Machine Learning

(https://leonpalafox.github.io/mlclase/)

Leon F. Palafox PhD

Noticias

The Arrival of Artificially Intelligent Beer

WRITTEN BY ALASDAIR ALLAN

29 August 2016 // 02:37 PM CET

Recap – Regresion Lineal

- Imagine you want to sell you car:
  - How much do you ask for it:
    - Mileage
    - Year
    - Color
    - Options
    - Condition

Valores de autos?

Model	Year	Brand	Price	Options	Condition	Mileage
Corvette	1961	Chevrolette	100K	Standard	As New	100,000
Corvette	1961	Chevrolette	10K	Standard	Rust	100,000
Corvette	1961	Chevrolette	120K	Standard	Used	20,000

Price = Year + Options + Condition + Mileage

This doesn't work

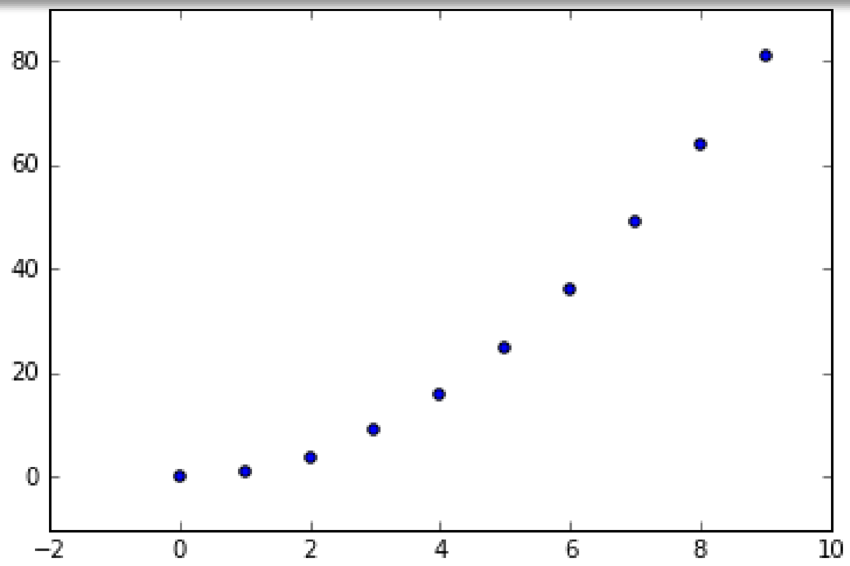
Price = A*Year+ B*Options + C*Condition+ D*Mileage

