



Pontificia Universidad  
**JAVERIANA**  
Bogotá

# Maquinas Soporte Vectorial SVM

Educación **Continua**

Generamos experiencias educativas

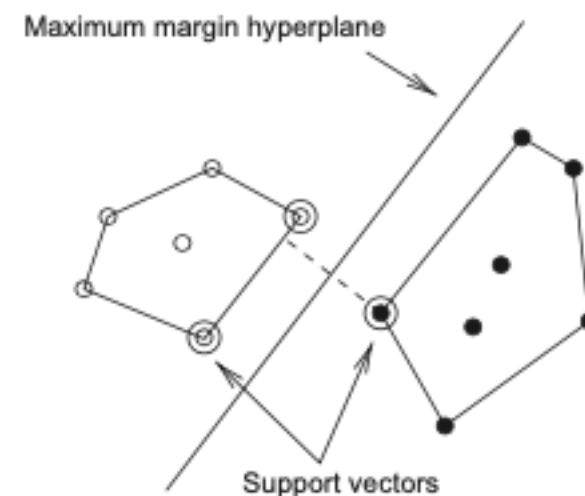
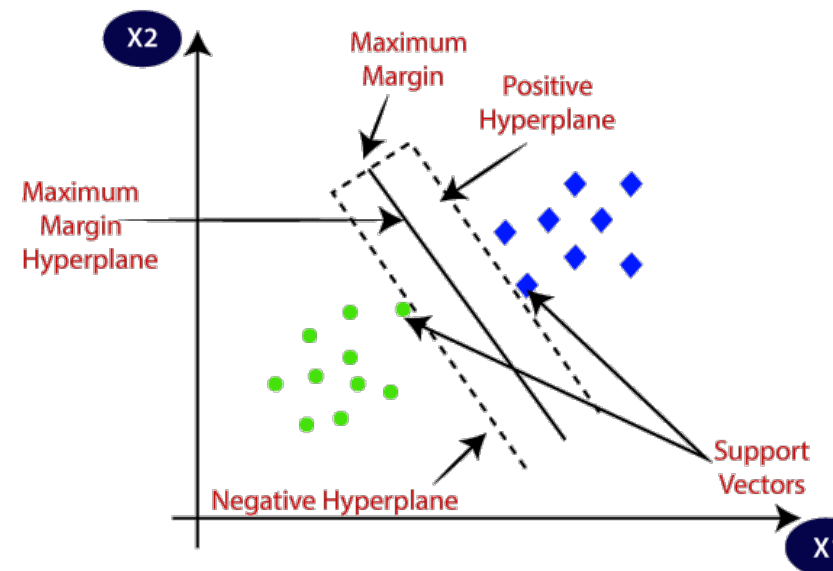
# Maquinas de Soporte Vectorial (SVM)

## Contexto:

- **SVM es un modelo de aprendizaje automático versátil:** SVM es capaz de realizar clasificación, regresión y detectar valores atípicos, tanto en problemas lineales como no lineales.
- **Amplia popularidad:** SVM es uno de los modelos más populares en el campo del aprendizaje automático.

