

Data Science & Machine Learning



red.es

Centro de
Búsqueda y
Análisis de
Datos
CRN Digital



"El FSE invierte en tu futuro"
Fondo Social Europeo

Data Science & Machine Learning



Data Science & Machine Learning



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es

Centro de
Bases de Datos
en Ciencia, Tecnología
y Planificación

CRN
Digital

GARANTÍA
JUVENIL



UNIÓN EUROPEA

Barrabés

The Valley

"El FSE invierte en tu futuro"
Fondo Social Europeo

Índice

1. Regresión

1.1 Regresión Simple vs Regresión Multiple

1.2 Regresión Lineal vs Regresión No Lineal

1.2 Regresión Lineal Simple



2. Prophet



3. Neural Prophet

Clasificación

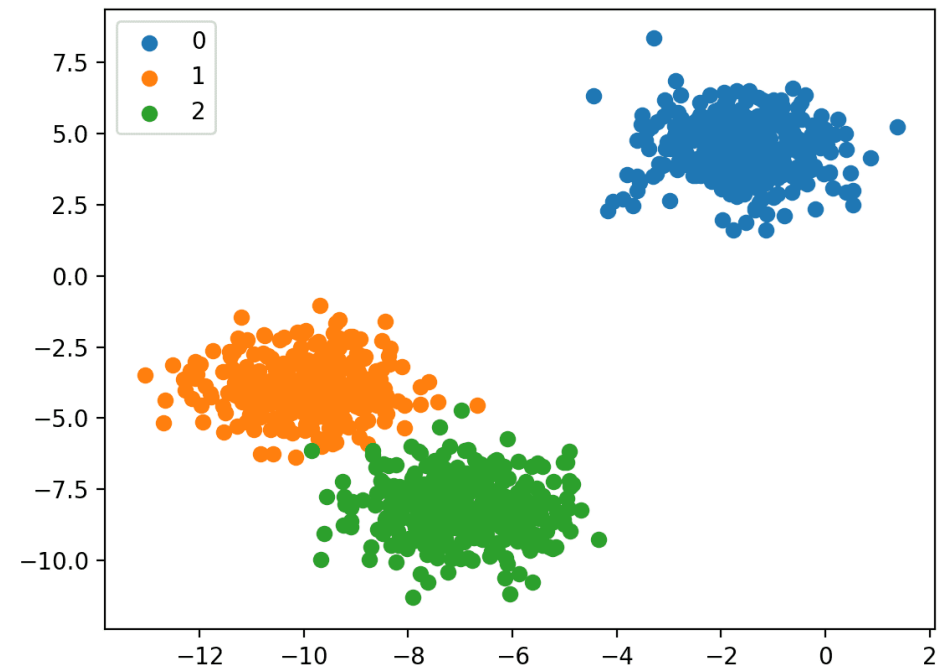
La regresión modela la relación entre una variable independiente (Y) y una o más variables independientes (X)

Y es una variable cualitativa.

Regresión vs Clasificación:

En clasificación Y es nominal u ordinal (NO importa la escala)

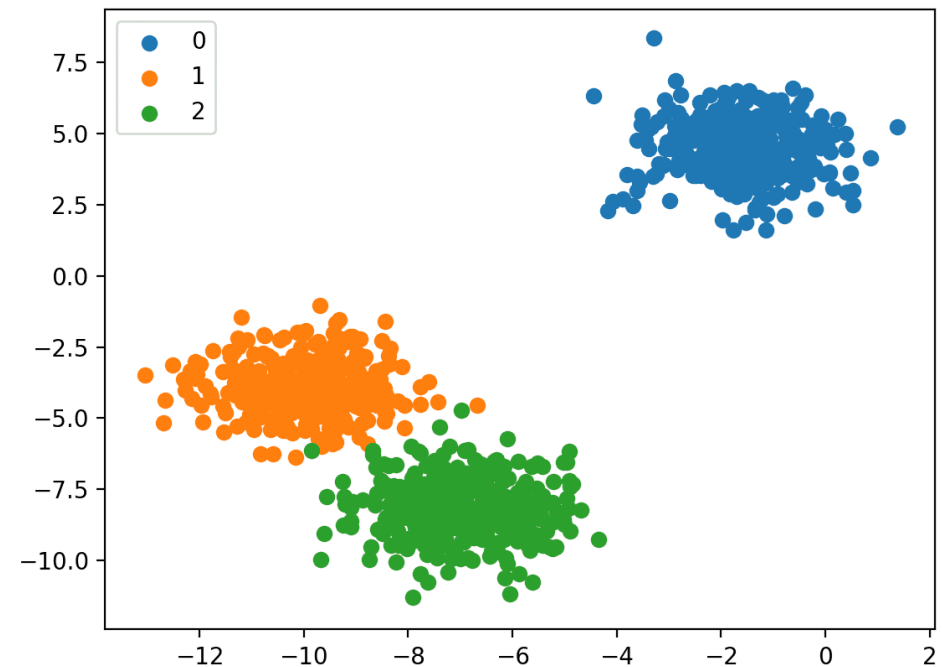
En regresión Y es continua o discreta. (Sí importa la escala)



Clasificación

Ejemplos:

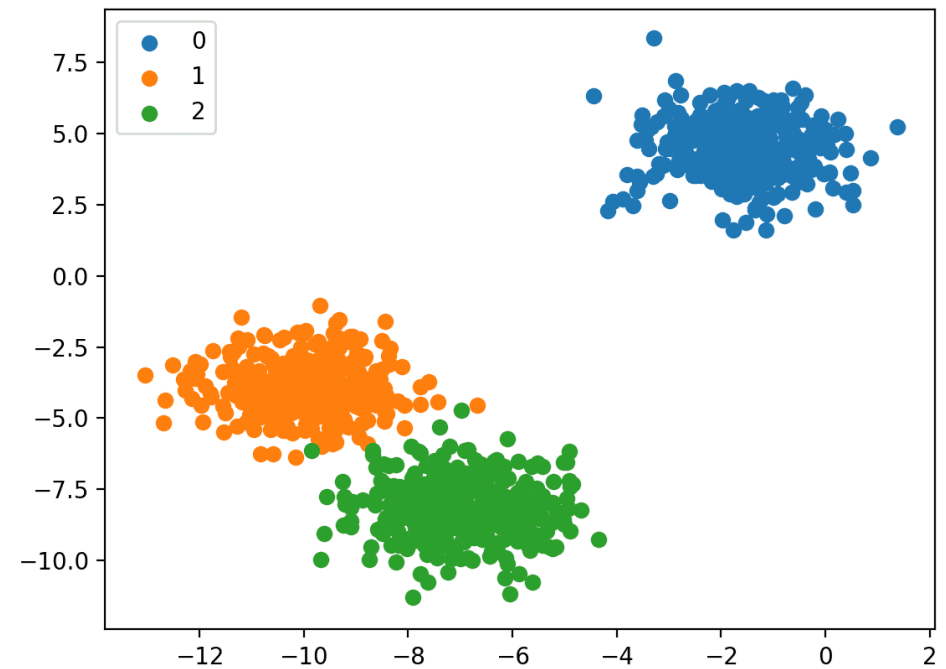
- Transacciones online (Fraudulentas o no)
- Hipotecas (Aprobadas o no)
- Tipo de cliente (Alto, medio y bajo potencial)
- Predicción de enfermedades (Riesgo alto, medio, bajo e inexistente)



