Objectivos

- Adquirir a noção de classificador linear
- Adquirir a noção de máquina de vectores de suporte

- Sumário:
  - Máquinas de vectores de suporte
    - Introdução
    - Classificação com máquinas de vectores de suporte

Máquinas de vectores de suporte (SVM)

- Foram introduzidas nos anos 90 por Vladimir Vapnik (1936 -)
- Se deriva da *Teoria de Aprendizagem Estatística*
- Técnica com muito êxito desde a sua introdução
  - Resultados práticos na resolução de diferentes tipos de problemas
  - Fundamentos teóricos sólidos

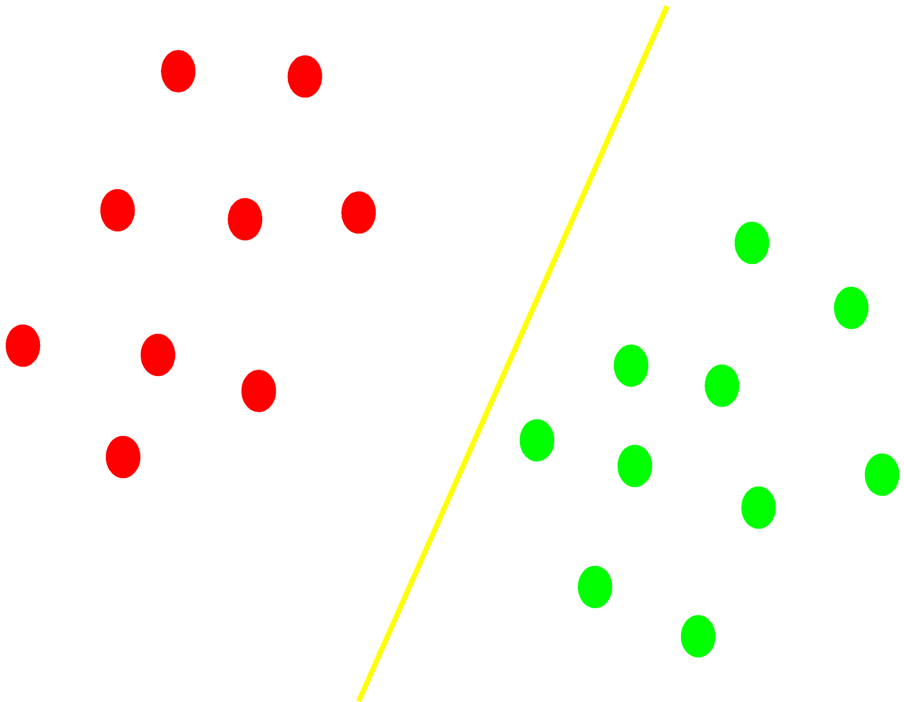
Máquinas de vectores de suporte (SVM)

- Se enquadram na categoria dos classificadores lineares

Classificadores lineares

- Dado um conjunto de treino $(\overrightarrow{x_1}, y_1), (\overrightarrow{x_2}, y_2), \ldots, (\overrightarrow{x_l}, y_l), \ \overrightarrow{x_i} \in R^n, y_i \in \{-1, 1\}$
- O objectivo é encontrar um hiperplano *h*, de dimensão n – 1, que separe os exemplos da classe -1 dos exemplos da classe +1

