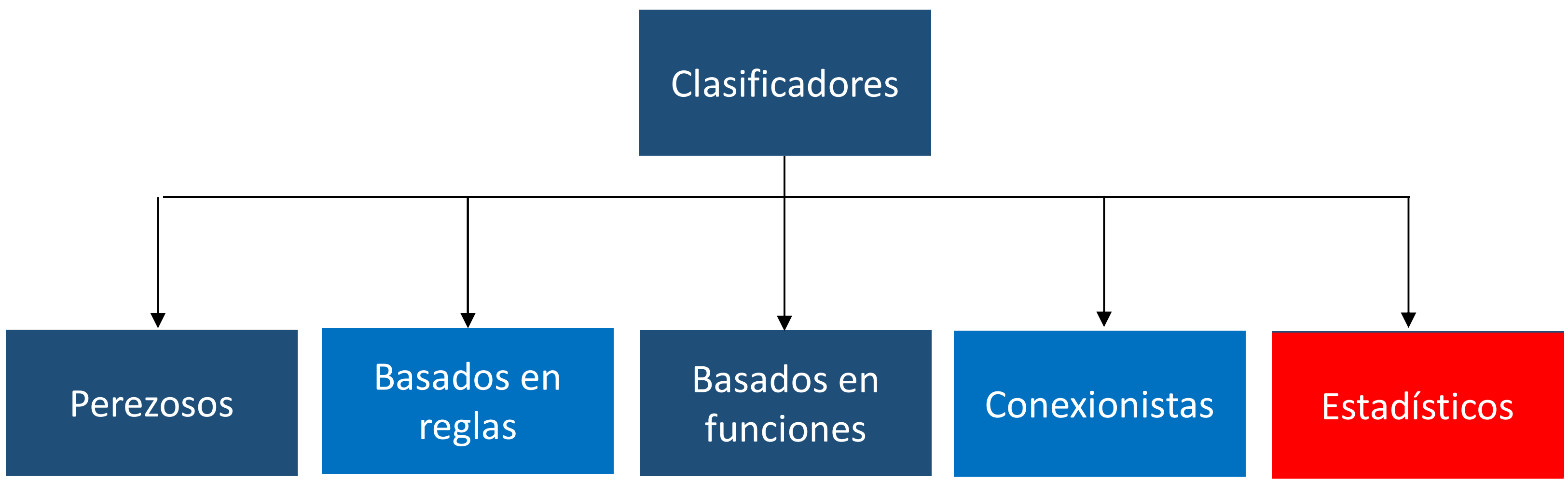
Aprendizaje Maquinal

Dr. Rubén Acevedo
ruben.acevedo@uner.edu.ar

Tecnicatura Universitaria
Procesamiento y Explotación de Datos

Clasificación

Tipos



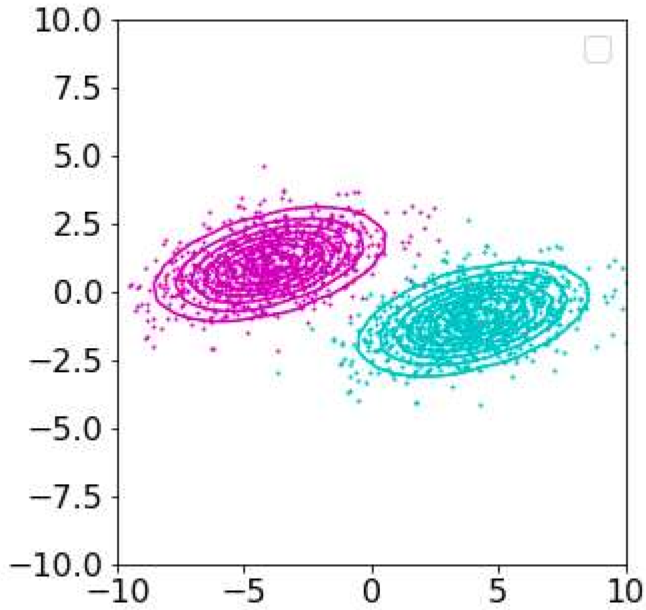
Clasificación

Discriminante lineal

El análisis por discriminante lineal se puede utilizar como tanto como clasificador así como también como estrategia para reducción de la dimensionalidad.

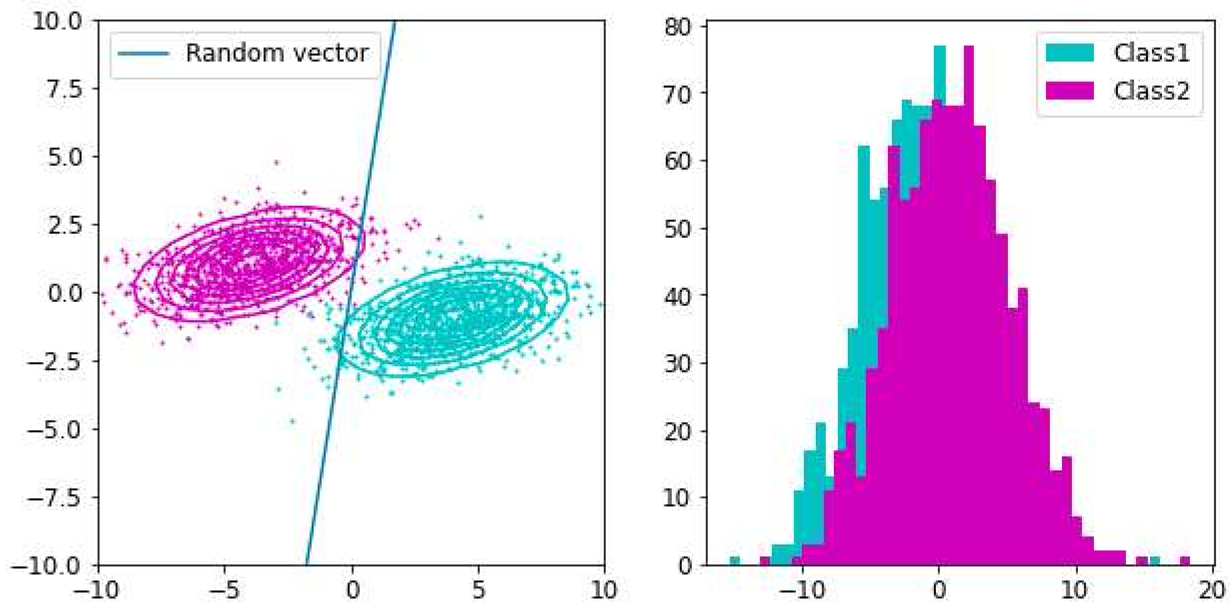
Clasificadores estadísticos

Discriminante lineal



Funciones gaussianas bivariadas con matrices de covarianza idénticas y medias distintas.

