Introducción a Machine Learning

Aprendizaje de Máquina – *Support Vector Machines*

MSc. Marco Sobrevilla

Objetivo

- Aprender conceptos sobre *Support Vector Machines*

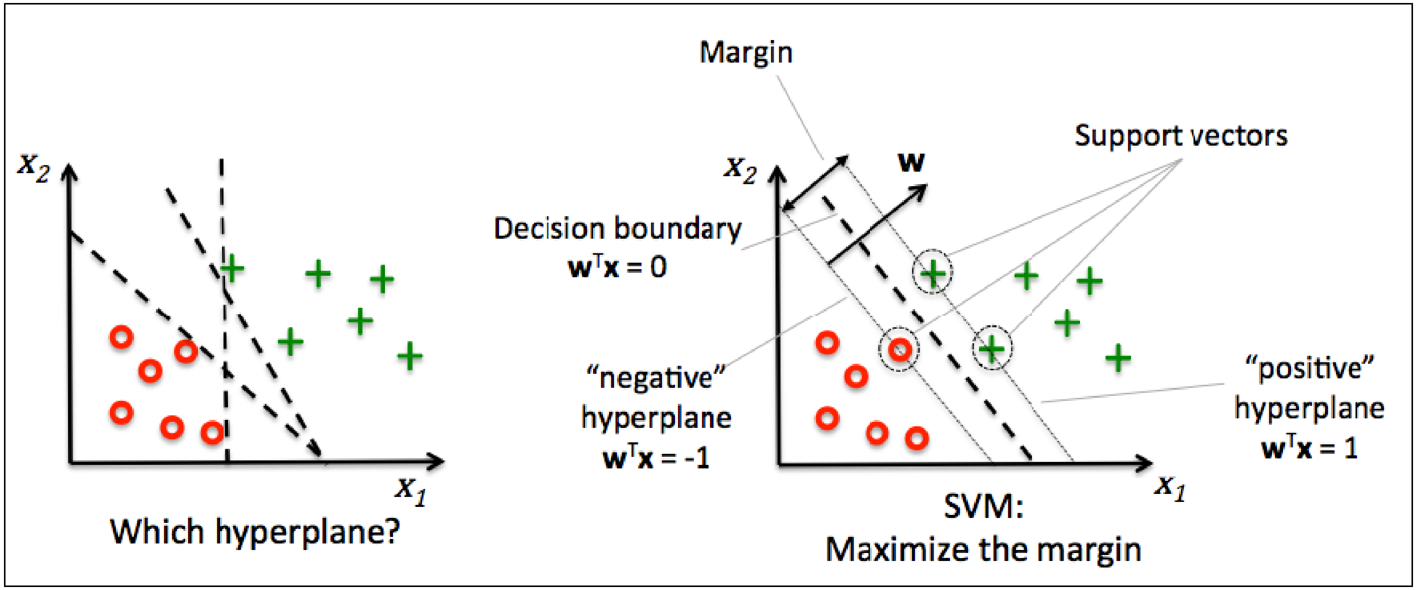
Agenda

- ***Introducción***
- ***Support Vector Machines***
- ***Kernel***

Support Vector Machines

- Propuesto por **Vladimir Vapnik**
- También conocido como: ***Large margin*** *classifier*
- Busca una función de decisión óptima (*optimal* ***hyperplane***)
- En la clasificación lineal no tiene problemas con óptimos locales (identifica el global).
- Aprovecha las funciones **kernel** para problemas de clasificación no lineales.

Support Vector Machines



- La idea es que margen grande posee bajo error de generalización
- Margen cortos son propensos a *overfitting*

Support Vector Machines

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}_{pos} = 1$$

$$w_0 + \boldsymbol{w}^T \boldsymbol{x}_{neg} = -1$$

$$\Rightarrow \boldsymbol{w}^T \left(\boldsymbol{x}_{pos} - \boldsymbol{x}_{neg} \right) = 2$$

- Se busca minimizar:

$$\frac{1}{2} \|\boldsymbol{w}\|^2$$

$$\|\boldsymbol{w}\| = \sqrt{\sum_{j=1}^{m} w_j^2}$$

Maximizar el margen $\dfrac{\boldsymbol{w}^T \left(\boldsymbol{x}_{pos} - \boldsymbol{x}_{neg} \right)}{\|\boldsymbol{w}\|} = \dfrac{2}{\|\boldsymbol{w}\|}$