Classificadores lineares



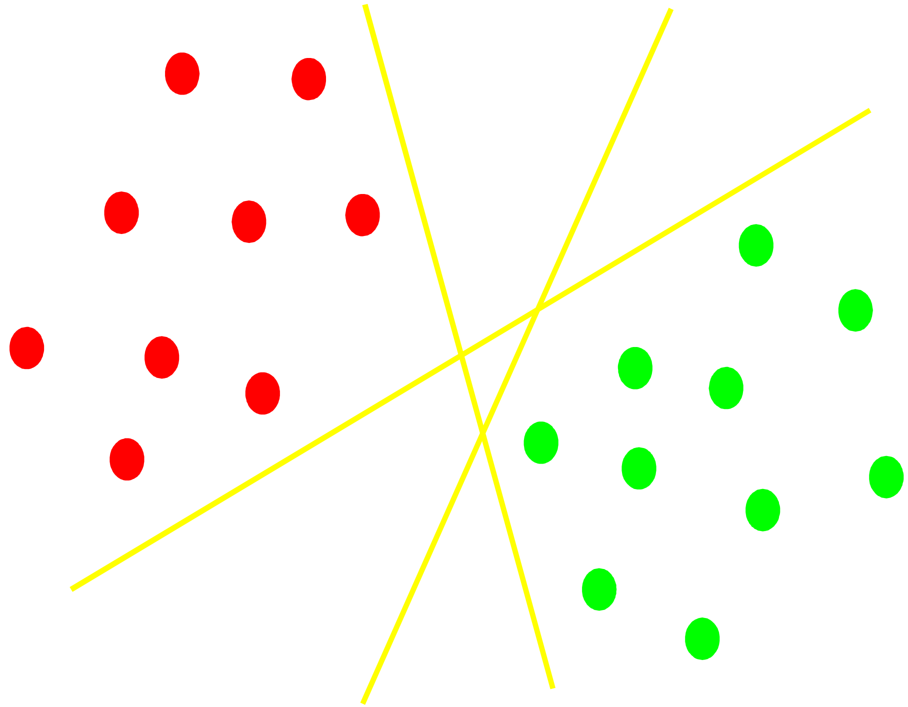
- O hiperplano de decisão é definido pela equação
  - $\vec{w} \cdot \vec{x} + b = 0$ ou
  - $\sum_{i=1}^{n} w_i x_i + b = 0$

Classificadores lineares

- Processo de aprendizagem consiste em determinar o vector de pesos ($\vec{w}$) e o umbral (b)
- Existem vários algoritmos que permitem aprender classificadores lineares
  - *Perceptron*
  - Regressão linear
  - Regressão logística…

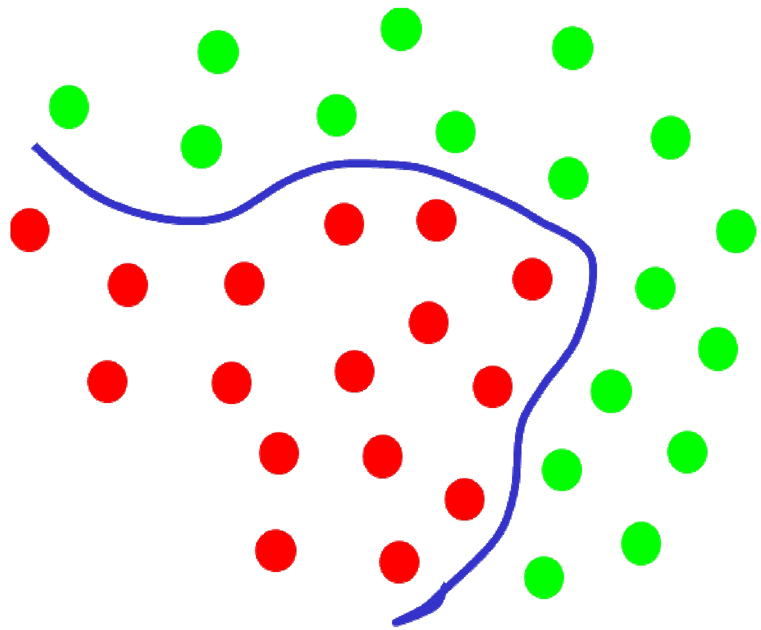
Classificadores lineares

- Para um conjunto de dados linearmente se- parável existem várias superfícies de decisão possíveis



