

Essential Machine Learning with Python

https://www.facebook.com/qslearningperu/?ref=page_internal



QS Learning

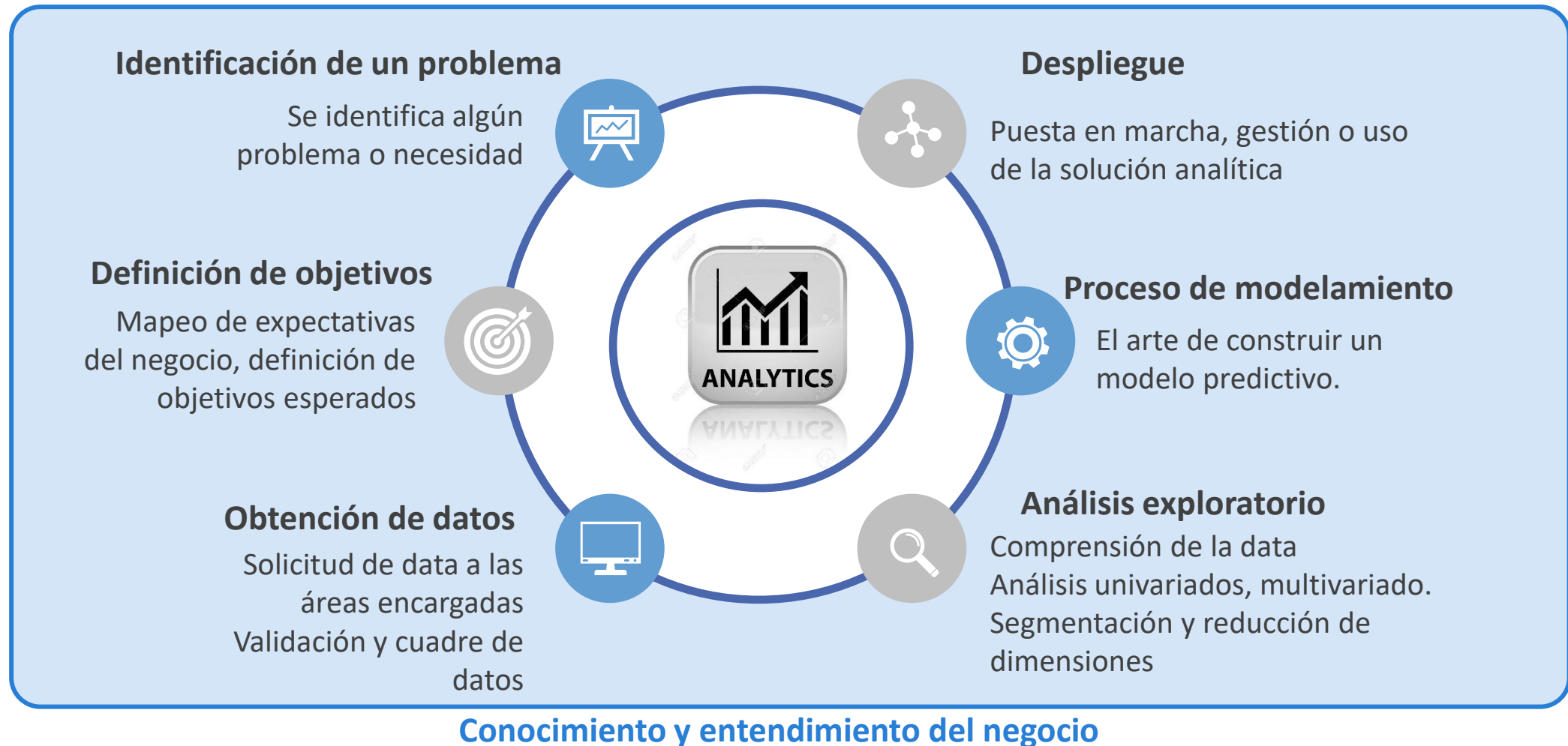


+51 937 012 707 / +51 915 111 457



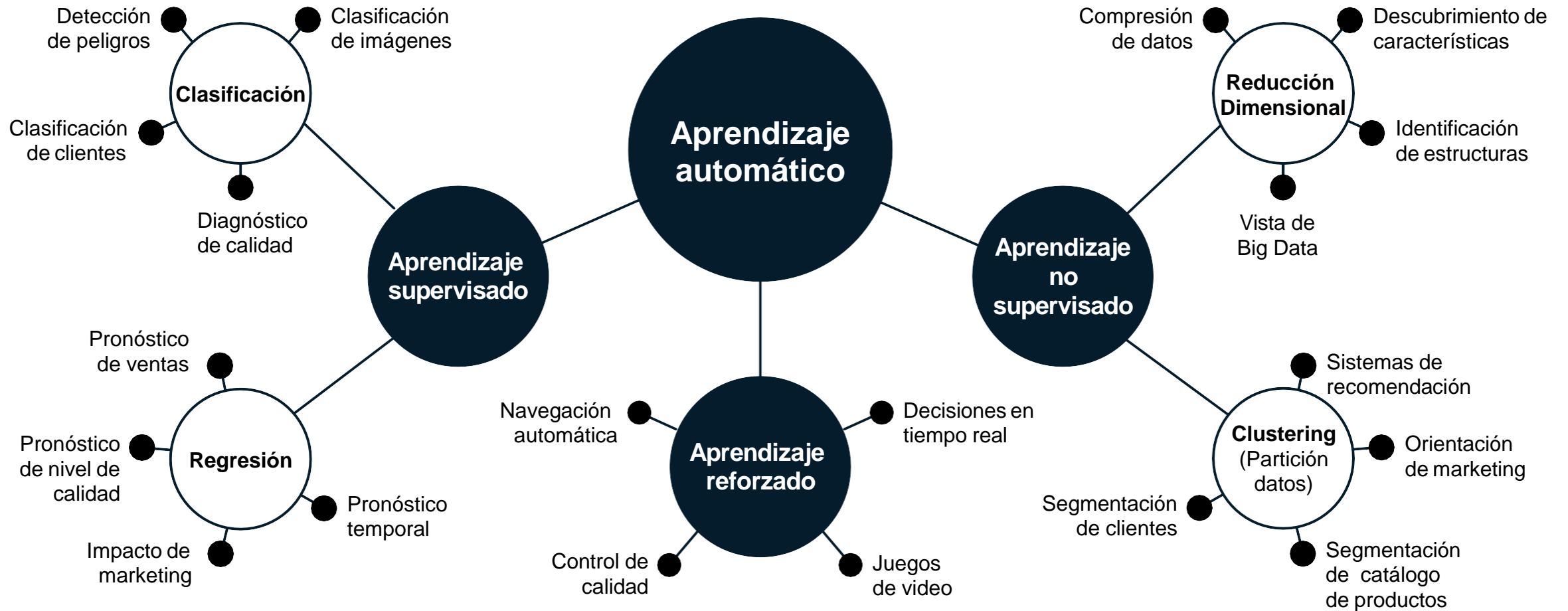
quants.admission@gmail.com

Proceso de modelamiento

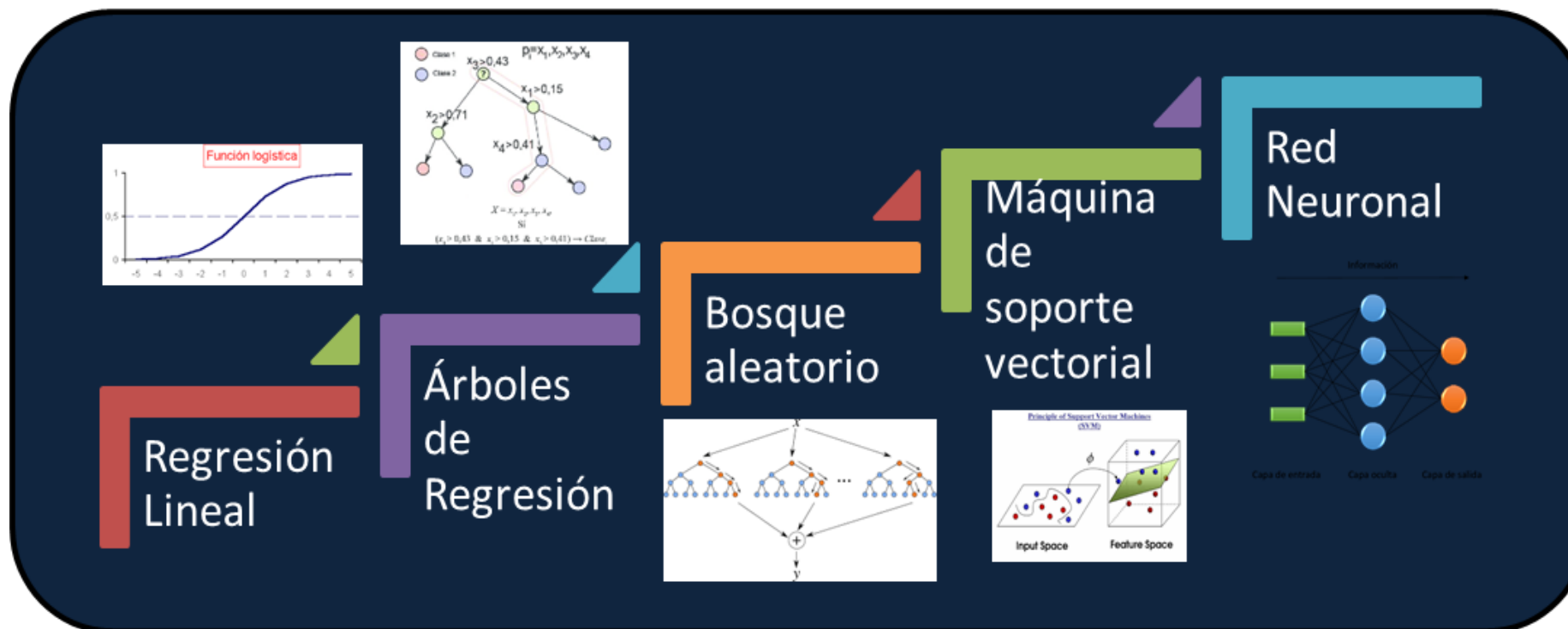


Aprendizaje automático

Aprendizaje automático



Modelos de Regresión



La exactitud y precisión de los modelos depende de siempre del set de datos con los que trabajamos, ellos definen quien es el mejor.

Ojo!!! No existe un modelo ideal, no existe un modelo único

Regresión Lineal

Modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (4)$$

Donde, p es la cantidad de predictores, X_j representa el j -ésimo predictor, β_j son los parámetros desconocidos a estimar que cuantifican la asociación entre la variable predictora y la respuesta, y ϵ es el error aleatorio.

Además, β_j se interpreta como el efecto promedio en Y de un incremento en una unidad de X_j manteniendo todos los demás predictores fijos.