Clases Linealmente no Separables

- SVM funciona bien para clases linealmente separables
- Qué pasa en otros casos?
  - *Soft-margin*
  - *C: Penalidad de clasificación errada*

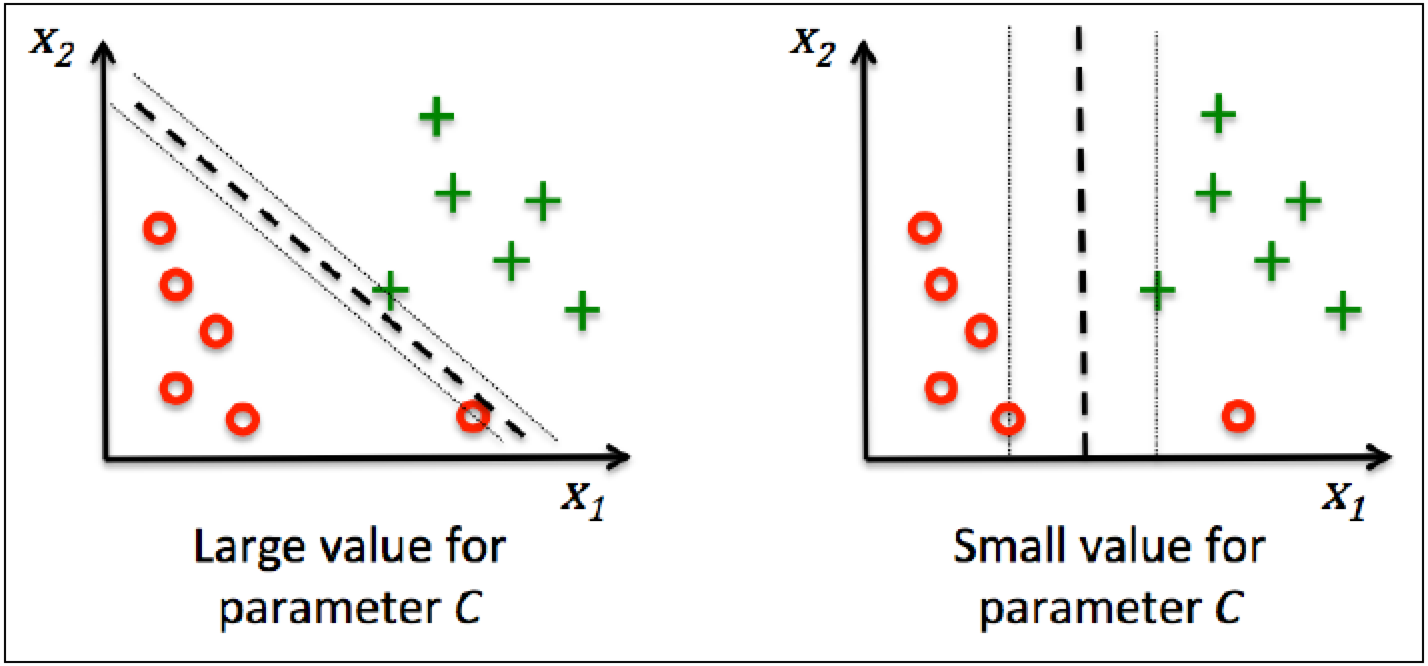
$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} \geq 1 \; if \;\; y^{(i)} = 1 - \xi^{(i)}$$

$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} < -1 \; if \;\; y^{(i)} = 1 + \xi^{(i)}$$

