

Uma Função Ativação para Redes Neurais Artificiais Mais Flexível e Poderosa e Mais Rápida

Adilson Elias Xavier

Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ, 21945-970

adilson@cos.ufrj.br

Abstract

Uma nova função de ativação para redes neurais é apresentada. Além de possuir a propriedade de ser completamente diferenciável, a existência de dois parâmetros oferece maiores flexibilidade e potencialidade a essa função, que ademais, é computacionalmente mais rápida que a função logística, a mais amplamente utilizada.

1 Introdução

Uma rede neural é formada por um conjunto de neurônios interligados. Assim, o funcionamento de um neurônio constitui um elemento fundamental na operação dessas redes.

Haykin (1994), por exemplo, descreve muito claramente um modelo para um neurônio, vide figura 1.

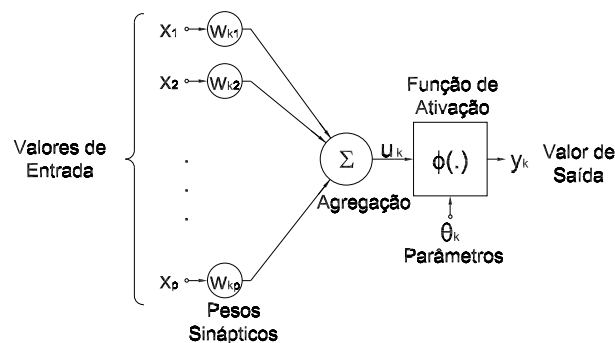


Figure 1: Modelo de um Neurônio

Dado um vetor de entradas $x = (x_1, x_2, \dots, x_p)$, a resposta y_k , saída de um neurônio genérico k , é calculada segundo três procedimentos básicos:

- Um conjunto de pesos associados a esse neurônio k , denominados também por sinapses, $w_k = (w_{k1}, w_{k2}, \dots, w_{kp})$, é aplicado às componentes de entrada;
- A seguir, é feita a agregação das componentes ponderadas pelas sinapses;
- Finalmente, esse valor agregado, transladado por um parâmetro θ_k , é fornecido a uma função de ativação $\varphi(\cdot)$, que produz a resposta do neurônio.

Tais procedimentos compreendem a seguinte seqüência de cálculos:

$$u_k = \sum_{j=1}^p w_{kj} x_j$$

$$v_k = u_k - \theta_k$$

$$y_k = \varphi(v_k)$$

As funções de ativação $\varphi(\cdot)$ tipicamente fornecem a resposta dentro do intervalo $[0,1]$, que, sem perda de generalidade, será tomado, neste trabalho, como referencial.

Abaixo, são apresentadas as principais alternativas de funções de ativação e suas correspondentes derivadas.

1. Função Degrau

$$\varphi_1(v) = \begin{cases} 1, & \text{se } v \geq 0. \\ 0, & \text{se } v < 0. \end{cases}$$

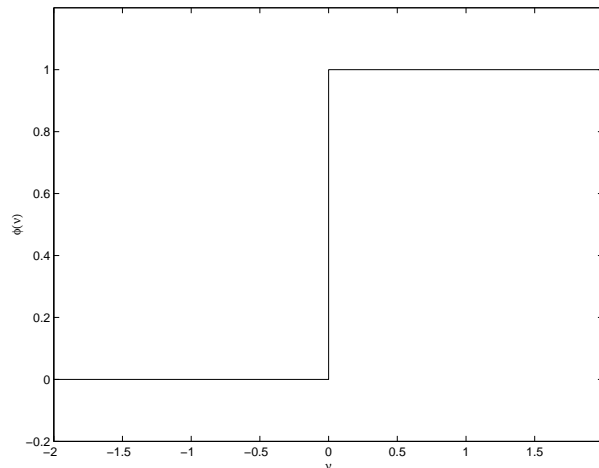


Figure 2: Função Degrau

A sua derivada é $\varphi'_1(v) = 0$ para $\forall v \neq 0$ e não é definida em $v = 0$.

