

penúltima camada, e conseqüentemente as modificações dos pesos de todas as conexões que a alimentam. A computação recursiva continua, camada por camada, propagando as modificações para todos os pesos sinápticos da rede.

Note que para a apresentação de cada exemplo de treinamento, o padrão de entrada está fixo ("preso") durante todo o ciclo, englobando o passo de propagação seguido pelo passo de retropropagação.

Função de Ativação

O cálculo do δ para cada neurônio do perceptron de múltiplas camadas requer o conhecimento da derivada da função de ativação $\varphi(\cdot)$ associada àquele neurônio. Para esta derivação existir, necessitamos que a função $\varphi(\cdot)$ seja contínua. Em termos básicos, a *diferenciabilidade* é a única exigência que a função de ativação deve satisfazer. Um exemplo de uma função de ativação não-linear, continuamente diferenciável normalmente utilizada nos perceptrons de múltiplas camadas é a *não-linearidade sigmóide*; descrevemos duas formas desta função:

1. *Função Logística*. Esta forma de não-linearidade sigmóide na sua forma geral é definida por

$$\varphi_j(v_j(n)) = \frac{1}{1 + \exp(-av_j(n))} \quad a > 0 \quad \text{e} \quad -\infty < v_j(n) < \infty \quad (4.30)$$

onde $v_j(n)$ é o campo local induzido do neurônio j . De acordo com esta não-linearidade, a amplitude da saída se encontra dentro do intervalo $0 \leq y_j \leq 1$. Diferenciando a Eq. (4.30) em relação a $v_j(n)$, obtemos

$$\varphi'_j(v_j(n)) = \frac{a \exp(-av_j(n))}{[1 + \exp(-av_j(n))]^2} \quad (4.31)$$

Com $y_j(n) = \varphi_j(v_j(n))$, podemos eliminar o termo exponencial $\exp(-av_j(n))$ da Eq. (4.31), e assim expressar a derivada $\varphi'_j(v_j(n))$ como

$$\varphi'_j(v_j(n)) = a y_j(n) [1 - y_j(n)] \quad (4.32)$$

Para um neurônio j localizado na camada de saída, $y_j(n) = o_j(n)$. Assim, podemos expressar o gradiente local para o neurônio j como

$$\begin{aligned} \delta_j(n) &= e_j(n) \varphi'_j(v_j(n)) \\ &= a [d_j(n) - o_j(n)] o_j(n) [1 - o_j(n)], \quad \text{o neurônio } j \text{ é um nó de saída} \end{aligned} \quad (4.33)$$

onde $o_j(n)$ é o sinal funcional na saída do neurônio j , e $d_j(n)$ é a resposta desejada para ele. Por outro lado, para um neurônio oculto arbitrário j , podemos expressar o gradiente local como

$$\begin{aligned} \delta_j(n) &= \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \\ &= a y_j(n) [1 - y_j(n)] \sum_k \delta_k(n) w_{kj}(n), \quad \text{o neurônio } j \text{ é oculto} \end{aligned} \quad (4.34)$$

Note da Eq. (4.32) que a derivada $\varphi'_j(v_j(n))$ alcança o seu valor máximo em $y_j(n) = 0,5$, e o seu valor mínimo (zero) em $y_j(n) = 0$, ou $y_j(n) = 1,0$. Como o valor da modificação do peso sináptico da rede

$$f(x) = \frac{1}{1 + e^{-x}}$$

Abordagem 01: regra do quociente $y = \frac{g(x)}{h(x)}$

$$\text{então } \frac{dy}{dx} = \frac{h(x) \cdot g'(x) - g(x) h'(x)}{(h(x))^2}$$

No nosso caso:

$$\rightarrow g(x) = 1 \Rightarrow g'(x) = 0$$

$$\rightarrow h(x) = 1 + e^{-x} \Rightarrow h'(x) = -e^{-x}$$

Substituindo:

$$\frac{dy}{dx} = \frac{(1 + e^{-x}) \cdot 0 - 1 \cdot (-e^{-x})}{(1 + e^{-x})^2}$$

$$\frac{dy}{dx} = \frac{0 + e^{-x}}{(1 + e^{-x})^2}$$

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \quad \text{h.}$$

Abordagem 2: Considerando $f(x) = (1 + e^{-x})^{-1}$

Nesse caso, usar regra da cadeia

$$\frac{dy}{dx} = -1 \cdot (1 + e^{-x})^{-2} \cdot [(1 + e^{-x})]'$$

$$\frac{dy}{dx} = -1 \cdot (1 + e^{-x})^{-2} \cdot [0 - e^{-x}]$$

$$\frac{dy}{dx} = \frac{-1 \cdot (-e^{-x})}{(1 + e^{-x})^2}$$

$$\frac{dy}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2}$$

Conclusão:

tanto pela abordagem 1 como 2,

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

$$\text{para } f(x) = \frac{1}{1 + e^{-x}}$$

* Seguindo:

se definirmos $f(x) = \frac{1}{1 + e^{-x}}$, podemos

reescrever $f(x)$ usando $f'(x)$

Assim:

$$f'(x) = f(x) * (1 - f(x))$$

* Prova

$$f'(x) = \frac{1}{(1 + e^{-x})} * \left(1 - \frac{1}{1 + e^{-x}} \right)$$

$$= \frac{1}{(1 + e^{-x})} * \left(\frac{\cancel{1 + e^{-x}} - \cancel{1}}{1 + e^{-x}} \right)$$

$$= \frac{1 * e^{-x}}{(1 + e^{-x})^2}$$

$$= \frac{e^{-x}}{(1 + e^{-x})^2}$$

• Como chegar então nesta relação?

Sabemos que:

$$f(x) = \frac{1}{1 + \exp^{-x}}$$

$$\text{e } f'(x) = \frac{\exp^{-x}}{(1 + \exp^{-x})^2}$$

* Vamos obter $f'(x)$:

$$f'(x) = \frac{1 \cdot \exp^{-x} \cdot 1}{(1 + \exp^{-x}) \cdot (1 + \exp^{-x})}$$

e $y = f(x)$, logo:

$$f(x) = y \cdot \exp^{-x} \cdot y$$

$$f(x) = y^2 \cdot \exp^{-x}$$

* Como $y = f(x)$ então $y = \frac{1}{1 + \exp^{-x}}$, assim

$$y \cdot (1 + \exp^{-x}) = 1 \Rightarrow y + y \exp^{-x} = 1$$

* Isolando o \exp^{-x} :

$$\exp^{-x} = \frac{1 - y}{y}$$

* Substituindo $\exp^{-x} = \frac{1-y}{y}$ em $f'(x)$, temos:

$$f'(x) = y^2 \cdot \exp^{-x}$$

$$f'(x) = y^2 \cdot \frac{(1-y)}{y}$$

$$\boxed{f'(x) = y \cdot (1-y)} \quad *$$

* Lembrando que: $y = f(x)$

$$f'(x) = f(x) \cdot (1 - f(x))$$

✓