GRADIENT-BASED OPTIMIZATION (CH. 5.2.4)

$$\underline{w}^{(\tau+1)} = \underline{w}^{(\tau)} - \eta \nabla E(\underline{w}^{(\tau)})$$

$\eta > 0$

LEARNING RATE

GRADIENT DESCENT
STEEPEST DESCENT

SIMPLE CONDITIONS

INVOLVING $\underline{\underline{H}}$

CONJUGATE GRADIENTS
QUASI-NEWTON, ETC.

(MORE INITIALIZATION RUNS)

ON-LINE GRADIENT DESCENT:
(SEQUENTIAL)

$$E(\underline{w}) = \sum_{n=1}^{N} E_n(\underline{w})$$

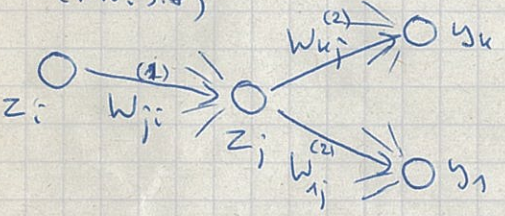
$$\underline{w}^{(\tau+1)} = \underline{w}^{(\tau)} - \eta \nabla E_n(\underline{w}^{(\tau)})$$

ONE DATA POINT AT A TIME

(BETTER ESCAPING FROM LOCAL MINIMA)

BACKPROPAGATION (CH. 5.3) - EFFICIENT GRADIENT COMPUTATION IN FEED-FORWARD NETWORK.

(FIG. 5.7)



$$\frac{\partial E}{\partial w_{kj}} = (y_k - t_k)\,\sigma'(a_k)\cdot z_j$$

$$\underbrace{\frac{\partial E}{\partial y_k}\quad\frac{\partial y_k}{\partial a_k}}_{\delta_k}\quad\frac{\partial a_k}{\partial w_{kj}}$$

$$\cdot z_j$$

$$E_n(\underline{w}) = \frac{1}{2}\sum_k (y_{kn} - t_{kn})^2$$

$$\frac{\partial E}{\partial w_{ji}} = \sum_k (y_k - t_k)\,\sigma'(a_k)\, w_{kj}\,\sigma'(a_j)\cdot z_i$$

$$\sum_k \frac{\partial E}{\partial y_k}\quad\frac{\partial y_k}{\partial a_k}\quad\frac{\partial a_k}{\partial w_{ji}}\quad\frac{\partial z_j}{\partial a_j}\quad\frac{\partial a_j}{\partial w_{ji}}$$

$$\sum_k \delta_k\, w_{kj}\,\sigma'(a_j)\cdot z_i$$

$$\underbrace{\sigma'(a_j)\sum_k w_{kj}\,\delta_k}_{\delta_j}\cdot z_i = \delta_j \cdot z_i$$

$$\boxed{\delta_j = \sigma'(a_j)\sum_k w_{kj}\,\delta_k}$$

DELTA-RULE