Máquinas de Vetores de Suporte (SVM)

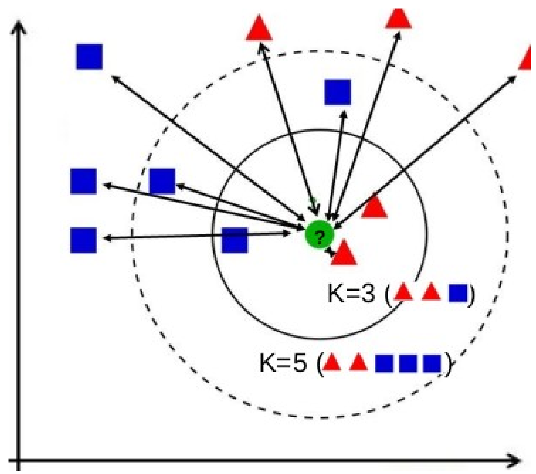
Prof. André Gustavo Hochuli

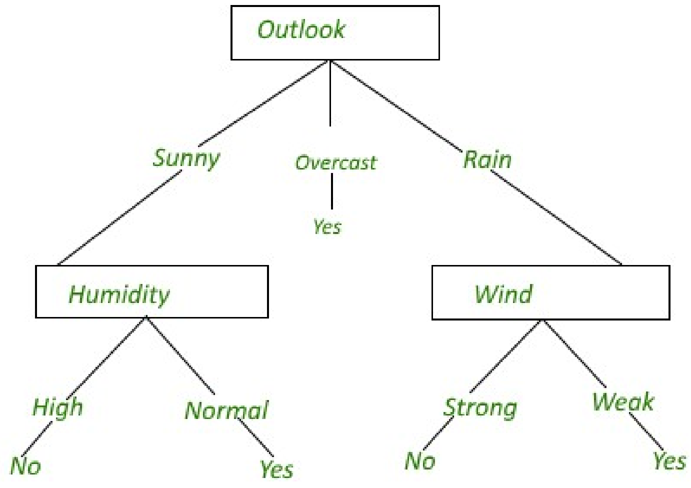
gustavo.hochuli@pucpr.br
aghochuli@ppgia.pucpr.br
github.com/andrehochuli/teaching