Minimizar:

$$\frac{1}{2}\left\|\boldsymbol{w}\right\|^2 + C\left(\sum_i \xi^{(i)}\right)$$

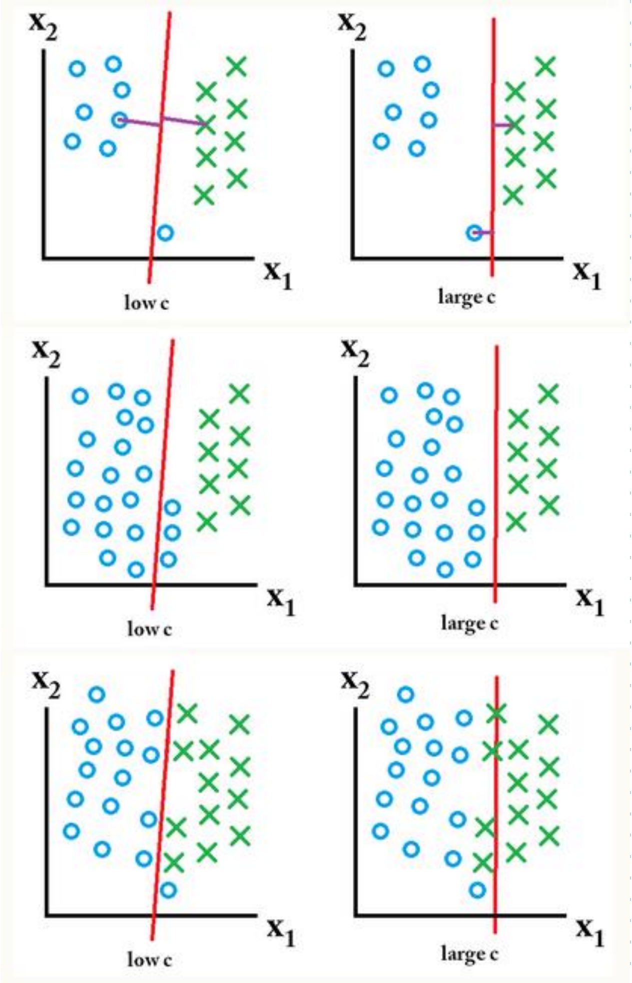
Clases Linealmente no Separables



Disminuye bias
Aumenta la varianza

