

Maximização de Margem

Victor Ruela

Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais

victorspruela@ufmg.br

7 de fevereiro de 2021

Agenda

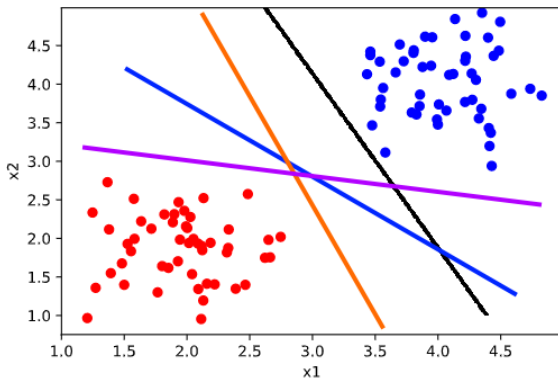
- 1 Introdução
- 2 Maximização de Margem
 - O Hiperplano Ótimo
 - Otimização
 - Padrões Não Separáveis
- 3 Problemas não-linearmente separáveis
 - Exemplos

Agenda

- 1 Introdução
- 2 Maximização de Margem
 - O Hiperplano Ótimo
 - Otimização
 - Padrões Não Separáveis
- 3 Problemas não-linearmente separáveis
 - Exemplos

Definição do Problema

- Dados de treinamento: $\mathcal{T} = \{(\mathbf{x}_1, d_1), \dots, (\mathbf{x}_N, d_N)\}$
- Problema de classificação binário: $d_i \in \{-1, 1\}$ **linearmente separável**
- Existem infinitos hiperplanos separadores!



Definição do Problema

- O uso do Perceptron, por exemplo, encontrará um hiperplano que separa estes padrões:

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (1)$$

- Entretanto, considerar a distância de todos os pontos para treinamento não é uma garantia de otimalidade
- E se fossem usados somente aqueles mais difíceis de serem separados e próximos da superfície de separação?

Maximização de Margem

Abordagem introduzida por Vapnik [1] para encontrar uma superfície de separação ótima com boa generalização, aplicável a diferentes modelos lineares

Agenda

- 1 Introdução
- 2 Maximização de Margem
 - O Hiperplano Ótimo
 - Otimização
 - Padrões Não Separáveis
- 3 Problemas não-linearmente separáveis
 - Exemplos

O Hiperplano Ótimo

- Quando a escolha de \mathbf{w} e b maximizam a margem de separação (ρ), o hiperplano é dito **ótimo** [Haykin, 2007]:

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0 \quad (2)$$

- Para otimalidade, o par (\mathbf{w}_o, b_o) deve satisfazer:

$$\begin{cases} \mathbf{w}_o^T \mathbf{x}_i \geq 1, & d_i = +1 \\ \mathbf{w}_o^T \mathbf{x}_i \leq -1, & d_i = -1 \end{cases} \quad (3)$$

- Os pontos (\mathbf{x}_i, d_i) que satisfazem com igualdade estas restrições são chamados de **vetores de suporte**
- Eles são os pontos mais próximos do hiperplano e consequentemente mais difíceis de classificar [Haykin, 2007].

O Hiperplano Ótimo

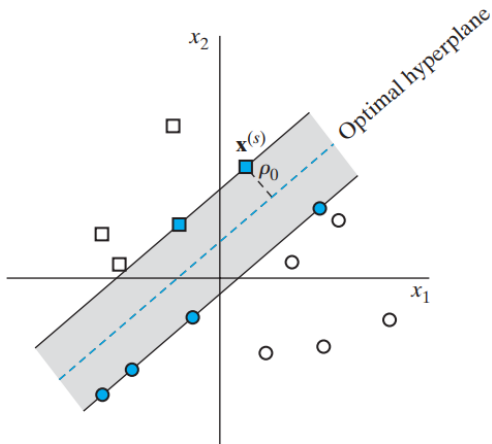


Figura: Representação do hiperplano ótimo e seus elementos. Extraído de [Haykin, 2007]

O Hiperplano Ótimo

- Por definição:

$$\mathbf{w}_o^T \mathbf{x}^{(s)} + b_o = \mp 1 \quad (4)$$

- Logo, a distância do vetor de suporte $\mathbf{x}^{(s)}$ ao hiperplano separador é dada por:

$$r = \begin{cases} \frac{1}{\|\mathbf{w}_o\|} & \text{se } d^{(s)} = +1 \\ \frac{-1}{\|\mathbf{w}_o\|} & \text{se } d^{(s)} = -1 \end{cases} \quad (5)$$

- Finalmente, a margem ótima ρ é:

$$\rho = \frac{2}{\|\mathbf{w}_o\|} \quad (6)$$

Conclusão

Maximizar a margem de separação entre classes binárias é equivalente a minimizar a norma Euclidiana do vetor de pesos \mathbf{w}