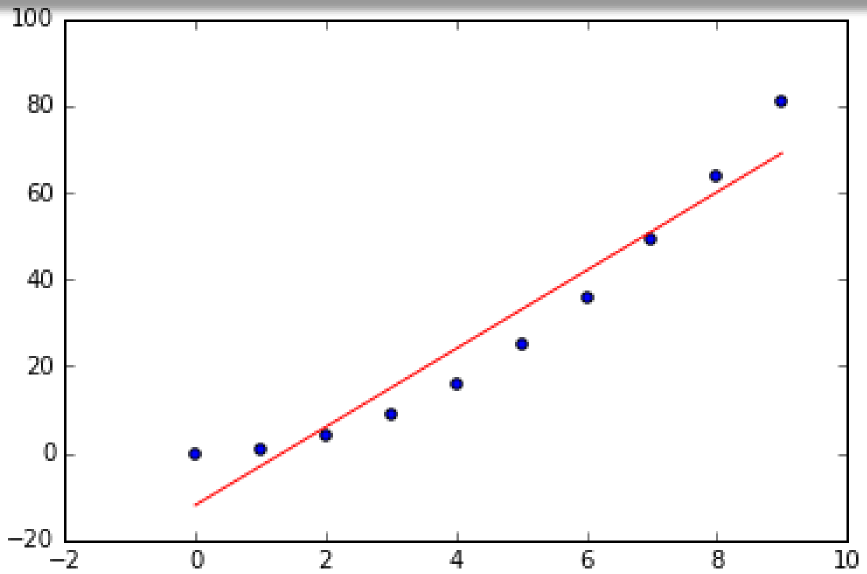
Forma Matricial

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix},$$

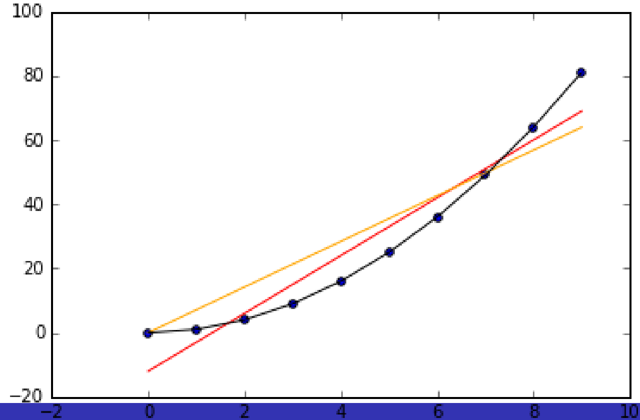
$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \mathbf{y} = \mathbf{X}\beta$$

Ingenieria de Features





By adding an extra squared factor, we got a perfect fit



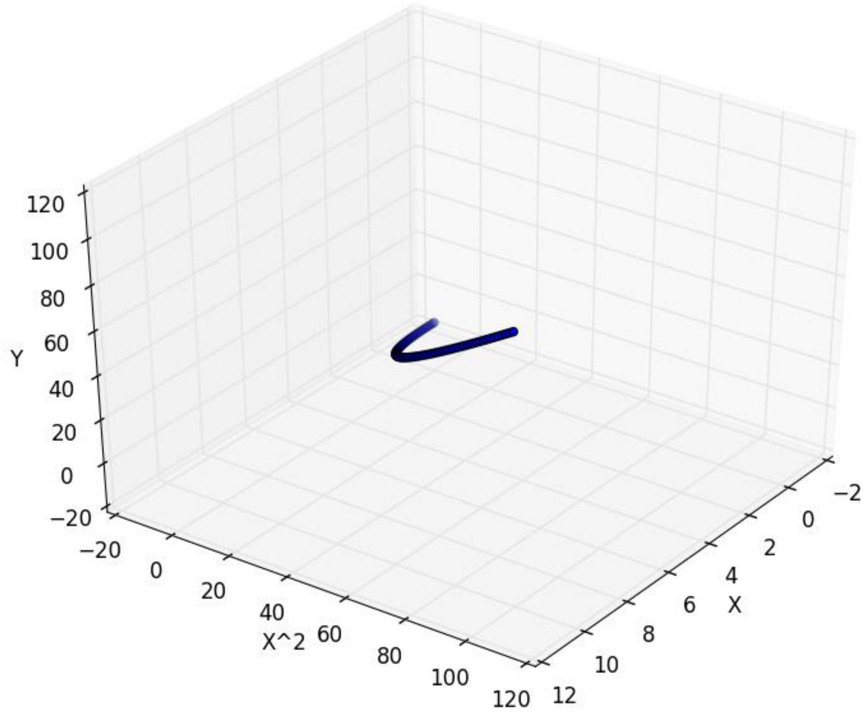
Ingeniería de Features

- https://eng.uber.com/cota/

- https://medium.com/netflix-techblog/distributed-time-travel-for-feature-generation-389cccdd3907

Expansión de espacios

- Al poner mas features, nos movemos a dimensiones superiors.



Problema de Optimización

- Optimize:

$$\mathbf{y} = \mathbf{X}\beta$$

- Score:

$$X\beta - y$$

$$(X\beta - y)^2$$

$$\boxed{\frac{1}{2}(X\beta - y)^2}$$

This is considered a cost!

