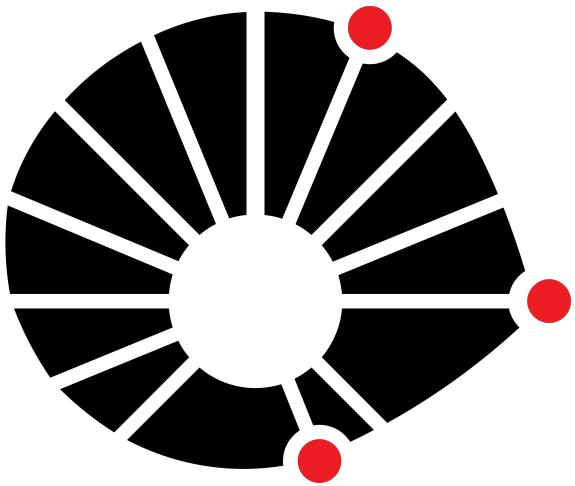
Aprendizado de Máquinas

Aula 15 – Redes Neurais Artificiais



UNICAMP

Marcos Eduardo Valle
Matemática Aplicada
IMECC - Unicamp

Introdução

Redes neurais articiais, ou simplesmente redes neurais, representam uma grande classe de modelos e técnicas de aprendizado de máquina que foram desenvolvidas em diferentes áreas, incluindo matemática aplicada, estatística e inteligência artificial.

Os estudos em redes neurais artificiais iniciaram com o trabalho de (McCulloch and Pitts, 1943). Resumidamente, eles mostraram que um neurônio artificial $\eta : \mathbb{R}^n \to \{0, 1\}$ descrito pela equação

$$\eta(x) = \chi_{\geq 0}\left[\mathbf{w}^T\mathbf{x} - b\right], \tag{1}$$

é capaz de implementar conectivos básicos (como “e” e “ou”) da lógica e, portanto, poderia resolver problemas da lógica clássica.

Neurônio Artificial

De um modo geral, um neurônio artificial é uma função $\eta : \mathbb{R}^n \to \mathbb{R}$ dada por

$$\eta(\mathbf{x}) = \varphi(\mathbf{w}^T\mathbf{x} - b), \tag{2}$$

em que $\varphi : \mathbb{R} \to \mathbb{R}$ é a função de ativação ou transferência, $\mathbf{w} = [w_1, \ldots, w_n]^T \in \mathbb{R}^n$ e $b \in \mathbb{R}$ denota o vetor dos pesos sinápticos e o viés, respectivamente.

Exemplos de função de ativação incluem

- Função limiar: $\chi_{\geq 0}(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$
- Função logística: $\sigma(t) = 1/(1 + e^{-t})$.
- Função ReLU: $\texttt{relu}(t) = \max\{0, t\}$.

Arquitetura da Rede Neural

Os neurônios artificiais são as unidades básicas de processamento de uma rede neural artificial.

Uma rede neural pode ser vista como um grafo direcionado em que os vértices correspondem aos neurônios e as arestas indicam que a saída de um neurônio alimenta outro neurônio.

A estrutura do grafo é chamada arquitetura ou topologia da rede neural.

Uma rede neural é recorrente quando está associada à um grafo com ciclos. Dizemos que a rede é progressiva quando o fluxo segue numa única direção.

Nesse curso iremos considerar apenas redes progressivas.

O Perceptron

Em 1958, Rodenblatt desvenvolveu um algoritmo de treinamento para o neurônio artificial de McCulloch e Pitts, resultando no modelo chamado "*perceptron*".

Do ponto de vista matemático, o neurônio de McCulloch e Pitts dado por (1) presenta um classificador linear.

A regra de treinamento desenvolvida por Rosenblatt fornece, em um número finito de passos, pesos e viés que classificam corretamente dados linearmente separáveis (Rosenblatt, 1958).

Apesar dos resultados promissores de Rosenblatt, o perceptron possui possui aplicações limitadas, não conseguindo resolver problemas mais complexos como o problema do ou-exclusivo (Minsky and Papert, 1969).

Multi-layer Perceptron (MLP)

As limitações do perceptron colaborou para o desenvolvimento de redes neurais de multiplas camadas como o Adaline e as redes MLP (do inglês *multi-layer perceptron*) (Rumelhart and McClelland, 1986; Widrow and Hoff, 1960).

Referimos como uma camada um conjunto de m neurônios artificiais em paralelo. Do ponto de vista matemático, uma camada com m neurônios define uma função $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ dada pela equação

$$\mathcal{L}(\mathbf{x}) = \varphi(\boldsymbol{W}\mathbf{x} - \mathbf{b}), \tag{3}$$

em que a função de ativação φ é aplicada componente-a-componente, $\boldsymbol{W} \in \mathbb{R}^{m \times n}$ e $\mathbf{b} \in \mathbb{R}^m$ é a matriz dos pesos sinápticos e o vetor viés, respectivamente.

