

# Fundamentos de Redes Neurais Artificiais

Victor São Paulo Ruela  
Programa de Pós-Graduação em Engenharia Elétrica  
Universidade Federal de Minas Gerais  
Belo Horizonte, Brasil  
Email: victorspruela@ufmg.br

**Resumo**—Este trabalho tem como objetivo apresentar uma revisão da literatura de redes neurais artificiais, com enfoque na evoluções das principais técnicas clássicas.

## I. INTRODUÇÃO

Redes neurais artificiais (RNA) é uma classe de modelos muito popular em problemas de classificação, reconhecimento de padrões, regressão e predição, sendo aplicado em diversas disciplinas.

## II. APRENDIZADO SUPERVISIONADO

RNAs de aprendizado supervisionado são responsáveis pela inferência de uma função desconhecida  $f$  que realiza o mapeamento de entre uma saída e um conjunto de entradas medidas, a partir de  $N$  pares de dados  $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ . Problemas de classificação, regressão e predição são comumente solucionados com estas técnicas, para os quais as primeiras estruturas de rede neural em camadas e algoritmos de treinamento descritos na literatura são o *Adaline*, em 1960, e o Perceptron simples, em 1957.

### A. Adaline

### B. Perceptron simples

Proposto inicialmente por Rosenblatt [1], este é um modelo capaz de aprender superfícies de decisão lineares para problemas de classificação. No seu trabalho original, Rosenblatt descreve formas de adaptação dos parâmetros da rede com o objetivo de reduzir a discrepância entre as saídas esperadas e estimadas e aprender associações entre os neurônios, o que é a base da indução para diversos algoritmos atuais.

Embora descrito como uma rede de duas camadas, originalmente seu treinamento só considerava uma destas. Por esse motivo, o Perceptron simples é comumente descrito na forma de somente um neurônio MCP [2].

### C. Máquinas de aprendizado extremo

### D. Redes RBF

### E. Perceptron de múltiplas camadas

### F. Máquinas de vetores suporte

### G. Aprendizado multiobjetivo

## III. APRENDIZADO NÃO-SUPERVISIONADO

### A. Aprendizado Hebbiano

### B. SOM

## IV. GENERALIZAÇÃO

Uma das suas principais características do modelo RNA é sua capacidade de generalização. Em geral, algoritmos de aprendizado supervisionado possuem como objetivo minimizar o erro quadrático dos valores previstos pelo modelo em relação às saídas em estudo:

$$\sum_{i=1}^N [y_i - f(\mathbf{x}_i)]^2 \quad (1)$$

onde  $y_i$  é uma resposta desejada para uma entrada  $\mathbf{x}_i$ , e  $f$  é a função que aproxima a resposta desejada. Ou seja, estamos interessados em encontrar o conjunto de pesos  $\mathbf{w}$  da rede a partir dos pares de dados de entrada-saída  $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$  que melhor aproxima a função desconhecida  $f$ .

Entretanto, se os dados a serem modelados são ruidosos o uso deste único objetivo pode levar a um overfitting sobre o conjunto de dados de treinamento, de forma que este não consiga generalizar bem para novos valores observados. Estatisticamente, podemos definir a efetividade de  $f$  como um estimador de  $y$  como [3]:

$$E[(y - f(\mathbf{x}; \mathcal{D}))^2 | \mathbf{x}, \mathcal{D}] = E[(y - E[y | \mathbf{x}])^2 | \mathbf{x}, \mathcal{D}] + (f(\mathbf{x}; \mathcal{D}) - E[y | \mathbf{x}])^2 \quad (2)$$

É importante notar neste indicador que o primeiro termo representa a variância de  $y$  dado  $\mathbf{x}$ , não dependendo dos dados. Já o segundo termo mede a distância entre o estimador e a regressão. Logo, podemos definir o erro quadrático médio de  $f$  como um estimador da regressão  $E[y | \mathbf{x}]$  para um conjunto de dados  $\mathcal{D}$  como:

$$\begin{aligned}
E_{\mathcal{D}}[(f(\mathbf{x}; \mathcal{D}) - E[y|\mathbf{x}])^2] = \\
(E_{\mathcal{D}}[f(\mathbf{x}; \mathcal{D})] - E[y|\mathbf{x}])^2 \quad \text{“viés”} \\
+ E_{\mathcal{D}}[f(\mathbf{x}; \mathcal{D}) - E_{\mathcal{D}}[f(\mathbf{x}; \mathcal{D})]] \quad \text{“variância”}
\end{aligned} \tag{3}$$

A derivação completa da relação acima pode ser encontrada em [3]. Logo é fácil notar que o aprendizado de RNAs é um problema multi-objetivo, no qual precisamos encontrar uma solução de compromisso entre o viés e a variância do modelo. Portanto, em um dos extremos teremos um conjunto de pesos que resultam em um viés máximo (*underfitting*) e no outro variância máxima (*overfitting*).

A. *Máquinas de Vetores Suporte*

B. *Aprendizado Multiobjetivo*

#### REFERÊNCIAS

- [1] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [2] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [3] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.