Clasificación

Modelos de clasificación:

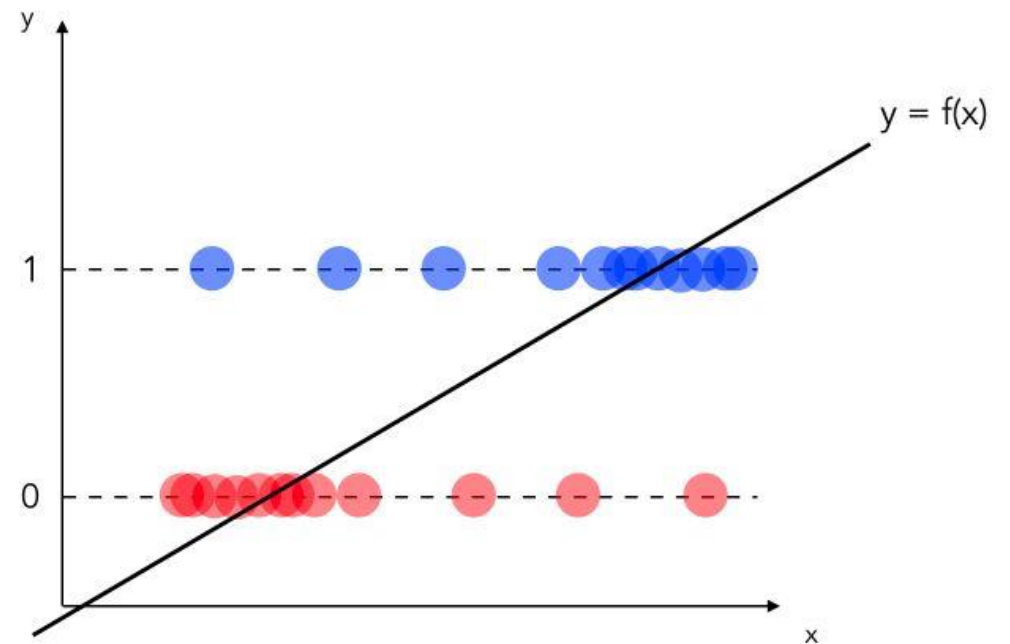
- LPM
- Regresión logística
- KNN
- SVM
- Decission Tree
- Random Forest



Clasificación (LPM)

Para variables de 2 categorías (0 y 1).

Podemos usar una regresión lineal para calcular la probabilidad de que pertenezca a cada grupo.

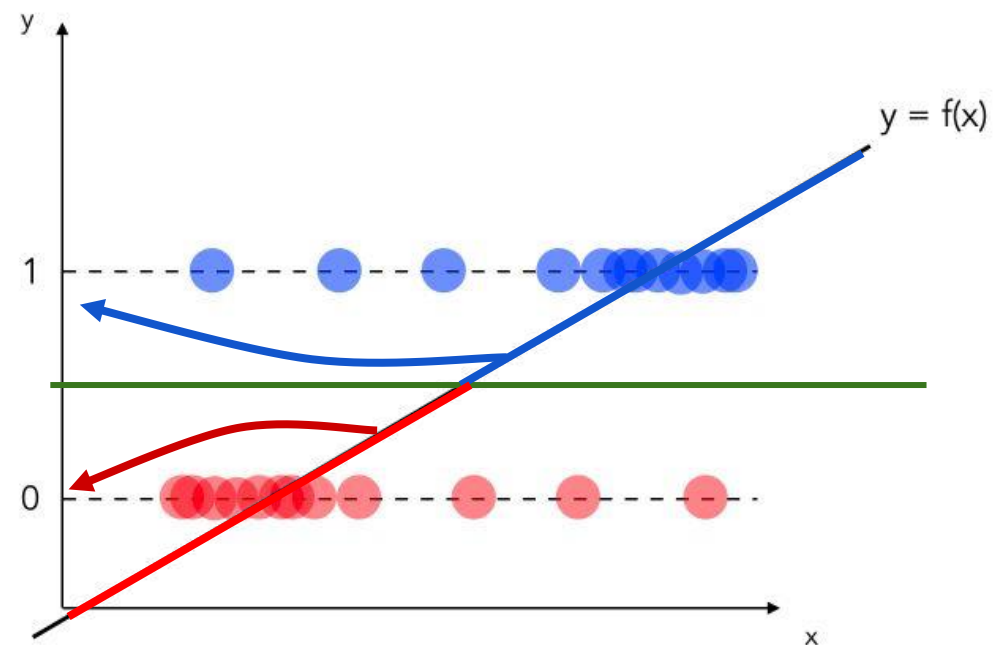


Clasificación (LPM)

Para variables de 2 categorías (0 y 1).

Podemos usar una regresión lineal para calcular la probabilidad de que pertenezca a cada grupo.

Definimos una probabilidad límite.



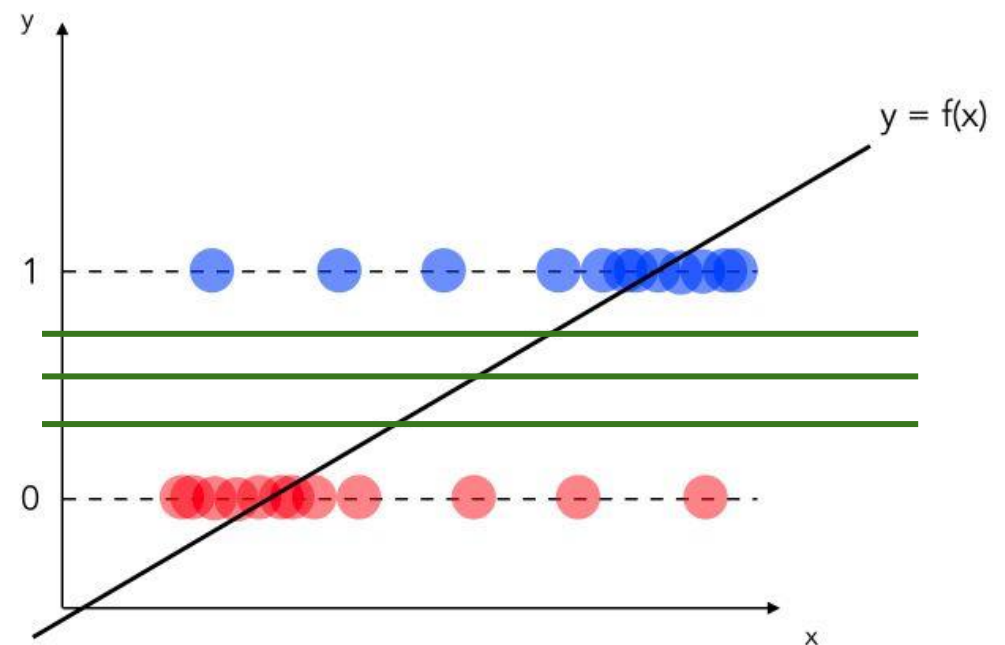
Clasificación (LPM)

Para variables de 2 categorías (0 y 1).