La estimación de los coeficientes $\hat{\beta}_0$ y $\hat{\beta}_1$ se realiza comúnmente mediante el método de Mínimos Cuadrados Ordinarios, el cual se enfoca en encontrar el valor de los coeficientes de regresión de modo tal que la suma de los cuadrados de las diferencias entre los valores observados y la línea de regresión sea mínima. Matemáticamente, esto es, minimizar la suma de cuadrados de los residuales (RSS):

$$RSS = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \dots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2 \quad (2)$$

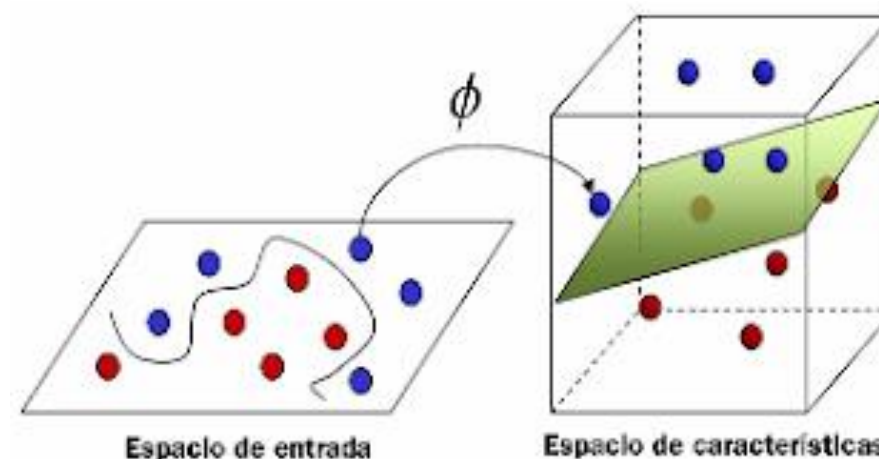
De donde se obtiene:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

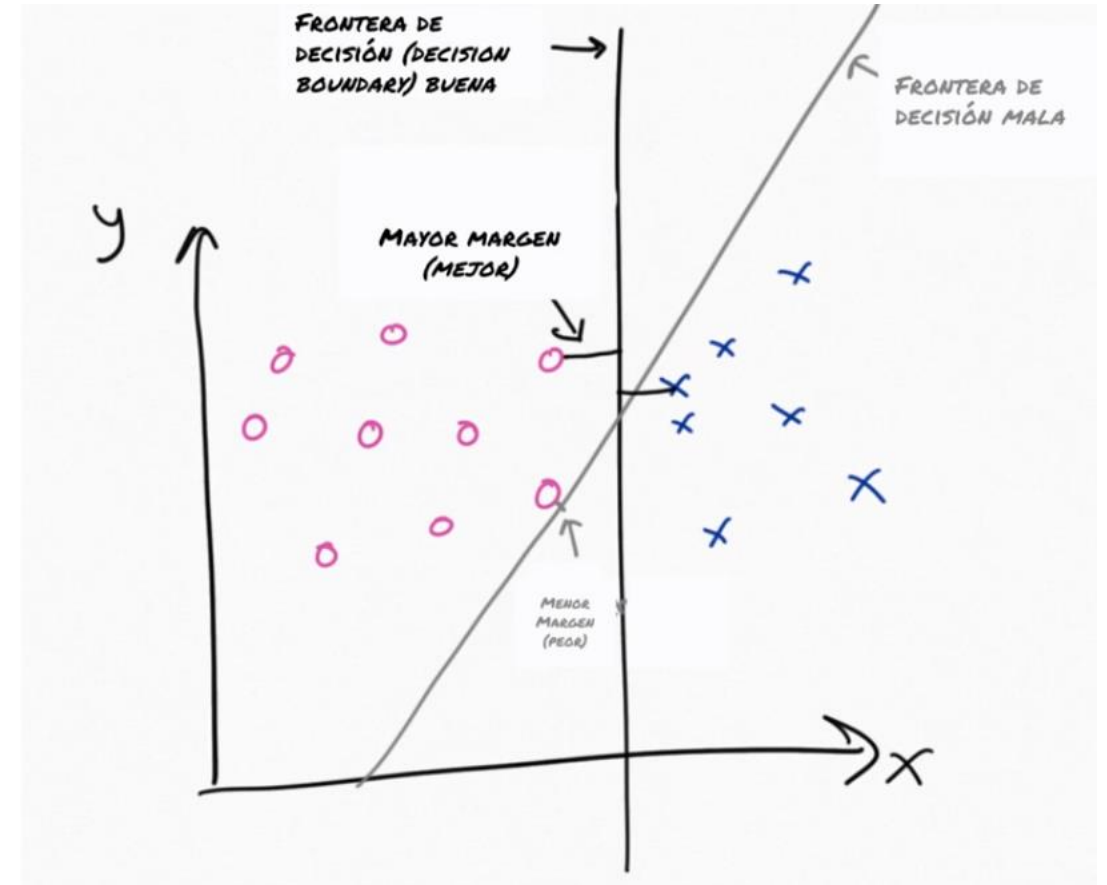
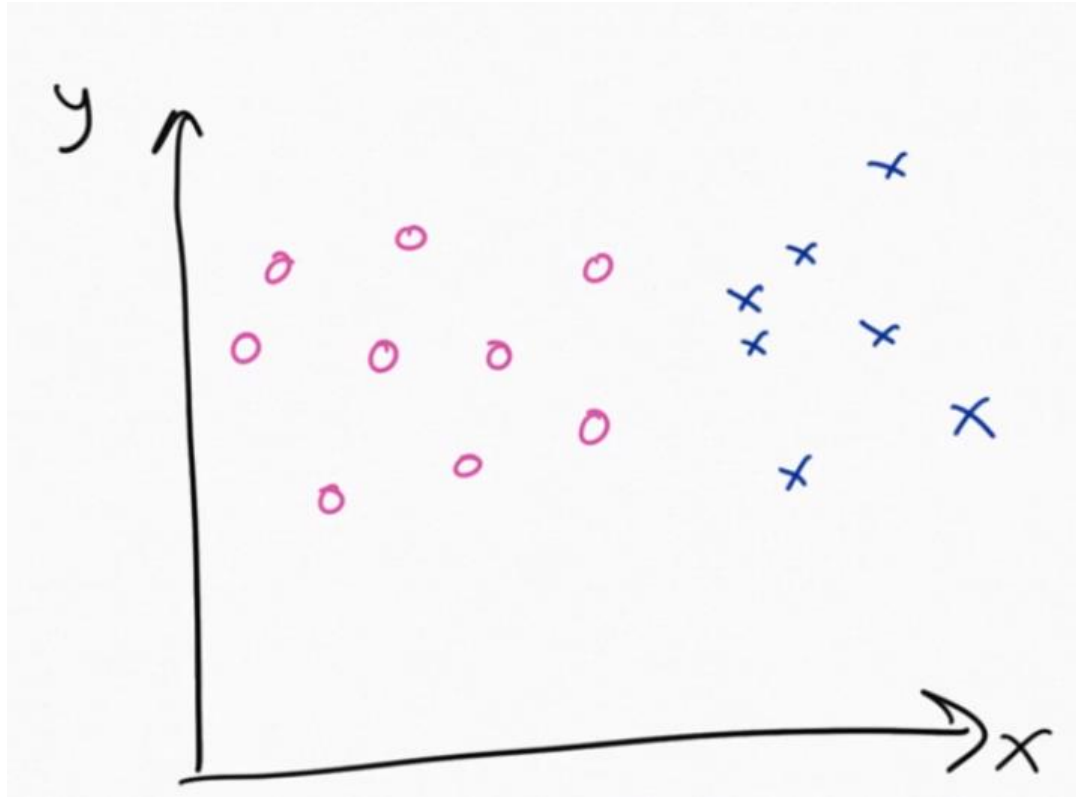
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Máquina de soporte vectorial

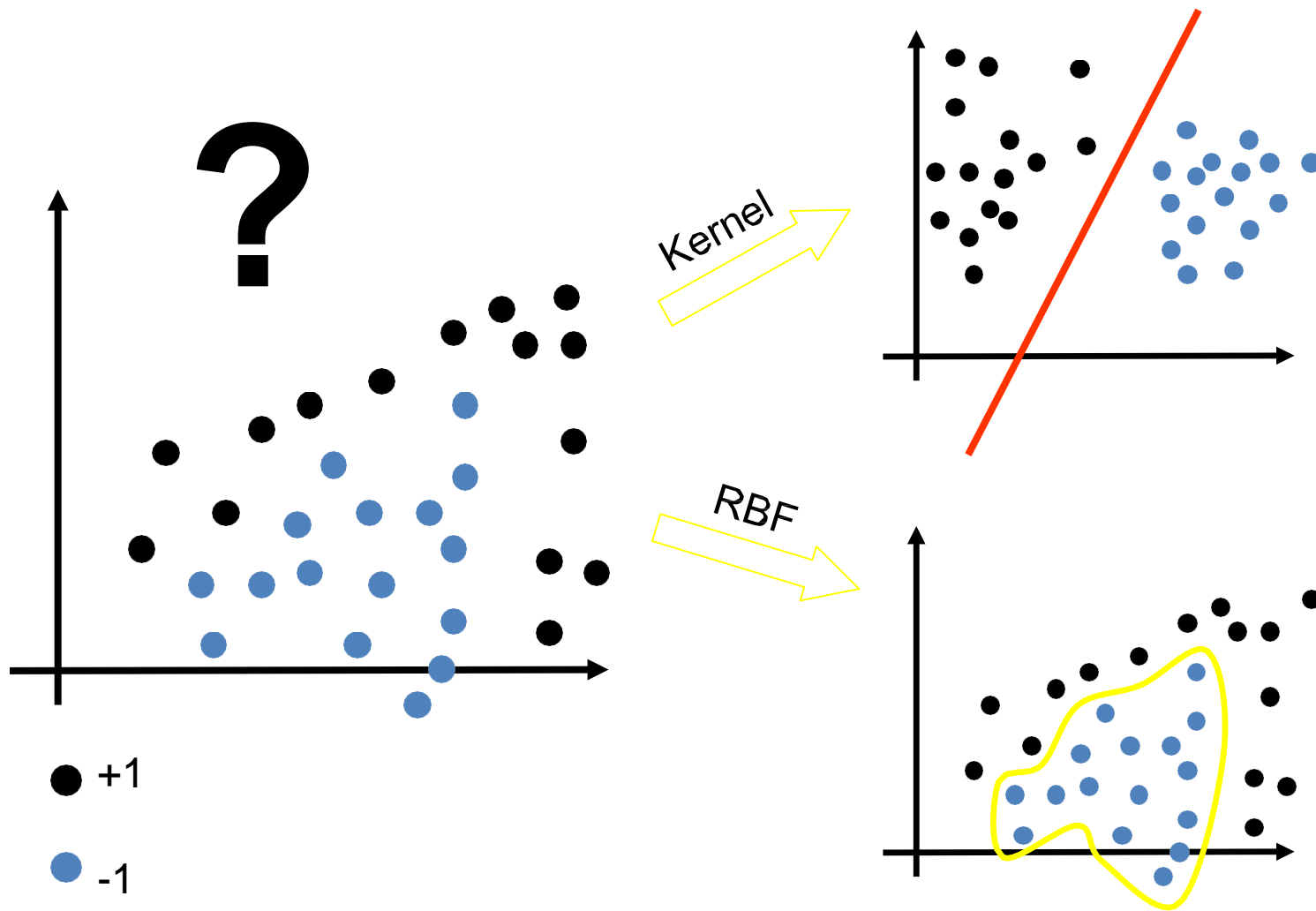
Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases, más cercanos al que se llama vector soporte. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o la otra clase.