## Conceptos claves:

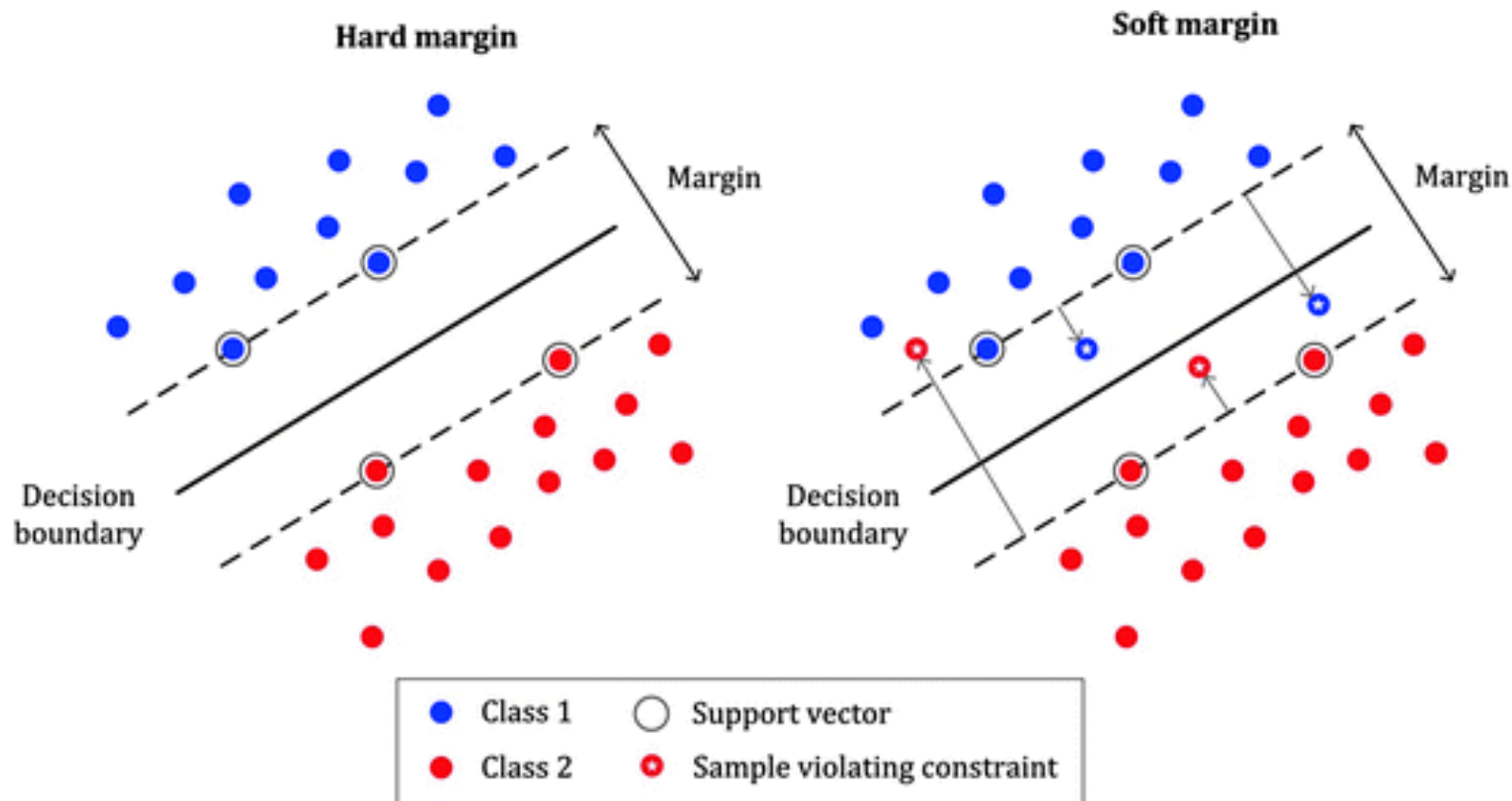
- **Búsqueda del hiperplano óptimo:** SVM se basa en encontrar un hiperplano (una superficie de decisión) que separe de la mejor manera posible las clases en un conjunto de datos.
- **Hiperplano de margen máximo:** El objetivo de SVM es identificar el hiperplano que maximiza el margen entre las clases. Esto implica encontrar un hiperplano que esté lo más lejos posible de los cascos convexos de las clases.
- **Vectores de soporte:** Las instancias de datos más cercanas al hiperplano de margen máximo se llaman "vectores de soporte". Son fundamentales para la definición del hiperplano y la toma de decisiones.



# Soft Margin

**Hard Margin:** Busca una separación perfecta sin errores de clasificación. Asume datos completamente linealmente separables.

**Soft Margin:** Permite errores de clasificación y puntos dentro del margen. Introduce un hiperparámetro "C" para controlar la tolerancia a errores. Adecuado para datos ruidosos o no perfectamente separables.



# Clasificador de Margen Máximo



Pontificia Universidad  
JAVERIANA  
Bogotá

- El objetivo es encontrar un buen equilibrio entre mantener la calle lo más grande posible y limitar las violaciones de los márgenes
- Es decir, instancias que terminan en el medio de la calle o incluso en el lado equivocado.
- A continuación, se presenta el modelo de optimización a resolver:

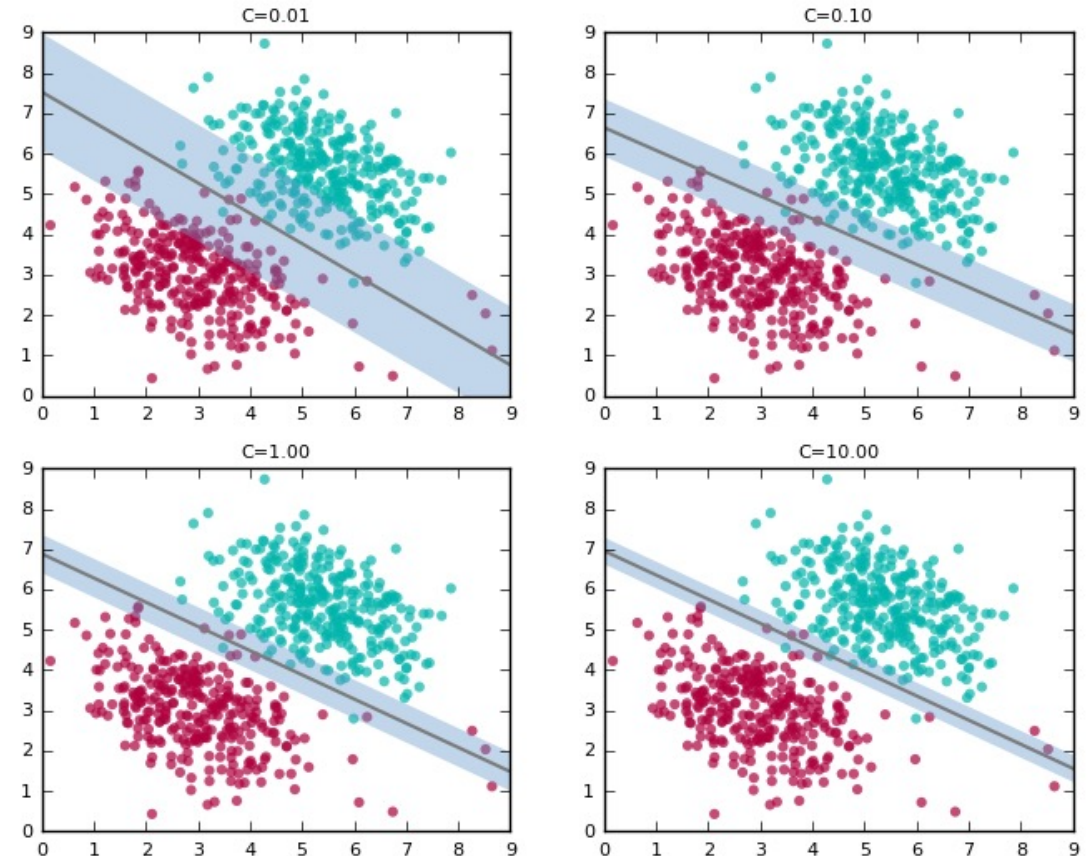
$$\begin{aligned} & \underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} && M \\ & \text{subject to} && \sum_{j=1}^p \beta_j^2 = 1, \\ & && y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i), \\ & && \epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C, \end{aligned}$$

$M$ : margen

$\beta_p$ : parámetros del hiperplano

$C$ : hiperparámetro de calibración

$\epsilon_i$ : variables de holgura

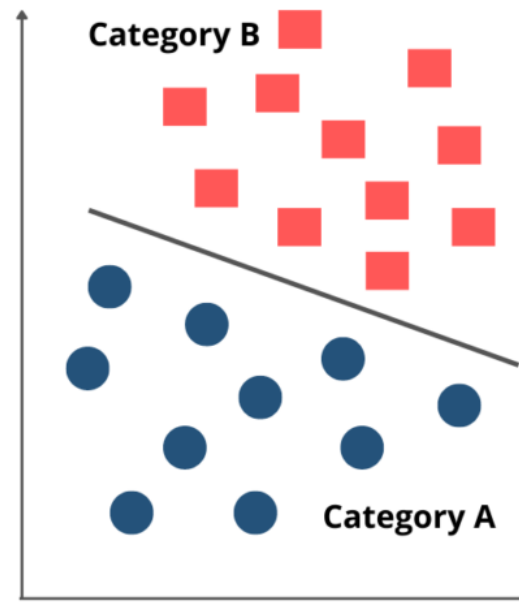


Clase 1:  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p > 0$ .

