Redes Neurais Artificiais

Exemplo de Aprendizado Supervisionado com o Neurônio Perceptron

Elloá B. Guedes

Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus, AM
ebgcosta@uea.edu.br

30 de agosto de 2020

1. Apresentação

Este documento é inspirado na Seção 2.3.5 do trabalho de Braga et al. [1] com o intuito de detalhar os passos do Aprendizado Supervisionado de um neurônio Perceptron perante um conjunto de dados de treinamento. Foram adicionadas figuras de autoria própria e outras observações advindas de fontes referenciadas ao longo do documento.

2. Conjunto de Treinamento

Considerando que neste exemplo o aprendizado ocorre de maneira supervisionada, o conjunto de treinamento Γ possui pares na forma $(\mathbf{x},y)$, em que $\mathbf{x}$ é o vetor de entrada dos atributos e y é a saída para esta entrada, obtida junto a um especialista no problema. Assim, tem-se um conjunto de treinamento com dois pares de exemplos e nomenclatura adotada destacada na cor azul:

$$\Gamma = \left\{ \left(\underbrace{\begin{bmatrix} 2 & 2 \end{bmatrix}}_{x(1)}, \underbrace{1}_{y_d^1} \right), \left(\underbrace{\begin{bmatrix} 4 & 4 \end{bmatrix}}_{x(2)}, \underbrace{0}_{y_d^2} \right) \right\} \tag{1}$$

Observe que o problema em questão é linearmente separável a partir da representação dos exemplos disponíveis no plano cartesiano, conforme denotado na Figura 1. Consideram-se os atributos como x_1 e x_2 e as classes como sendo denotadas em cores e formas distintas.

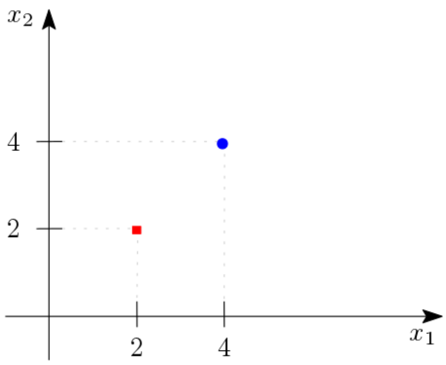

Figura 1: Ilustração dos exemplos do conjunto de treinamento no plano cartesiano. A classe 1 é representada por um quadrado vermelho e a classe 0 é representada por um círculo azul.

3. Parâmetros e Hiperparâmetros

Para resolver este problema, será utilizado o modelo de neurônio artificial *Perceptron de Rosenblatt* [2], o qual adiciona um viés à entrada em contraste ao modelo de neurônio artificial de Mcculloch & Pitts [3]. Considerado o problema em questão, tem-se a seguinte instância ilustrada na Figura 2.

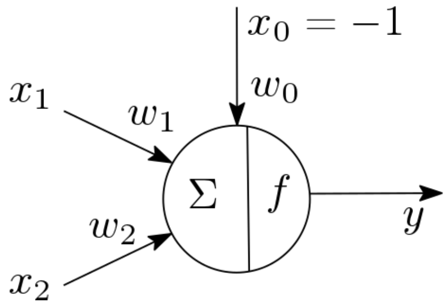

Figura 2: Neurônio perceptron de Rosenblatt instanciado para o problema.

Assume-se que o vetor de pesos inicial $w(0)$ foi inicializado de maneira aleatória, com os seguintes valores:

$$w(0) = \begin{bmatrix} -0{,}5441 \\ 0{,}5562 \\ -0{,}4074 \end{bmatrix} \tag{2}$$

Outro parâmetro que precisa ser definido neste neurônio é a função de ativação. Neste caso será utilizada a função de ativação degrau com limiar $\vartheta = 0$, dada na Eq. (3) e ilustrada graficamente na Figura 3.

$$f(u) = \begin{cases} 1, & u \geq \vartheta = 0 \\ 0, & \text{c.c.} \end{cases} \tag{3}$$

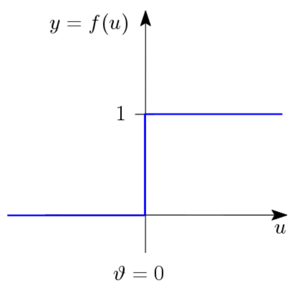

Figura 3: Função de ativação degrau.