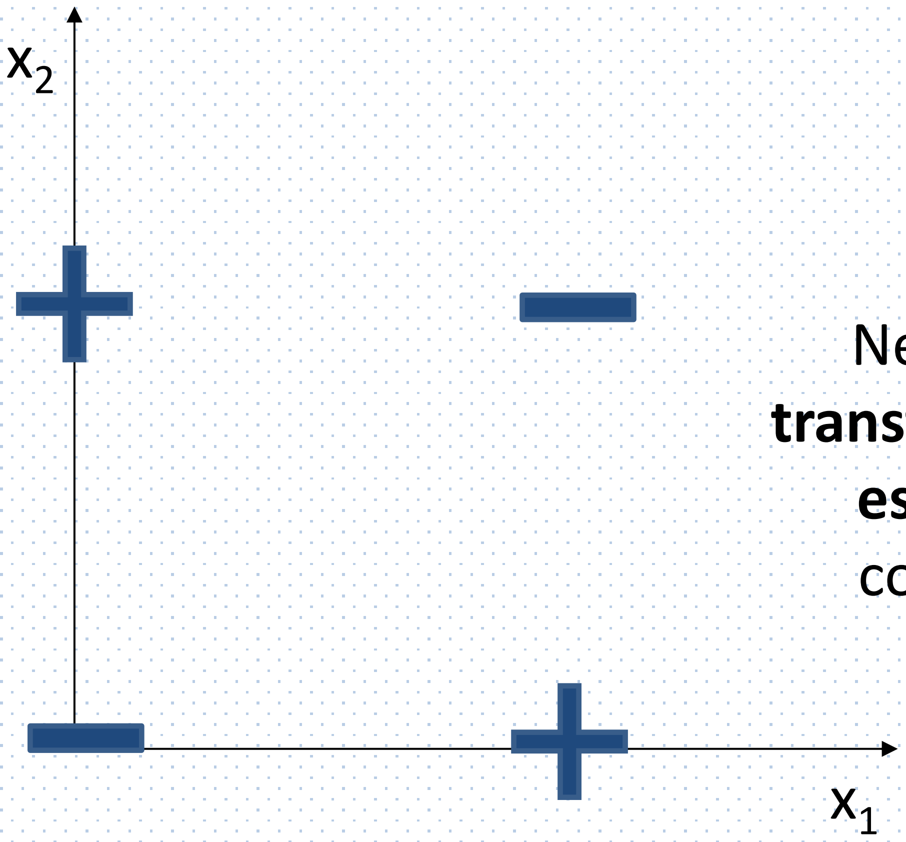
Aumenta bias
Disminuye varianza

Parámetro C



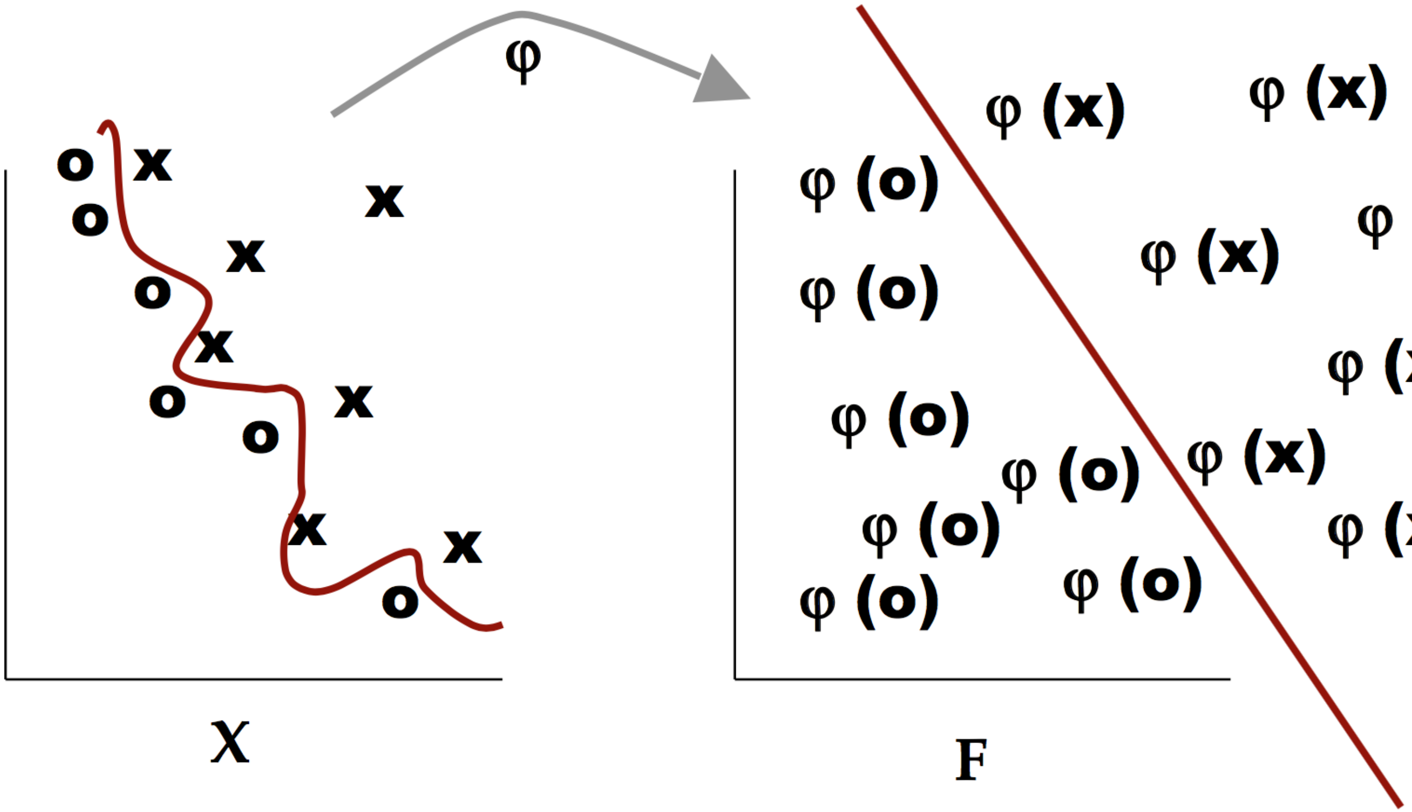
x2
x1
low c
x2
x1
large c
x2
x1
low c
x2
x1
large c
x2
x1
low c
x2
x1
large c