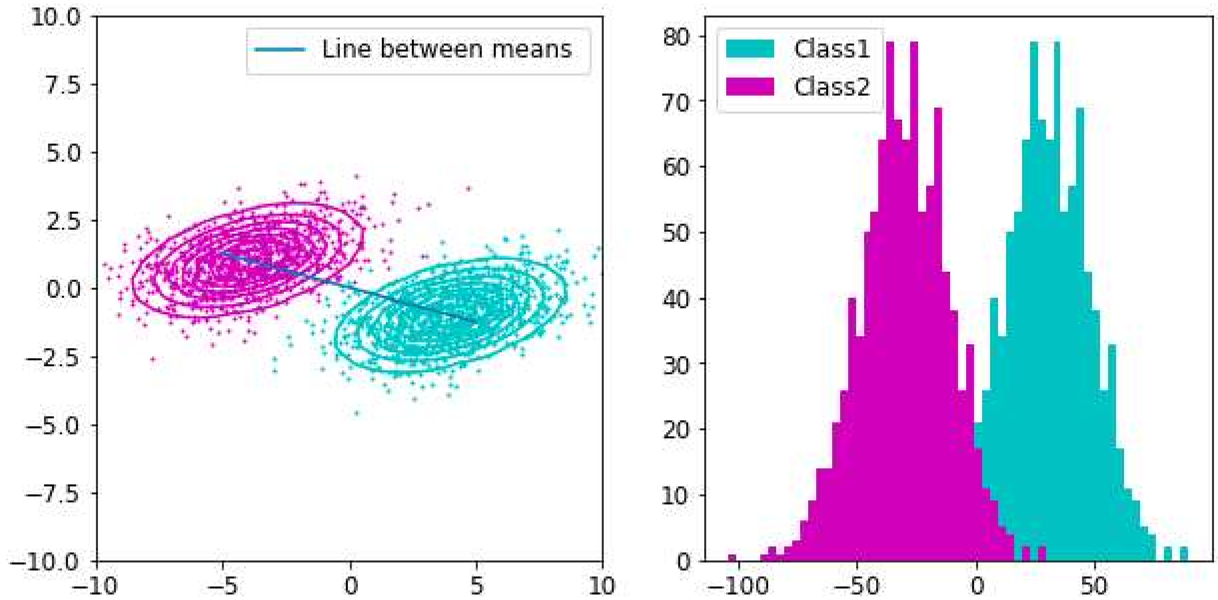
Discriminante lineal



Proyecciones de los datos sobre el vector (el producto escalar del vector de pesos y la matriz de datos).

Clasificadores estadísticos

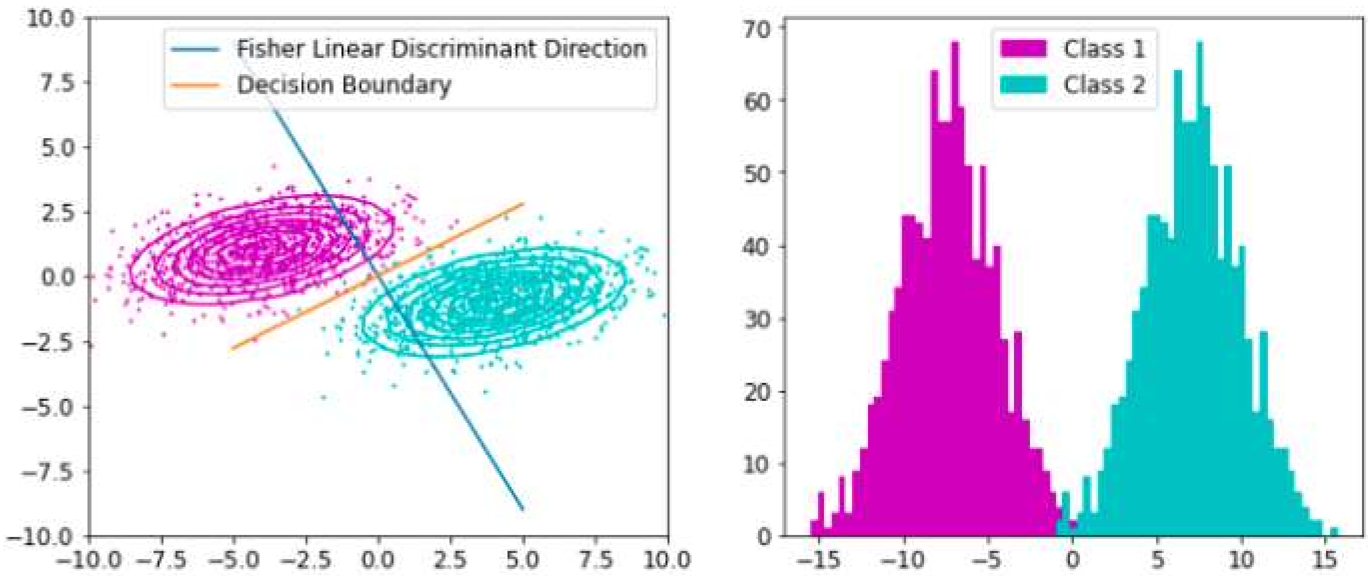
Discriminante lineal



Maximización de la distancia entre las medias proyectadas (las distribuciones están, en promedio, lo más alejadas posible entre sí).

Clasificadores estadísticos

Discriminante lineal



El discriminante lineal de Fisher minimiza la varianza intra-clase de las proyecciones, al mismo tiempo que maximiza las proyecciones entre las medias.