Podemos usar una regresión lineal para calcular la probabilidad de que pertenezca a cada grupo.

Definimos una probabilidad límite.

PROBLEMAS:
Difícil decidir dónde ajustar la probabilidad límite.



Clasificación (LPM)

Para variables de 2 categorías (0 y 1).

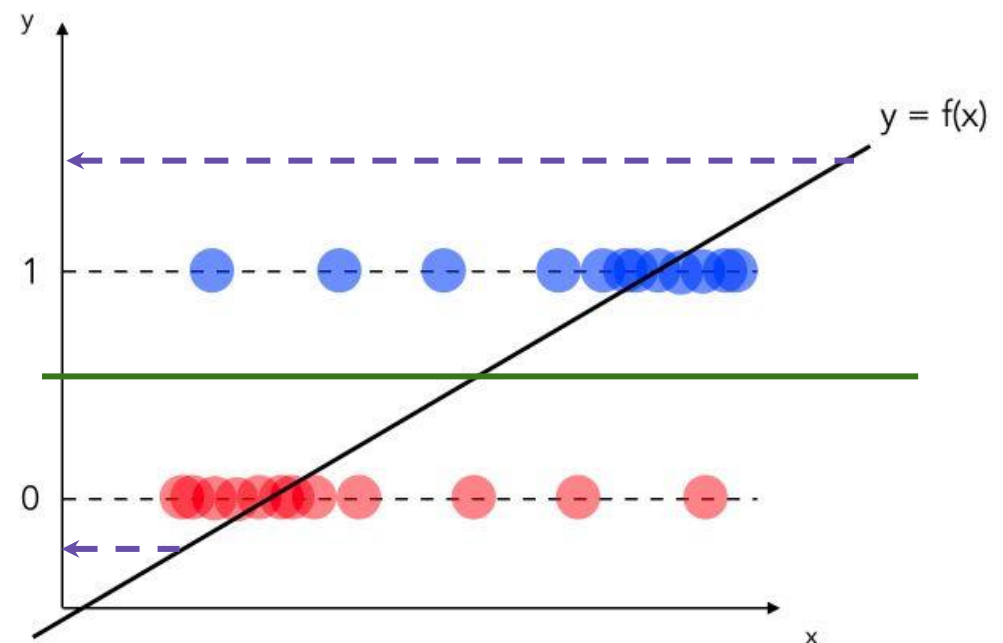
Podemos usar una regresión lineal para calcular la probabilidad de que pertenezca a cada grupo.

Definimos una probabilidad límite.

PROBLEMAS:

Difícil decidir dónde ajustar la probabilidad límite.

Predicciones por fuera del $[0,1]$



Clasificación (LPM)

Para variables de 2 categorías (0 y 1).

Podemos usar una regresión lineal para calcular la probabilidad de que pertenezca a cada grupo.

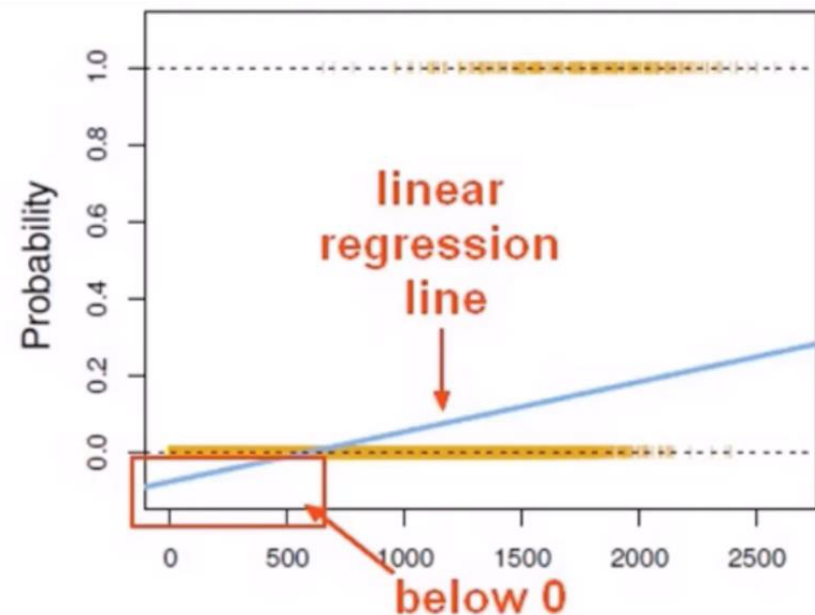
Definimos una probabilidad límite.

PROBLEMAS:

Difícil decidir dónde ajustar la probabilidad límite.