Neste problema, o único hiperparâmetro a ser definido é a taxa de aprendizado η, a qual determina a taxa em que os pesos serão ajustados. Um valor muito pequeno de η pode implicar em demora para convergência, ao passo que um valor muito grande pode gerar instabilidade no treinamento. Em termos práticos, para este tipo de problema, recomenda-se que o valor desta taxa esteja no intervalo $0{,}1 \leq \eta \leq 0{,}4$. Neste exemplo, será adotado o valor $\eta = 0{,}1$.

4. Necessidade de Aprendizado

Para justificar a necessidade de ajustes nos pesos do neurônio em questão, basta verificar que tais valores não permitem uma distinção clara entre os exemplos das classes existentes.

Para o exemplo $x(1)$, tem-se:

$$\begin{aligned} u &= \sum_{i=0}^{2} x_i w_i && (4)\\ &= -1 \cdot (-0{,}5441) + 2 \cdot 0{,}5562 + 2 \cdot (-0{,}4074) && (5)\\ &= 0{,}5441 + 1{,}1124 - 0{,}8148 && (6)\\ &= 0{,}8417. && (7) \end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$\begin{aligned} y &= f(u) && (8)\\ &= f(0{,}8417) && (9)\\ &= 1. && (10) \end{aligned}$$

Como $y = y_d^1$, tem-se que o neurônio em questão prevê corretamente esta entrada.

De maneira análoga, tem-se o que segue para o exemplo $x(2)$:

$$
\begin{aligned}
u &= \sum_{i=0}^{2} x_i w_i &(11)\\
&= -1 \cdot (-0{,}5441) + 4 \cdot 0{,}5562 + 4 \cdot (-0{,}4074) &(12)\\
&= 0{,}5441 + 2{,}2248 - 1{,}6296 &(13)\\
&= 1{,}1459. &(14)
\end{aligned}
$$

Aplicando-se a função de ativação, tem-se:

$$
\begin{aligned}
y &= f(u) &(15)\\
&= f(1{,}1459) &(16)\\
&= 1. &(17)
\end{aligned}
$$

Como $y \neq y_d^2$, tem-se que o neurônio em questão não prevê corretamente a saída esperada esta entrada.

O aprendizado em neurônios Perceptron dá-se pelo ajuste de pesos, conforme estabelecido pela Regra Delta:

$$
\begin{aligned}
w(n+1) &= w(n) + \Delta w(n) &(18)\\
&= w(n) + \eta \cdot e \cdot x(n), &(19)
\end{aligned}
$$

, em que η é a taxa de aprendizado, e denota o erro, dado pela diferença entre os valores previstos e desejados, isto é, $e = (y_d - y)$, e $x(n)$ corresponde à entrada para a qual a respectiva previsão deu-se de maneira incorreta, ajudando a determinar a direção correta do ajuste do vetor de pesos. De maneira geométrica, a Regra Delta pode ser interpretada como mostrado na Figura 4.

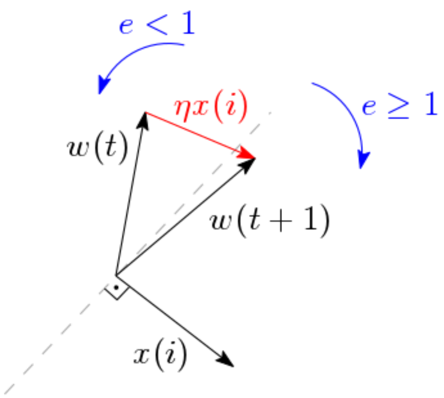

Figura 4: Interpretação geométrica da Regra Delta.