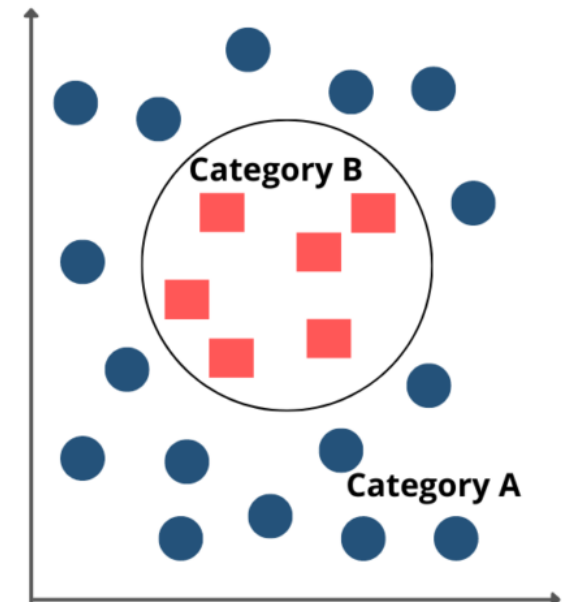
Clase 2:  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p < 0$

- El clasificador de vectores de soporte es un enfoque natural para la clasificación en el entorno de dos clases, si el límite entre las dos clases es lineal.
- Sin embargo, en la práctica a veces nos enfrentamos a fronteras de clase no lineales.
- Por ejemplo, consideremos los datos de la gráfica.
- Está claro que un clasificador de vectores de soporte o cualquier clasificador lineal no funcionará para el caso del lado derecho.

**Linear SVM**



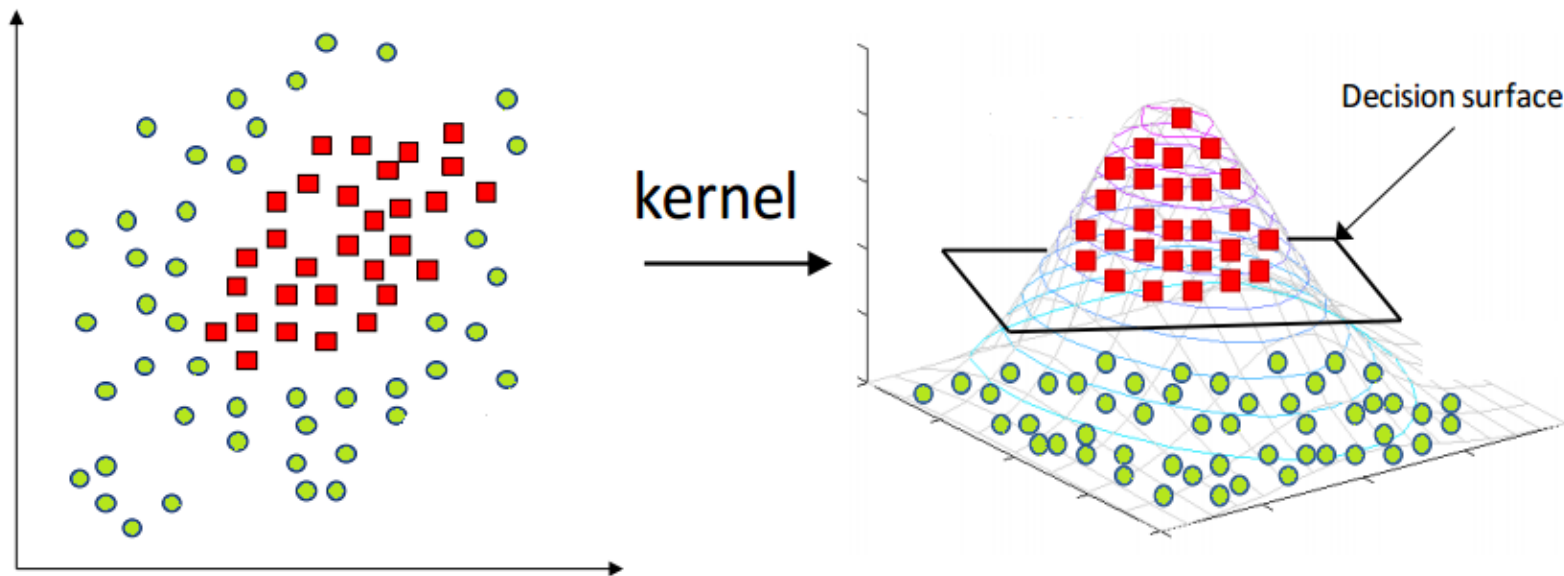
**Non Linear SVM**



Para manejar límites no lineales entre clases, podemos expandir el espacio de variables utilizando funciones. *Ejemplo funciones polinómicas de grado superior (como cuadráticas o cúbicas) de los predictores.*