Classificadores lineares

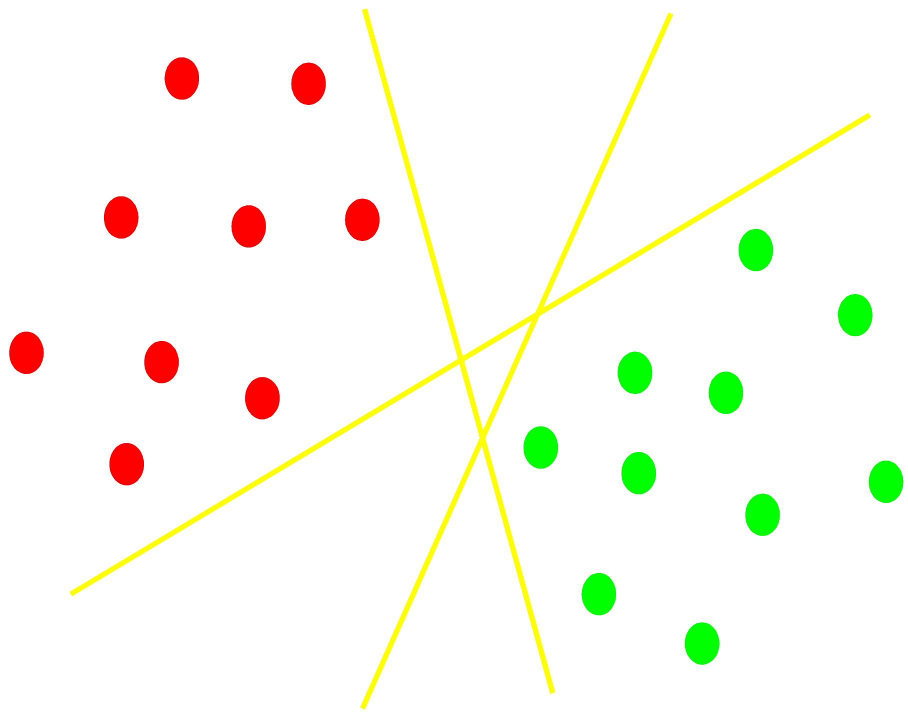
- E se o conjunto de dados não for linearmente separável?



Máquinas de vectores de suporte

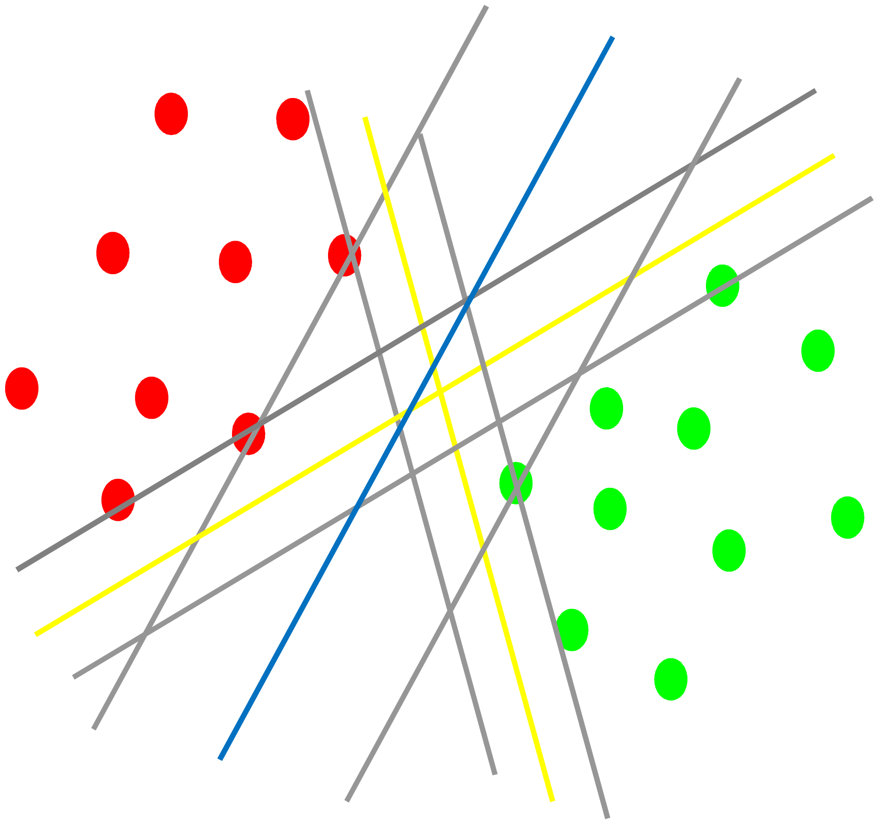
- Constituem uma extensão dos modelos lineares
- Têm como princípios fundamentais
  - A definição de superfícies de decisão com margem máxima
  - O uso de funções kernel
    - Permite construir *superfícies de decisão não lineares* usando *algoritmos lineares*

