Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

A geometria do perceptron de Rosenblatt

José Jorge de Souza Silva[1]
Instituto Federal de Educação, Ciência e Tecnologia - IFPB (Campus Cajazeiras)
Larissa Soares de Sousa[2]
Instituto Federal de Educação, Ciência e Tecnologia - IFPB (Campus Cajazeiras)
Vinicius Martins Teodosio Rocha[3]
Instituto Federal de Educação, Ciência e Tecnologia - IFPB (Campus Cajazeiras)

Resumo. Os algoritmos de Redes Neurais se tornaram praticamente onipresentes na discussões sobre Inteligência Artifical. Apresentamos aqui o Perceptron de Rosenblatt, um dos algoritmos elementares que desencadearam a construção das redes neurais multicamadas, e exploramos a interpretação geométrica que fundamenta o funcionamento do algoritmo.

Palavras-chave. Inteligência Artificial, Perceptron, Geometria Analítica

1 O perceptron de Rosenblatt

O princípio do perceptron de Rosenblatt, introduzido em [3], tem como objetivo, dado um conjunto finito $X \subset \mathbb{R}^n$ particionado em dois conjuntos X_+ e X_-, encontrar um hiperplano π de $\mathbb{R}^n$ que separa os conjuntos X_+ e X_-. A existência de tal hiperplano é garantida sempre que os fechos convexos de X_+ e X_- são disjuntos, veja [2]. Nesse caso diremos que tais conjuntos são *linearmente separáveis*. A partir disso, pode-se fazer previsões quanto a classe de pontos não previamente classificados (isto é, fora de X) de acordo com a posição relativa ao hiperplano π. Descrevemos aqui o Perceptron e apresentamos sua interpretação geométrica, bem como o teorema que garante sua convergência.

O algoritmo Após uma escolha de ordem para os pontos em X, denote por $x^{(i)} = (x_1, x_2, \ldots, x_n)$ seu i-ésimo elemento. Defina $y^{(i)} = \begin{cases} +1 & \text{se } x^{(i)} \in X_+ \\ -1 & \text{se } x^{(i)} \in X_- \end{cases}$. Tomamos inicialmente $w \in \mathbb{R}^n$ e $b \in \mathbb{R}$ arbitrários, chamados de *vetor de pesos* e *bias*(usualmente tomados ambos nulos). O objetivo é atualizar w e b a medida que percorremos o conjunto X (possivelmente mais de uma vez) de forma que, ao final do processo, se X_+ e X_- são linearmente separáveis, w e b determinam um hiperplano $\pi_{w,b}$ que separa tais conjuntos. Para a verificação, calculamos, para cada $i \in \{0, \ldots, |X|\}$, o valor de $\alpha = x^{(i)} \cdot w + b$. O sinal de α determina a posição relativa entre o hiperplano $\pi_{w,b}$ e o ponto $x^{(i)}$ (veja a interpretação geométrica abaixo). Comparando o sinal de α com o valor de $y^{(i)}$ podemos concluir quando o hiperplano classifica corretamente o ponto $x^{(i)}$ dado. Se a classificação for correta seguimos para o ponto seguinte $x^{(i+1)}$, caso contrário, atualizamos w e b da seguinte forma: $w \leftarrow (w + y^{(i)} x^{(i)})$ e $b \leftarrow b + y^{(i)}$. Dessa forma, com o novo vetor de pesos, a classificação se aproxima da real. De fato, se $x^{(i)} \in X$ foi classificado incorretamente, para o novo peso e o novo bias temos

[1]jorge.souza@academico.ifpb.edu.br
[2]larissa.soares@academico.ifpb.edu.br
[3]vinicius.rocha@ifpb.edu.br

$$(w + y^{(i)}x^{(i)}) \cdot x^{(i)} + (b + y^{(i)}) = \alpha + y^{(i)}\|x^{(i)}\|^2 + y^{(i)} \begin{cases} > \alpha & \text{se } y^{(i)} = +1 \\ < \alpha & \text{se } y^{(i)} = -1 \end{cases}$$

O comportamento do Perceptron em cada iteração está resumido em 1. Destacamos que o hiperplano $\pi_{w,b}$ pode não classificar corretamente todos os pontos após o conjunto X ser percorrido uma única vez. Em geral, é necessário que esse ciclo — chamado de *época* — seja feito diversas vezes. O Teorema 2.1 fornece uma estimativa para essas repetições.

