Maximização de Margem

Victor Ruela

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

victorspruela@ufmg.br

1 de fevereiro de 2021

Agenda

1. Introdução
   - Aprendizado Supervisionado
   - Minimização do Erro
2. Maximização de Margem
   - Padrões Não-linearmente Separáveis
   - Otimização
3. Exemplos
   - SVM

Problema de Classificação

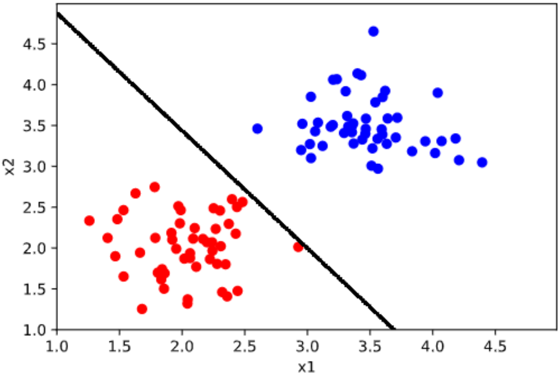
- Dados de treinamento: $\mathcal{T} = \{(\mathbf{x}_1, d_1), \ldots, (\mathbf{x}_N, d_N)\}$
- Para um problema de classificação binário: $d_i = \{-1, 1\}$
- Assume-se que os padrões são linearmente separáveis
- Em geral, algoritmos para treinamento de RNAs objetivam minimizar o erro quadrático:

$$\min \sum_{i=1}^{N} [d_i - sign(\mathbf{w}^T \mathbf{x}_i + b)]^2 \tag{1}$$

Minimização do Erro

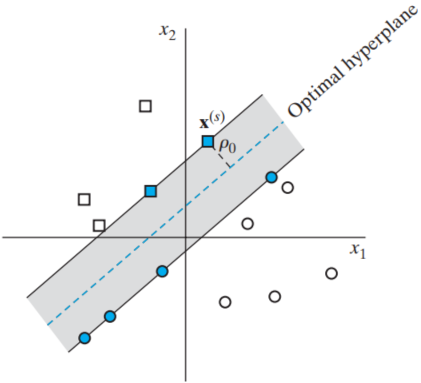
- O resultado da solução do Problema 1 será um conjunto de pesos representando um hiperplano que separa estes padrões:

$$\mathbf{w}^T\mathbf{x} + b = 0 \qquad (2)$$



O Hiperplano Ótimo

- Margem de separação: ρ
- Quando a escolha de **w** e b maximizam ρ, o hiperplano é dito **ótimo** [Haykin, 2007].



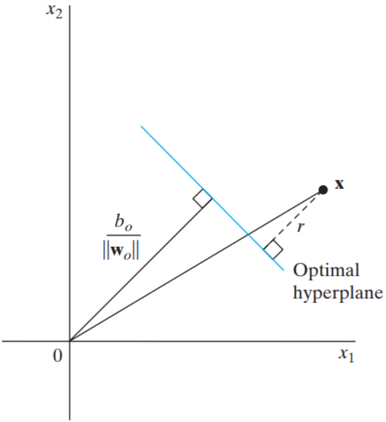
O Hiperplano Ótimo

- A função discriminante ótima é definida como:

$$g(\mathbf{x}) = \mathbf{w}_o^T \mathbf{x} + b_o \qquad (3)$$

- E distância deste hiperplano por:

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}_o\|} \qquad (4)$$



O Hiperplano Ótimo

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

$$\begin{cases} \mathbf{w}_o^T \mathbf{x}_i \geq -1, d_i = +1 \\ \mathbf{w}_o^T \mathbf{x}_i \leq -1, d_i = -1 \end{cases} \tag{5}$$

- 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

- Por definição:

$$g(\mathbf{x}^{(s)}) = \mathbf{w}_o^T \mathbf{x}^{(s)} + b_o = \mp 1 \tag{6}$$

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

$$\begin{cases} \frac{1}{\|\mathbf{w}_o\|} & \text{se } d^{(s)} = +1 \\ \frac{-1}{\|\mathbf{w}_o\|} & \text{se } d^{(s)} = -1 \end{cases} \tag{7}$$

- Finalmente, a margem ótima ρ é:

$$\rho = \frac{2}{\|\mathbf{w}_o\|} \tag{8}$$

Conclusão

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

Referências

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

Obrigado!