$$X_1, X_2, \dots, X_p$$

$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2.$$



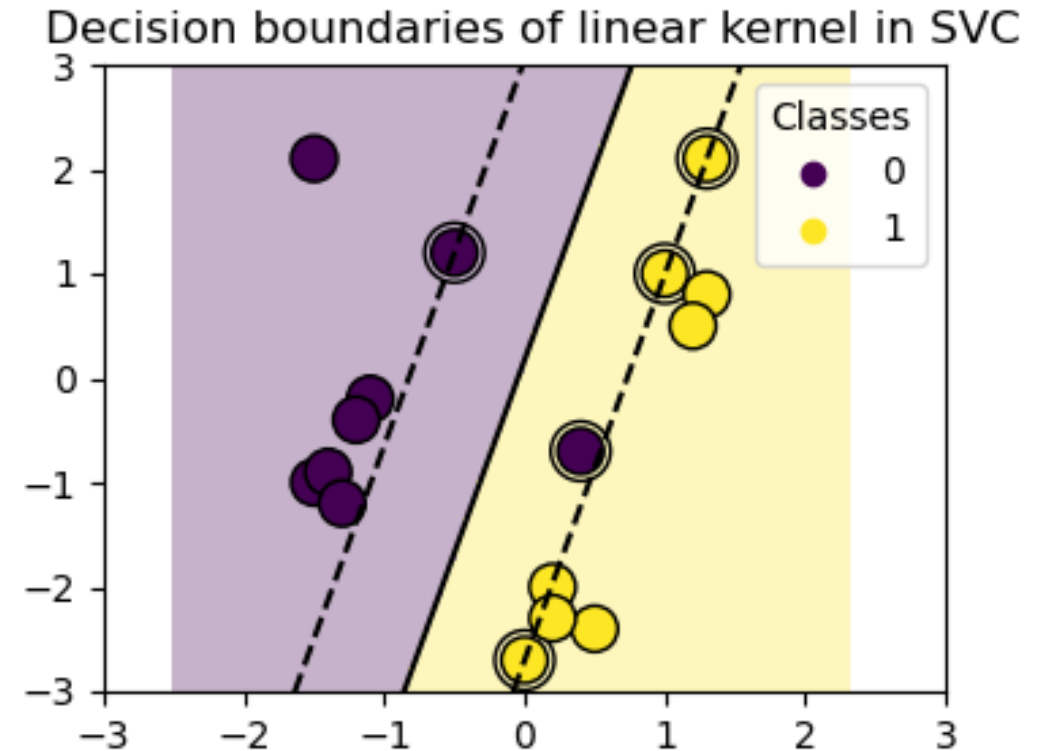
Las funciones que generan esa transformación se denominan funciones kernel



El kernel lineal es el producto escalar de las muestras de entrada:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^\top \mathbf{x}_2$$

Entrenar un SVC en un núcleo lineal da como resultado un espacio de características no transformado, donde el hiperplano y los márgenes son líneas rectas.