Usando algebra matricial

$$\frac{1}{2}(X\beta - y)^2 = \frac{1}{2}(X\beta - y)^T(X\beta - y)$$

$$= J(\beta)$$

$$\nabla_\beta J(\beta) = 0$$

Usando algebra matricial

$$\nabla_\beta J(\beta) = X^T X \beta - X^T y = 0$$

$$X^T X \beta = X^T y$$

$$\beta = (X^T X)^{-1} X^T y$$

Gradient descent

$$h_\beta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$J(\beta) = \frac{1}{2}\sum_{i=1}^{m}(h_\beta(x^{(i)}) - y^{(i)})^2$$

$$\boxed{\beta_{j+1} = \beta_j - \alpha \frac{\delta}{\delta \beta_j} J(\beta)}$$

Gradient descent

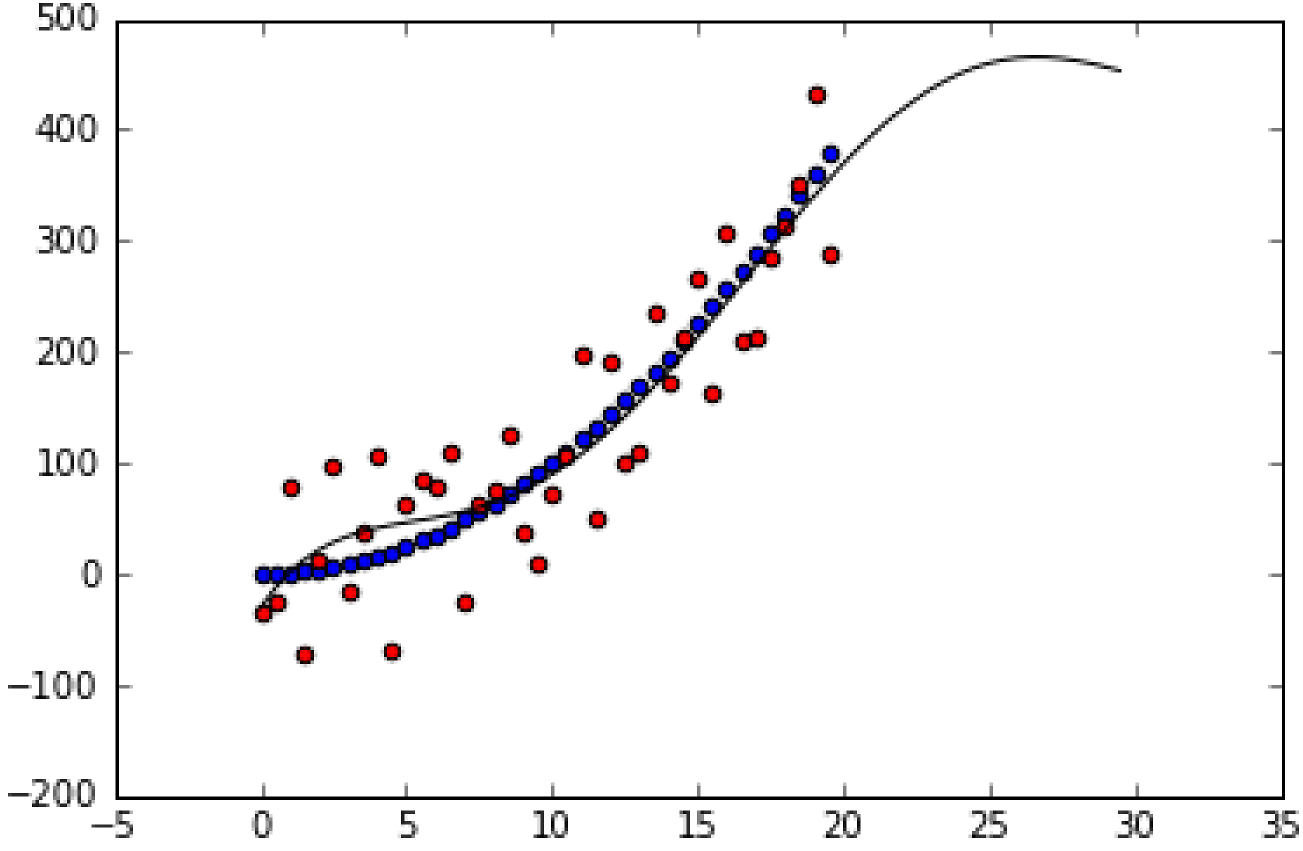
$$\beta_{j+1} = \beta_j - \alpha(y - h_\beta(x^{(i)}))x_j^{(i)}$$

