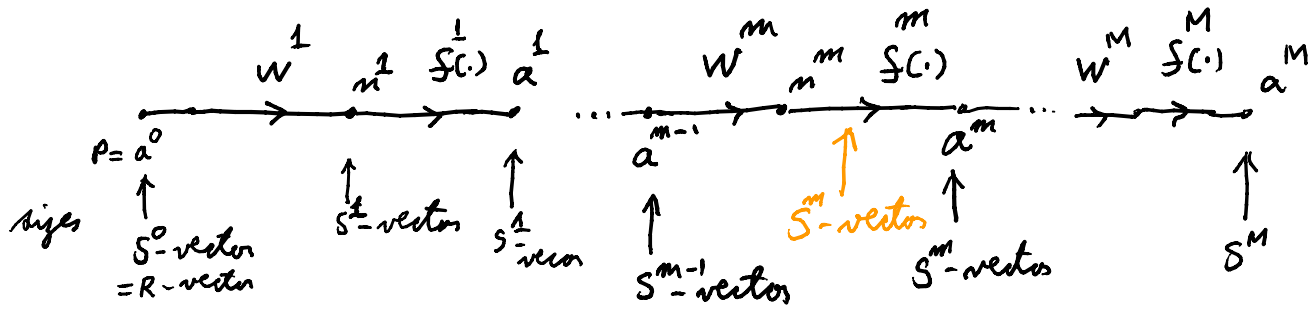


pp. 11-7 - 11-25 main results 11-13, 14

EE 434
02/07/07

p. 11-8



examples

W is an $S^m \times S^{m-1}$ matrix $m = \{0, 1, \dots, M\}$

$$[p_0, t_0], [p_1, t_1], \dots, [p_Q, t_Q]$$

k indexes the k th training time:

Measure of "goodness" $\Rightarrow$ how close are actual outputs a_i^M versus desired outputs t_i for examples input p_i .

$$e(k) = t(k) - a^M(k) ; \quad e^T(k) \cdot e(k) \text{ is a "quadratic form" = sum of squares}$$

$$\hat{F} = e^T(k) e(k) \quad \text{desire this to be zero by changing weights}$$

End result (back propagation): p. 11-14; $\alpha = \text{learning rate}$
 $0 < \alpha \leq 1$

$$W^m(k+1) = W^m(k) - \alpha \Delta^m(k) (a^{m-1}(k))^T$$

$\Delta^m(k) = \text{sensitivity vector at } m^{\text{th}} \text{ layer at } k^{\text{th}} \text{ time.}$

sizes

$$S^m \times S^{m-1}$$

column vector $\uparrow$ row vector $\uparrow$

$$S^m\text{-vector} \quad S^{m-1}\text{-vector} \Rightarrow \begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} \quad \end{bmatrix}$$

Δ^m from (11.35)

$$\Delta^m = \underbrace{\Delta^m(n^m)}_{S^m \times S^m \text{ matrix}} \cdot \underbrace{(W^{m+1})^T}_{S^m \times S^{m+1} \text{ matrix}} \cdot \underbrace{\Delta^{m+1}}_{\text{an } S^{m+1}\text{-vector}}$$

$m = 1, \dots, M-1$
gives $\Delta^1, \dots, \Delta^M$

$$\dot{F}^m(n^m) = \text{diag} \left[\frac{dF_i^m(n_i^m)}{dn_i} \right]$$

Ex: where $f(x) = \tanh(x)$;

$$\frac{d \tanh x}{dx} = \frac{e^x + e^{-x}}{e^x + e^x} - \frac{(e^x - e^{-x})}{(e^x + e^x)^2} \cdot (e^x - e^{-x})$$
$$= \frac{\sinh(x)}{\cosh(x)} = 1 - \tanh^2(x)$$
$$= \frac{(e^x - e^{-x})/2}{(e^x + e^{-x})/2}$$

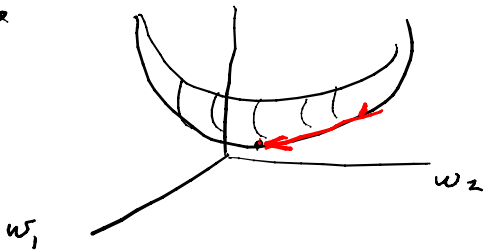
then $\dot{F}^m_{(n^m)} = \text{diag}(1 - (f^m_{(n^m_i)})^2) = \text{diag}([1 - (a^m_i)^2])$

$\underline{\underline{a^M}}$: eq (11.44): $a^M = -2 \dot{F}^M(n^M) \cdot (\underline{t} - \underline{a^M})$ is a S^M -vector
 $\therefore$ know a^M after $\uparrow$ run the network $\uparrow$ S^M -vector
 $\uparrow$ $\uparrow$
 an S^M -vector is an $S^M \times S^M$ matrix



$$W^m(k+1) = W^m(k) - \underbrace{\alpha}_{\text{learning rate}} \underbrace{A^m(k)}_{\text{error}} \underbrace{\left(a^{m-1}(k) \right)^T}_{\text{input}}$$

here



as biases all have 1's as inputs

$$a^{m-1}(n)^T \rightarrow 1 \text{ so}$$

$$b^{(m)}(k+1) = b^{(m)}(k) - \alpha \Delta b^{(m)}(k) \quad (11.47)$$

$$\text{If } f(x) = \frac{1}{1-e^{-x}}, \quad \frac{df}{dx} = \left(\frac{-1}{(1-e^{-x})^2} \cdot (-1)(-1)e^{-x} \right) = \frac{-e^{-x}}{(1-e^{-x})^2}$$

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