Predicciones por encima del $[0,1]$

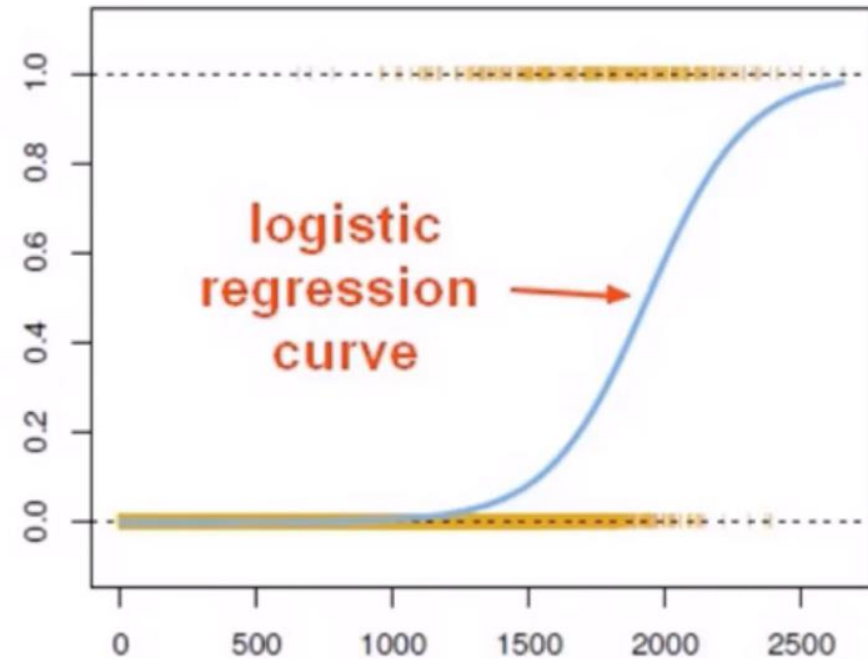
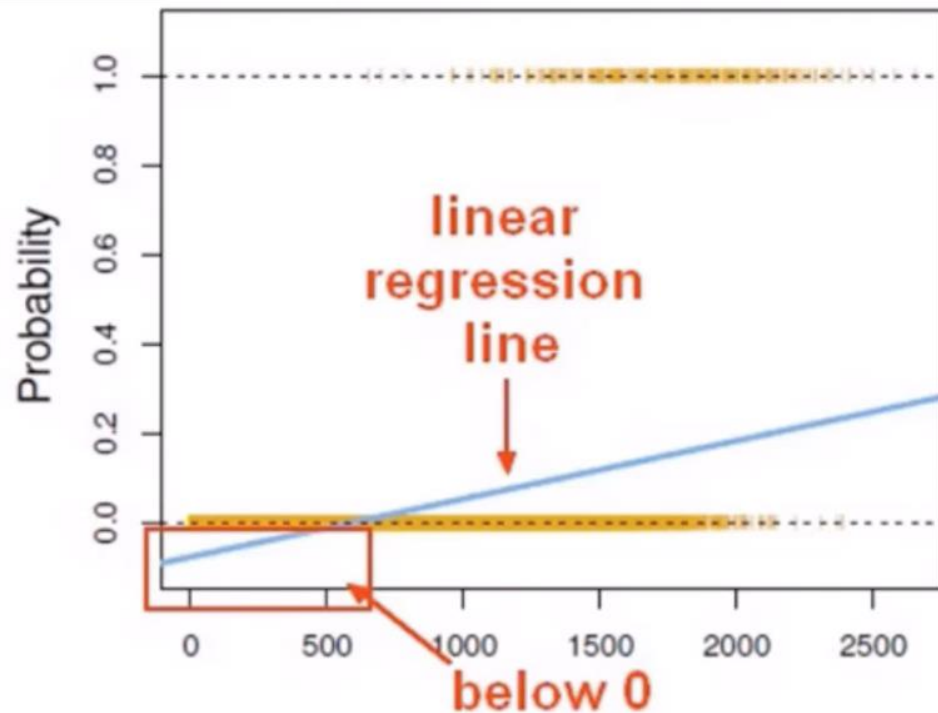
Sensible a la cantidad de datos



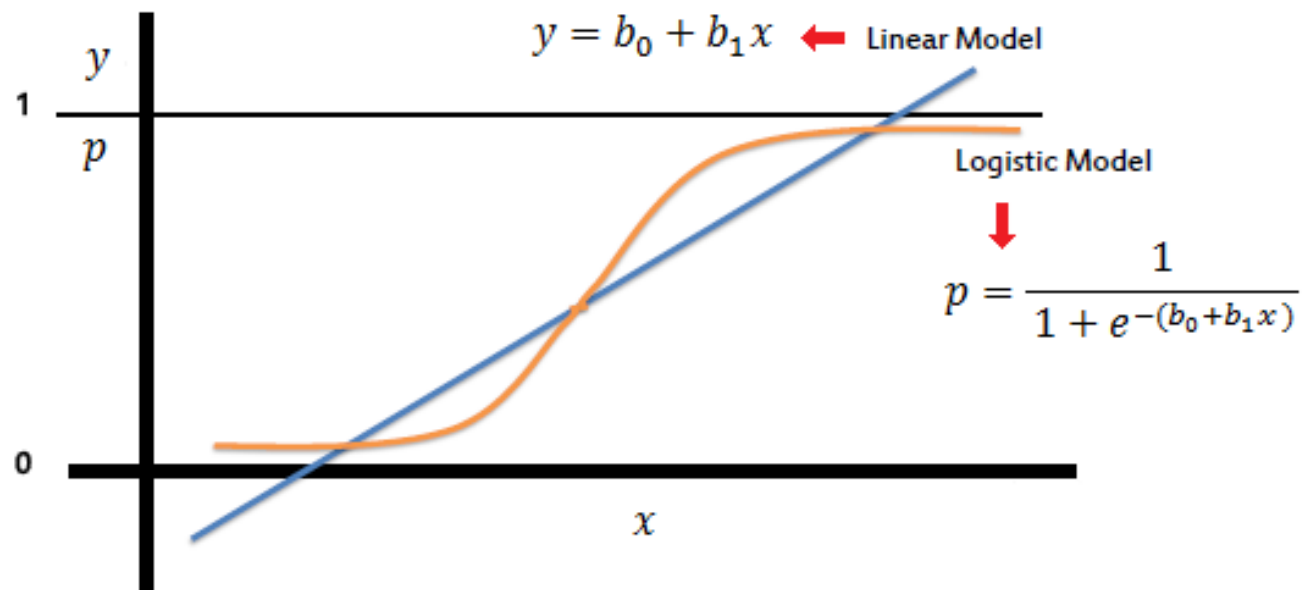
Clasificación (LPM)

Por los anteriores motivos el modelo de probabilidad lineal no se utiliza.

El modelo de Regresión Logística es mucho más efectivo.



