernel Sigmoídeo)

$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$



Polynomial Kernel

(Kernel Polinomial)

$$K(X, Y) = (\gamma \cdot X^T Y + r)^d, \gamma > 0$$

Tunning de Parámetros

- ④ **Kernel:** el parámetro Kernel selecciona el tipo de hiperplano usado para separar la data. Utilizando kernel ‘linear’ va a usar un hiperplano lineal, ‘rbf’ y ‘poly’ usa un no-linear hiperplano.
- ④ **Gamma:** Gamma es un parámetro para los hiperplanos no-lineales. Define la influencia que tienen los puntos lejanos y cercanos para la determinación del hiperplano. Mayor gamma apunta a que los puntos cercanos tienen más influencia en la determinación del hiperplano, menor gamma apunta a que puntos más lejanos tienen mayor influencia. Gamma alto podría causar sobreajuste.
- ④ **Regularización:** el parámetro C es el parámetro de regularización, el cual penaliza el término de error en la función de costo. Controla el trade-off entre un límite de decisión suave y la correcta clasificación de los puntos.



Ventajas

- 1. SVM puede manejar eficazmente conjuntos de datos de alta dimensionalidad.
- 1. Es efectivo en espacios de características con muchos datos faltantes.
- 1. SVM es una técnica muy potente para problemas de clasificación binaria y multi-clase.
- 1. Es robusto en presencia de datos atípicos y puede manejar grandes conjuntos de datos.



Desventajas

- 1. SVM puede ser computacionalmente costoso en grandes conjuntos de datos, especialmente si se utiliza un kernel no lineal.
- 1. Es difícil de interpretar y explicar cómo funciona el modelo.
- 1. Si los datos de entrada tienen ruido o solapamiento, SVM puede ser sensible a la selección de parámetros, lo que puede requerir ajustes cuidadosos.
- 1. En la práctica, el ajuste del modelo SVM puede ser difícil y requerir experiencia y conocimientos especializados para obtener un buen rendimiento.



Dudas y consultas



Fin Presentación



KIBERNUM