Plano de Aula

- Discussões Iniciais
- Classificação Binária vs Multi-classe
- SVM
- Exercícios

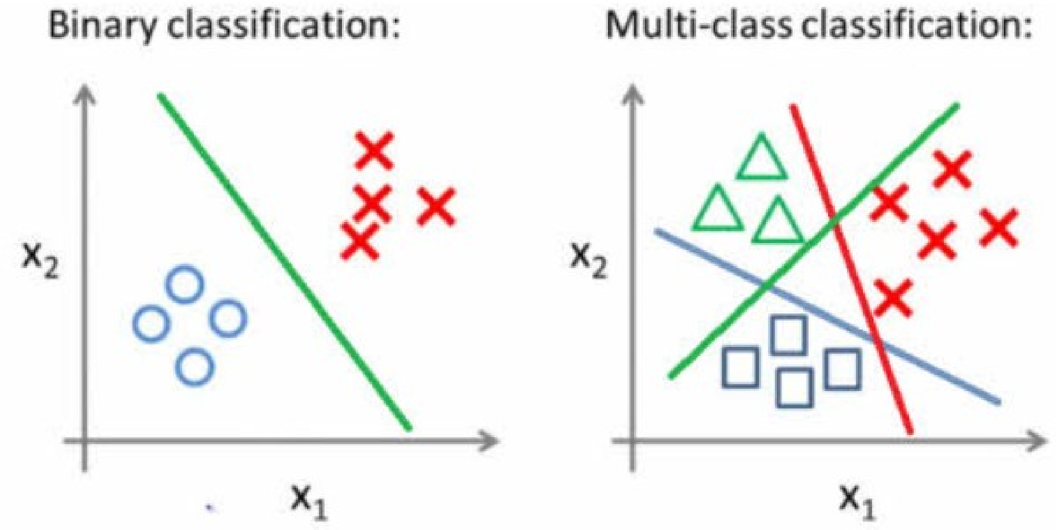
Discussões Iniciais





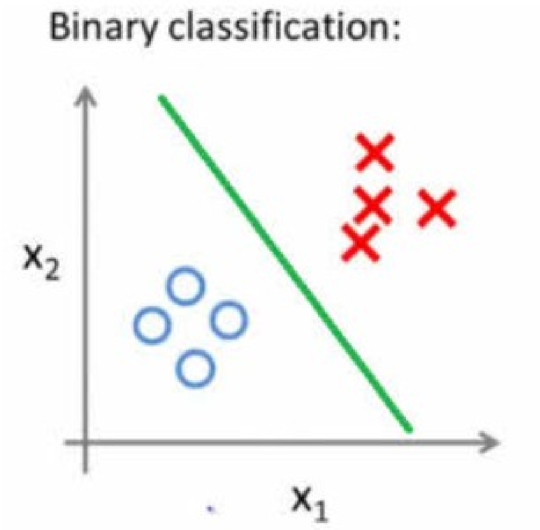
Classificação Binária vs Multi-Classes

- Os modelos vistos até agora, trabalham implicitamente com problemas multi-classes exclusivamente pela natureza de seus algoritmos
  - Vizinhança
  - Probabilísticos