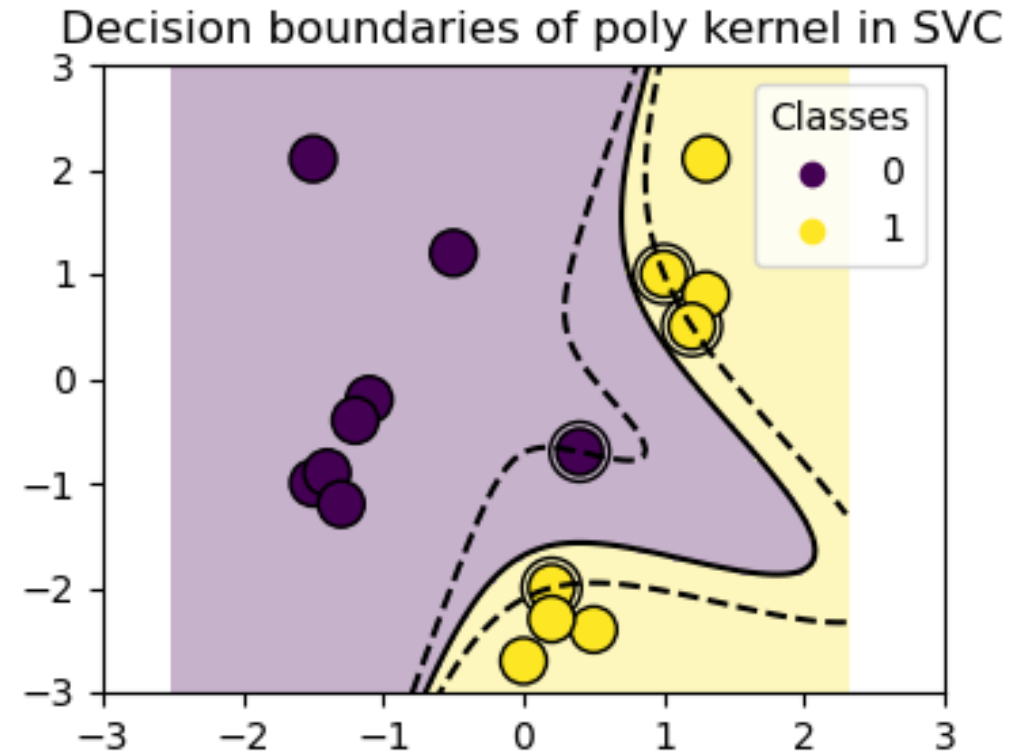


El kernel polinomial cambia la noción de similitud.  
La función del kernel se define como:

$$K(\mathbf{x}_1, \mathbf{x}_2) = (\gamma \cdot \mathbf{x}_1^\top \mathbf{x}_2 + r)^d$$

Donde:

- $d$  es el grado (grado) del polinomio,
- $\gamma$ : (gamma) controla la influencia de cada muestra de entrenamiento individual en el límite de decisión
- $r$ : es el término de sesgo (coef0) que desplaza los datos hacia arriba o hacia abajo



Usar un núcleo polinómico equivale a crear variables polinómicas y luego ajustar un SVC con un núcleo lineal a los datos transformados.



El kernel de base radial (RBF), también conocido como núcleo gaussiano

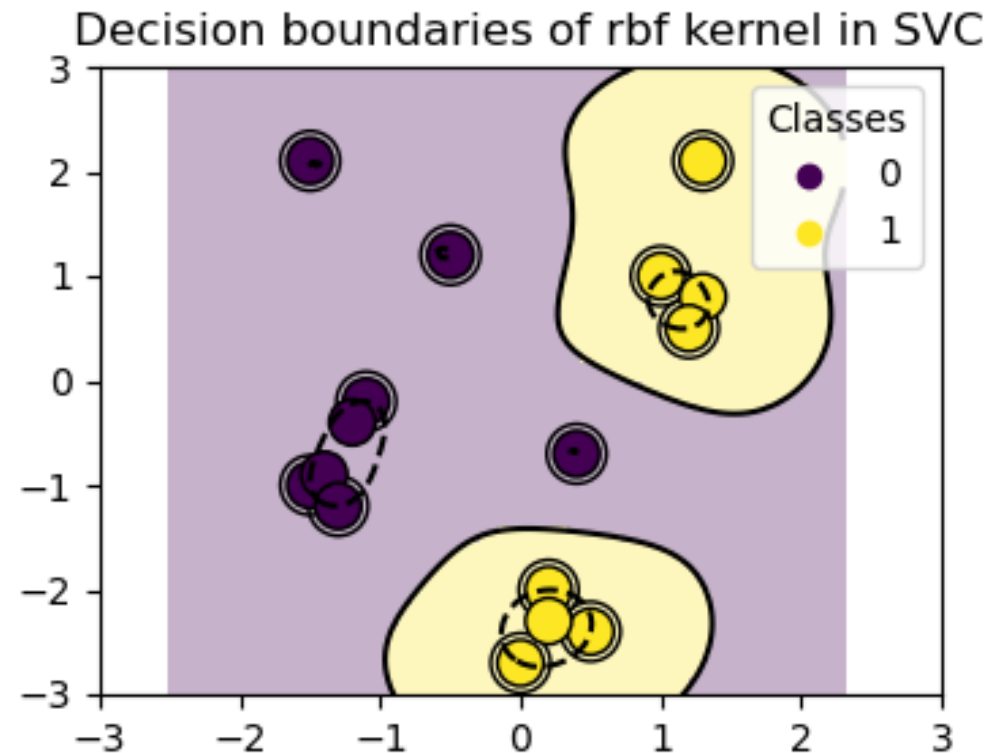
Mide la similitud entre dos puntos de datos en infinitas dimensiones y luego se acerca a la clasificación por mayoría de votos.