Cual es la diferencia

- Grandes cantidades de datos:
  - Gradiente descente es el rey!
  - Algebra Matricial es terrible (por que?).
- Pequeños datos
  - Algebra Matricial es lo mejor.

- En práctica se usa gradient descendente por que las bases de datos se hacen grandes de manera muy rapida.

Como controlamos el problema de ayer



Regularizacion

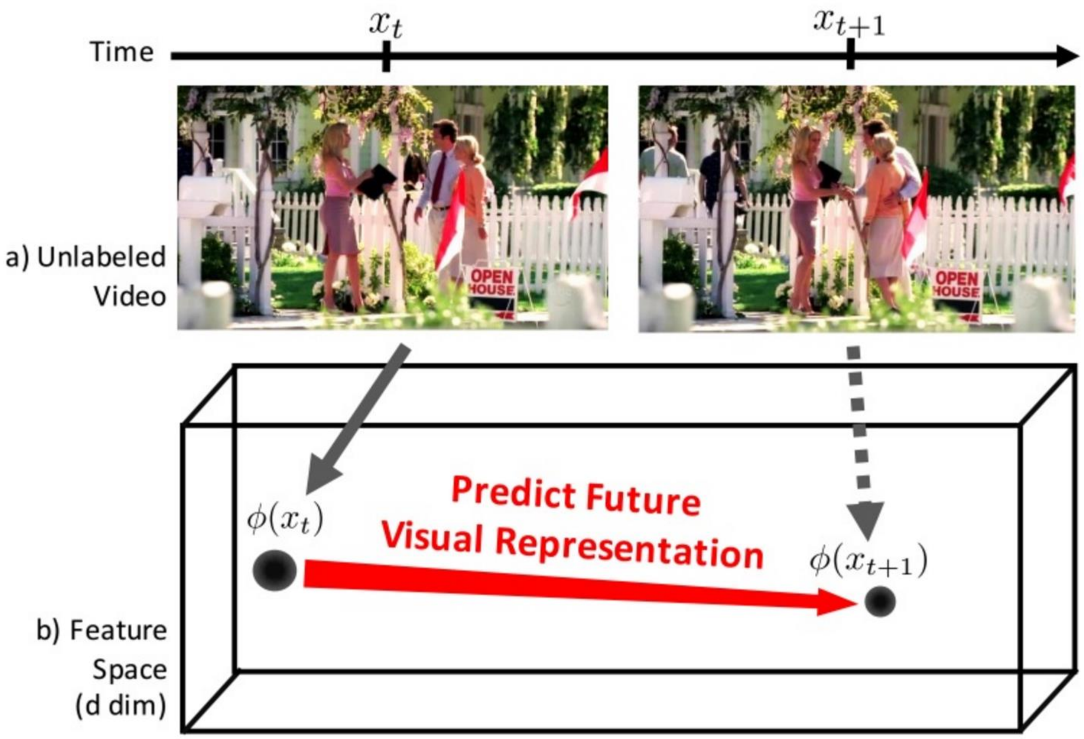
- Trata de mantener los valores de los pesos bajos (por que?).