Máquina de soporte vectorial



Máquina de soporte vectorial



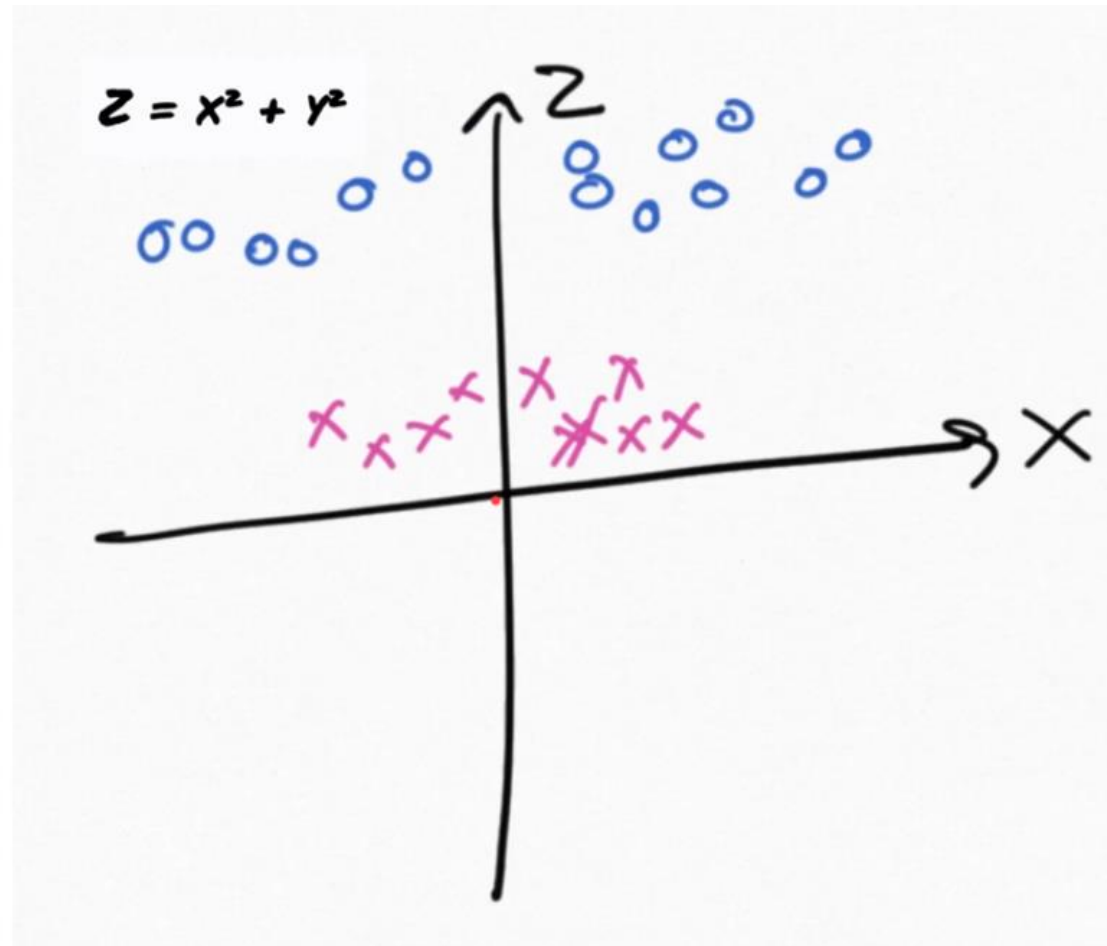
Cuando los datos no se pueden separar linealmente se hace un cambio de espacio mediante una función que transforme los datos de manera que se puedan separar linealmente. Tal función se llama **Kernel**.

También hay métodos para separar los datos (x_i, y_i) directamente aún no siendo separables linealmente, mediante **funciones polinómicas** y otro tipo de funciones, las **Funciones de Base Radial** (RBF).

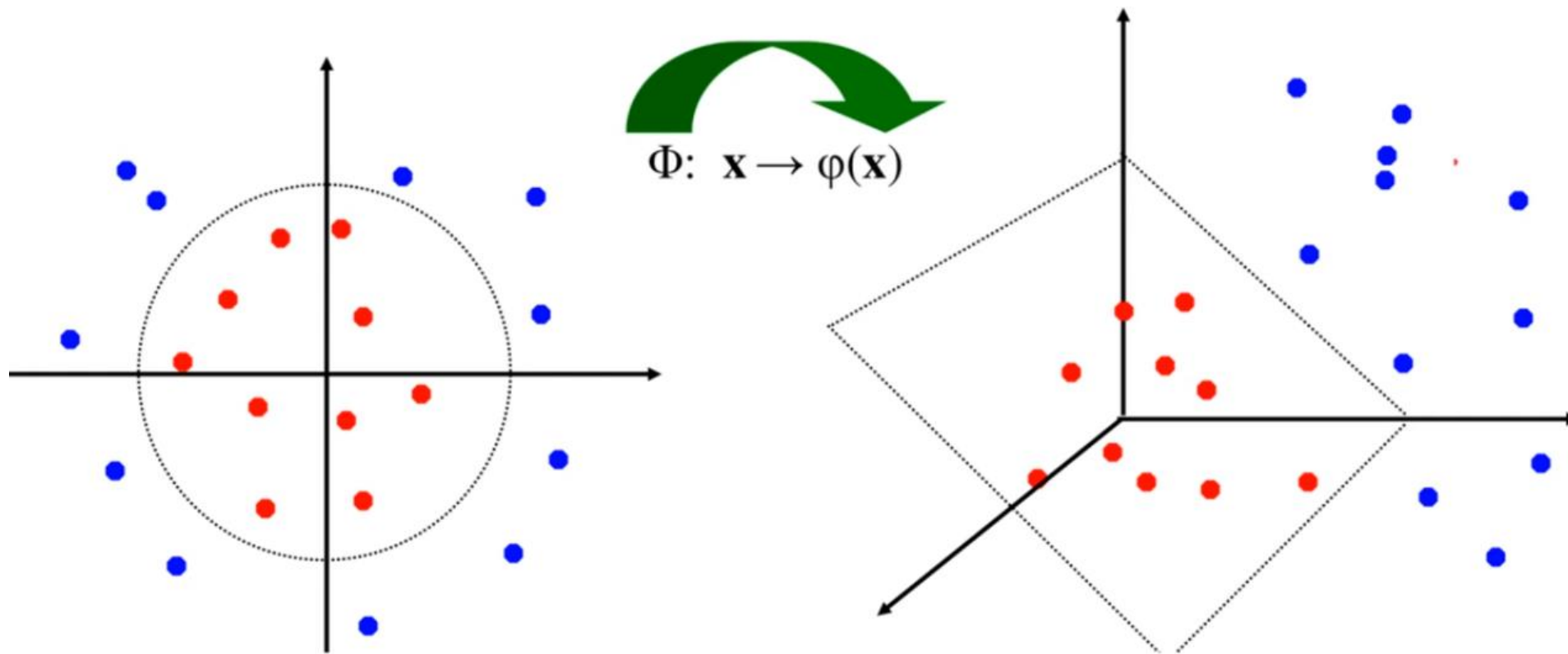
Máquina de soporte vectorial

Aplicamos una función que permite separar los puntos en un hiperplano.

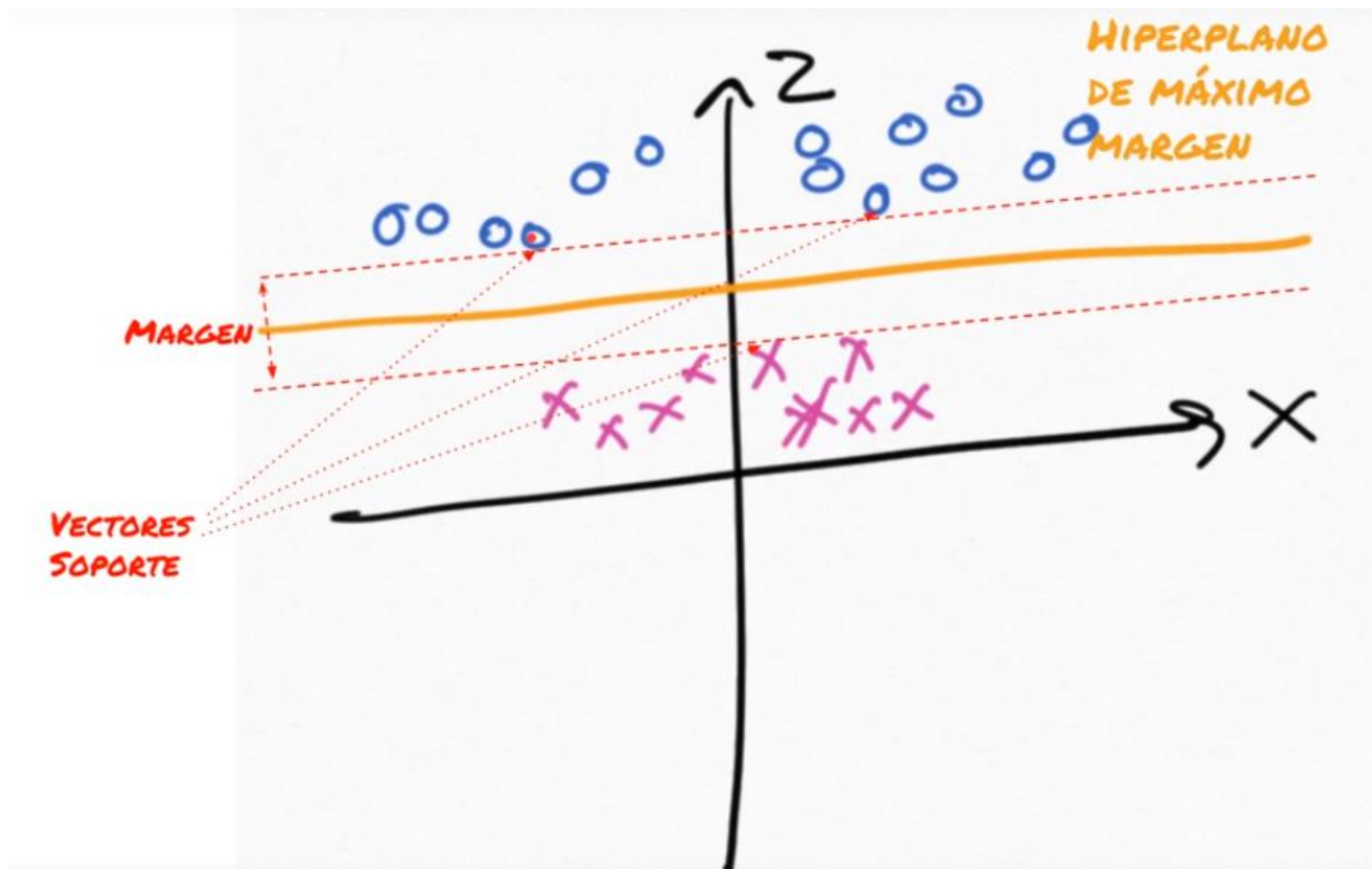
Para este caso particular en 2 dimensiones, la función $z = x^2 + y^2$ consigue separar linealmente las clases.