Clasificación (Regresión Logística)



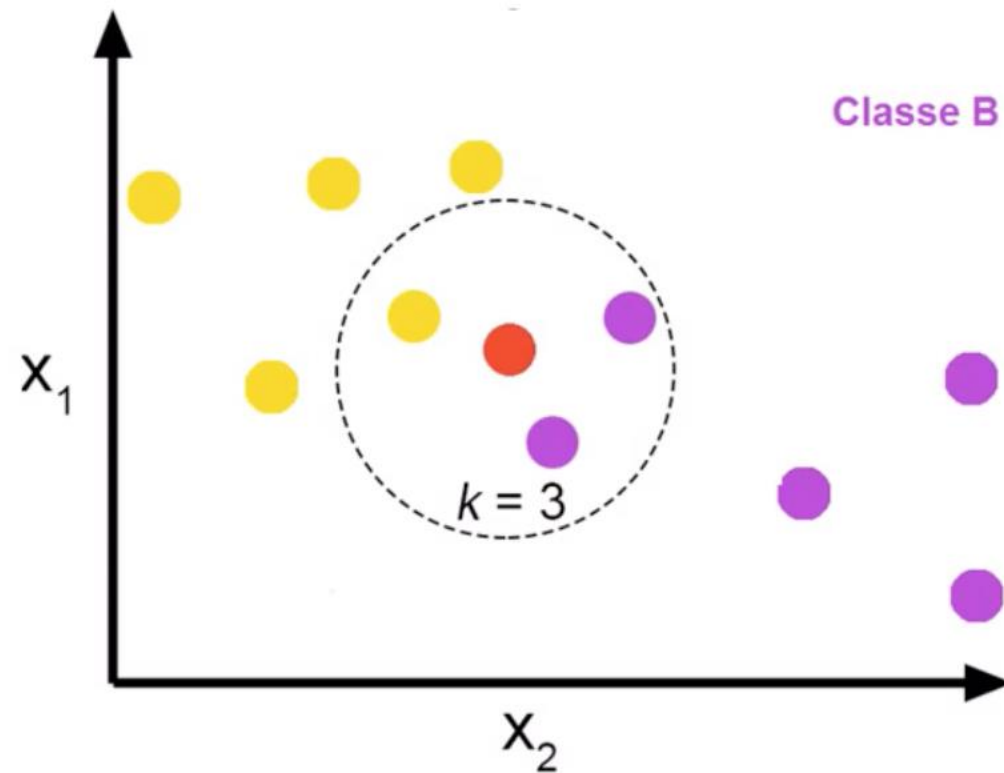
$$\text{Min}_{w,b} \text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - f_{w,b}(X_i))^2$$
$$f_{w,b}(X) = \frac{1}{1 + e^{-(wX + b)}}$$

Clasificación (KNN)

Modelo sencillo pero realmente potente

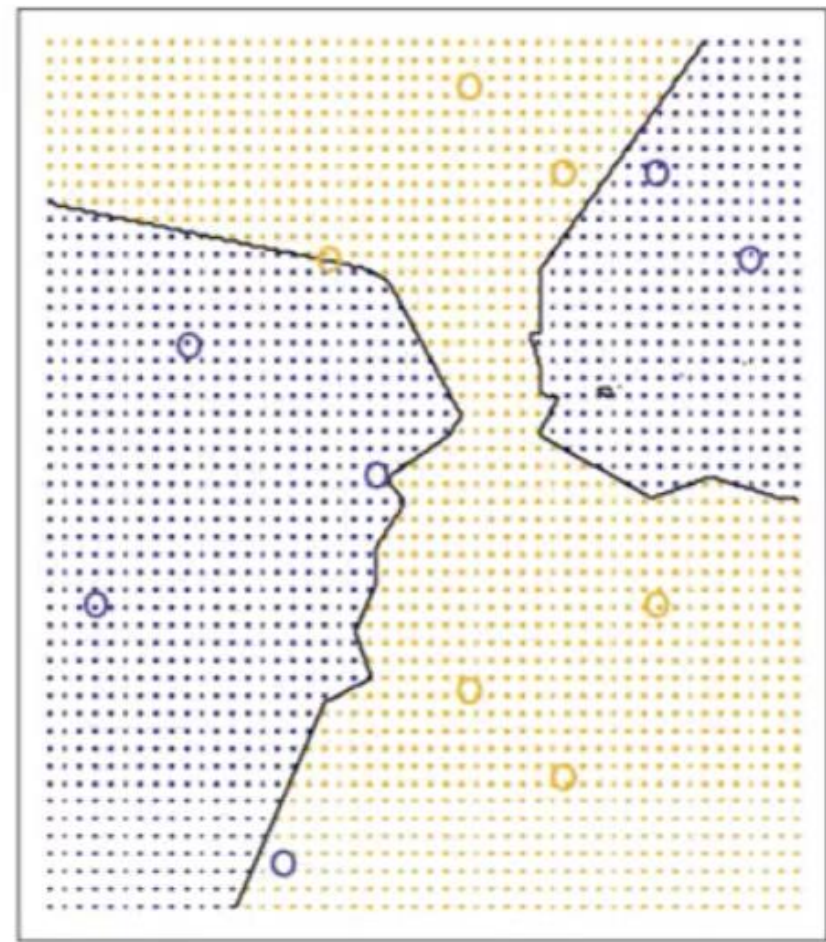
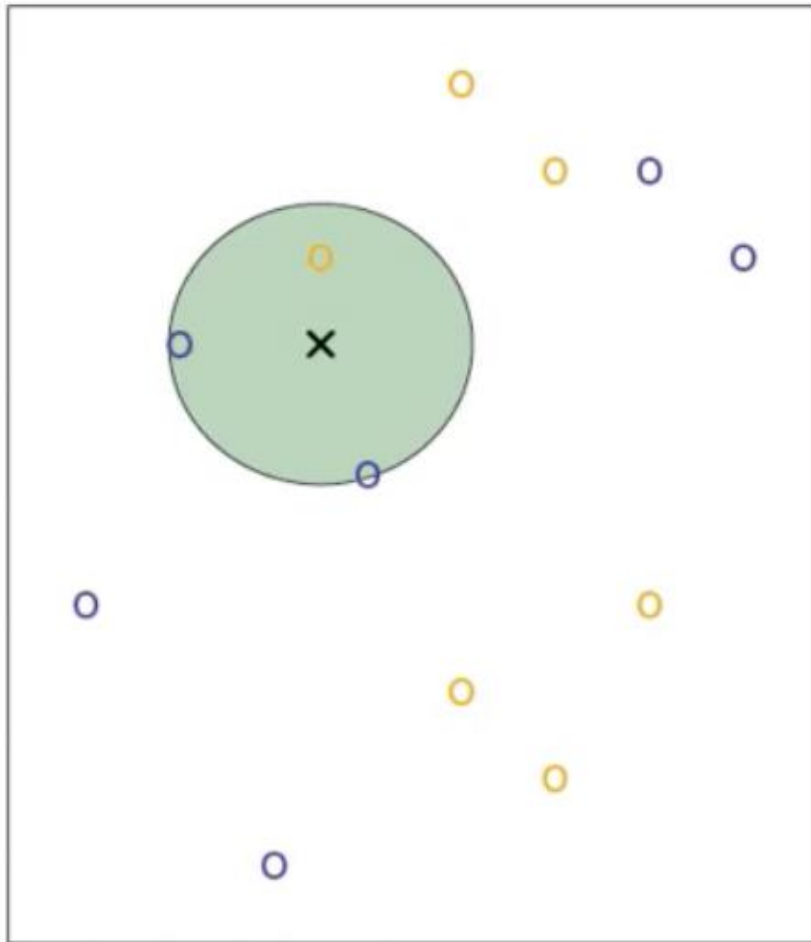
Se puede modificar K y la métrica utilizada.