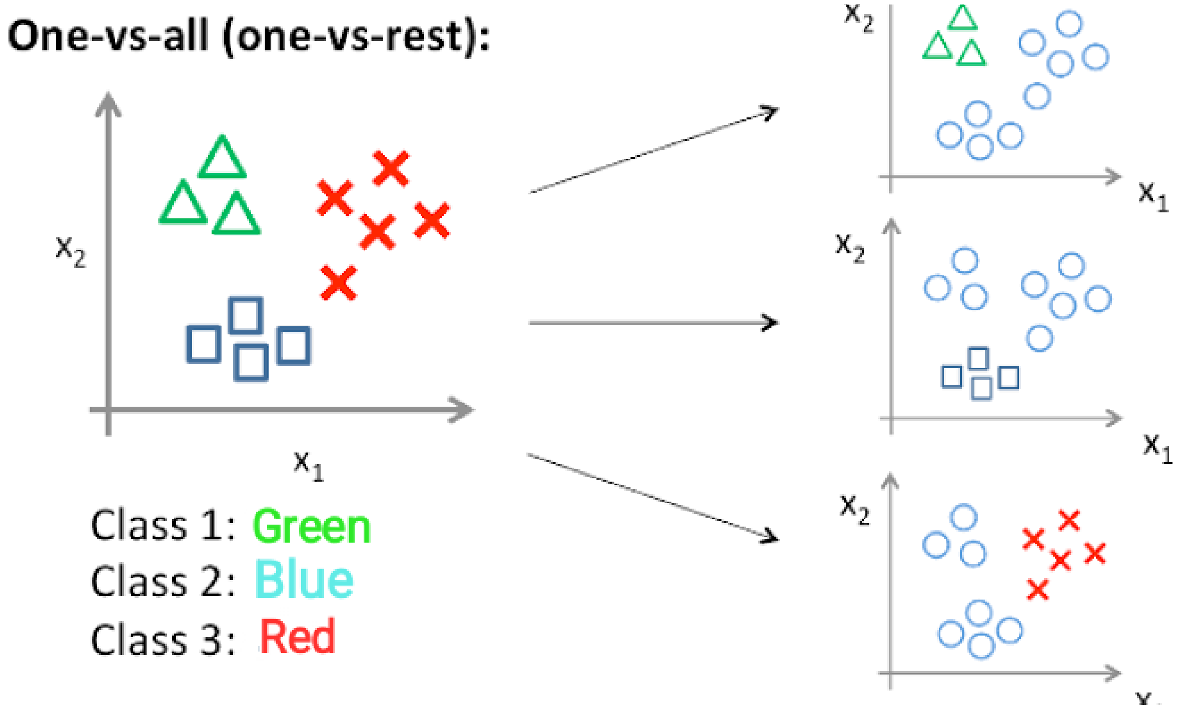
Classificação Binária vs Multi-Classes

- Mas e quando o modelo é naturalmente binário?
  - SVM
  - RNA



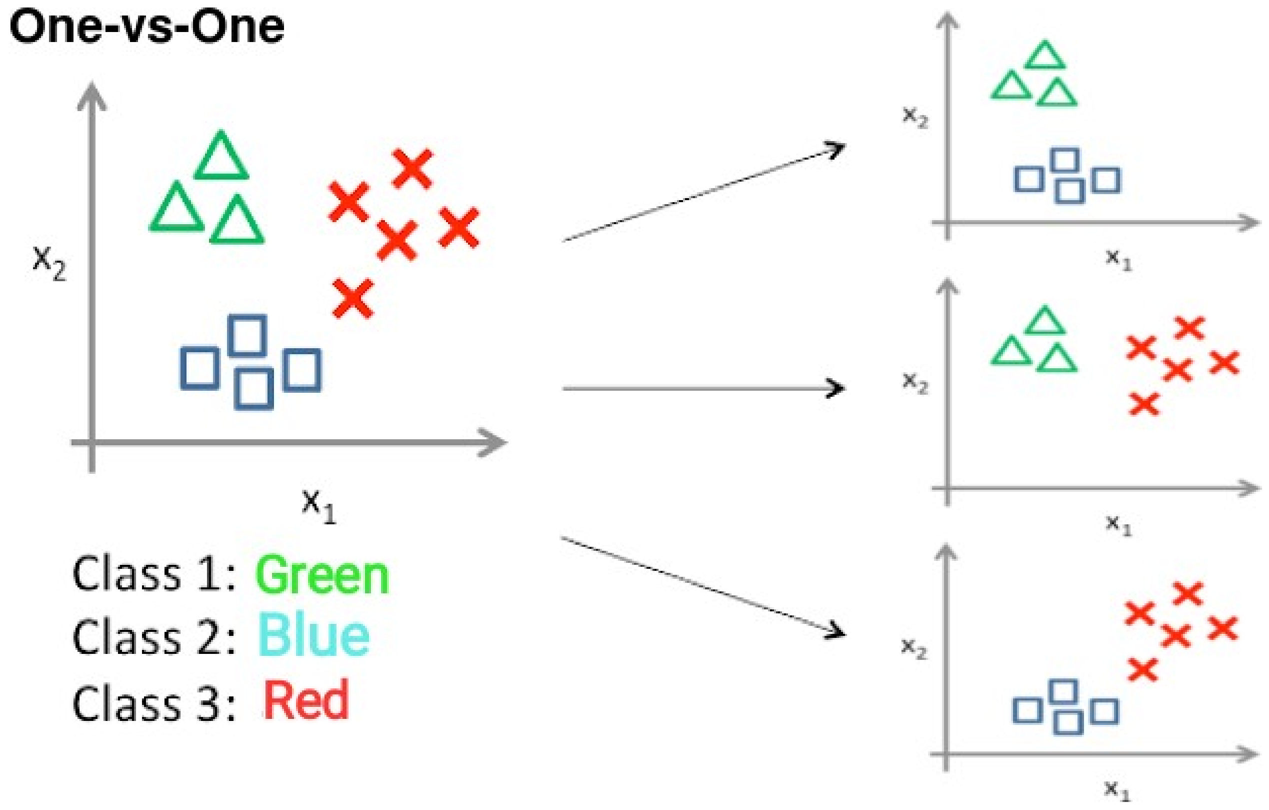
Classificação Binária vs Multi-Classes

- One-vs-All (OVA) ou One-vs-Rest (OVR)
  - Modelo 1:- [Green] vs [Red, Blue]
  - Modelo 2:- [Blue] vs [Green, Red]
  - Modelo 3:- [Red] vs [Blue, Green]
- Predict: Max(Modelo 1, Modelo 2, Modelo 3)