$$J(\beta) = \frac{1}{2}(y - h_\beta(x^{(i)}))^2 + \lambda||\beta||$$

$$J(\beta) = \frac{1}{2}(y - h_\beta(x^{(i)}))^2 + \alpha||\beta||$$

- Esto se denomina Ridge Regression

Un ejemplo de uso

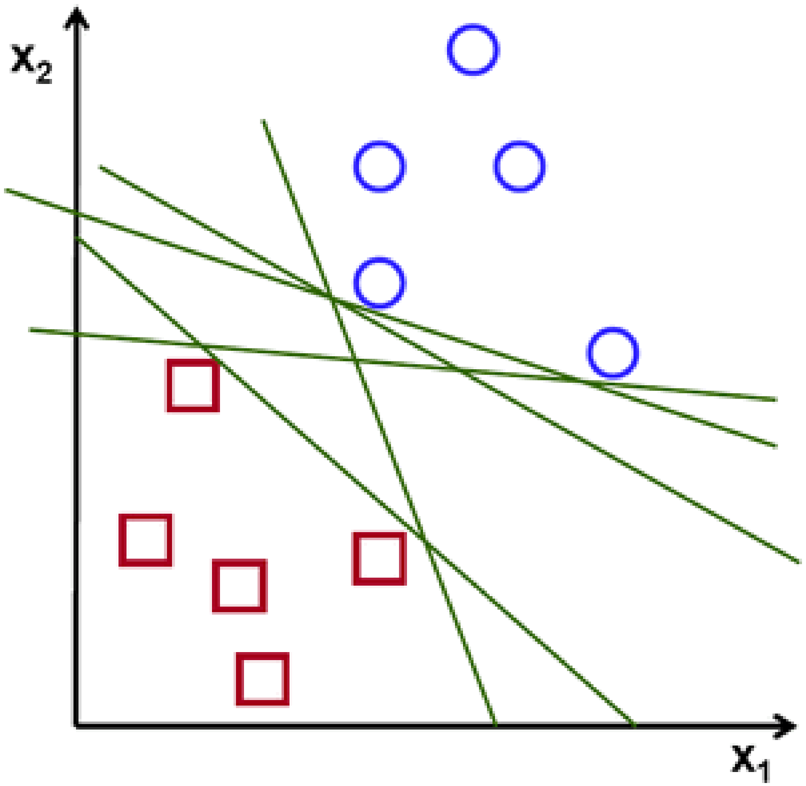


https://www.youtube.com/watch?v=AR3hY9iB5-I

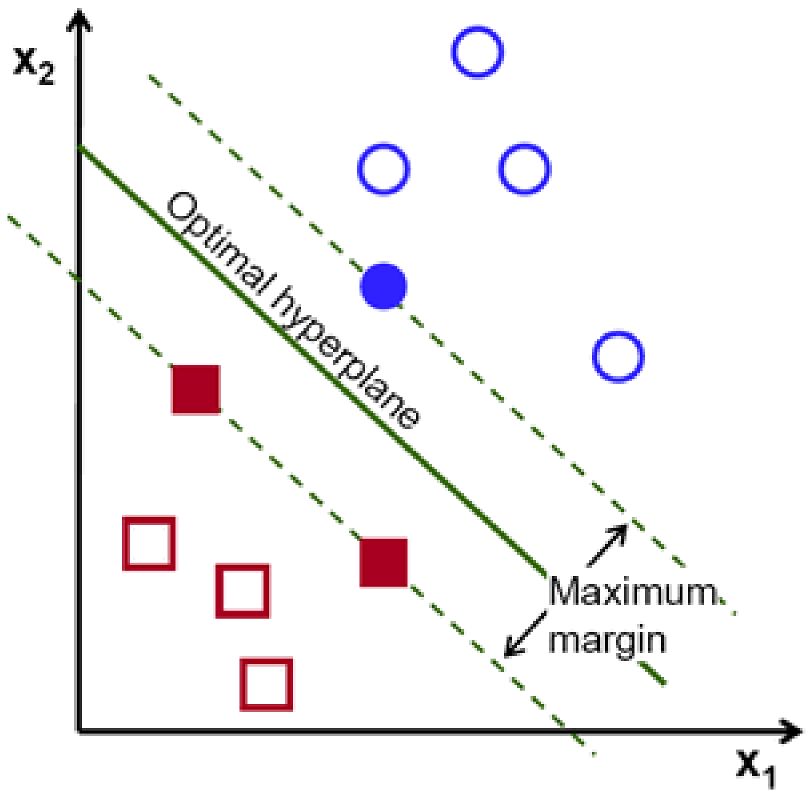
Nuestro super martillo– Support Vector Machines