Máquina de soporte vectorial

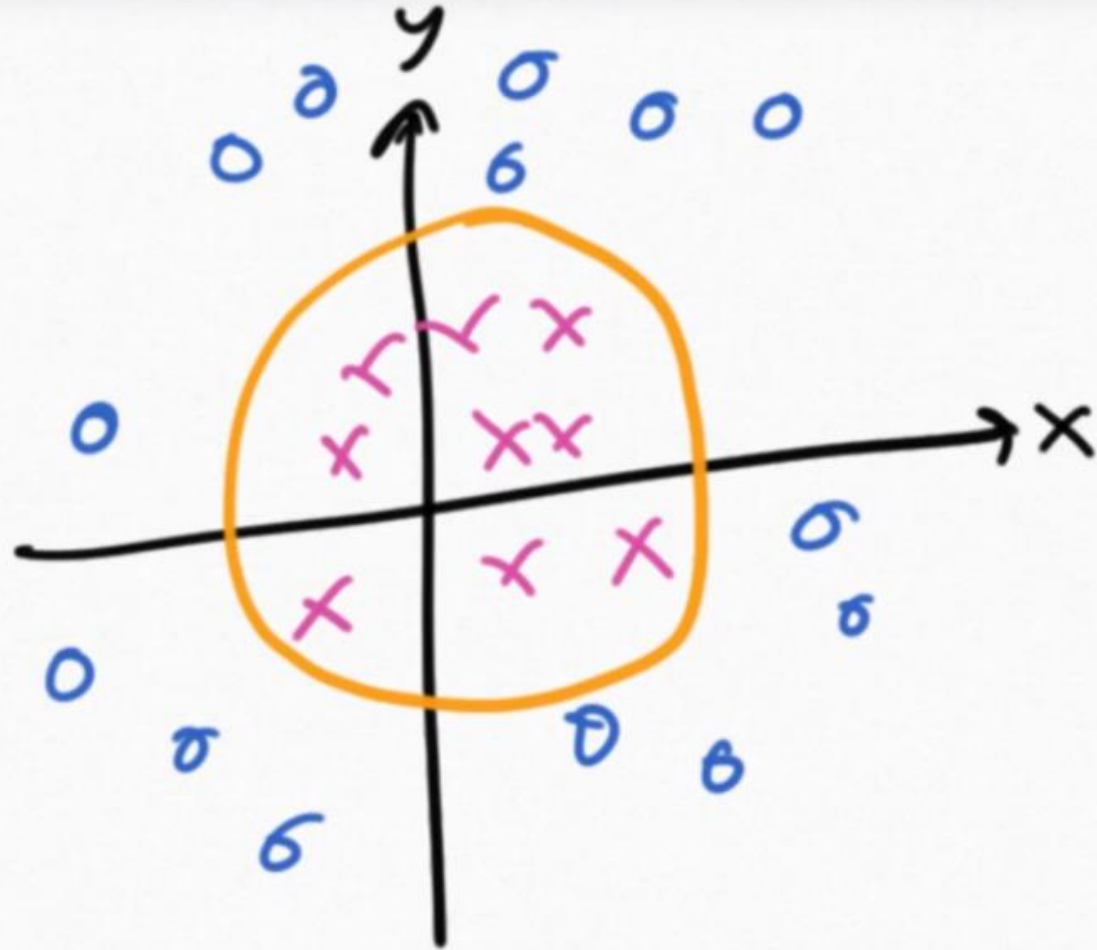


Máquina de soporte vectorial

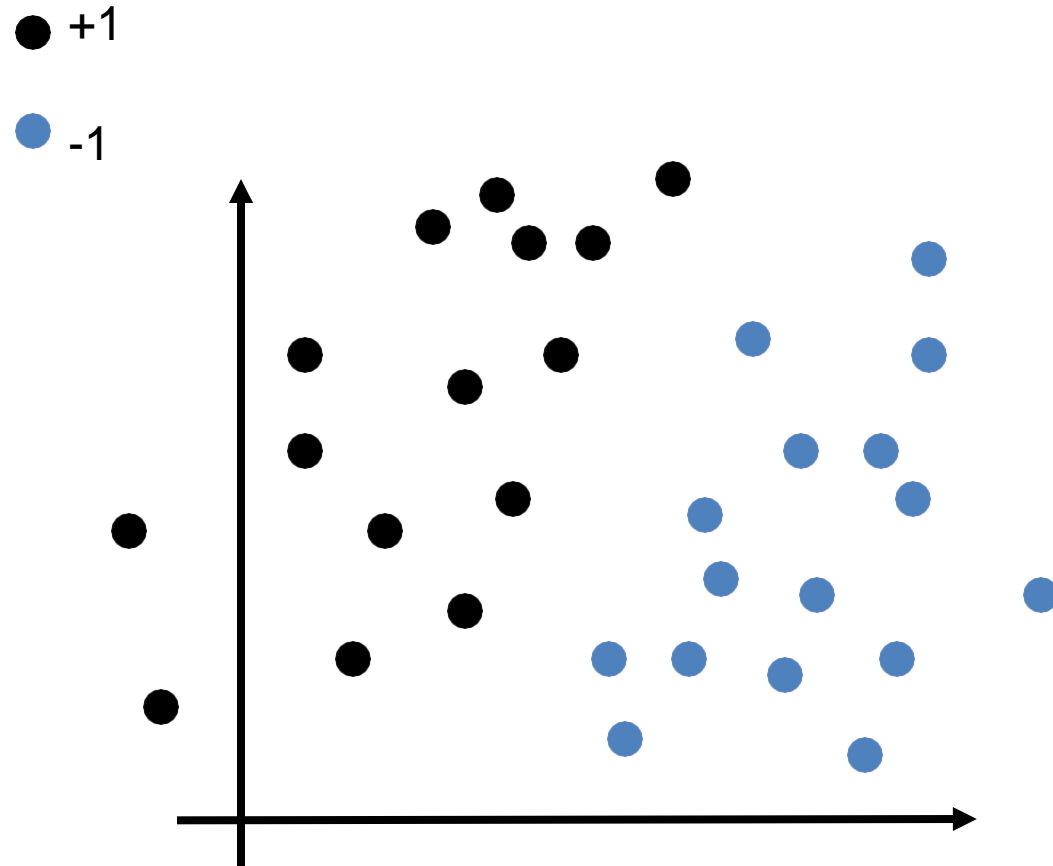


Máquina de soporte vectorial

Una vez tenemos la función de decisión en el hiperplano, podemos aplicarle una transformación y volverla a las dimensiones originales de los datos

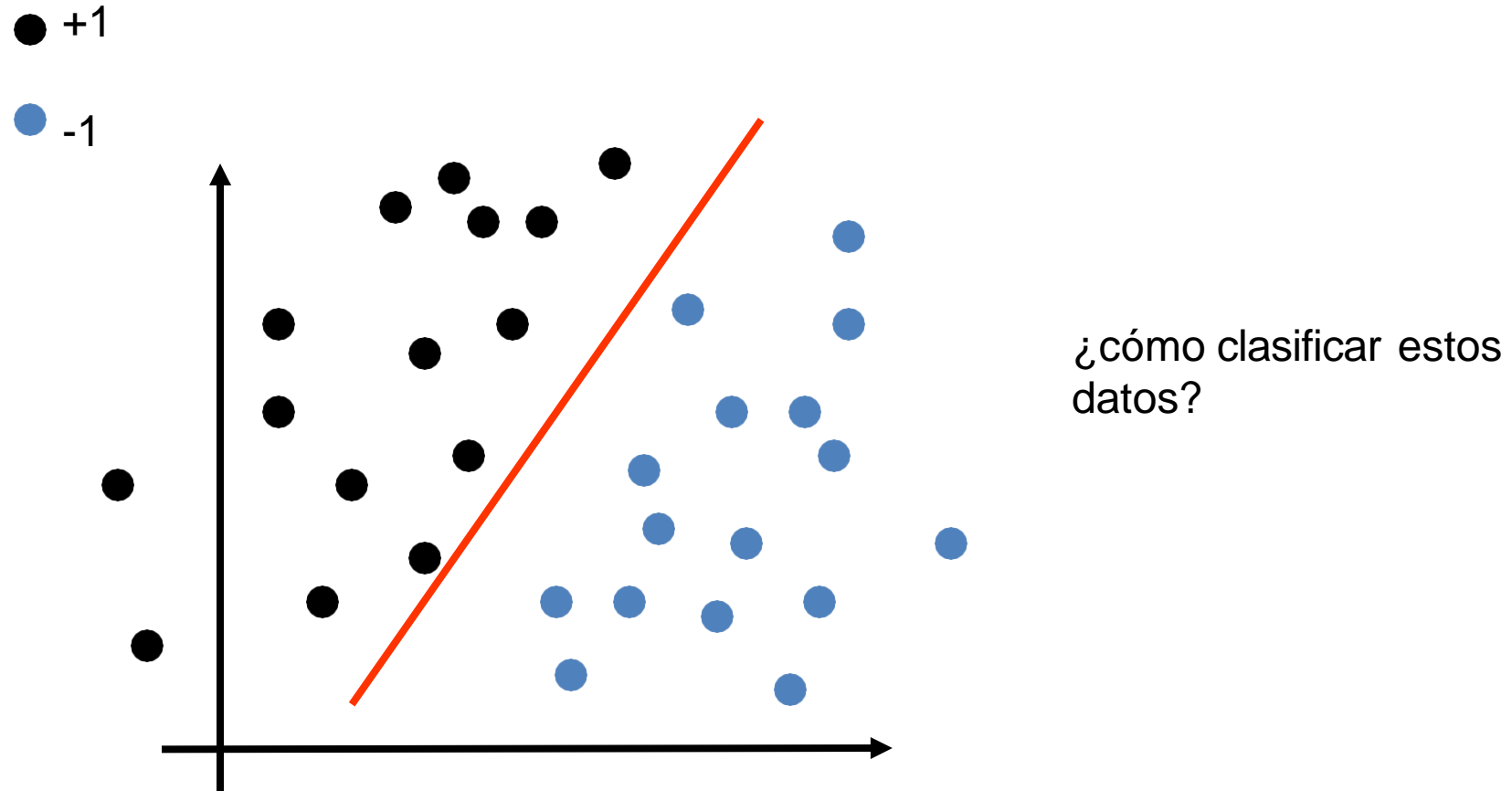


Máquina de soporte vectorial - Interpretación geométrica

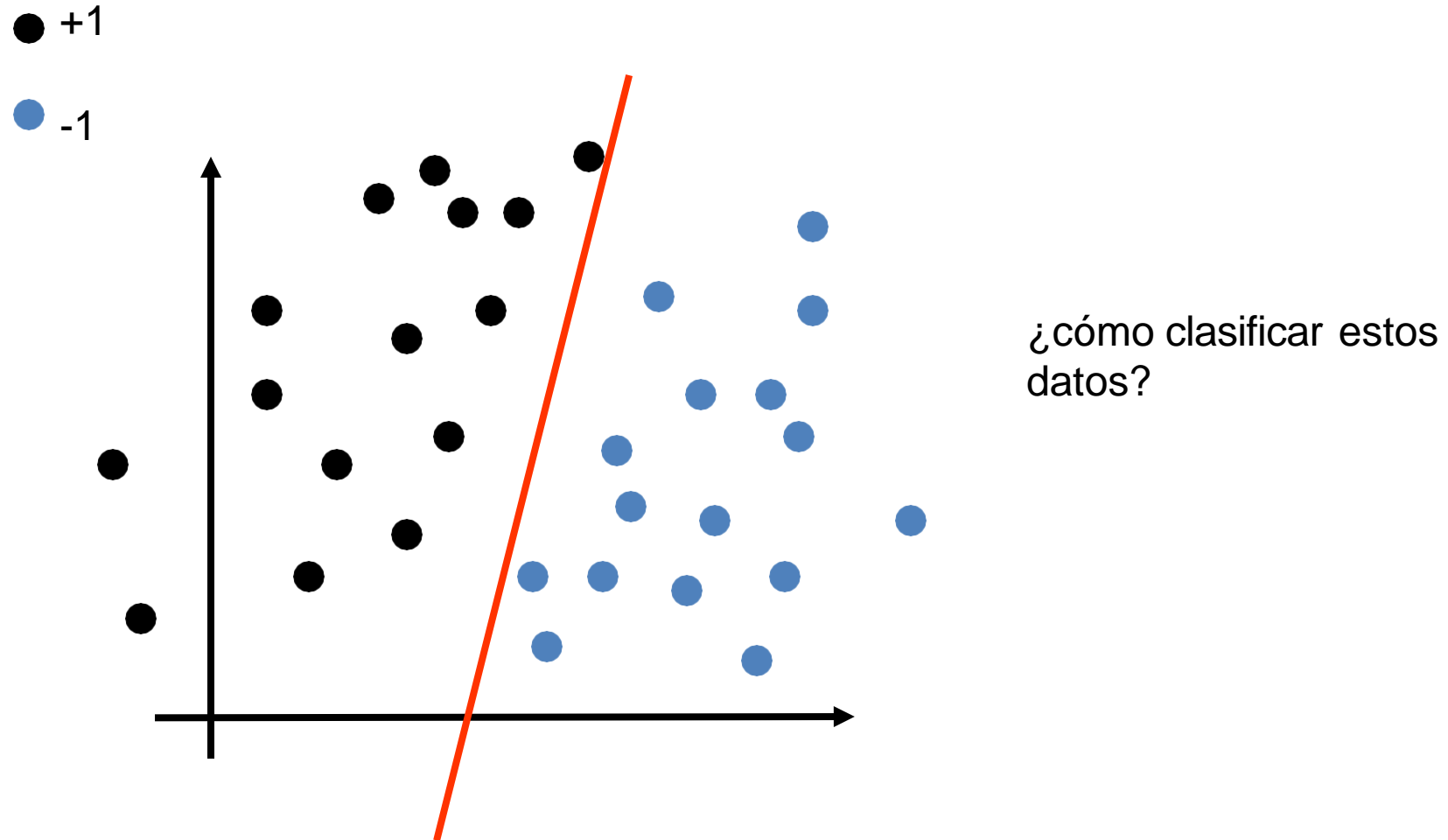


¿cómo clasificar estos datos?