A descontinuidade na origem e o valor nulo da derivada nos demais pontos praticamente anulam a utilidade prática dessa função $\varphi_1(\cdot)$.

2. Função Patamar

$$\varphi_2(v, b) = \begin{cases} 0, & \text{se } v \leq -b; \\ (v + b) / 2b, & \text{se } -b < v < b; \\ 1, & \text{se } v > b. \end{cases}$$

sendo $b = 1/(2 \tan \alpha)$, onde α é o ângulo de inclinação.

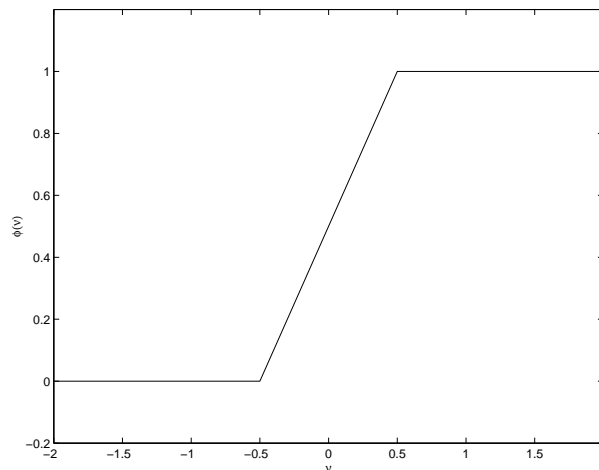


Figure 3: Função Patamar

A sua derivada não é definida nos pontos $v = -1/2$ e $v = 1/2$, e nos demais valores assume:

$$\varphi'_2(v, b) = \begin{cases} 0, & \text{para } v < -b; \\ 1/2b, & \text{para } -b < v < b; \\ 0, & \text{para } v > b. \end{cases}$$

A insensibilidade da derivada fora do intervalo $(-b, b)$ limita consideravelmente o uso prático dessa função de ativação $\varphi_2(\cdot)$.

3. Função Logística

A função de ativação mais amplamente utilizada é a função logística.

$$\varphi_3(v, a) = \frac{1}{1 + e^{-av}}$$

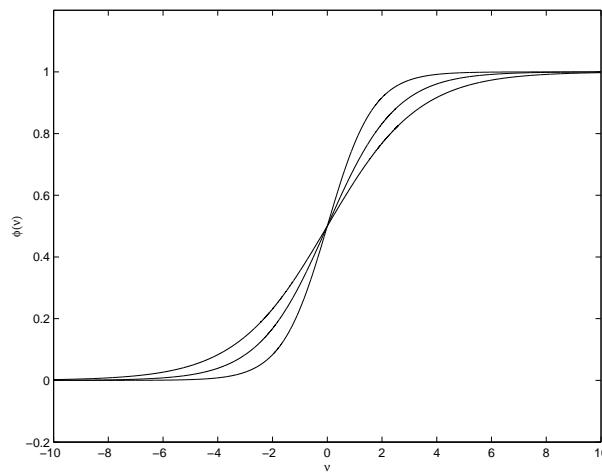


Figure 4: Função Logística - Efeito da variação do parâmetro a

$$\varphi'_3(v, a) = a \varphi_3(v, a) (1 - \varphi_3(v, a))$$

A função tangente hiperbólica, também muito utilizada como função de ativação, corresponde unicamente a um deslocamento da função logística.

A função logística oferece a importante flexibilidade dada por sua inclinação na origem, $\varphi'_3(0, a) = a/4$, ser variável com o parâmetro a , conforme ilustra a figura 4.

Além disso, a função logística apresenta propriedades de simetria e de completa diferenciabilidade, ou seja, pertence à classe de funções C^∞ .