Tabela 1: Comportamento do perceptron

Classe	sinal($\alpha y^{(i)}$)	Classificação	Atualização
X_+	$\alpha y^{(i)} \geq 0$	Correta	não ocorre
X_+	$\alpha y^{(i)} < 0$	Incorreta	$w \leftarrow (w + y^{(i)}x^{(i)})$, $b \leftarrow b + y^{(i)}$

2 Interpretação geométrica

Dado $w = (w_1, ..., w_n) \in \mathbb{R}^n$ e $b \in \mathbb{R}$. Definimos o *hiperplano* associado a w e b, denotado por $\pi_{w,b}$, como o conjunto dos pontos $x = (x_1, \ldots, x_n) \in \mathbb{R}$ satisfazendo $\sum_{i=0}^{n} w_i x_i + b = 0$.

Lembremos que podemos classificar a posição relativa entre dois vetores em $\mathbb{R}^n$, de acordo com o ângulo θ formado entre eles, em três casos: $\theta < \frac{\pi}{2}$, $\theta = \frac{\pi}{2}$ e $\frac{\pi}{2} < \theta < \pi$, para seu produto interno positivo, negativo ou zero, resp. (veja [4]). O sinal de α calculado no algoritmo fornece portanto a posição relativa entre o ponto $x^{(i)}$ e o hiperplano $\pi_{w,b}$, como ilustrado nas figuras a seguir ($n = 2$).

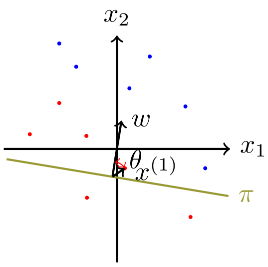

Figura 1: Inicialização

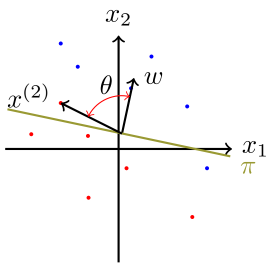

Figura 2: Quinta atualização

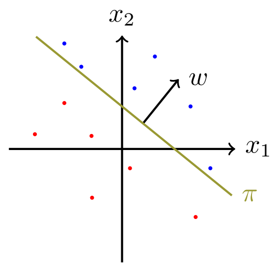

Figura 3: Convergência

A convergência do perceptron é garantida pelo Teorema de Rosenblatt (veja [1, Sec. 1.3]).

Teorema 2.1 (Teorema da Convergência do Perceptron) Seja $X = X_+ \cup X_- \subset \mathbb{R}^n$ finito com X_+ e X_- linearmente separáveis. Então Perceptron de Rosenblatt converge (no sentido de não haver mais atualizações) após uma quantidade $n_{\max}$ de iterações, onde $n_{\max}$ é uma constante dependente do comprimento máximo dos vetores em X e do máximo das margens (distância mínima entre um hiperplano separador de X_+ e X_- e os pontos de X).

Referências

[1] Haykin, S. *Neural networks and learnig machines, 3a. edição.* Hamilton, Pearson, 1999.

[2] Hyperplane separation theorem. *Wikipedia*, 2021. Disponível em: `https://en.wikipedia.org/wiki/Hyperplane_separation_theorem`. Acesso em: 27 de abr. de 2021.

[3] Rosenblatt, F. *The Perceptron - a perceiving and recognizing automaton*(1957) Report 85-460-1. Cornell Aeronautical Laboratory.

[4] Winterle, P. *Vetores e geometria analitica, 1a. edição.* São Paulo, Pearson Makron Books, 2000.