Problemas No Lineales



Necesitamos **transformarlo** a un **espacio** más conveniente

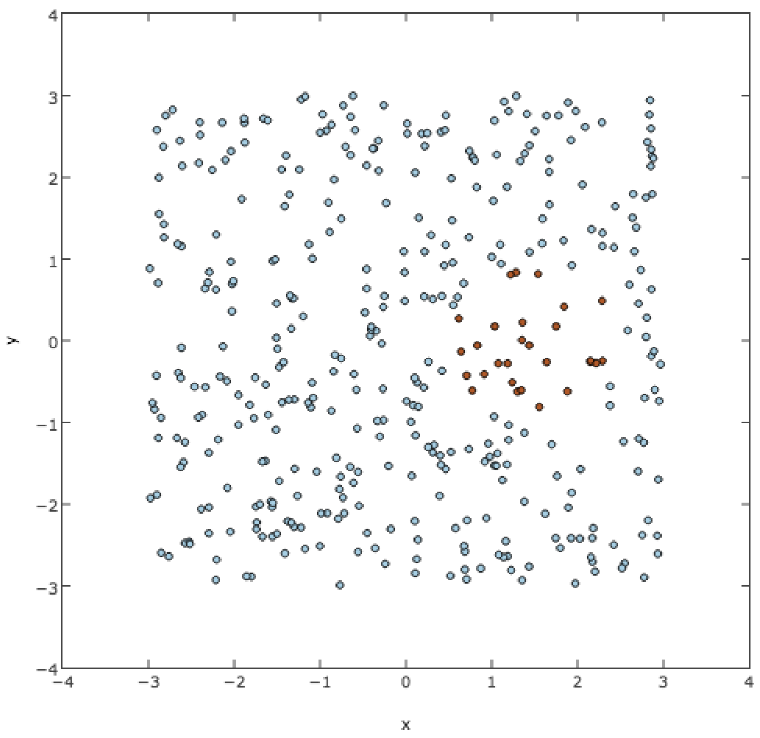
Uso de Kernels

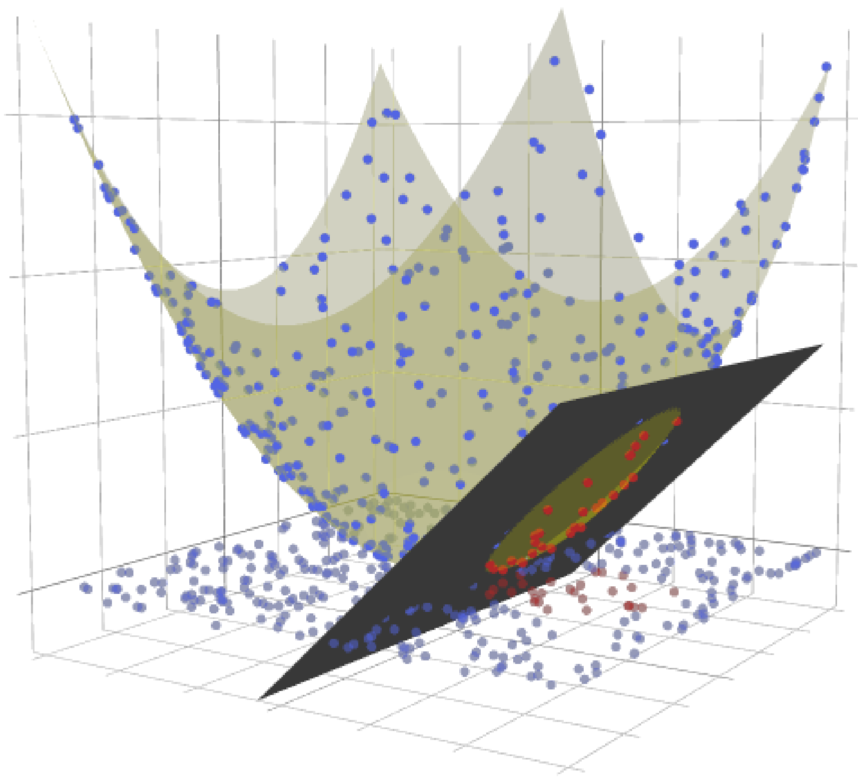


Funciones Conocidas

- linear: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.
- polynomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i{}^T \mathbf{x}_j + r)^d$, $\gamma > 0$.
- radial basis function (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $\gamma > 0$.
- sigmoid: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i{}^T \mathbf{x}_j + r)$.

Kernel SVM





Fin ☺