4. Função Ativação de Elliot (1993)

$$\varphi_4(v) = \left(\frac{v}{1 + |v|} + 1 \right) / 2$$

$$\varphi'_4(v) = \frac{1}{2(1 + |v|)^2}$$

A função ativação de Elliot apresenta a inclinação de sua derivada na origem invariante, $\varphi'_4(0) = 1/2$, independente de qualquer transformação de escala, fato que limita fortemente a flexibilidade dessa função e seu decorrente uso prático.

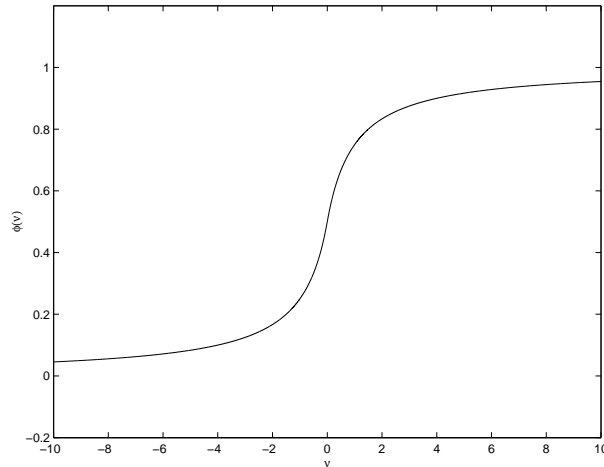


Figure 5: Função de Elliot

2 Função de Ativação Proposta

A função de ativação proposta e sua derivada, na sua forma mais geral, assumem as seguintes expressões:

$$\phi_X(v, \lambda, \tau_1, \tau_2) = \sqrt{\lambda^2 \left(v + \frac{1}{4\lambda}\right)^2 + \tau_1^2} - \sqrt{\lambda^2 \left(v - \frac{1}{4\lambda}\right)^2 + \tau_2^2} + \frac{1}{2}$$

$$\phi'_X(v, \lambda, \tau_1, \tau_2) = \frac{\lambda^2(v + 1/4\lambda)}{\sqrt{\lambda^2(v + 1/4\lambda)^2 + \tau_1^2}} - \frac{\lambda^2(v - 1/4\lambda)}{\sqrt{\lambda^2(v - 1/4\lambda)^2 + \tau_2^2}}$$

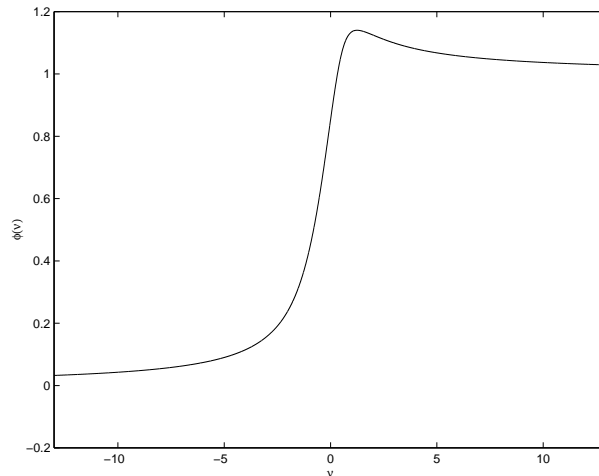


Figure 6: Função Proposta

A função $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ apresenta a desejada propriedade de possuir diferenciabilidade infinita, ou seja, pertence à classe de funções C^∞ .

A função $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ apresenta as seguintes propriedades triviais, consentâneas as demais funções de ativação:

$$\lim_{v \rightarrow -\infty} \phi_X(v, \lambda, \tau_1, \tau_2) = 0$$

$$\lim_{v \rightarrow \infty} \phi_X(v, \lambda, \tau_1, \tau_2) = 1$$

$$\lim_{v \rightarrow -\infty} \phi'_X(v, \lambda, \tau_1, \tau_2) = 0$$

$$\lim_{v \rightarrow \infty} \phi'_X(v, \lambda, \tau_1, \tau_2) = 0$$

Diversamente das funções de ativação ortodoxas, quando $\tau_1 \neq \tau_2$, a função $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ apresenta as diferenciadas propriedades de não oferecer simetria, bem como, de possuir uma imagem que ultrapassa o tradicional espaço definido pelo intervalo $[0, 1]$. Assim, quando $\tau_1 > \tau_2$, o valor da função $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ ultrapassa o valor 1, situação retratada pelo gráfico da figura 6. Inversamente, quando $\tau_1 < \tau_2$, o valor da função $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ assume valores menores do que 0.