La función kernel se define como:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \cdot \|\mathbf{x}_1 - \mathbf{x}_2\|^2)$$

Donde:

- $\gamma$ : (gamma) controla la influencia de cada muestra de entrenamiento individual en el límite de decisión
- Cuanto mayor sea la distancia euclidiana entre dos puntos  $\|\mathbf{x}_1 - \mathbf{x}_2\|^2$ , más cerca estará la función kernel de cero. Esto significa que es más probable que dos puntos lejanos sean diferentes.



Es el kernel predeterminado para  
Support Vector Machines en scikit-learn

La función kernel se define como:

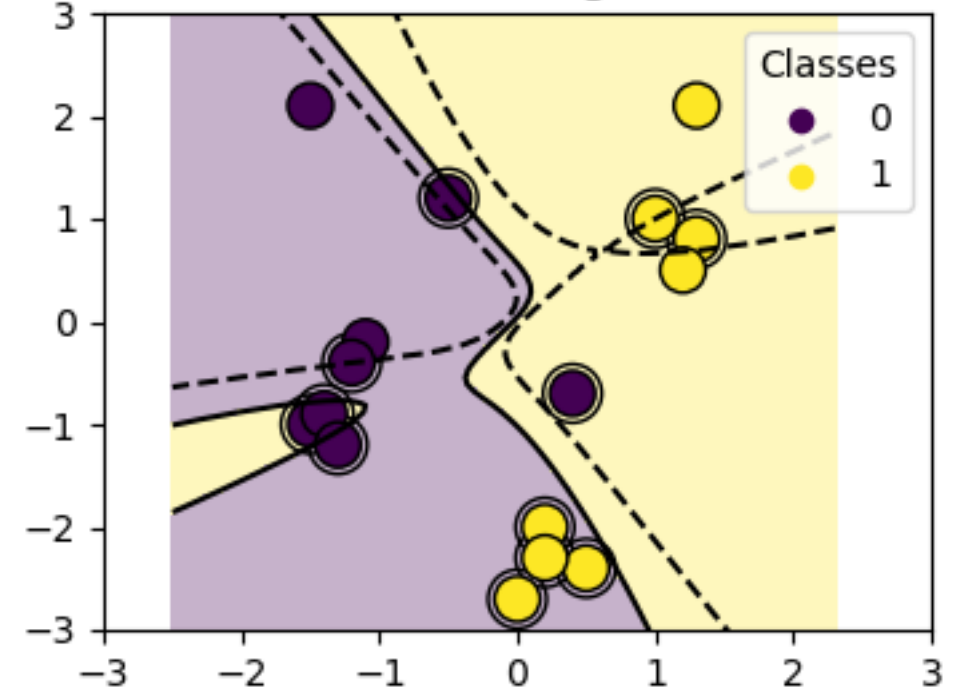
$$K(\mathbf{x}_1, \mathbf{x}_2) = \tanh(\gamma \cdot \mathbf{x}_1^\top \mathbf{x}_2 + r)$$

Donde:

- $\gamma$ : (gamma) controla la influencia de cada muestra de entrenamiento individual en el límite de decisión
- $r$ : es el término de sesgo (coef0) que desplaza los datos hacia arriba o hacia abajo

En el kernel sigmoido, la similitud entre dos puntos de datos se calcula utilizando la función tangente hiperbólica.

Decision boundaries of sigmoid kernel in SVC



Podemos ver que los límites de decisión obtenidos con el núcleo sigmoido aparecen curvados e irregulares. El límite de decisión intenta separar las clases ajustando una curva en forma de sigmoide