Hiperplano com margem máxima



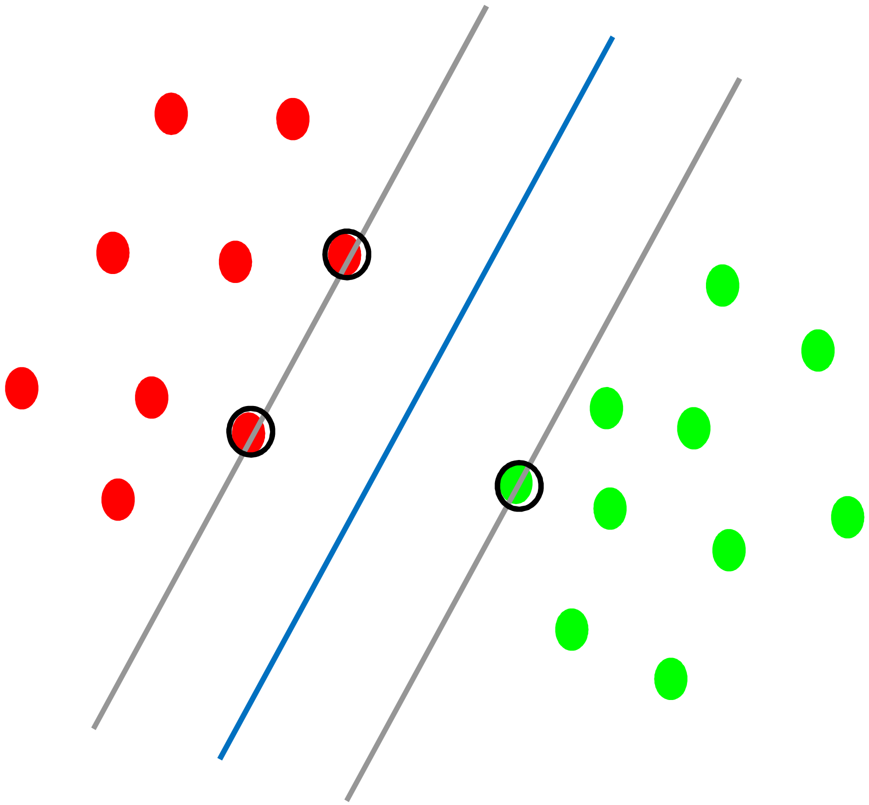
- Existem diferentes hiper-planos de separação
- É possível definir algum critério de qualidade?

Hiperplano com margem máxima



- Relacionado com o conceito de margem
- O hiperplano que classifica melhor os dados é o correspondente à maior margem

Hiperplano com margem máxima



- Os vectores de dados mais próximos do hiper-plano se denominam vectores de suporte
- Definem de forma única o hiperplano de separação
  - Todos os outros vectores são irrelevantes