Clasificación

Discriminante lineal

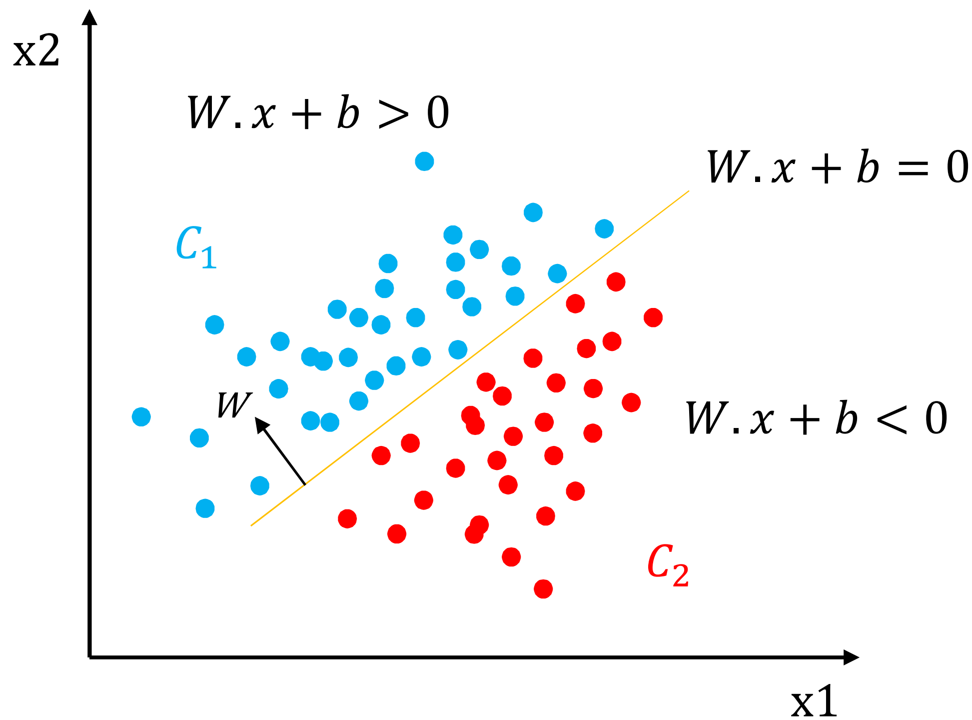
Criterio de Fisher

Concepto: La potencia de discriminación se maximiza cuando la distribución espacial de los datos (instancias) de clases distintas se encuentra *tan distantes como sea posible* y la distribución espacial de las muestras dentro de la misma clase se encuentra *tan cerca como sea posible.*

Clasificación

Discriminante lineal

Criterio de Fisher



Clasificación

Discriminante lineal | **Criterio de Fisher**

$$w_1.x_1 + w_2.x2_1 + b = W.x + b = 0$$ Ecuación del plano

$$W = \hat{\Sigma}_C^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$$ Vector de proyección

$\hat{\mu}_i$: media estimada de la clase C_i

$\sum_C = \frac{1}{2}.\left(\hat{\Sigma}_1 + \hat{\Sigma}_2\right)$: matriz de covarianza estimada

Clasificación

Discriminante lineal | **Criterio de Fisher**

$$\boldsymbol{X} = [\bar{x}_1, \bar{x}_2, .., \bar{x}_K]$$

Conjunto de K datos

$$\hat{\mu}_i = \frac{1}{K_i} . \sum_{\bar{x}_i \epsilon C_i} \bar{x}_i$$

Media de la clase C_i

$$\hat{\sigma}_i{}^2 = \frac{1}{K_i} . \sum_{\bar{x}_i \epsilon C_i} \|\bar{x}_i - \hat{\mu}_i\|^2$$

Varianza de la clase C_i

Clasificación

Discriminante lineal

Criterio de Fisher

$$\hat{\mu} = \frac{1}{K} . \sum_{k=1}^{K} \bar{x}_i$$

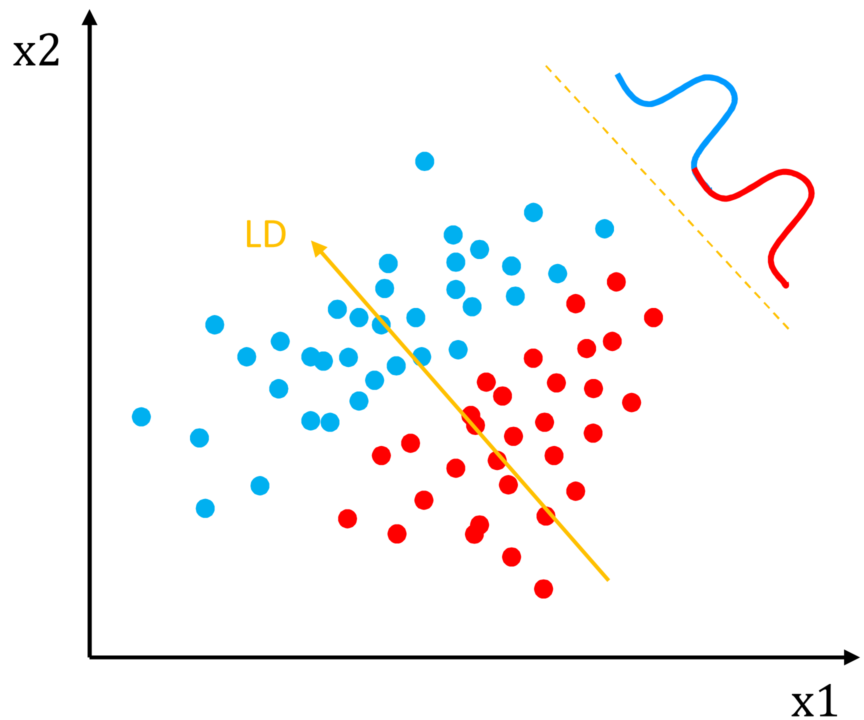
Media de todos los datos

$$F(\boldsymbol{X}) = \frac{SB}{SW} = \frac{\left\| \sum_{k=1}^{K} K_i . (\bar{\mu}_i - \mu)(\bar{\mu}_i - \mu)^T \right\|^2}{\sum_{i=1}^{c} \bar{\sigma}_i{}^2}$$