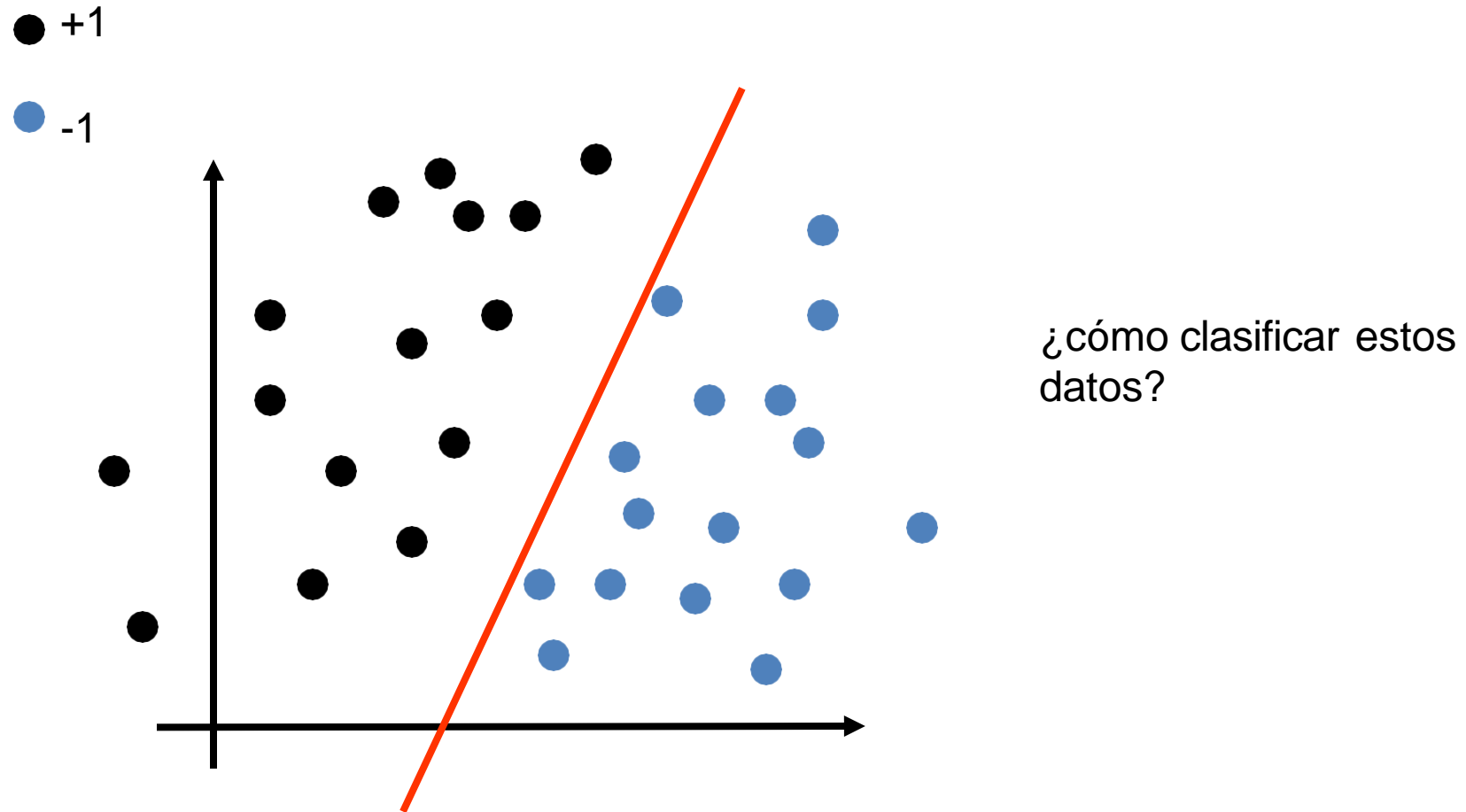
Máquina de soporte vectorial - Interpretación geométrica



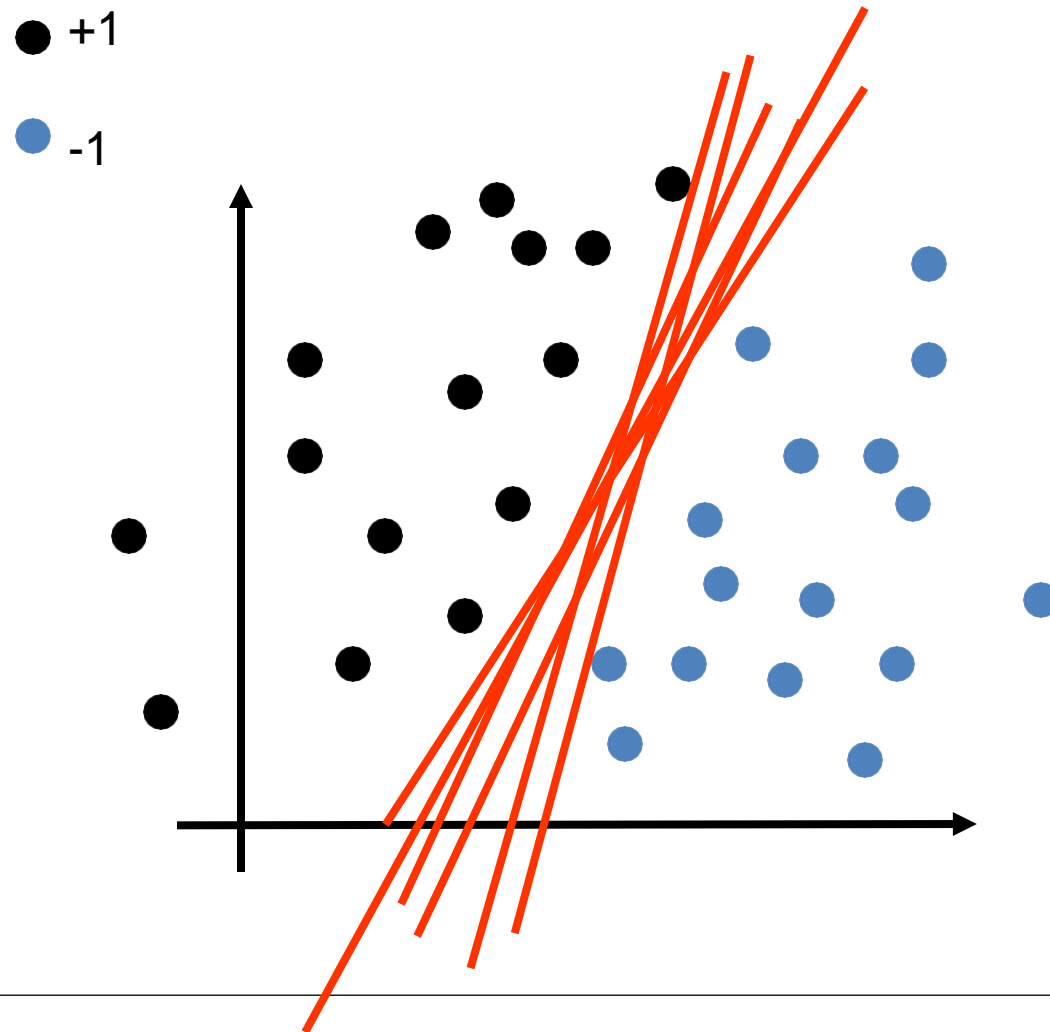
Máquina de soporte vectorial - Interpretación geométrica