Dizemos que $\mathcal{L}$ dada por (3) é densa ou totalmente conectada se $\boldsymbol{W}$ é uma matriz cheia.

Uma rede neural de múltiplas camadas é dada pela composição de camadas de neurônios $\mathcal{L}_k$, para $k = 1, \ldots, K$.

Formalmente, uma rede progressiva com K camadas define uma função $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}^m$ dada pela composição

$$\mathcal{N} = \mathcal{L}_K \circ \ldots \circ \mathcal{L}_2 \circ \mathcal{L}_1, \tag{4}$$

em que $\mathcal{L}_k : \mathbb{R}^{n_{k-1}} \to \mathbb{R}^{n_k}$ são camadas de neurônios para $k = 1, \ldots, K$, com $n = n_0$ e $m = n_K$. Em alguns casos, incluímos também uma camada $\mathcal{L}_0$ igual a identidade.

As primeiras camadas podem ser vistas como um extrator de características enquanto que as últimas camadas realizam a tarefa de aprendizado de máquina (classificação, regressão, etc.).

Dizemos que uma rede neural é profunda (*deep*) quando $K \geq 3$, ou seja, a rede possui três ou mais camadas (Aurelien Géron, 2019).

Uma rede MLP de camada única é definida pela composição de duas camadas densas, ou seja, $\mathcal{N} = \mathcal{L}_2 \circ \mathcal{L}_1$.

Em termos matemáticos, a saída $\mathbf{y} = \mathcal{N}(\mathbf{x})$ de uma rede MLP de camada única é dada por

$$\mathbf{y} = \varphi_2(\mathbf{W}_2\mathbf{x}_1 + \mathbf{b}_2) \quad \text{com} \quad \mathbf{x}_1 = \varphi_2(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1). \tag{5}$$

Pode-se mostrar que as redes MLP são, em geral, aproximadores universais. Portanto, dada uma função contínua $f : \mathbf{X} \to \mathbb{R}$, em que $X \subset \mathbb{R}^n$ é um compacto, e uma tolerância $\epsilon > 0$, existe uma rede MLP com uma única camada oculta $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}$ tal que $|f(\mathbf{x}) - \mathcal{N}(\mathbf{x})| < \epsilon$ para todo $\mathbf{x} \in \mathbf{X}$ (Cybenko, 1989; Hornik, 1991; Pinkus, 1999).

A capacidade de aproximação universal das redes MLP justificam, do ponto de vista teórico, suas aplicações em problemas de aprendizado de máquina.

Treinamento de Redes Neurais Artificias

O treinamento de uma rede neural artificial (*shallow* ou *deep*) consiste em minimizar uma função perda como o erro quadrático médio (MSE) em problemas de regressão e a entropia cruzada em problemas de classificação.

Dessa forma, o treinamento de uma rede neural é formulado como um problema de otimização irrestrito nos parâmetros da rede (pesos sinápticos e vieses) (Luenberger, 1984).

Devido ao grande número de parâmetros treináveis, o problema de otimização é geralmente abordado usando um método baseado no gradiente (primeira ordem).

O cálculo do gradiente é efetuado usando a regra da cadeia, resultando no algoritmo de retropropagação (*backpropagation*).

Considerações Finais

Na aula de hoje apresentamos o conceito de redes neurais artificias, com foco nas redes progressivas.

Em termos gerais, as redes progressivas são dadas por composições de camadas. Uma rede MLP, em particur, é composta por camadas densas de neurônios.

O treinamento de uma rede neural é efetuado minimizando uma função perda.

Como as principais bibliotecas de redes neurais tratam do calculo do gradiente, nessa disciplina não entraremos nos detalhes da resolução do problema de otimização envolvido no treinamento.

Muito grato pela atenção!

References (1)

Aurelien Géron. *Hands–On Machine Learning with Scikit–Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O Reilly, Sebastopol, California, USA., 2nd edition, 10 2019. ISBN 1492032646.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems 1989 2:4*, 2(4):303–314, 12 1989. ISSN 1435-568X. doi: 10.1007/BF02551274. URL `https://link.springer.com/article/10.1007/BF02551274`.

K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1 1991. ISSN 08936080. doi: 10.1016/0893-6080(91)90009-T.

D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2 edition, 1984.

References (2)

W. S. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

M. L. Minsky and S. A. Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1969.

A. Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1 1999. ISSN 0962-4929. doi: 10.1017/S0962492900002919.

F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65: 386–408, 1958.

D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstrutures of Cognition*, volume 1. MIT Press, Cambridge, MA, 1986.

References (3)

W. B. Widrow and M. E. Hoff. Adaptive Switching Circuits. *WESCON Convention Record*, pages 96–104, 1960.