Clasificación

Discriminante lineal | **Criterio de Fisher**

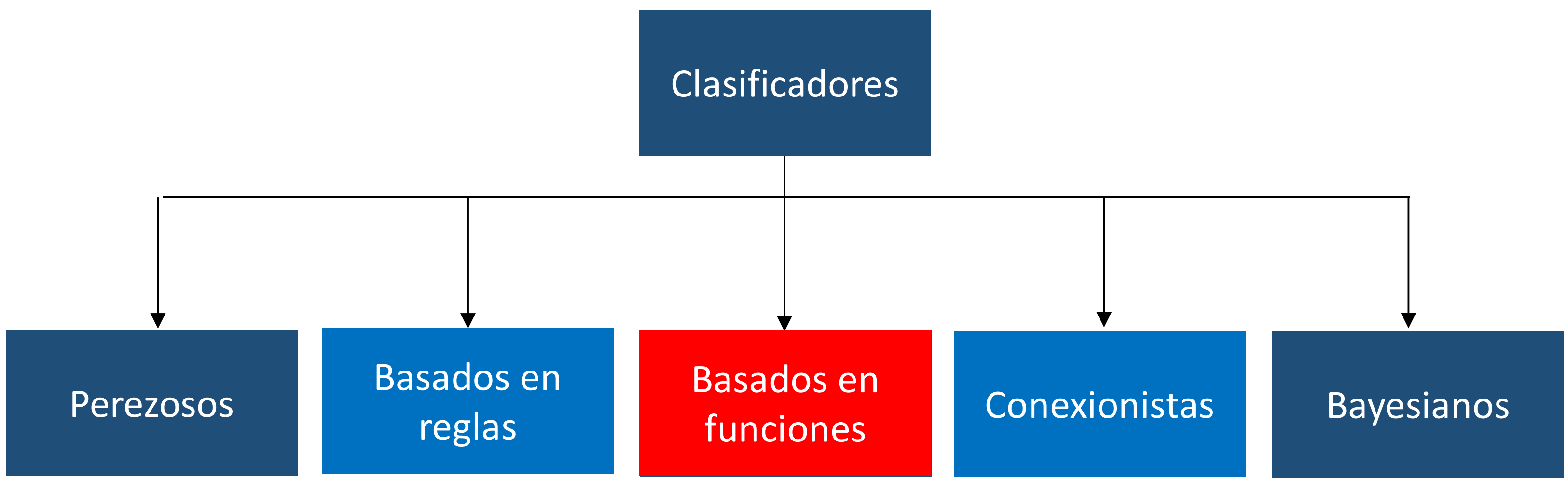
LD: dirección de máxima discriminibilidad.



Proyección en un espacio de menor dimensión.

Clasificación

Tipos



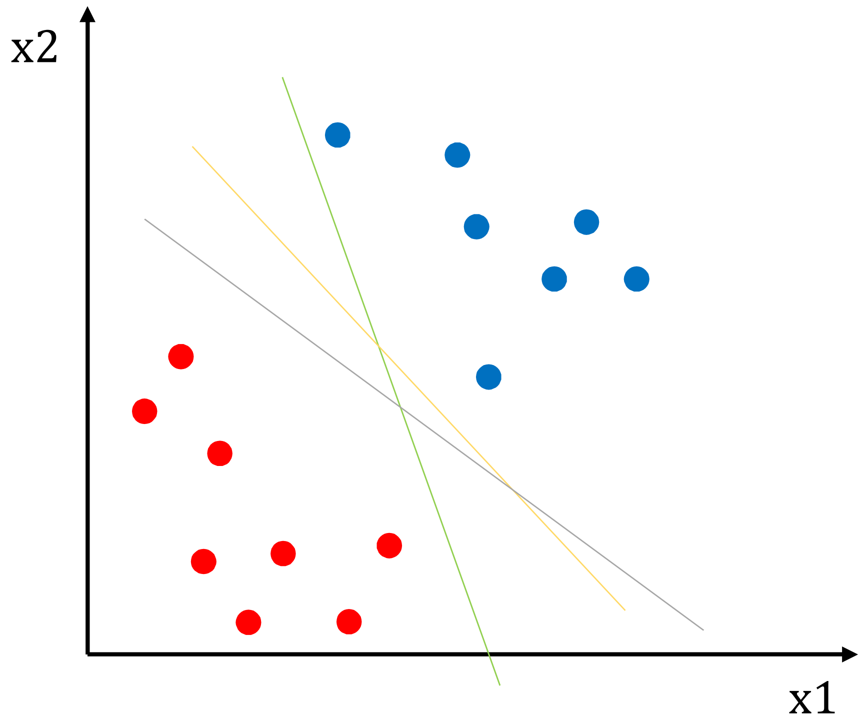
Clasificadores basados en funciones

SVM

Las Máquinas de Soporte Vectorial (SVM, del inglés Support Vector Machines) son clasificadores lineales derivados de la teoría del aprendizaje estadístico. También pueden resolver problemas no lineales mediante la utilización de funciones (*kernels*).