Máquina de soporte vectorial - Interpretación geométrica

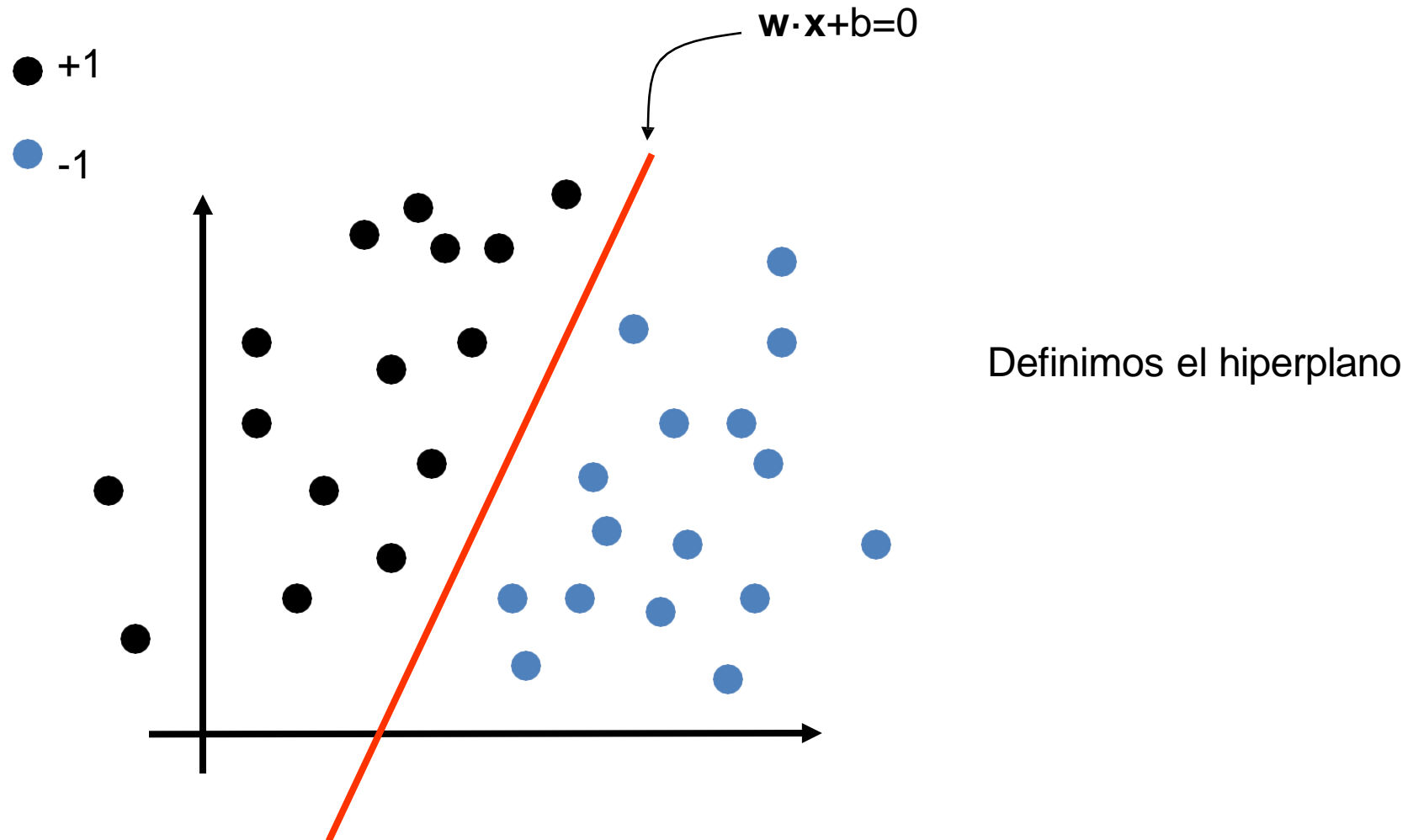


Máquina de soporte vectorial - Interpretación geométrica

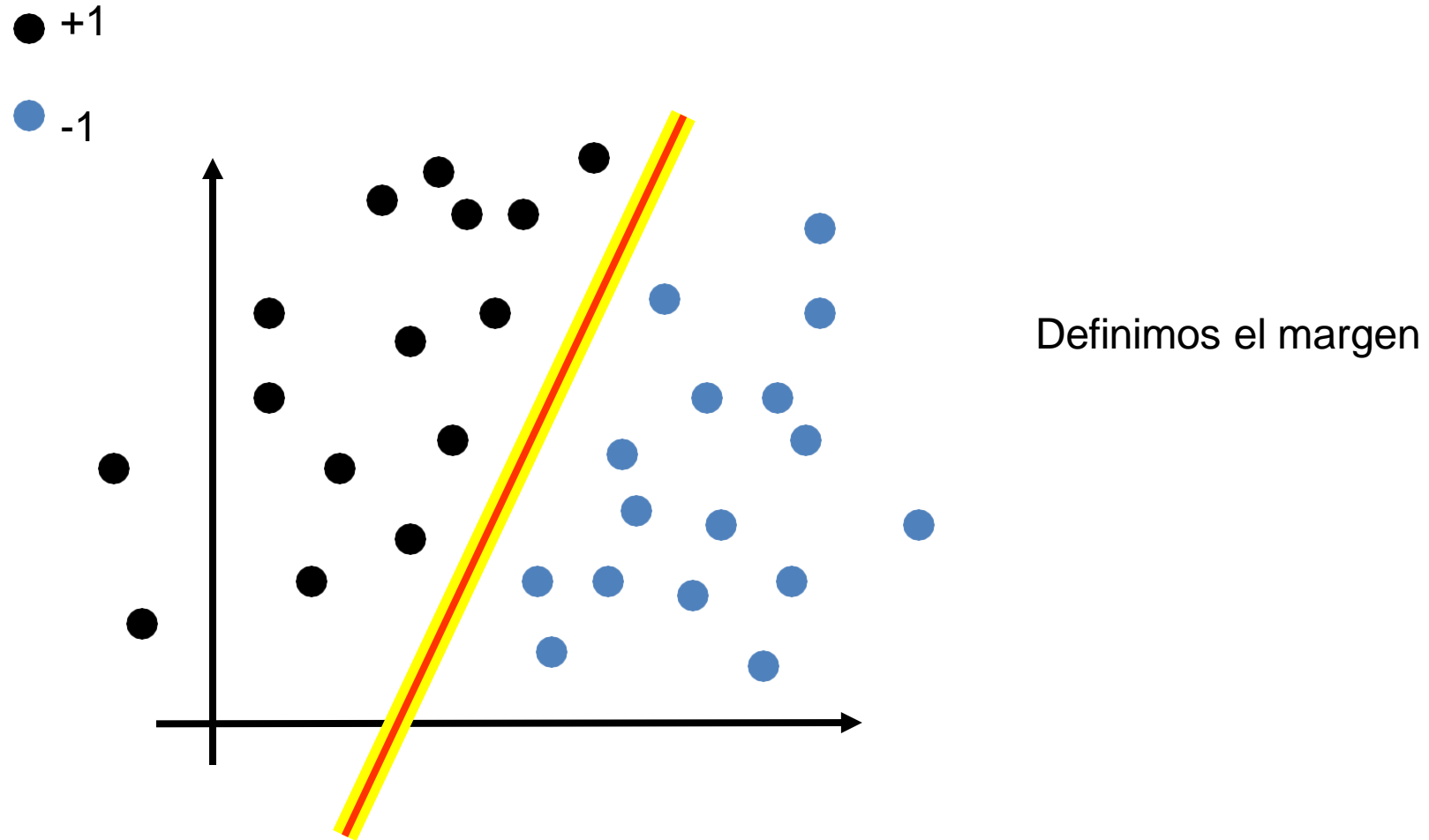


Cualquiera puede ser buena, ¿pero cuál es la mejor?

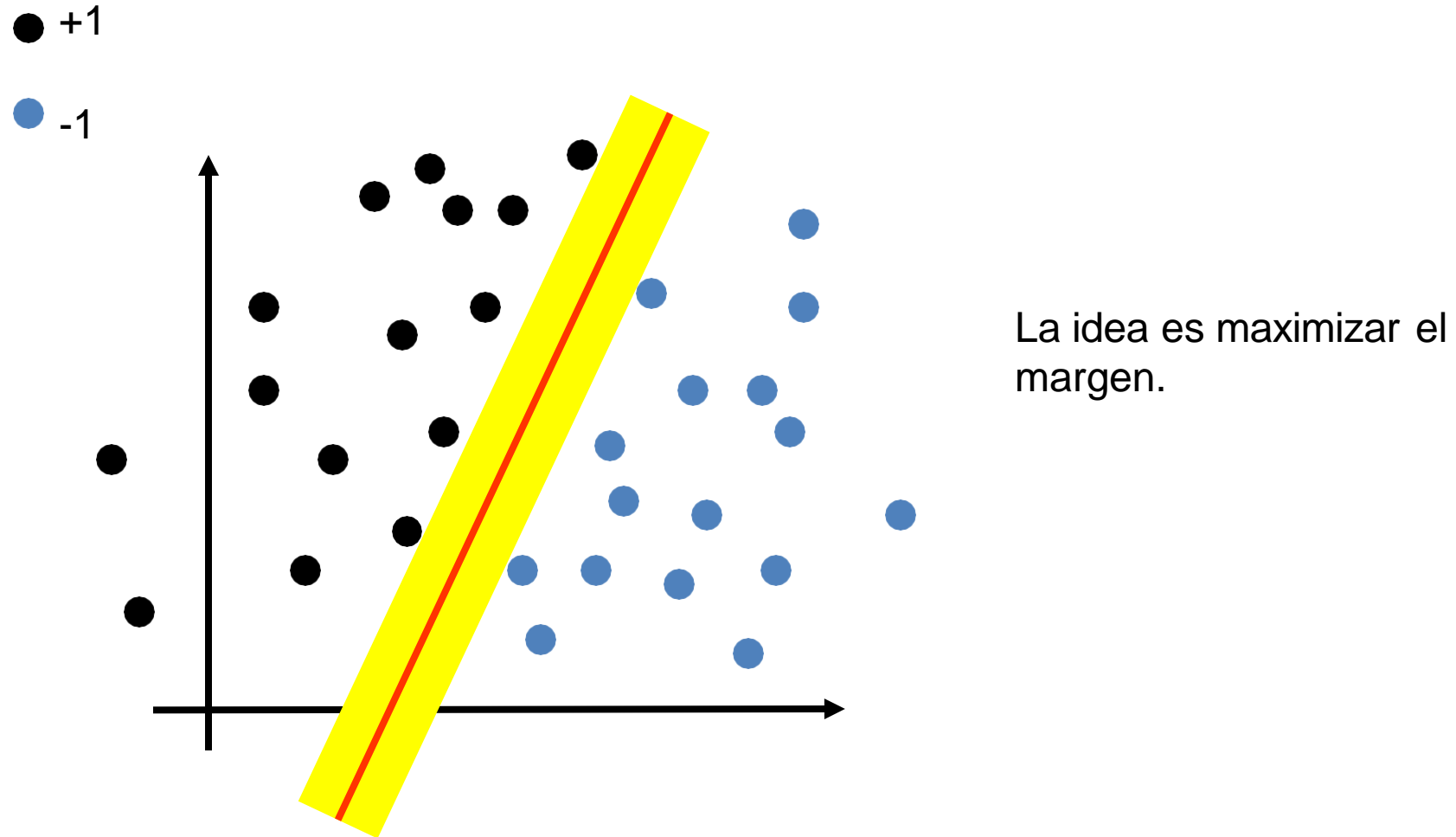
Máquina de soporte vectorial - Interpretación geométrica



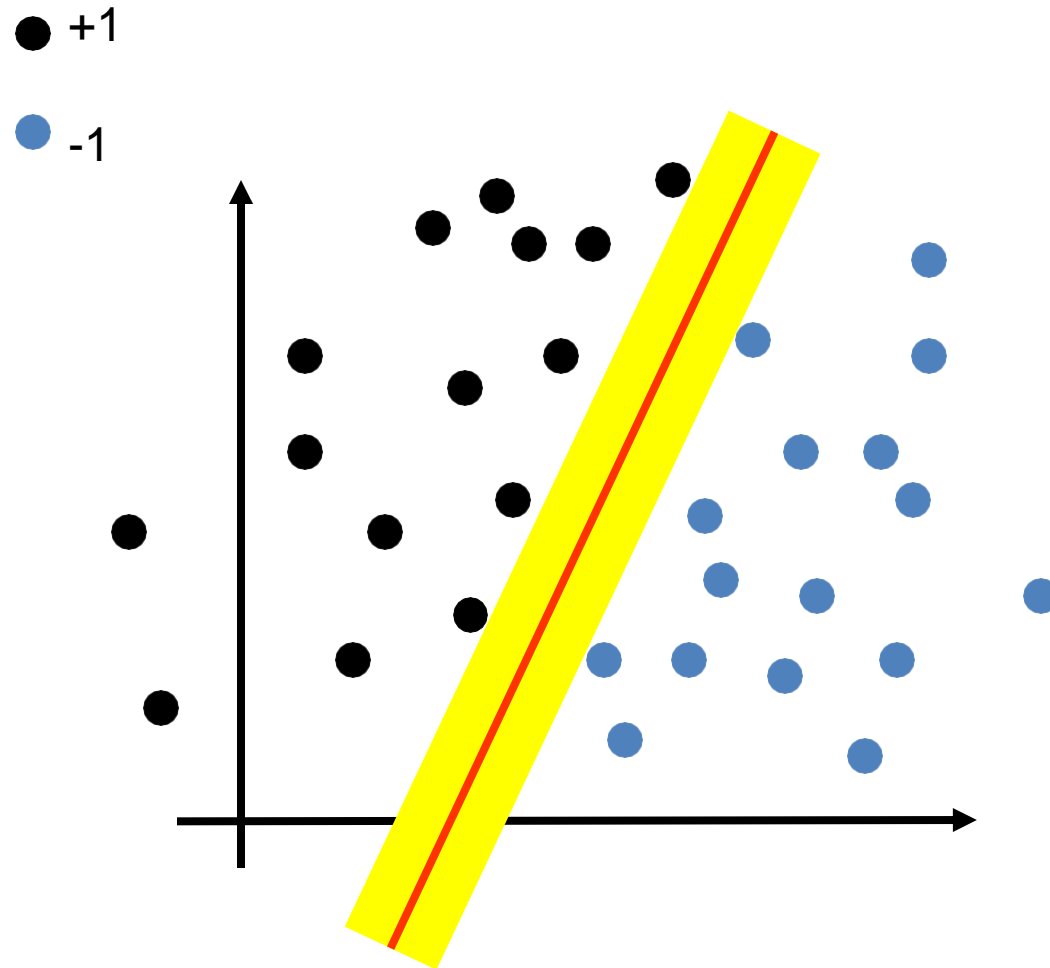
Máquina de soporte vectorial - Interpretación geométrica