Se puede emplear para regresión



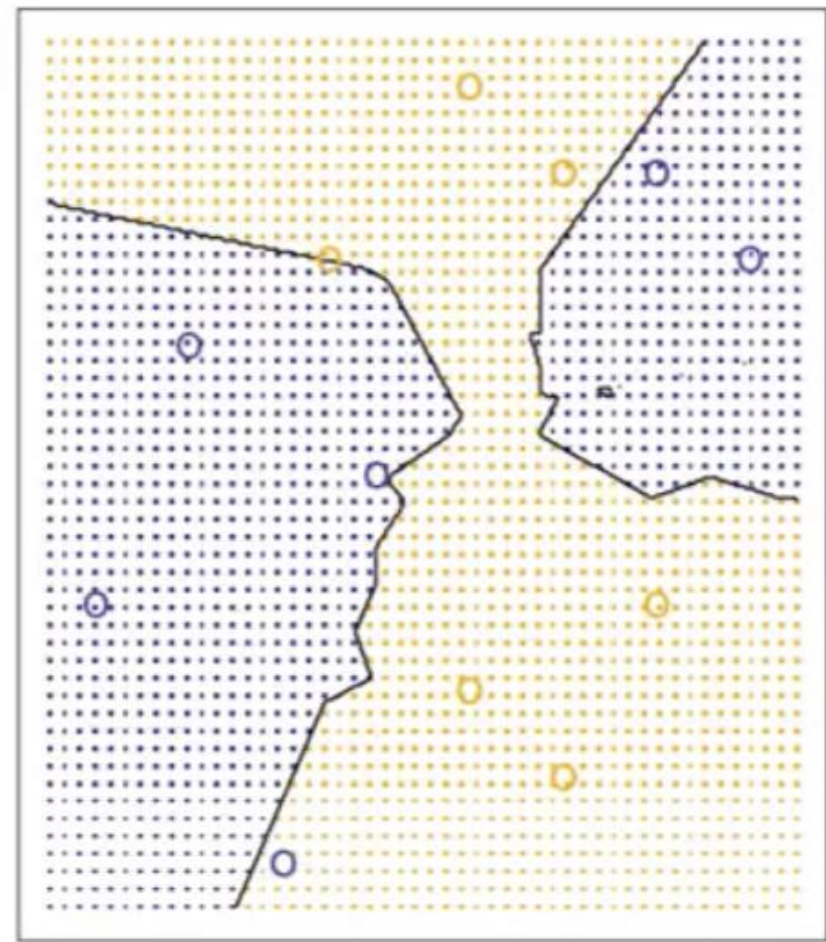
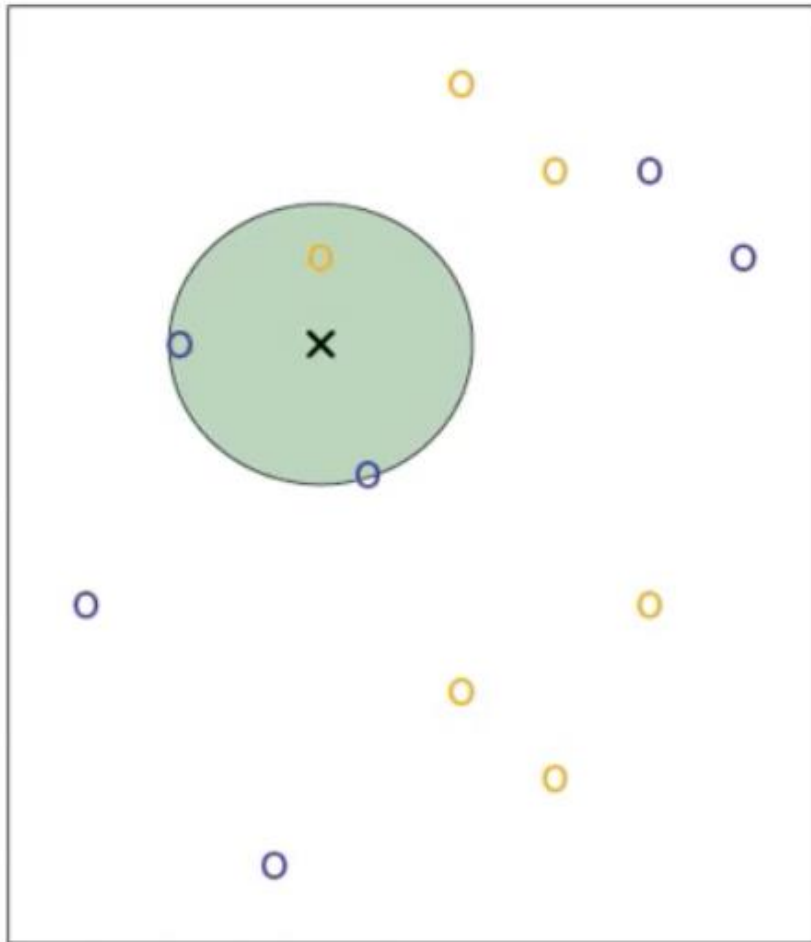
Clasificación (KNN)

Frontera de clasificación (Decision Boundary)



Clasificación (KNN)

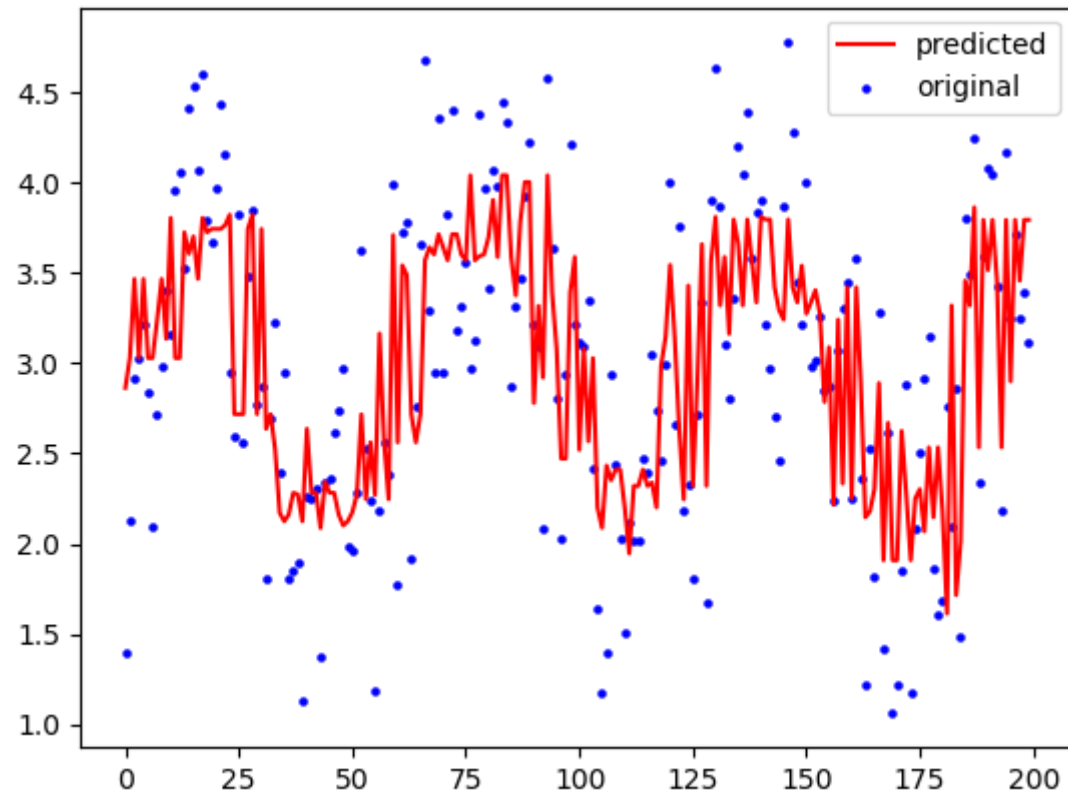
Frontera de clasificación (Decision Boundary)