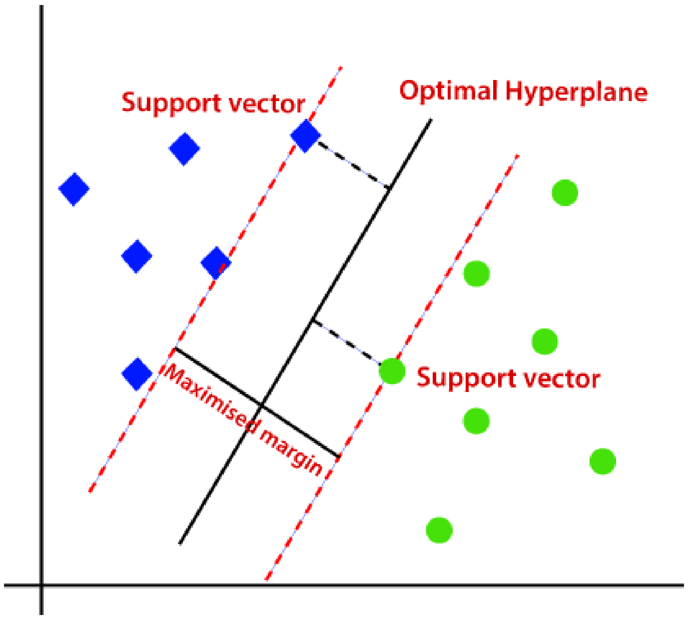
Classificação Binária vs Multi-Classes

- One-vs-One
  - Número de Modelos: N* (N-1)/2



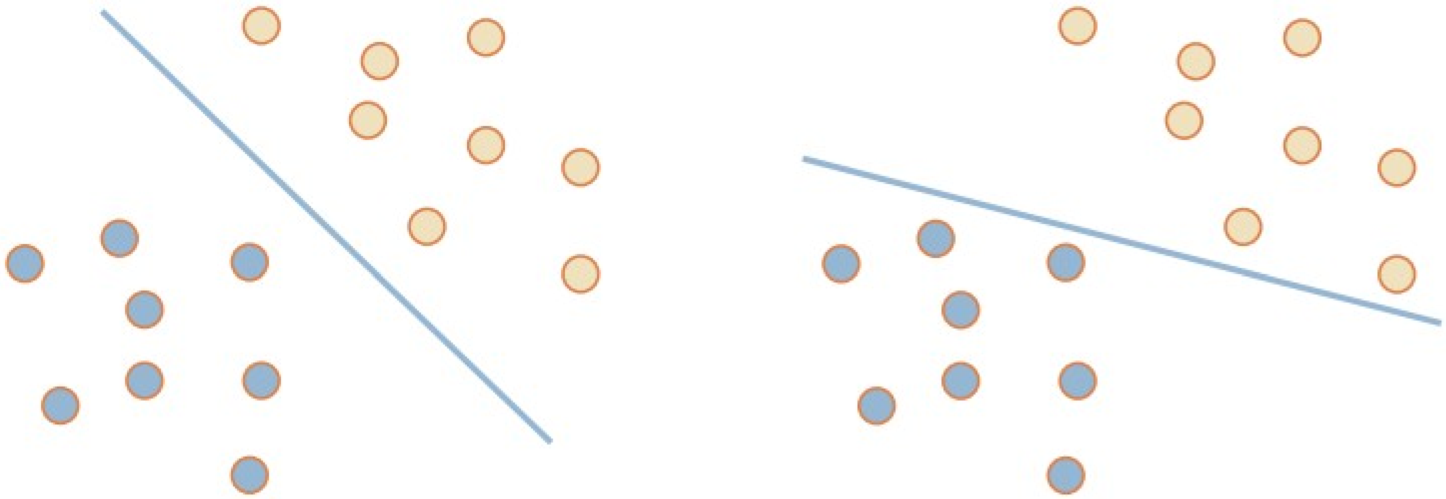
SVM Linear

- Vladimir Vapnik (1979)
- Binário – Não Probabilístico
- Define um hiperplano de separação das classes
  - +1 e -1
- Parâmetros: C (Regularização) e Kernel.