Máquina de soporte vectorial - Interpretación geométrica



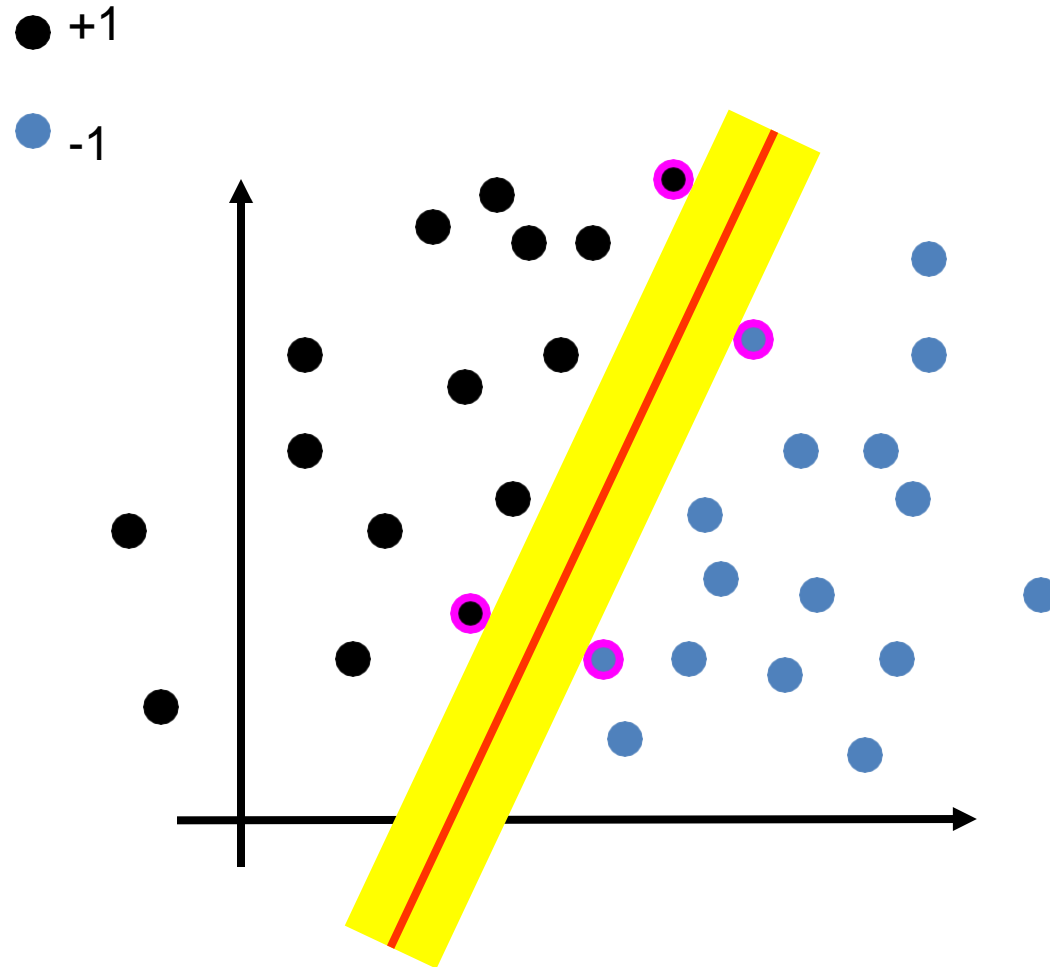
Máquina de soporte vectorial - Interpretación geométrica



El hiperplano que tenga el mayor margen es el mejor clasificador de los datos.

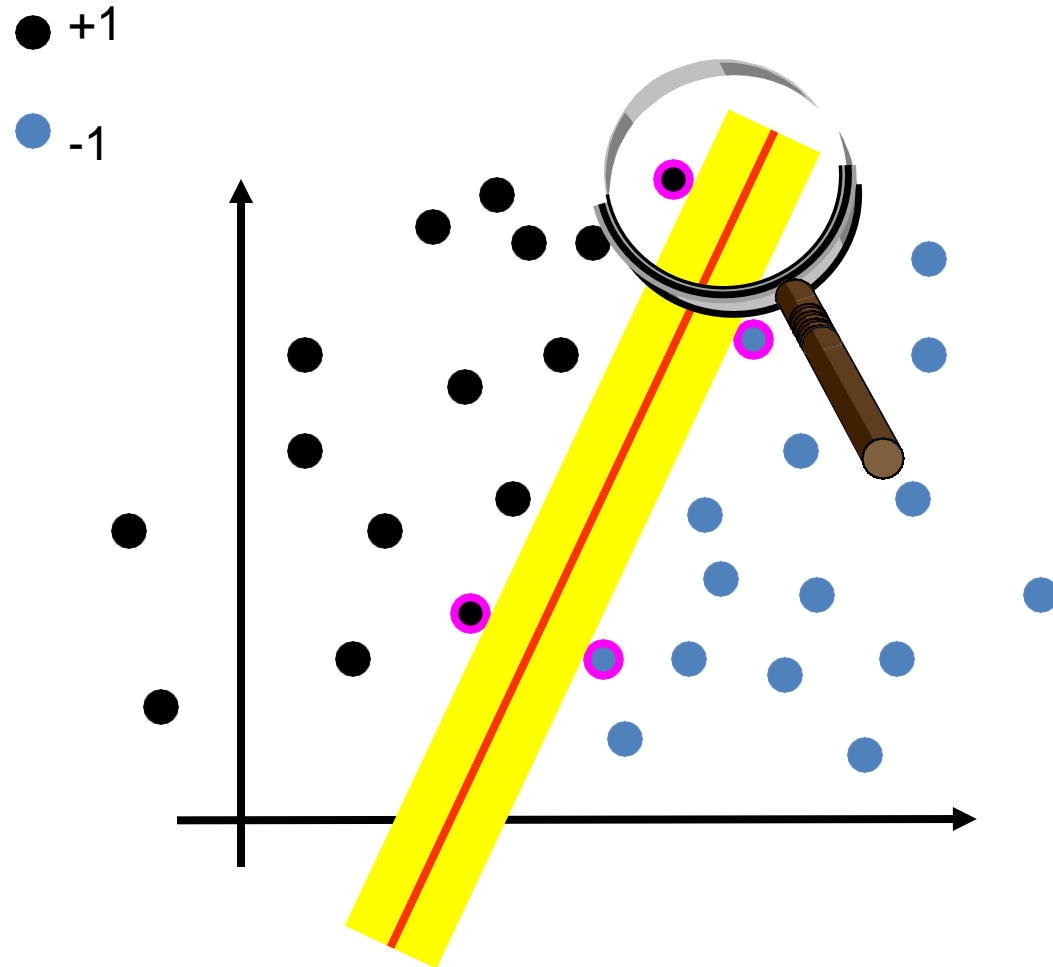
Esta es la clase más simple de SVM, la LSVM.

Máquina de soporte vectorial - Interpretación geométrica



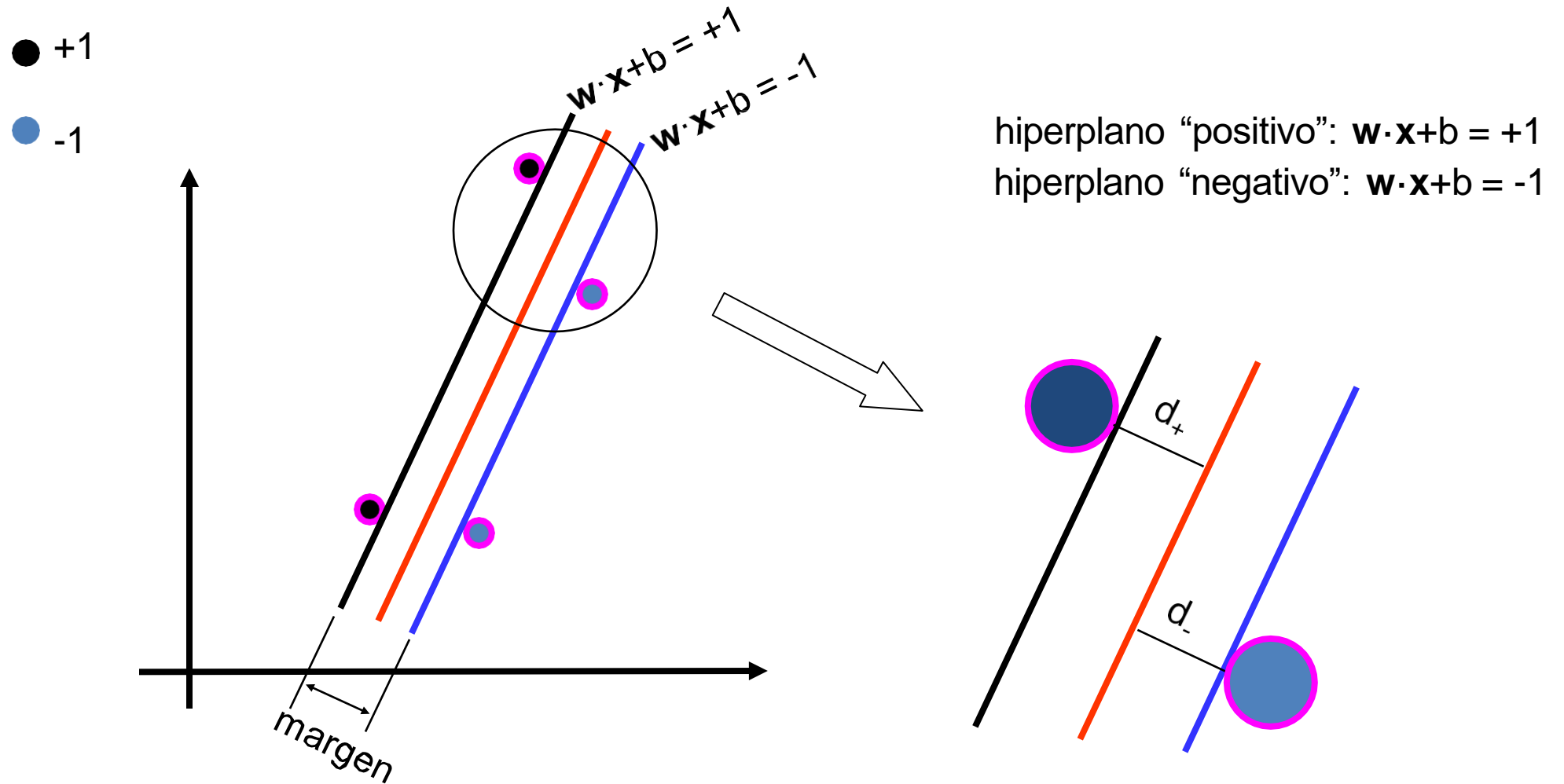
Los vectores de soporte son los puntos que tocan el límite del margen.