Hiperplano com margem máxima

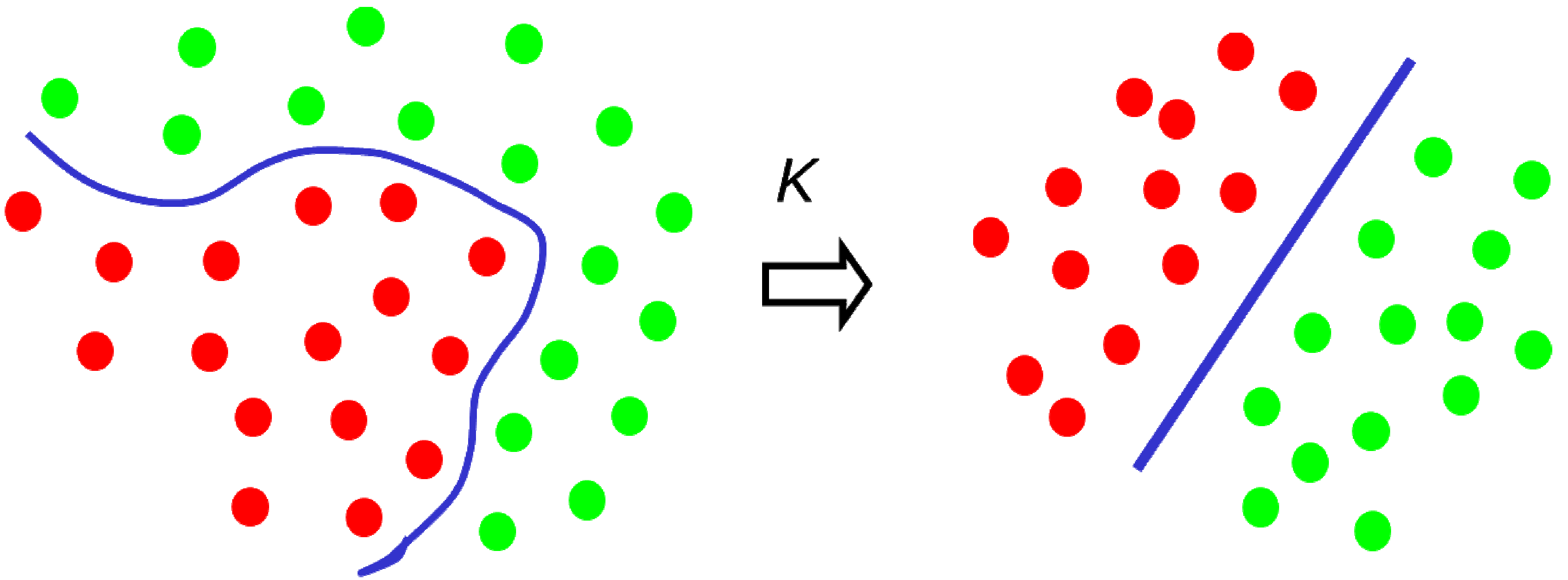
- A busca do hiperplano de margem máxima corresponde a um problema de optimização quadrática (Quadratic Programming, QP)
  - Se trata de encontrar os parâmetros que optimizam a uma equação de segundo grau, sujeita a uma série de restrições lineares sobre os referidos parâmetros
- A sua solução é dada na forma
  - $h(\vec{x}) : \vec{w} . \vec{x} + b = \sum_{i=1}^{l} \alpha_i y_i (\vec{x_i} . \vec{x}) + b$
    - $\alpha_i > 0$, se $\vec{x_i}$ é um *vector de suporte*
    - $\alpha_i = 0$, se $\vec{x_i}$ não é um *vector de suporte*

Dados não separáveis linearmente

- Nestes casos se aplica uma transformação não linear de forma a realizar a optimização num espaço em que os dados são linearmente separáveis
- Designado *espaço de características*

Dados não separáveis linearmente

Espaço de entrada **Espaço de características**



Inconvenientes potenciais

- Difícil definir uma função de transformação adequada a cada problema
- Custo computacional da conversão dos vectores do espaço de entrada para o espaço de características
- Custo computacional do cálculo de produtos escalares no espaço de características