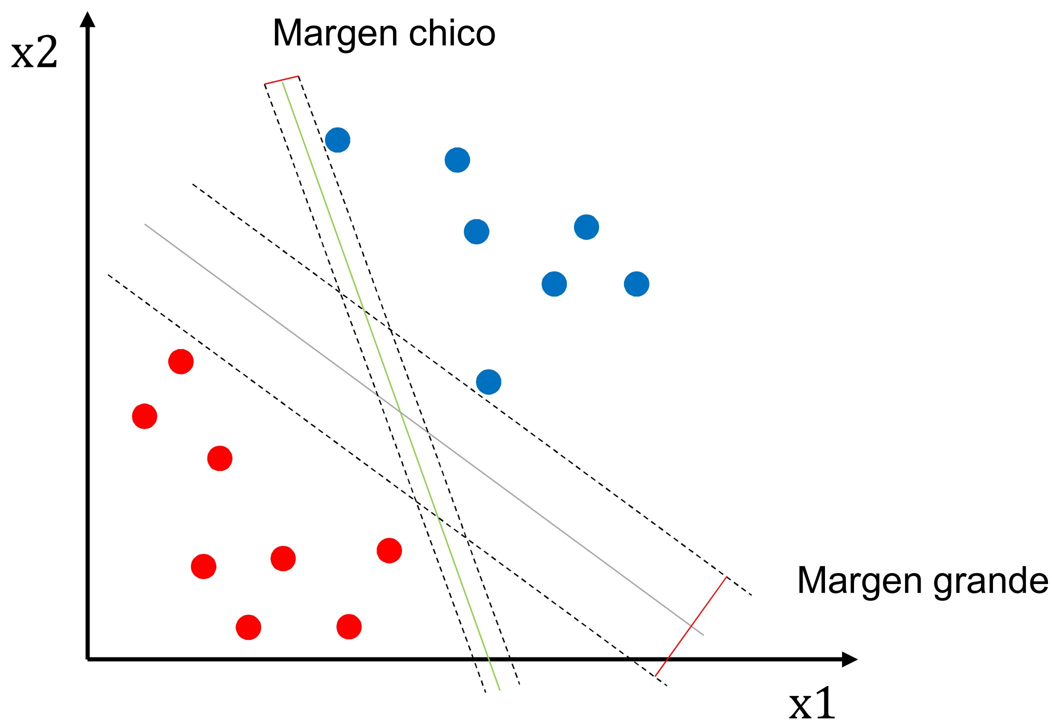
Clasificadores basados en funciones

SVM



Clasificadores basados en funciones

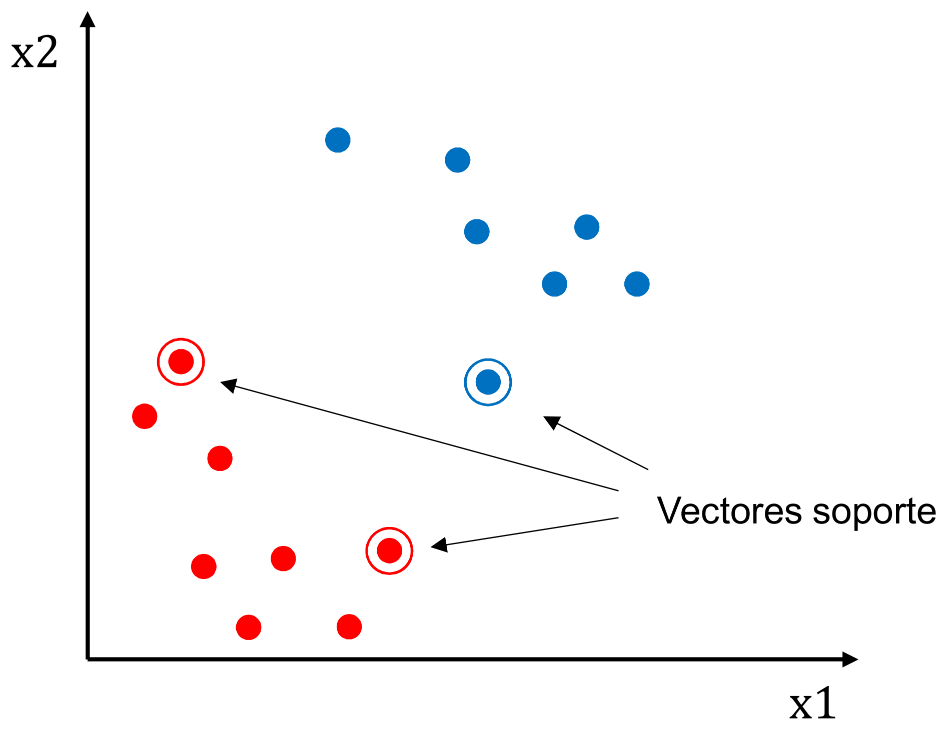
SVM



Criterio para elección del hiperplano → maximizar el margen

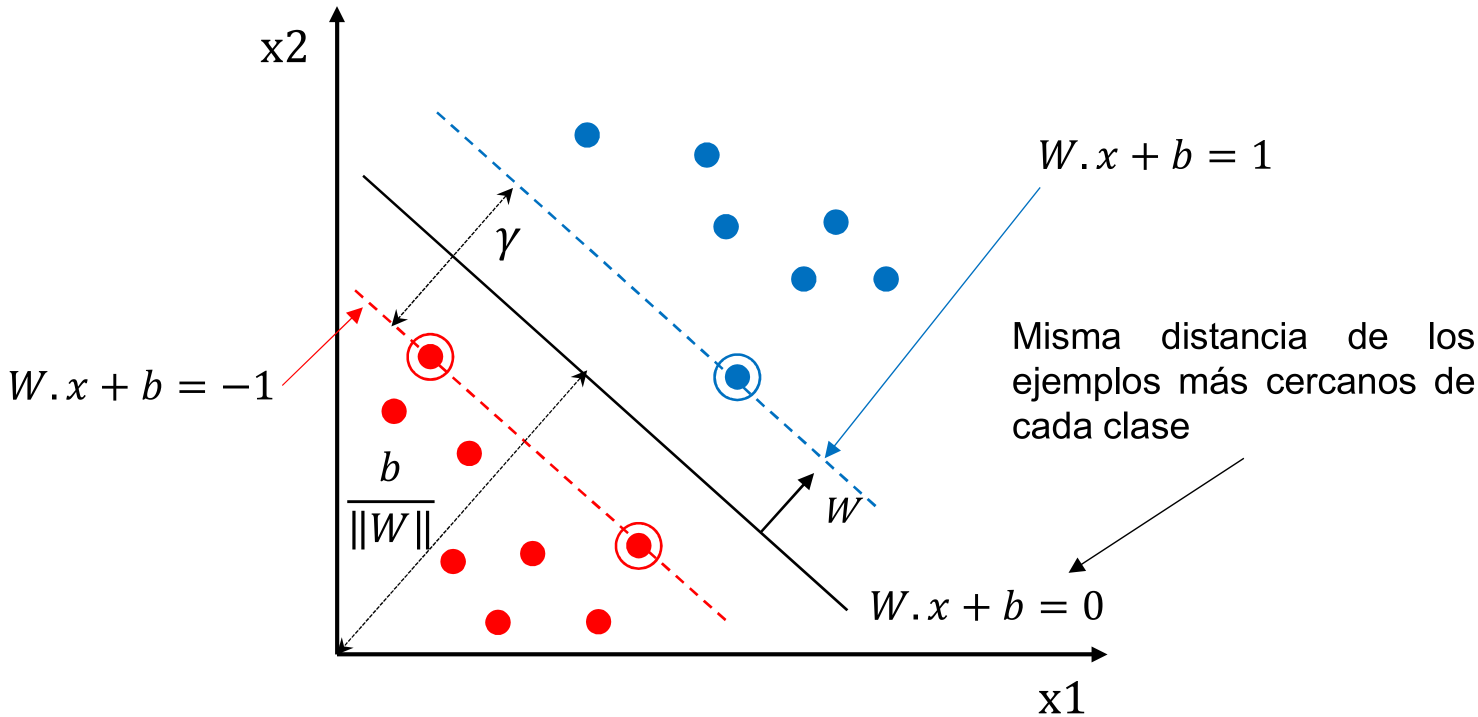
Clasificadores basados en funciones

SVM



Clasificadores basados en funciones

SVM



(W,b) define el hiperplano equidistante a las clases, γ es el margen geométrico.