Visto que o neurônio em questão é incapaz de prever corretamente a saída para a entrada $x(2)$, segue-se então o processo de Aprendizado Supervisionado.

5. Ilustração do Processo de Aprendizado

O algoritmo de treinamento do Perceptron é composto pelos seguintes passos:

1. Inicialize η;
2. Inicialize o vetor de pesos $w(0)$ com valores aleatórios;
3. Aplique a regra de atualização dos pesos:

$$w(n+1) = w(n) + \eta \cdot e \cdot x(n)$$

   para todos os p pares $(x(i), y_d^i)$ do conjunto de treinamento Γ;
4. Repita o passo anterior até que $e = 0$ para todos os p elementos de Γ.

Denomina-se e enumera-se em épocas a cada vez que o Passo 3 é aplicado, sem que a condição de parada estabelecida pelo Passo 4 seja alcançada. Assim, o processo de aprendizado deste neurônio será organizado sob a forma de épocas.

5.1 Primeira Época

Usando o par $(x(1), y_d^1)$:

$$\begin{aligned} u &= \sum_{i=0}^{2} x_i w_i && (20) \\ &= -1 \cdot (-0{,}5441) + 2 \cdot 0{,}5562 + 2 \cdot (-0{,}4074) && (21) \\ &= 0{,}5441 + 1{,}1124 - 0{,}8148 && (22) \\ &= 0{,}8417. && (23) \end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$\begin{aligned} y &= f(u) && (24) \\ &= f(0{,}8417) && (25) \\ &= 1. && (26) \end{aligned}$$

Observa-se que o valor y previsto pelo neurônio é igual ao valor de y_d^1. Em consequência, $e = (y_d^1 - y) = (1 - 1)$ e, portanto, não há necessidade de ajuste, o que é verificado quando aplica-se a Regra Delta:

$$\begin{aligned}
w(1) &= w(0) + \eta \cdot e \cdot x(1) = \begin{bmatrix} -0{,}5441 \\ 0{,}5562 \\ -0{,}4074 \end{bmatrix} + 0{,}1 \cdot \cancel{(1-1)}^{\,0} \cdot \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \quad (27)\\
&= \begin{bmatrix} -0{,}5441 \\ 0{,}5562 \\ -0{,}4074 \end{bmatrix} \quad (28)
\end{aligned}$$

De modo análogo, com o próximo exemplo $(x(2), y_d^2)$ disponível na base de dados:

$$\begin{aligned}
u &= \sum_{i=0}^{2} x_i w_i \quad (29)\\
&= -1 \cdot (-0{,}5441) + 4 \cdot 0{,}5562 + 4 \cdot (-0{,}4074) \quad (30)\\
&= 0{,}5441 + 2{,}2248 - 1{,}6296 \quad (31)\\
&= 1{,}1459. \quad (32)
\end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$\begin{aligned}
y &= f(u) \quad (33)\\
&= f(1{,}1459) \quad (34)\\
&= 1. \quad (35)
\end{aligned}$$

Note que houve erro, pois $y \neq y_d^2$. Assim, há necessidade de ajuste no vetor de pesos, o qual é dado por:

$$\begin{aligned}
w(2) &= w(1) + \eta \cdot e \cdot x(2) = \begin{bmatrix} -0{,}5441 \\ 0{,}5562 \\ -0{,}4071 \end{bmatrix} + 0{,}1 \cdot (0-1) \cdot \begin{bmatrix} -1 \\ 4 \\ 4 \end{bmatrix} \quad (36)\\
&= \begin{bmatrix} -0{,}5441 \\ 0{,}5562 \\ -0{,}4071 \end{bmatrix} + \begin{bmatrix} +0{,}1 \\ -0{,}4 \\ -0{,}4 \end{bmatrix} = \begin{bmatrix} -0{,}4441 \\ 0{,}1562 \\ -0{,}8074 \end{bmatrix} \quad (37)
\end{aligned}$$

Todos os exemplos de Γ foram apresentados ao neurônio. Conclui-se, portanto, a primeira época.