Formulação do Problema

- Combinando as Equações (4) e (6), podemos formular o problema de encontrar a margem ótima como:

$$\begin{aligned} &\underset{\mathbf{w}, b}{\text{minimizar}} && \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ &\text{sujeito a} && d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, N \end{aligned} \tag{7}$$

- Problema com objetivo quadrático e restrições lineares
- É reformulado e resolvido através da técnica de multiplicadores de Lagrange [Haykin, 2007]

Resultado

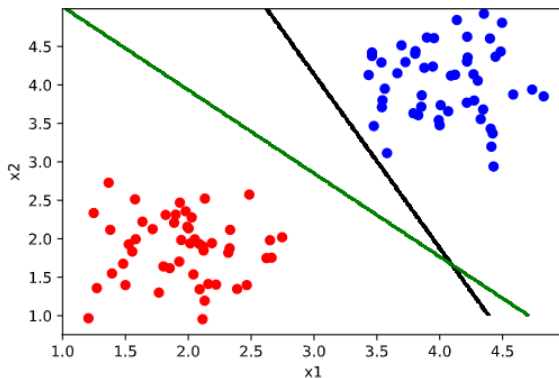


Figura: Superfícies de separação: MSE (preta) e margem máxima (verde)

Padrões Não Separáveis

- Em aplicações práticas, não podemos garantir que os dados sejam perfeitamente separáveis
- Logo, é adicionada uma variável de folga ξ para representar estas discrepâncias:

$$\begin{aligned} \underset{\mathbf{w}, b}{\text{minimizar}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{sujeito a} \quad & d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, N \\ & \xi_i \geq 0 \end{aligned} \tag{8}$$

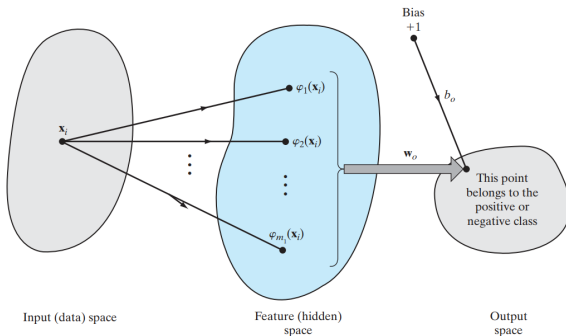
- É reformulado e resolvido de forma similar ao Problema (8)
- A constante C é especificada pelo usuário e controla a generalização do modelo, sendo conhecida como o parâmetro de regularização

Agenda

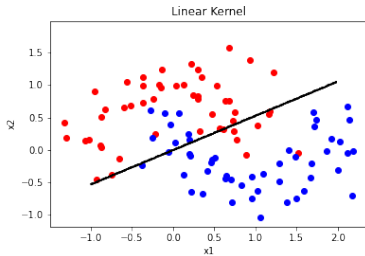
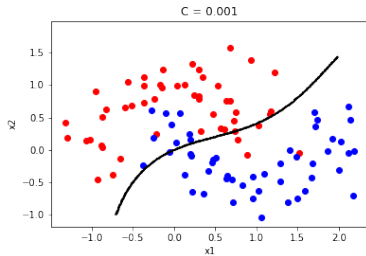
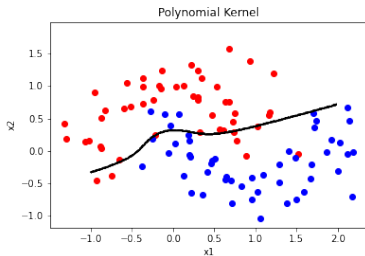
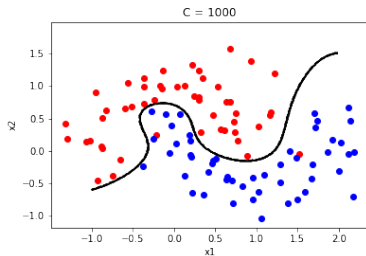
- 1 Introdução
- 2 Maximização de Margem
 - O Hiperplano Ótimo
 - Otimização
 - Padrões Não Separáveis
- 3 Problemas não-linearmente separáveis
 - Exemplos

Problemas não-linearmente separáveis

- Problemas reais nem sempre podem ser separados linearmente
- Entretanto, podemos mapeá-lo para um novo espaço de alta dimensão onde ele é mais provável de ser linearmente separável: **Teorema de Cover** [Cover, 1965]
- Este mapeamento é feito por funções não-lineares conhecidas como *Kernels*



Exemplo - SVM



Referências



Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152).



Haykin, S. (2007). Neural Networks: A Comprehensive Foundation (3rd Edition).



Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE transactions on electronic computers, (3), 326-334.

Obrigado!