Clasificadores basados en funciones

SVM

$$\left.\begin{aligned} w_0 + W^T . x_{pos} &= 1 \\ w_0 + W^T . x_{neg} &= -1 \end{aligned}\right\}$$

Ecuaciones de los planos

$$w_0 + W^T . x_{pos} - w_0 + W^T . x_{neg} = 2 \rightarrow W^T . (x_{pos} - x_{neg}) = 2$$

$$\|W\| = \left(\sum_{j=1}^{m} {w_j}^2 \right)^{\frac{1}{2}}$$

Norma del vector de pesos

$$\frac{W^T . (x_{pos} - x_{neg})}{\|W\|} = \frac{2}{\|W\|}$$

Normalización

Clasificadores basados en funciones

SVM

$$\frac{W^T.(x_{pos}-x_{neg})}{\|W\|} = \frac{2}{\|W\|}$$

Distancia entre los hiperplanos (margen).

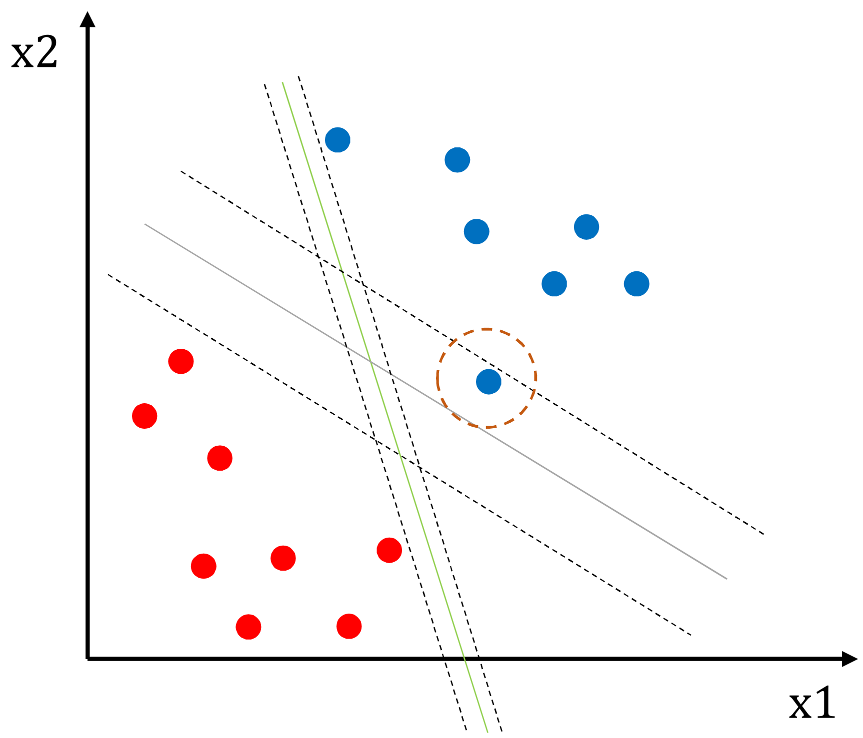
Término a maximizar

Restricciones
$$\begin{cases} w_0 + W^T.x^{(i)} \geq 1 \text{ si } y^{(i)} = 1 \\ w_0 + W^T.x^{(i)} \leq -1 \text{ si } y^{(i)} = -1 \end{cases}$$

Clasificadores basados en funciones

SVM

Margen duro



Clasificadores basados en funciones

SVM

Margen blando

ξ : variable de holgura

$$w_0 + W^T . x^{(i)} \geq 1 - \xi^{(i)} \; si \; y^{(i)} = 1$$

$$w_0 + W^T . x^{(i)} \leq -1 + \xi^{(i)} \; si \; y^{(i)} = -1$$

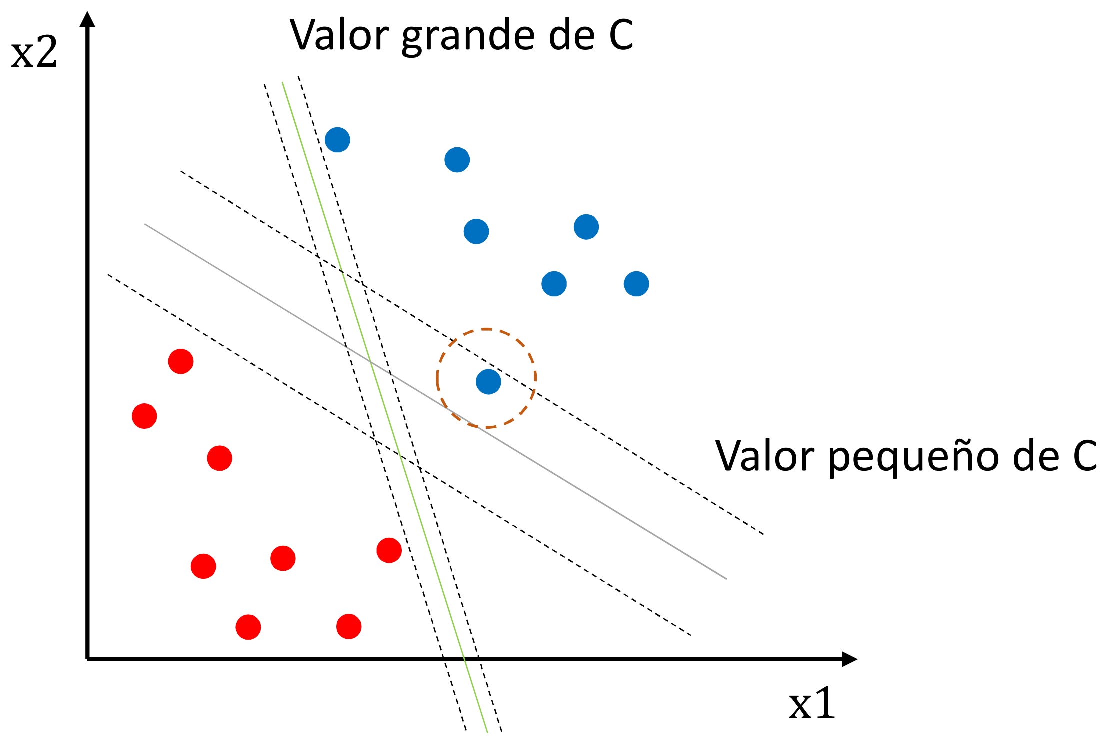
$$\frac{2}{\|W\|} + c . \left(\sum_{i=1}^{N} \xi^{(i)} \right)$$