5.2 Segunda Época

Dando prosseguimento ao treinamento, usando o par $(x(1), y_d^1)$ e os pesos de $w(2)$:

$$\begin{aligned} u &= \sum_{i=0}^{2} x_i w_i && (38)\\ &= -1 \cdot (-0{,}4441) + 2 \cdot 0{,}1562 + 2 \cdot (-0{,}8074) && (39)\\ &= 0{,}4441 + 0{,}3124 - 1{,}6148 && (40)\\ &= -0{,}8614. && (41) \end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$y = f(-0{,}8614) = 0. \quad (42)$$

Como $y \neq y_d^1$, tem-se o ajuste de pesos:

$$w(3) = w(2) + \eta \cdot e \cdot x(1) = \begin{bmatrix} -0{,}4441 \\ 0{,}1562 \\ -0{,}8074 \end{bmatrix} + 0{,}1 \cdot (1-0) \cdot \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \quad (43)$$

$$= \begin{bmatrix} -0{,}4441 \\ 0{,}1562 \\ -0{,}8074 \end{bmatrix} + \begin{bmatrix} -0{,}1 \\ 0{,}2 \\ 0{,}2 \end{bmatrix} = \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix} \quad (44)$$

Agora usando o exemplo $(x(2), y_d^2)$:

$$\begin{aligned} u &= \sum_{i=0}^{2} x_i w_i && (45)\\ &= -1 \cdot (-0{,}5441) + 4 \cdot 0{,}3562 + 4 \cdot (-0{,}6074) && (46)\\ &= 0{,}5441 + 1{,}4248 - 2{,}4296 && (47)\\ &= -0{,}4607. && (48) \end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$y = f(-0{,}4607) = 0. \quad (49)$$

Não houve erro, uma vez que $y = y_d^2$. Isto será refletido no vetor de pesos $w(4)$, o qual não sofrerá atualizações quando comparado ao vetor de pesos $w(3)$:

$$
\begin{aligned}
w(4) &= w(3) + \eta \cdot e \cdot x(2) && (50)\\
&= \begin{bmatrix} -0{,}4441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix} + 0{,}1 \cdot \cancelto{0}{(0-0)} \cdot \begin{bmatrix} -1 \\ 4 \\ 4 \end{bmatrix} && (51)\\
&= \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix} && (52)
\end{aligned}
$$

Todos os exemplos de Γ foram apresentados ao neurônio. Conclui-se, portanto, a segunda época.

5.3 Terceira Época

Usando o par $(x(1), y_d^1)$ e os pesos de $w(4)$:

$$
\begin{aligned}
u &= \sum_{i=0}^{2} x_i w_i && (53)\\
&= -1 \cdot (-0{,}5441) + 2 \cdot 0{,}3562 + 2 \cdot (-0{,}6074) && (54)\\
&= 0{,}5441 + 0{,}7124 - 1{,}2148 && (55)\\
&= 0{,}0417. && (56)
\end{aligned}
$$

Aplicando-se a função de ativação, tem-se:

$$y = f(0{,}0417) = 1. \qquad (57)$$

Note que não houve erro. Assim:

$$
\begin{aligned}
w(5) &= w(4) + \eta \cdot e \cdot x(1) && (58)\\
&= \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix} + 0{,}1 \cdot \cancelto{0}{(1-1)} \cdot \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} && (59)\\
&= \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix}. && (60)
\end{aligned}
$$