- SVMs son la mejor herramienta para hacer clasificacion que necesita pocas variaciones.
  - Es razonablemente facil de usar
  - Tiene pocos parametros para optimizar
  - Funciona increiblemente bien.
  - Es rapido

Support Vector Machines (SVMs)



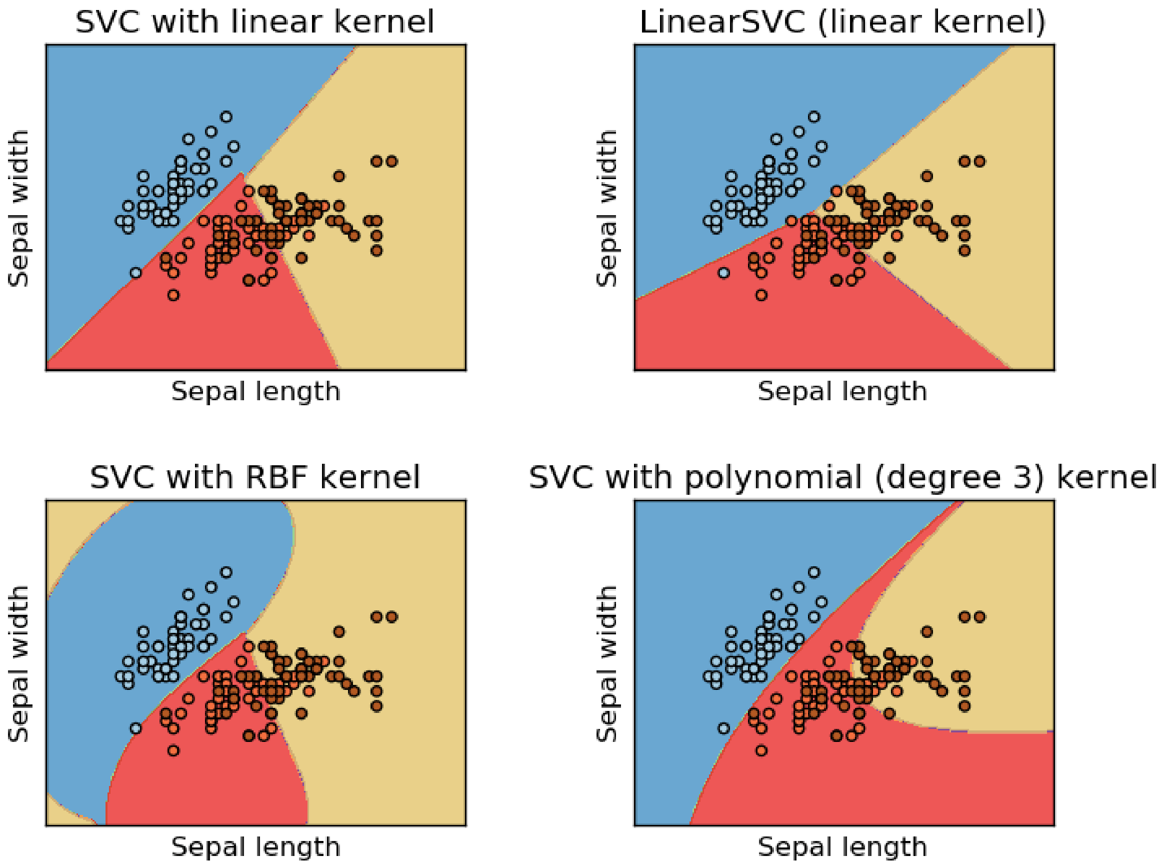
SVM Introduction



Kernels

- Utilizar lineas es impractico por que no necesariamente es la major solución.
- Usamos el concepto del Kernel (Distancia)
  - Diferentes Kernels nos permiten usar diferentes espacios
  - Los mas communes:
    - Linear
    - Polynomial (expanded powers) (power of polynomial)
    - RBF (Gaussian Kernel) (“variance”)