Regresión (KNN)

Valor en función de los K más cercanos:

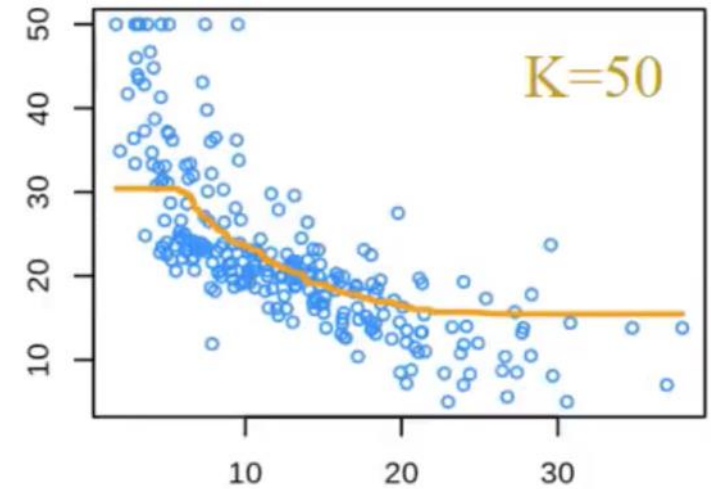
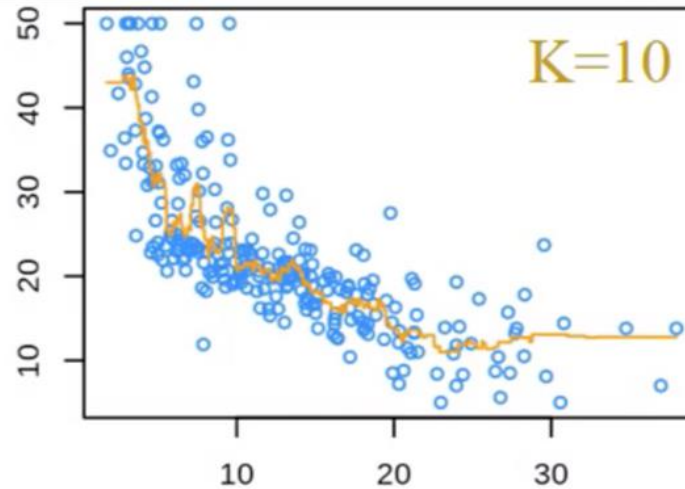
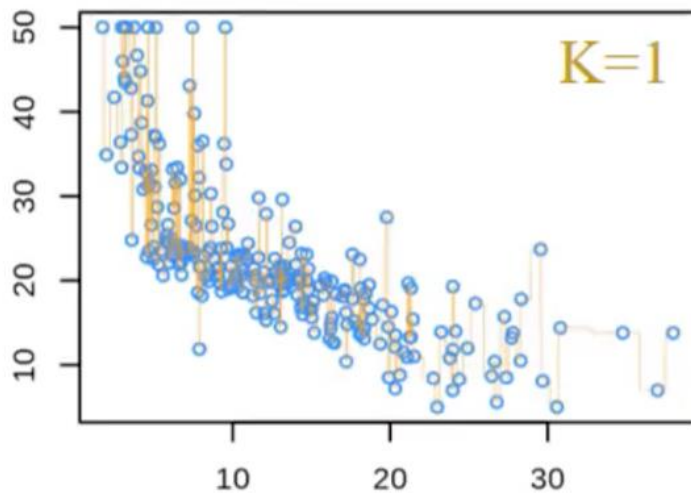
- Media
- Media penalizada por distancia
- Mediana



Regresión (KNN)

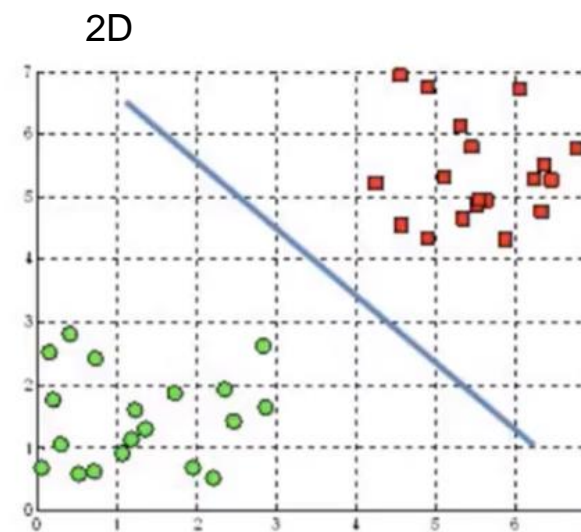
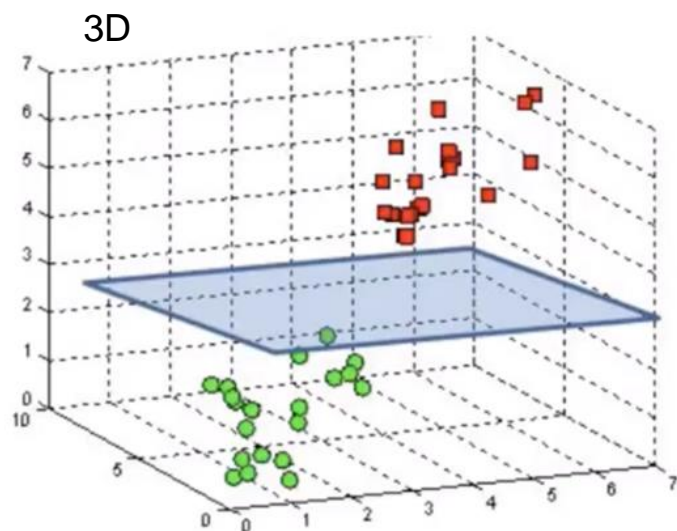
La importancia de la K

Underfitting vs Overfitting



SVM

¿Que es un hiperplano?