Seguindo para o próximo exemplo $(x(2), y_d^2)$, tem-se:

$$\begin{aligned} u &= \sum_{i=0}^{2} x_i w_i && (61)\\ &= -1 \cdot (-0{,}5441) + 4 \cdot 0{,}3562 + 4 \cdot (-0{,}6074) && (62)\\ &= 0{,}5441 + 1{,}4248 - 2{,}4296 && (63)\\ &= -0{,}4607 && (64)\\ & && (65)\end{aligned}$$

Aplicando-se a função de ativação, tem-se:

$$y = f(-0{,}4607) = 0. \tag{66}$$

Note que também não houve erro. Assim, tem-se:

$$\begin{aligned} w(6) &= w(5) + \eta \cdot e \cdot x(2) && (67)\\ &= \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix} + 0{,}1 \cdot \cancel{(0-0)}^{\,0} \cdot \begin{bmatrix} -1 \\ 4 \\ 4 \end{bmatrix} && (68)\\ &= \begin{bmatrix} -0{,}5441 \\ 0{,}3562 \\ -0{,}6074 \end{bmatrix}. && (69)\end{aligned}$$

Observe que $w(4) = w(5) = w(6)$ e, em consequência, o Passo 4 do algoritmo de treinamento do neurônio perceptron foi satisfeito. Encerra-se então o treinamento tendo o vetor de pesos $w(5)$ como solução para o problema proposto.

6. Interpretação do Resultado

Ao final do treinamento, identificou-se o vetor $w(6)$ como sendo capaz de prover a saída correta para todos os exemplos em Γ. Na Figura 1 foi possível obter a visualização de que o problema era linearmente separável, mas qual a relação dos pesos obtidos com a reta que separa as classes do problema?

A resposta para esta pergunta é obtida por meio de uma reta que separa as classes do problema advinda dos pesos do neurônio, a qual é expressa da seguinte forma:

$$x_2 = (w_0/w_2) - (w_1/w_2)x_1. \tag{70}$$

No exemplo em questão, tem-se:

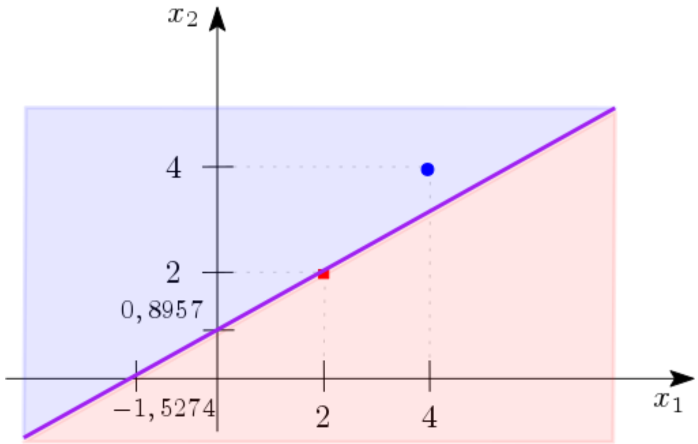

Figura 5: Resolução do problema linearmente separável proposto.

$$
\begin{aligned}
x_2 &= \left(\frac{-0{,}5441}{-0{,}6074}\right) - \left(\frac{0{,}3562}{-0{,}6074}\right) \cdot x_1 && (71)\\
&= 0{,}8957 + 0{,}5864 \cdot x_1. && (72)
\end{aligned}
$$

Ao traçar esta reta, denotada na Figura 5 na cor roxa, vê-se claramente as regiões de ativação e não-ativação do neurônio para o problema.

Referências

[1] DE PáDUA BRAGA, A.; DE LEON F. DE CARVALHO, A. P.; LUDERMIR, T. B. *Redes neurais artificias: Teoria e aplicações*. 2. ed. Rio de Janeiro, Brasil: LTC, 2012.

[2] ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, Washington, v. 6, n. 65, p. 386–408, 1958.

[3] MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 127–147, 1943.