Em sua forma mais particular, igualando os valores dos parâmetros $\tau_1 = \tau_2 = \tau$, a função, $\phi_X(\cdot, \lambda, \tau_1, \tau_2) \triangleq \varphi_X(\cdot, \lambda, \tau)$, assume uma forma mais consentânea a outras funções de ativação, tendo imagem no intervalo $[0, 1]$ e oferecendo a propriedade de simetria, conforme retratado pelos gráficos das figuras 7 e 8.

$$\varphi_X(v, \lambda, \tau) = \sqrt{\lambda^2 \left(v + \frac{1}{4\lambda}\right)^2 + \tau^2} - \sqrt{\lambda^2 \left(v - \frac{1}{4\lambda}\right)^2 + \tau^2} + \frac{1}{2}$$

$$\varphi'_X(v, \lambda, \tau) = \frac{\lambda^2(v + 1/4\lambda)}{\sqrt{\lambda^2(v + 1/4\lambda)^2 + \tau^2}} - \frac{\lambda^2(v - 1/4\lambda)}{\sqrt{\lambda^2(v - 1/4\lambda)^2 + \tau^2}}$$

A função $[\varphi_X(v, \lambda, \tau) - 1/2]$ é anti-simétrica, ou seja:

$$\varphi_X(v, \lambda, \tau) - 1/2 = -[\varphi_X(-v, \lambda, \tau) - 1/2]$$

No ponto $v = 0$, são observados os seguintes valores para a função φ_X e sua derivada:

$$\varphi_X(0, \lambda, \tau) = 1/2,$$

$$\varphi'_X(0, \lambda, \tau) = \frac{\lambda}{2\sqrt{1/16 + \tau^2}},$$

$$\lim_{\tau \rightarrow 0} \varphi'_X(0, \lambda, \tau) = 2\lambda.$$

Na figura 7, são mostrados gráficos para a forma simétrica da função proposta correspondentes a três diferentes valores para o parâmetro λ , mantendo-se o parâmetro τ constante. Pode-se ver um efeito similar àquele produzido pela variação do parâmetro a na função logística, conforme apresentado pela figura 4. Dessa forma, pode-se associar o parâmetro λ à inclinação da função na origem.

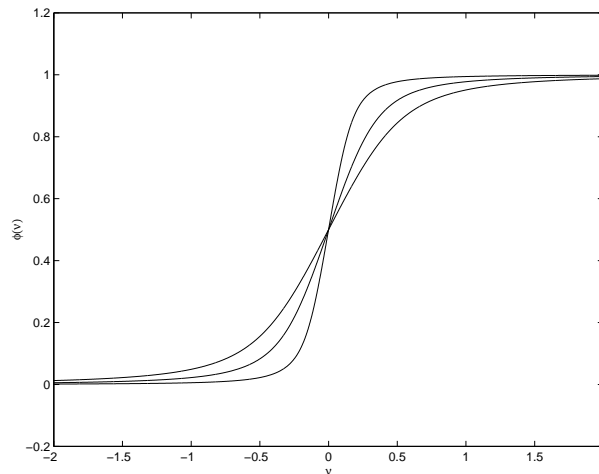
A função $\varphi_X(\cdot, \lambda, \tau)$ apresenta, ademais, os seguintes comportamentos assintóticos:

$$\lim_{\lambda \rightarrow \infty} \varphi_X(v, \lambda, \tau) = \varphi_1(v)$$

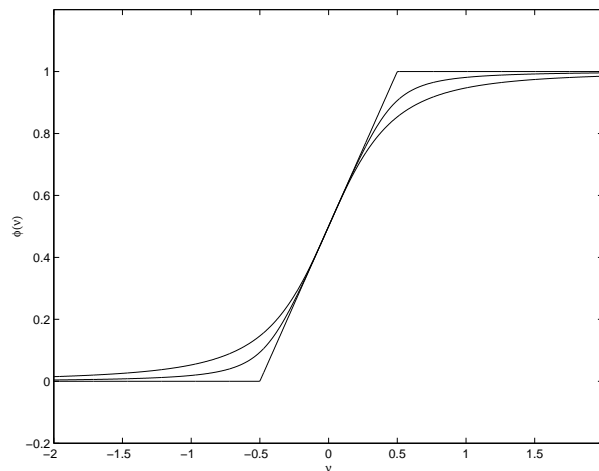
$$\lim_{\tau \rightarrow 0} \varphi_X(v, \lambda, \tau) = \varphi_2\left(v, \frac{1}{4\lambda}\right)$$

Podemos assim aproximar a função de ativação patamar $\varphi_2(\cdot, \cdot)$ por uma forma continuamente diferenciável usando a função $\varphi_X(\cdot, \cdot, \cdot)$, podendo essa aproximação ser tão próxima quanto se queira. Adicionalmente, pode-se efetuar essa aproximação assintótica mantendo constante uma inclinação com especificado ângulo α na origem, $v = 0$, bastando para isso se fazer a reparametrização

$$\lambda = 4\lambda' \sqrt{1/16 + \tau^2}$$

Figure 7: Função φ_X – Efeito da Variação do Parâmetro λ

sendo $\lambda' = \frac{1}{2} \tan \alpha$. Na figura 8, são mostrados 2 gráficos da função proposta com uma inclinação constante na origem. Aí pode-se ver, que o parâmetro τ está associado com o afastamento da curva às duas assíntotas horizontais.

Figure 8: Função φ_X – Efeito da Variação do Parâmetro τ

Em termos computacionais, através do cálculo de 100.000.000 valores aleatoriamente gerados, pode ser observado que a função $\phi_X(\cdot, \cdot, \cdot, \cdot)$, é 24,8 % mais rápida que a função logística. Esse resultado foi obtido através de simulação programada na linguagem FORTRAN, compilador COMPAQ, computador tipo IBM-PC, com 800Mhz de velocidade.

A existência de dois parâmetros, um a mais que as demais funções de ativação, enseja a essa função uma maior flexibilidade para representar mais adequadamente os fenômenos modelados com redes neurais.

Numa rede neural multicamadas, por exemplo, essa maior flexibilidade engendra, certamente à função de ativação ora proposta o poder de aproximar qualquer função de uma forma mais sintética, com menor número de neurônios.

Através da manipulação conveniente de seus parâmetros, a função $\varphi_X(\cdot, \lambda, \tau)$, oferece ademais a possibilidade de poder enfrentar mais convenientemente o desastroso fenômeno de saturação, que não raramente ocorre no treinamento de redes neurais, (vide, por exemplo, Looney (1997), pag 161 ou Haykin (1994), pag 165).

Igualmente, uma manipulação adequada dos parâmetros de $\varphi_X(\cdot, \lambda, \tau)$, em tese, pode ser usada para se evitar um indesejado mínimo local.

Em relação às questões de saturação e de mínimo local, acima referidas, a manipulação dos parâmetros τ_1 e τ_2 oferece à função assimétrica $\phi_X(\cdot, \lambda, \tau_1, \tau_2)$ possibilidades adicionais para enfrentá-las.

Referências

Elliot D. L., (1993), "A Better Activation Function for Artificial Neural Networks", Institute for Systems Research, Technical Report TR93-8

Haykin S., (1994), "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company

Looney, C.G. (1997), "Pattern Recognition Using Neural Networks", Oxford University Press