Nuevo término a maximizar

Clasificadores basados en funciones

SVM

Margen blando



Clasificadores basados en funciones

SVM

Problema no lineal

Clasificador no lineal

Para utilizar una SVM como un clasificador no lineal es necesario utilizar transformación no lineal del espacio de atributos de entrada (*input space*) en un espacio de características (*feature space*) de **dimensionalidad mucho mayor** y donde sí es posible separar linealmente los ejemplos.

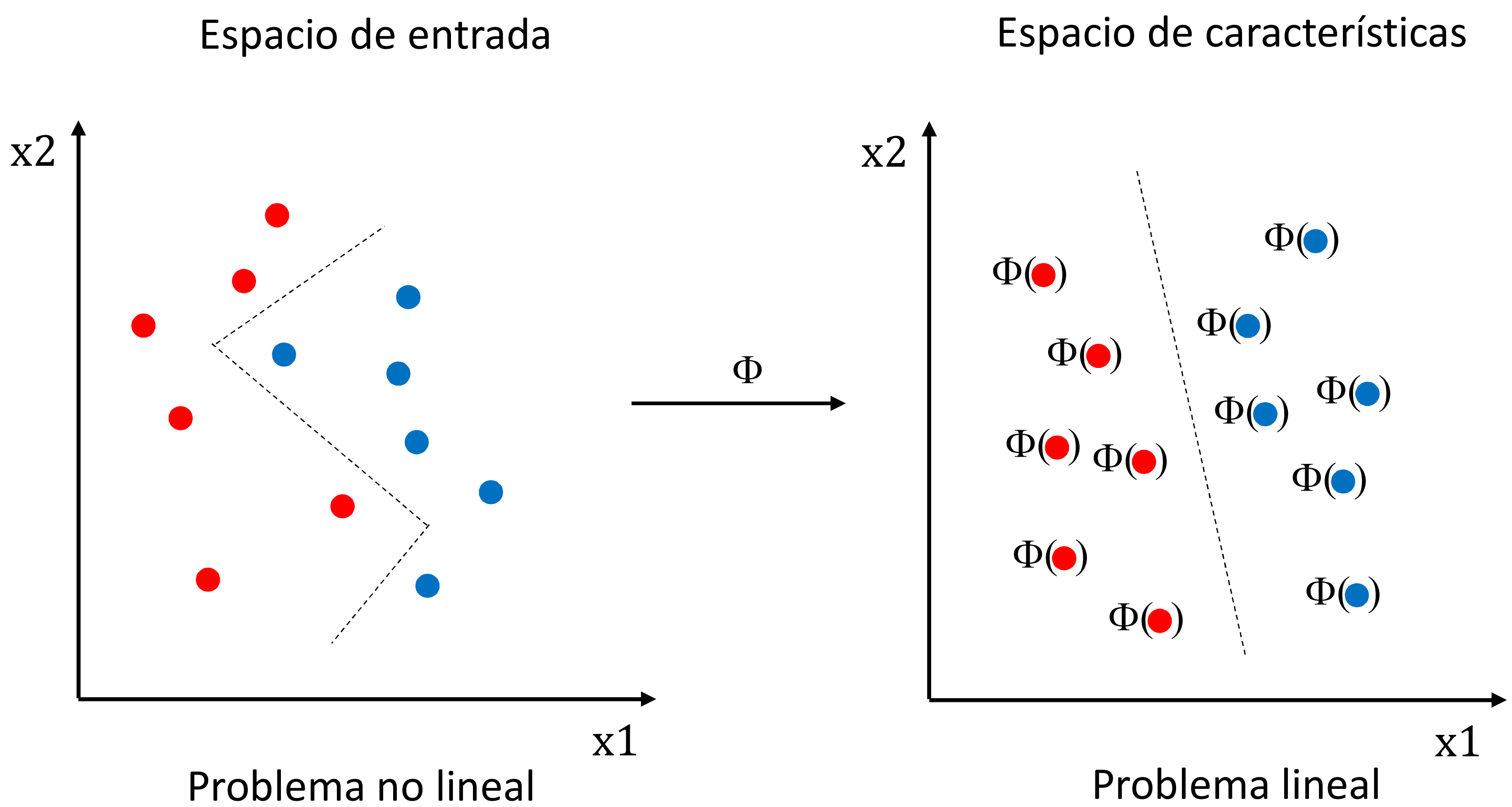
Función núcleo (*kernel*)

Las denominadas funciones núcleo (kernel) calculan el producto escalar de dos vectores en el espacio de características. Esto permite trabajar de manera eficiente en el espacio de características sin necesidad de calcular explícitamente las transformaciones de los ejemplos de aprendizaje