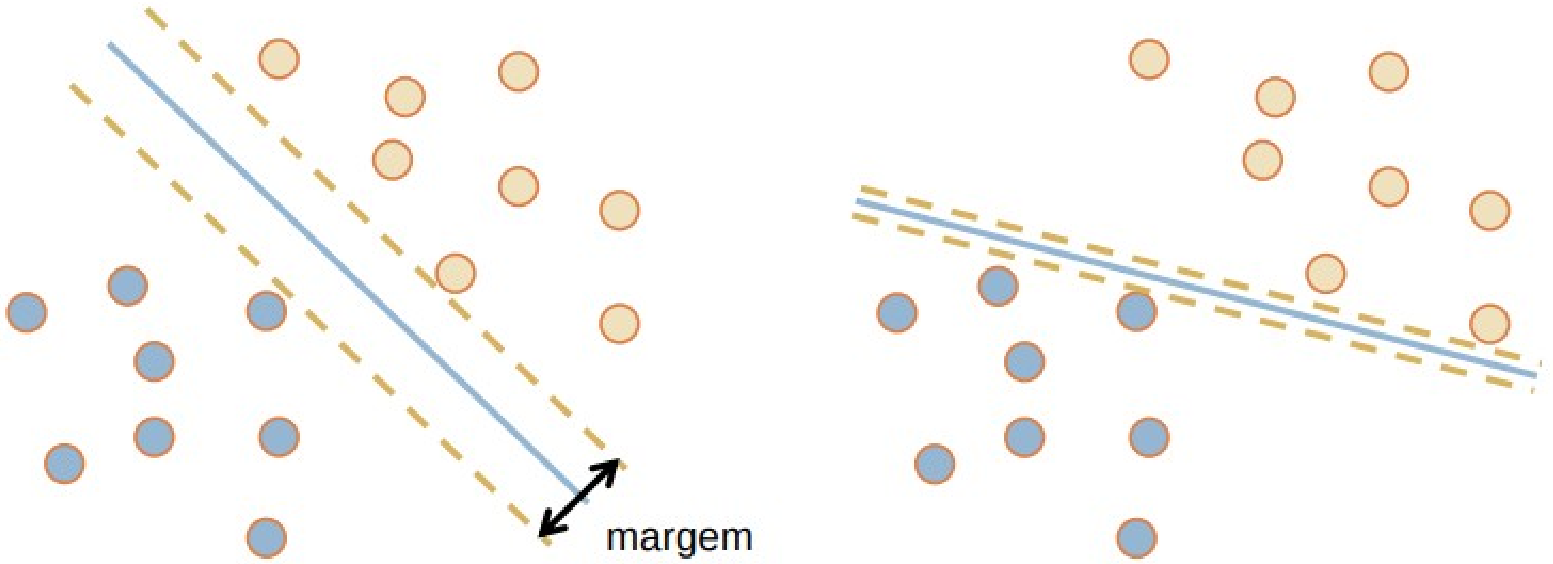
SVM Linear

- Qual o melhor Hiperplano ?