Máquina de soporte vectorial - Interpretación geométrica

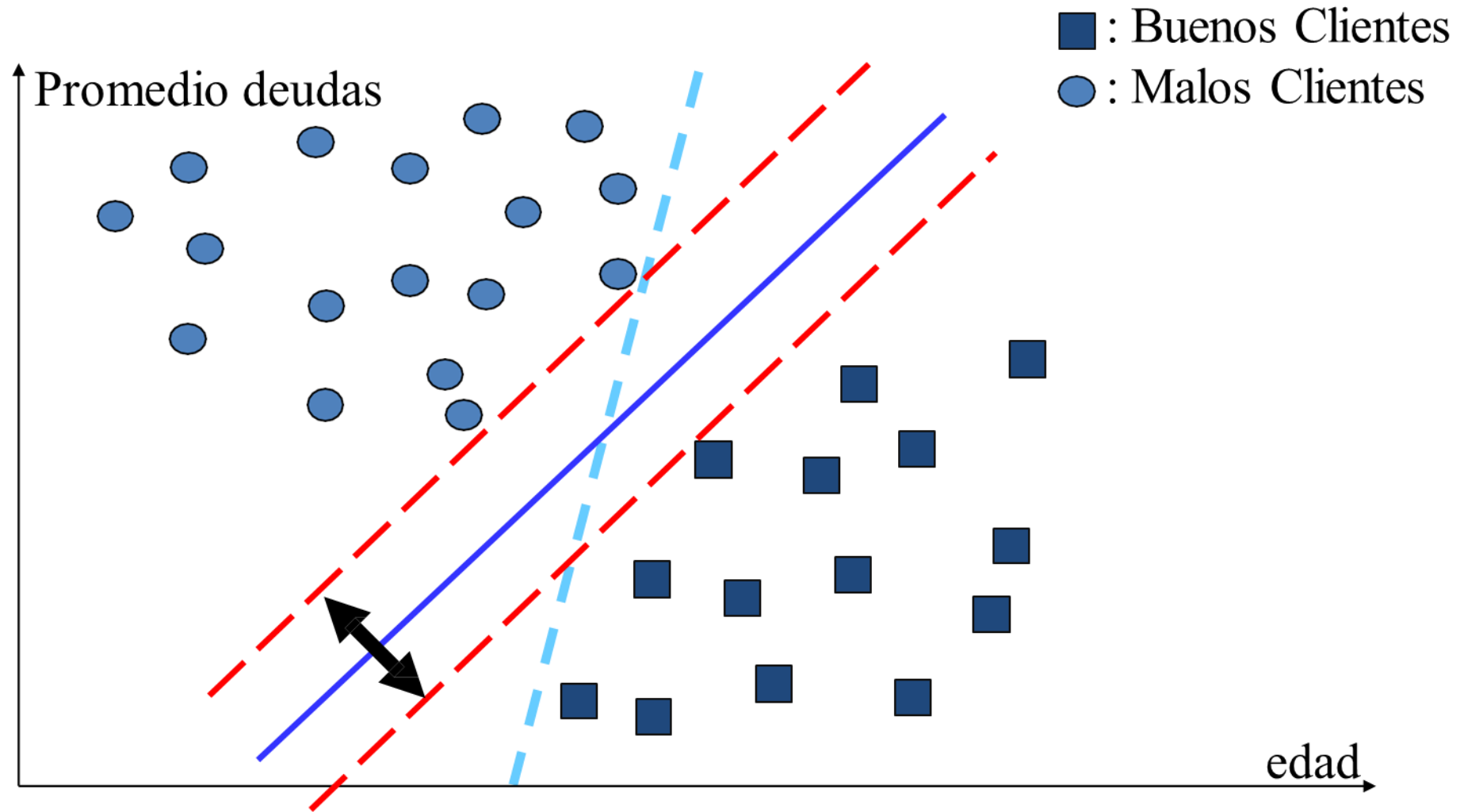


Veamos los hiperplanos
“positivo” y “negativo”

Máquina de soporte vectorial - Interpretación geométrica



Máquina de soporte vectorial - Interpretación geométrica



Árboles de decisión

Algoritmos: ID3, C4.5 (Quinlan), CART (Breiman et al.)

Construcción del árbol:

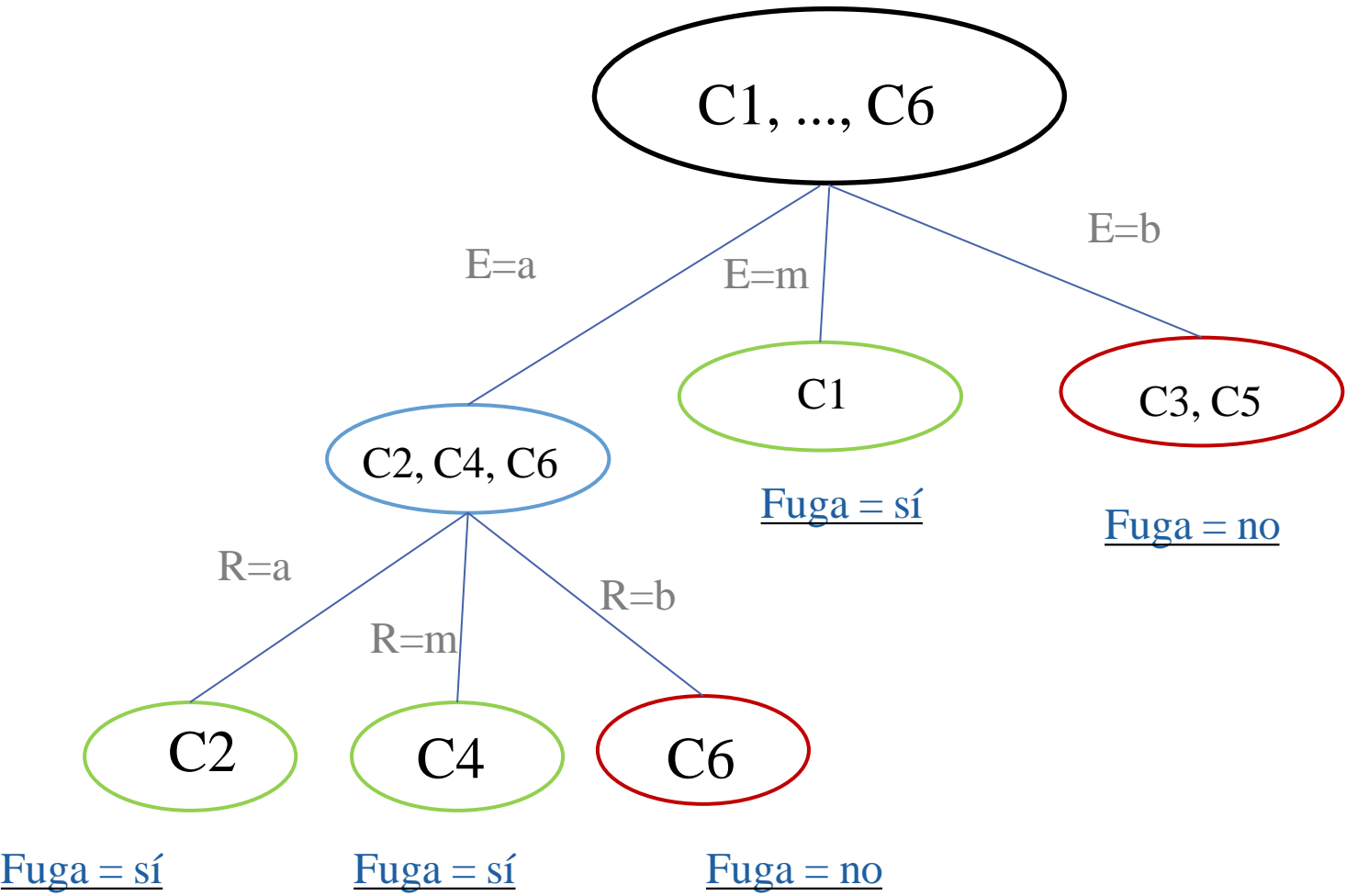
criterio de detención, criterio para seleccionar un atributo discriminante

Idea básica de ID3: (ejemplos tienen 2 clases: positivo, negativo)

Criterio de detención: Detiene la construcción del árbol si cada hoja del árbol tiene solamente ejemplos de una clase (pos. o neg.)

	Edad	Renta	Fuga?
C1	medio	alto	sí
C2	alto	alto	sí
C3	bajo	bajo	no
C4	alto	medio	sí
C5	bajo	medio	no
C6	alto	bajo	no

Árboles de decisión



	Edad	Renta	Fuga?
C1	medio	alto	sí
C2	alto	alto	sí
C3	bajo	bajo	no
C4	alto	medio	sí
C5	bajo	medio	no
C6	alto	bajo	no

Reglas a partir del árbol:

Si: E = a y R = a,
entonces Fuga = Si

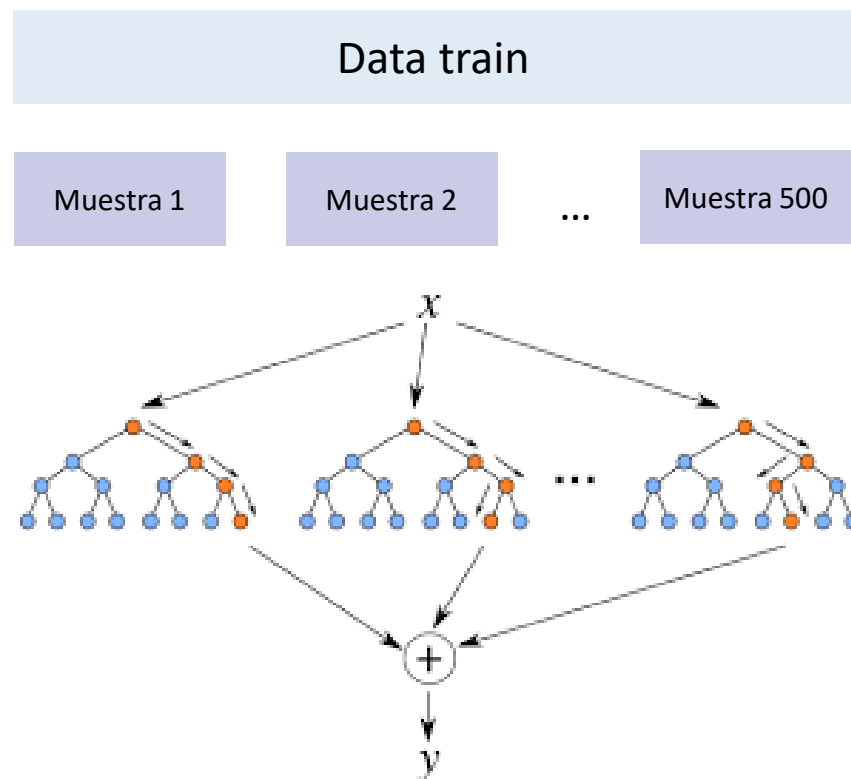
Si: E = a y R = b,
entonces Fuga = No

...

Bosque Aleatorio

El algoritmo de un árbol por separado no puede producir modelos exactos, porque la variedad provoca una inestabilidad que se puede observar al crear árboles de decisiones por separado.

El resultado de un bosque aleatorio es el promedio del resultado de un grupo de árboles de clasificación, donde los árboles son contruidos con diferentes muestras del train y seleccionando en cada nodo un subconjunto de las variables predictoras. Con esto se consigue resultados más robustos al cambio de los datos y con el subconjunto de predictores en cada nodo conseguimos que se obtenga más información de todas las variables. Funciona tanto para clasificación como regresión.



Métricas de performance


Las métricas de evaluación disponibles para los modelos de regresión son:

- Error cuadrático medio (MSE)



$$\frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2$$

- Raíz del error cuadrático medio (RMSE)

- Error absoluto promedio (Mean Absolute Error, MAE) 

$$\frac{1}{n} \sum_i^n |y_i - \hat{y}_i|$$

- Raíz del MAE (Root Mean Absolute Error, RMAE)

- Coeficiente de determinación



$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$