Clasificadores basados en funciones

SVM

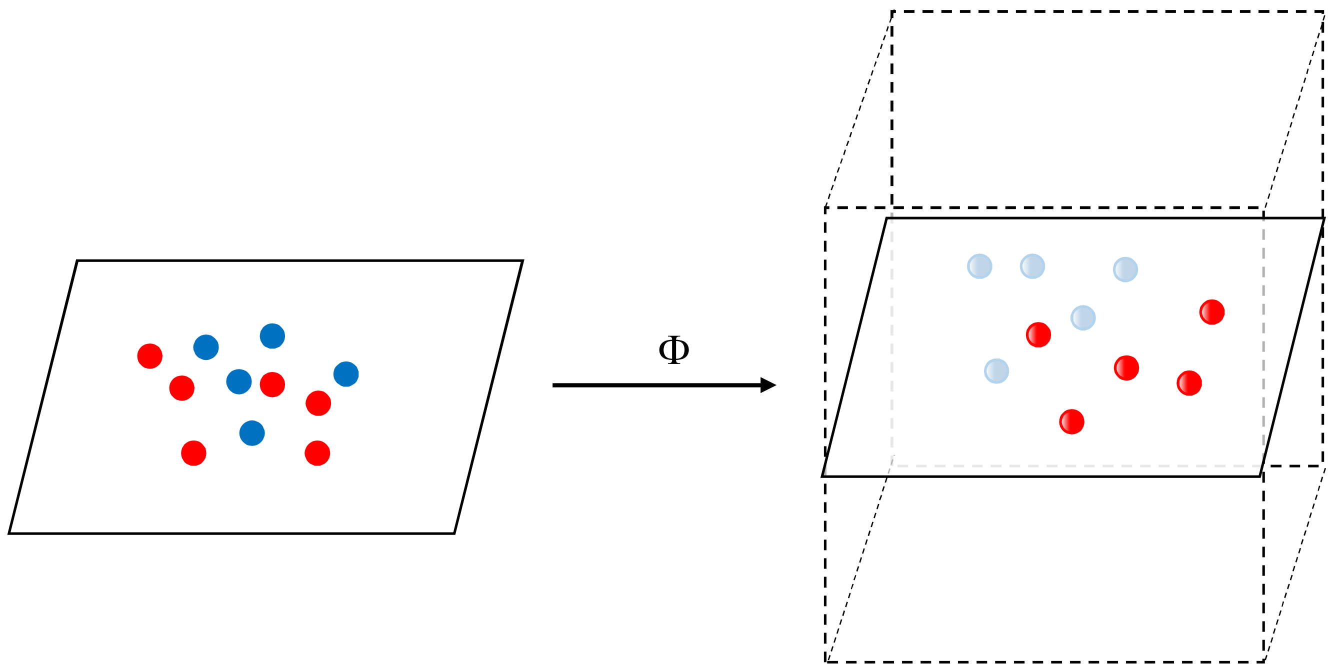
Problema no lineal



Clasificadores basados en funciones

SVM

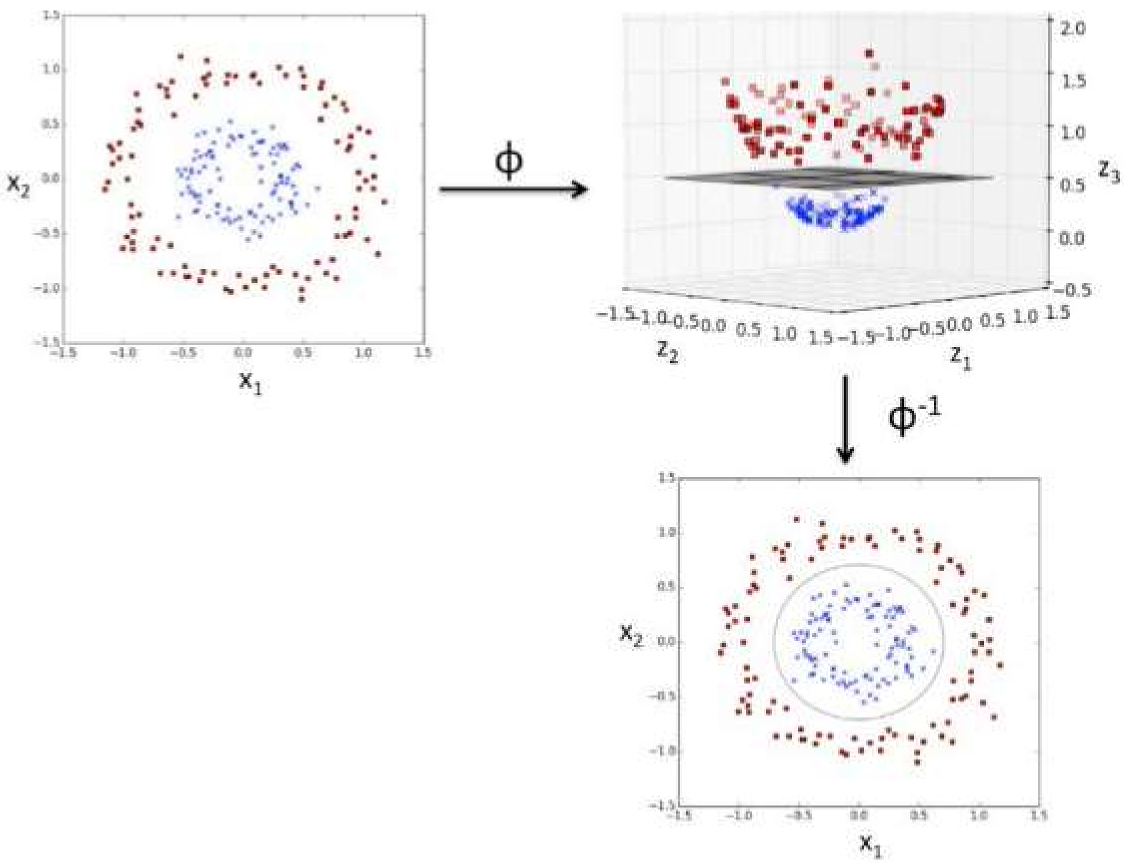
Problema no lineal



Clasificadores basados en funciones

SVM | **Problema no lineal**

$$\phi(x) = \phi(x_1, x_2) = (z_1, z_2, z_3) = (x_1, x_2, {x_1}^2 + {x_2}^2)$$



Algunas funciones núcleos

Polinómica	$(\langle x, y\rangle + c)^d, c \in \Re, d \in \aleph$
Gaussiana	$exp\left(\frac{-\lVert x-y\rVert^2}{\gamma}\right), \gamma > 0$
Sigmoidea	$tanh(s < x, y > r), s, r \in \Re$
Multicuadrática inversa	$\frac{1}{\sqrt{\lVert x-y\rVert^2 + c^2}}, c \geq 0$