SVM Linear

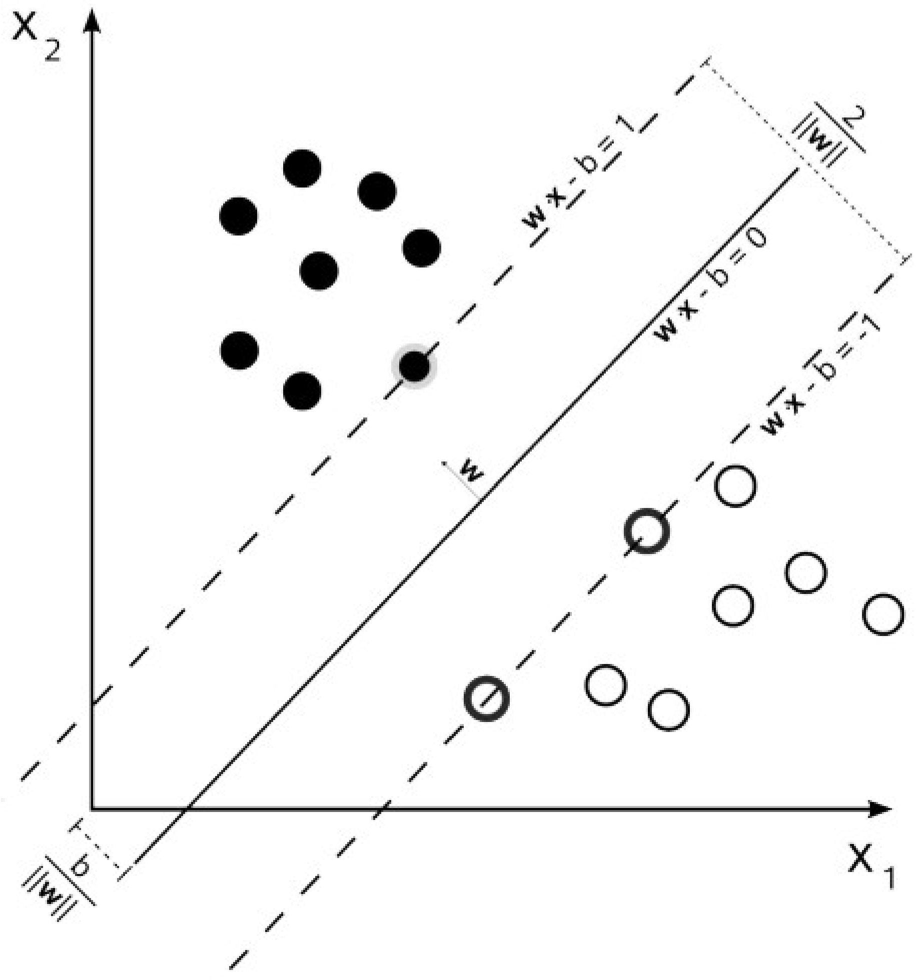
- Definição da Margem



SVM Linear

- Definição da Margem
  - f(x) = w.x + b
    - w = pesos
    - x = Amostras
    - b = bias
  - Logo y(x) =

$$y(x) = \begin{cases} +1, se\ wx + b > 0 \\ -1, se\ wx + b < 0 \end{cases}$$



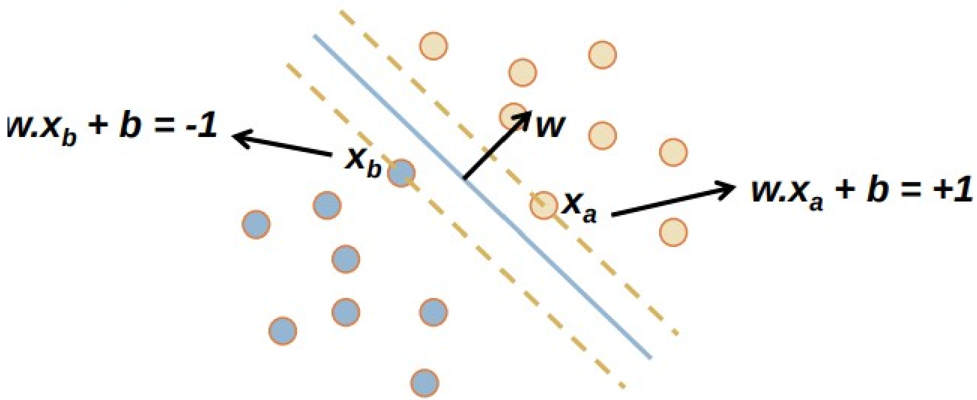
SVM Linear

- Treinamento: Otimizar ‘W’ e ‘b’

$$\begin{aligned} wx_a + b &= +1 \\ wx_b + b &= -1 \end{aligned} \Rightarrow w(x_a - x_b) = 2$$

$$w(x_a - x_b) = 2 \Rightarrow \|x_a - x_b\| = \frac{2}{\|w\|}$$

$$margem = \frac{2}{\|w\|} \Rightarrow \frac{1}{2}\|w\|^2$$



- Tem-se então a otimização de uma função quadrática, dada a restrição:

$$y_i(w.x_i + b) \geq 1, i = 1, \dots n$$

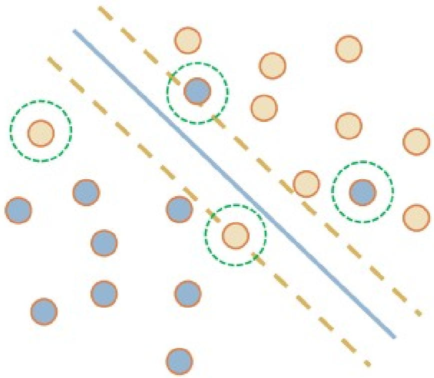
$x_i, i = 1, \dots n$, conjunto de padrões

$y_i = \{-1, +1\}, i = 1, \dots n$, respectivas classes

$$\xrightarrow{\text{Lagrange}} \quad L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(wx_i + b) - 1)$$

SVM Linear (Margens Suaves)