Kernels



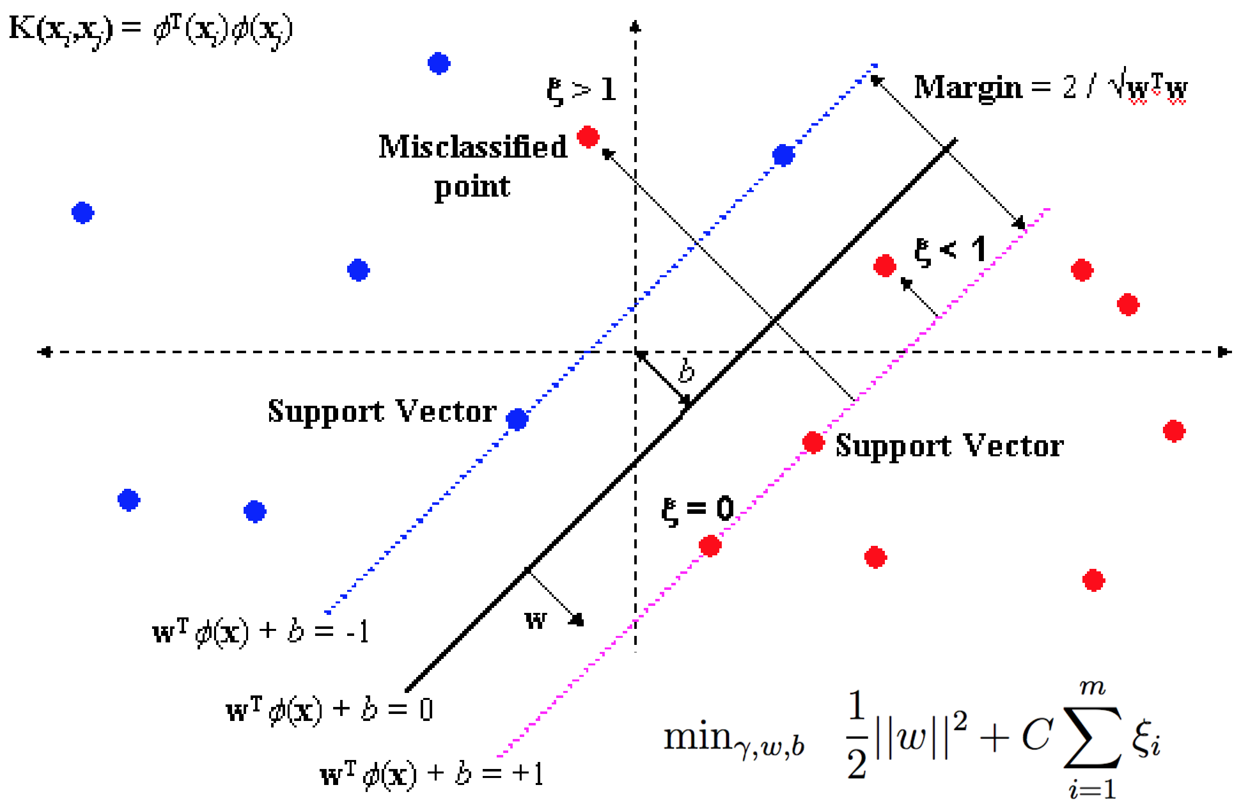
http://scikit-learn.org/stable/modules/svm.html#svm-kernels

Support Vector Machines (SVMs)

- Froman parte de la familia de los métodos de Kernels

- Estos métodos son muy versatiles
  - Mucha investigación se hace alrededor de cual es el major Kernel.
  - Es una lata de gusanos
  - La panacea es tener Kernels que se autodefinan

Parameter C



$$\min_{\gamma,w,b} \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1,\ldots,m$$

$$\xi_i \geq 0, \quad i = 1,\ldots,m.$$