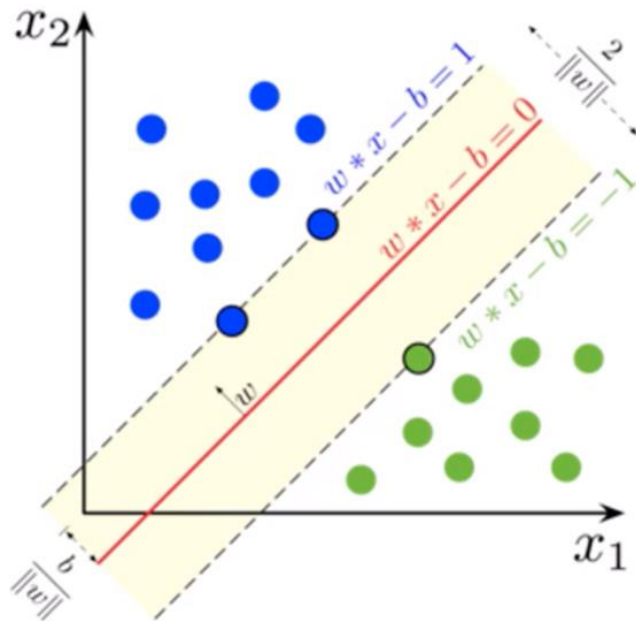
1D



SVM

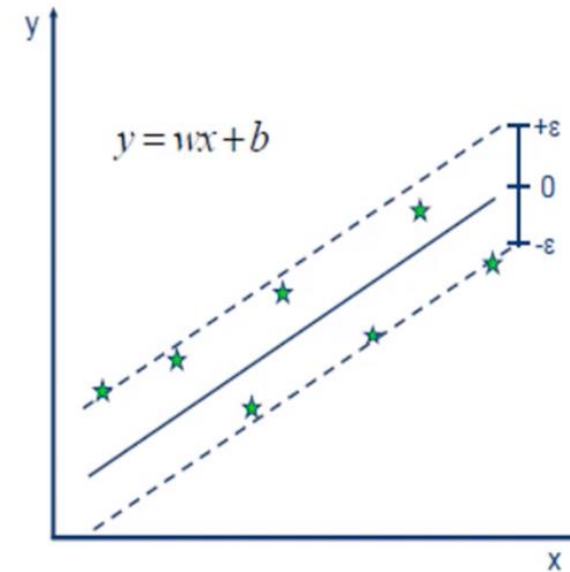
CLASIFICACIÓN:

El hiperplano mantiene la máxima distancia posible



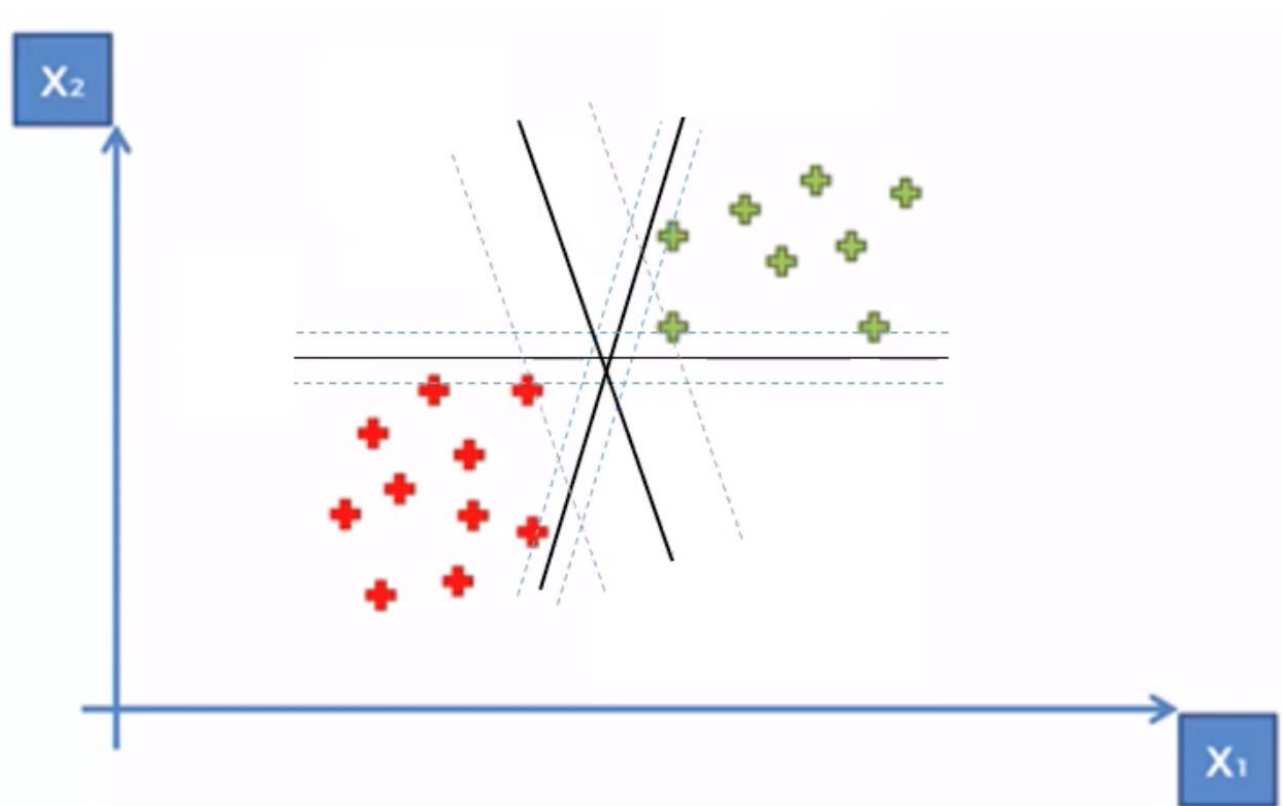
REGRESIÓN:

El hiperplano incluye a los puntos en la mínima distancia posible



Clasificación (SVM)

¿Qué hiperplano es mejor?





Contacto:

luis@lubay.es





red.es



UNIÓN EUROPEA

"El FSE invierte en tu futuro"

Fondo Social Europeo