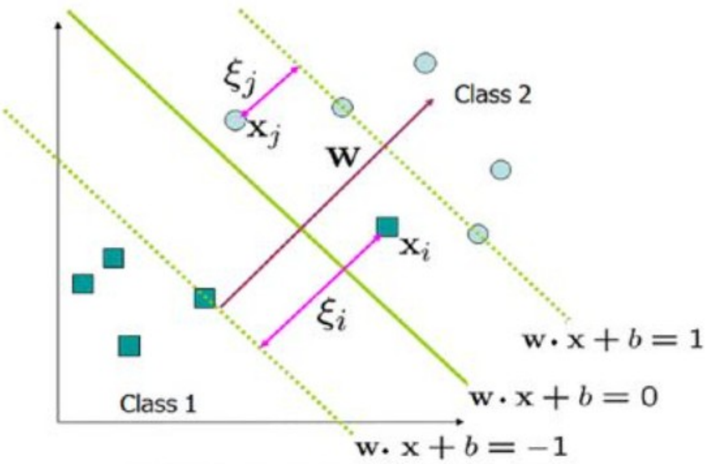
- Presença de ruídos ou outliers



- Solução: Suavização (Folga)
  - C: Define a folga (Definido experimentalmente)

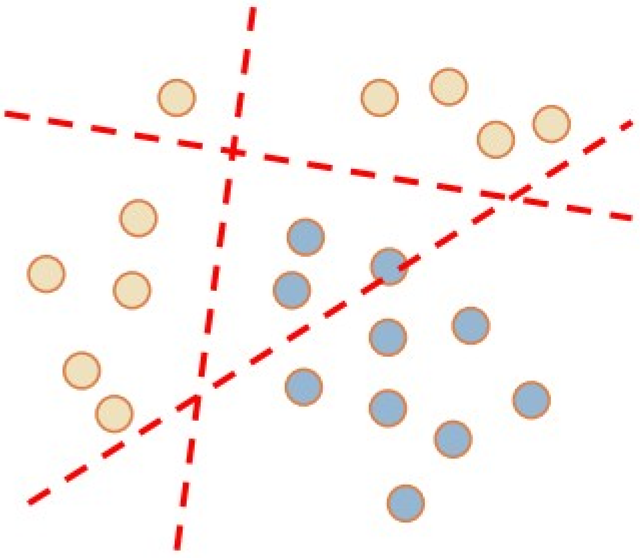
$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

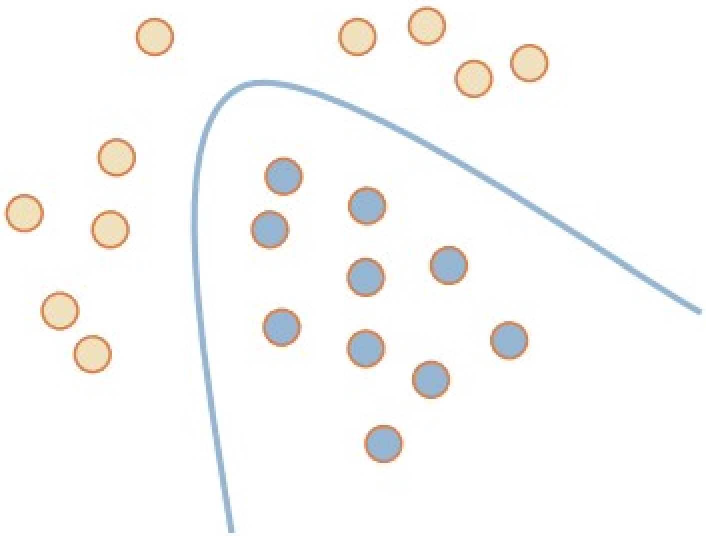
$$y_i(wx_i + b) \geq 1 - \xi_i$$



SVM Não Linear

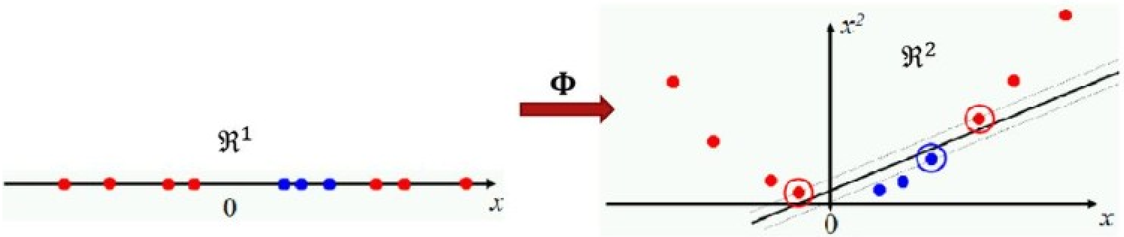
- E quando os dados não são linearmente separáveis?