Funções kernel

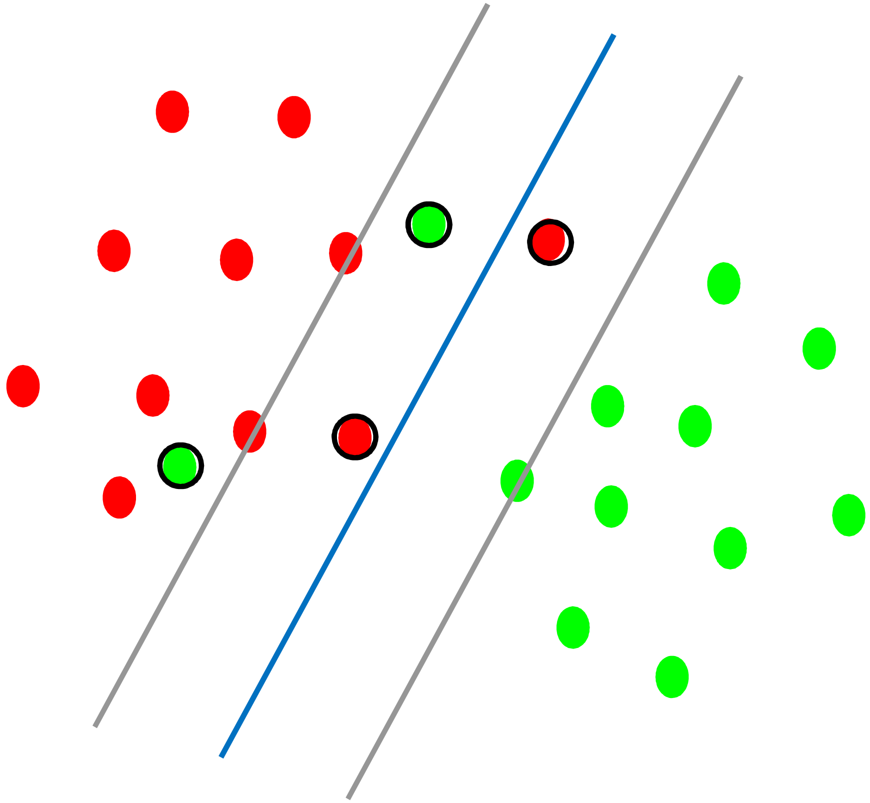
- Permitem definir a função de transformação de forma implícita (não é necessário calculá-la para todos os vectores de dados)
- Permite calcular os produtos escalares no espaço de características aplicando a função sobre os vectores no espaço de entrada

Funções kernel

- Exemplos de funções kernel
  - Kernel linear: $K(\vec{x},\vec{y}) = \vec{x}.\vec{y}$
  - Kernel polinómico: $K(\vec{x},\vec{y}) = (\vec{x}.\vec{y} + r)^n$
  - Funções de base radial: $K(\vec{x},\vec{y}) = e^{-\frac{\|\vec{x}-\vec{y}\|^2}{2\sigma}}$

Margem suave

- Quando a separabilidade linear não é possível (no *espaço de entradas* nem no *de características*) se relaxa o problema admitindo certa grau de erro no hiperplano óptimo
  - A quantidade de erro admitida é controlada por um parâmetro → parâmetro de penalização (C)
  - Valor de C é definido pelo usuário em cada problema



Exemplo 1

- Considere os seguintes exemplos de treino para um problema com duas classes:
  - Classe 1: (1, 1), (1, 2), (3, 2)
  - Classe 2: (-1, -1), (1, 0), (0, 1)
- Represente os seis pontos, construa a superfície de separação obtida através de SVM e mostre os vectores de suporte
- Conhecendo que a equação de uma recta é dada por w.x + b = 0, trate de escrever a equação do hiperplano obtido na alínea anterior

Exemplo 2

- Considere o seguinte conjunto de treino formado por pontos com uma dimensão
  - Classe 1: -5, 5
  - Classe 2: -2, 1
- Represente graficamente os pontos. São separáveis linearmente?
- Considere a transformação $f : R \rightarrow R^2$, $f(x) = (x, x^2)$. Transforme os dados e represente-os no novo espaço
- Qual é o hiperplano de separação óptimo no espaço transformado? Corresponde a uma superfície de separação não linear no espaço original?

Tarefa

- Leitura da semana
  - Hsu, C. W., Chang, C. C. e Lin, C. J., “A Practical Guide to Support Vector Classification”

Bibliografia

- Witten, pg. 124 – 131, 223 – 227