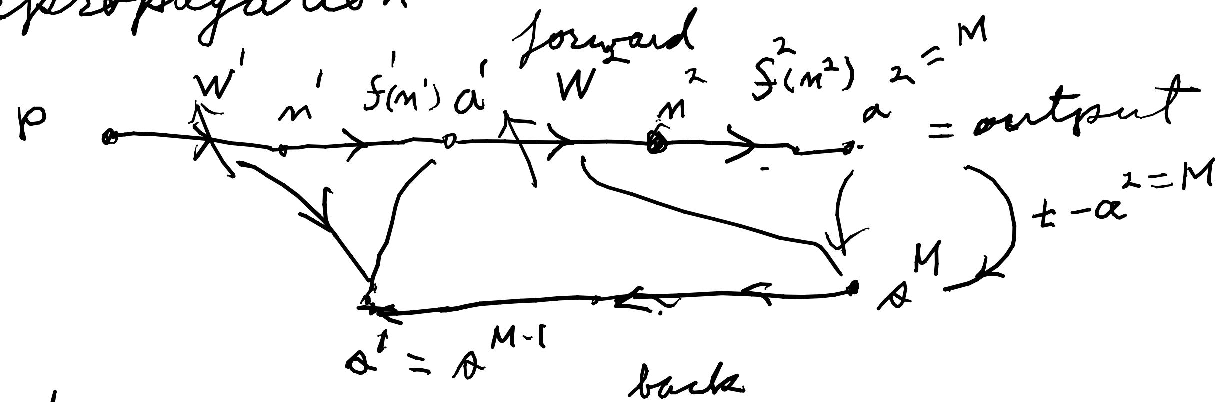
EE 434
02/13/06

Backpropagation



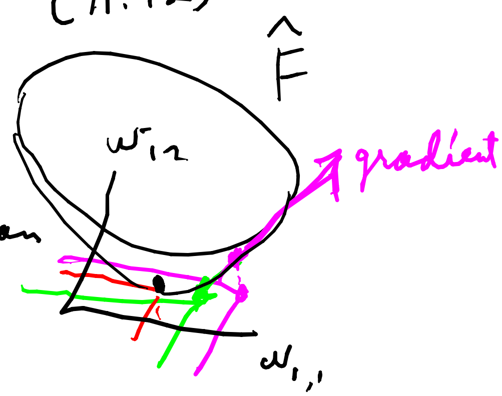
for training

$$\hat{F} = (t - a^M)^T (t - a^M) \qquad (11.12)$$

$$= \|t - a^M\|^2_E$$

E = Euclidian

desire $\min \hat{F}$ over weights



formula for weights

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \nabla \hat{F}_{w_{i,j}^m} \qquad (11.13)$$

$$\nabla \hat{F}_{w_{i,j}^m} = \partial \hat{F} / \partial w_{i,j}^m$$

$m = 1, \ldots, M$

$0 < \alpha < 1$

up to you to choose; convergence factor = learning rate

$$\frac{\partial \hat{F}}{\partial w_{i,j}^M} = \frac{\partial \left(t - f^M\left(\sum_{j=} w_{i,j}^M n_j^{M-1}\right)\right)^T \left(t - f^M\left(\sum_{j=1} w_{i,j}^M n_j^{M-1}\right)\right)}{\partial w_{i,j}^M}$$

can differentiate directly

get (11.27) $W^M(k+1) = W^M(k) - \alpha s^M (a^{M-1})^T$

with $s^M \Rightarrow$ (11.40) $s^M = -2\dot{F}(n^M)(t-a)$

$$\dot{F}(x) = \begin{bmatrix} \frac{df_1^M(x)}{dx} & & \\ & \ddots & \\ & & \frac{d}{dx} f^M_{\#\text{ neurons in Mth layer}}(x) \end{bmatrix}$$

we need $\frac{df(x)}{dx}$'s for useful activation functions

if $f(x) = \text{purelin}(x) = x$, $\frac{df}{dx}(x) = 1$

if $f(x) = \text{tansig}(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

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

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

return to $\frac{\partial \hat{F}}{\partial w_{ij}^m}$ use (11.18)

$$= \frac{\partial \hat{F}}{\partial n_{\cancel{j}\,i}^m} \cdot \frac{\partial n^m_{\cancel{j}\,i}}{\partial w_{i,j}^m} \qquad (11.18)$$

here $\frac{\partial n_j^m}{\partial w_{i,j}^m}$ comes from $n^m = W^m \cdot a^{m-1}$

n^m: vector, W^m: matrix, a^{m-1}: vector

$$n_i^m = \sum_k w_{ik}^m a_k^{m-1} \Longrightarrow \frac{\partial n_i^m}{\partial w_{i,j}^m} = a_j^{m-1} \quad (11.21)$$

need $\dfrac{\partial \hat{F}}{\partial n_i^m} = s_i^m$ (11.22) gives the name "sensitivity"

↑ a vector entry

use (11.35)

$$s_i^m = \frac{\partial \hat{F}}{\partial n_i^m} = \frac{\partial \hat{F}}{\partial n_j^{m+1}} \cdot \frac{\partial n_j^{m+1}}{\partial n_i^m} = s_j^{m+1} \cdot \frac{\partial n_j^{m+1}}{\partial n_i^m}$$

$$s^m = \left[\frac{\partial n_j^{m+1}}{\partial n_i^m}\right] s^{m+1}$$

∴ need

$$\frac{\partial n_i^{m+1}}{\partial n_j^m} \Rightarrow (11.31) \qquad n_i^{m+1} = \sum_{k=} w_{i,k}^{m+1} a_k^m$$

$$= \sum_k w_{i,k}^{m+1} f^m(n_k^m) \quad k=j$$

or

$$\frac{\partial n_i^{m+1}}{\partial n_j^m} = w_{i,j}^{m+1} \frac{d f_j(\cdot)}{d n_j^m} \cdot \underbrace{\frac{\partial n_j^m}{\partial n_j^m}}_{1} \Bigg|_{n_j^m} \quad (11.31)$$

$$= w_{i,j}^{m+1} \frac{d f_j^m(n_j^m)}{d n_j^m}$$

(11.33)

$$\left[\frac{\partial n^{m+1}}{\partial n^m}\right] = W^{m+1} \cdot \begin{bmatrix} df_1^m/dn_1^m & & 0 \\ & \ddots & \\ 0 & & df_{last}^m/dn_{last}^m \end{bmatrix}$$

$$= W^{m+1} \dot{F}^{m}(n^{m})$$

$$s^{m} = \left(W^{m+1} \dot{F}^{m}(n^{m})\right)^{T} s^{m+1} \qquad (11.35)$$

gives the new weights

$$\boxed{W^{m}(k+1) = W^{m}(k) - \alpha\, s^{m} (a^{m-1})^{T}} \qquad (11.46)$$

$$s^{M} = -2\, \dot{F}^{M}(n^{M})(t - a^{M}) \qquad (11.40)$$