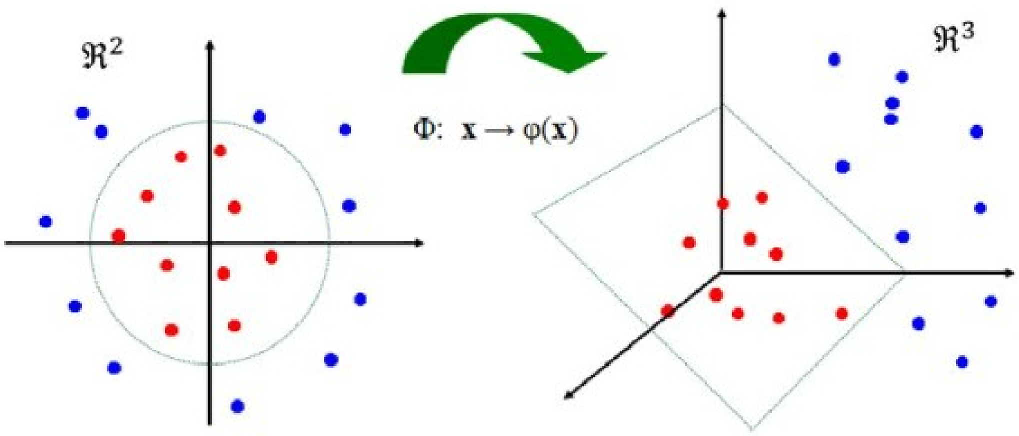




SVM Não Linear

- Encontrar uma transformação não linear ( ) Φl que $R^N \rightarrow R^M$ (M > N)
  - Teorema de Cover



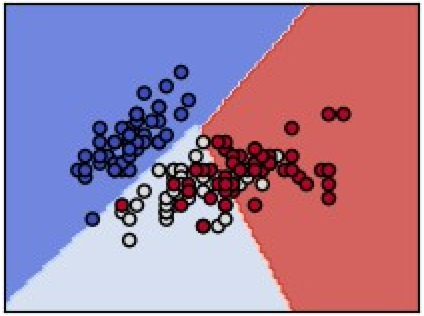


$$\frac{1}{2}\|w\|^2 + C\sum \xi_i$$

$$y_i(w \cdot \mathbf{\Phi}(x_i) + b) \geq 1 - \xi_i, \forall x_i$$
$$\xi_i \geq 0$$

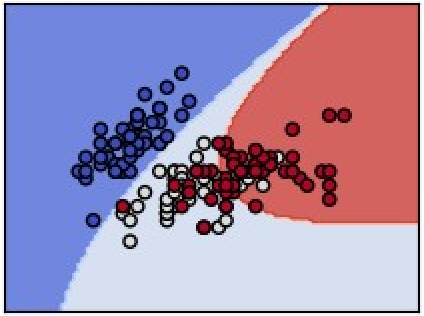
SVM Não Linear

- Kernel Linear



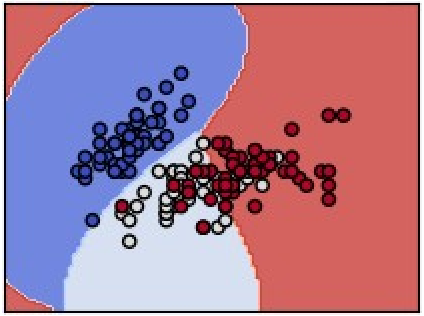
- Non-Linear Kernels
  - Polinomiais

$$(\delta(x_i . x_j) + k)^d$$



  - Gaussianos ou RBF

$$\exp(-\sigma . \|x_i - x_j\|^2)$$



Considerações Finais

- Vantagens
  - Se adaptam bem a problemas complexos
  - Pouca parametrização (‘C’)

- Desvantagens
  - Otimização pode ser demasiadamente complexa
    - ‘w’, ‘b’ e Kernel podem demorar a convergir
    - Bases Volumosas ou Muitas Classes
  - Modelo Caixa-preta (Interpretabilidade reduzida)