Pontificia Universidad  
**JAVERIANA**  
Bogotá

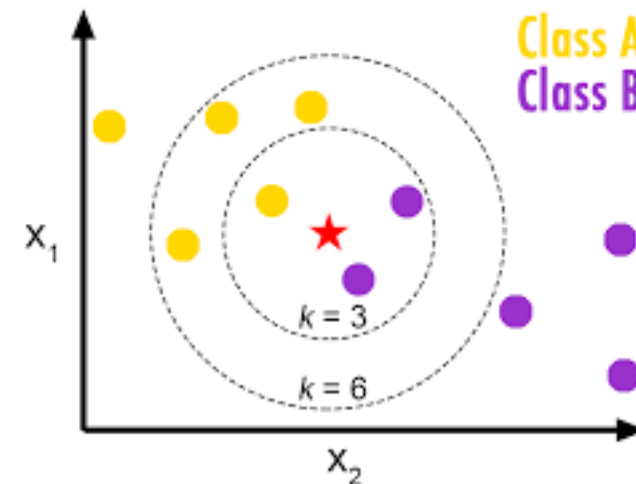
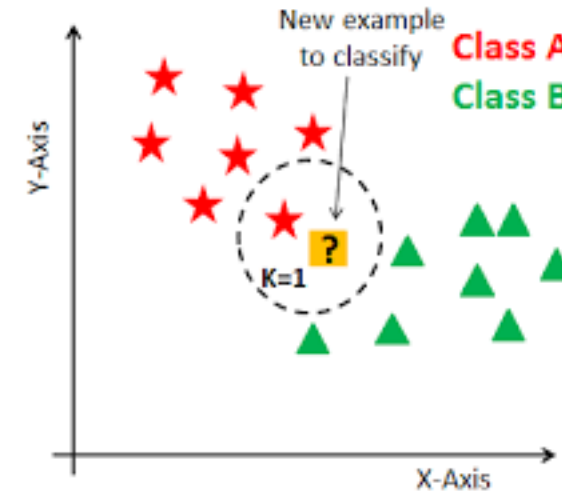
# **K Vecinos mas Cercanos**

Educación **Continua**

Generamos experiencias educativas

# La Regla del Vecino mas Próximo

- La regla del vecino más próximo simplemente asigna la clase del ejemplo más próximo, utilizando una función de distancia.
- Esta regla, conocida como 1-NN (one nearest neighbor), tiene bastantes problemas, ya que ignora la densidad o la región donde se encuentra el ejemplo.
- Una variante de este método son los k vecinos más próximos (kNN, k-nearest neighbors) en el que se asigna la clase mayoritaria entre los k vecinos más próximos.
- En un problema de regresión se asigna el promedio o mediana de los vecinos mas próximos.

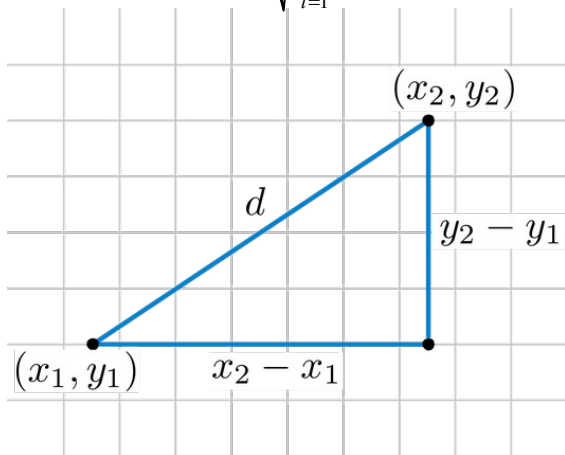


# Medidas de Distancia

Las medidas de distancia se utiliza para dimensionar la separación de dos instancias u observaciones.

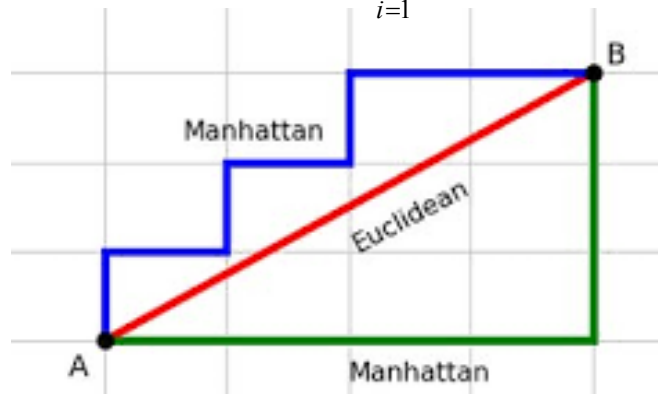
**Distancia Euclídeana.** Es la distancia clásica, como la longitud de la recta que une dos puntos en el espacio euclídeo:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$



**Distancia de Manhattan:** hace referencia a recorrer un camino no en diagonal o en zigzag.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$



**Distancia del coseno.** Si se considera que cada ejemplo es un vector, la distancia sería el coseno del ángulo que forman .

$$d(x, y) = \arccos\left(\frac{x^T y}{\|x\| \cdot \|y\|}\right)$$

