Derivation & Operation of Backpropagation of Error

Error function:

$$F = \sum_{p \in P} {}^pE \quad \text{with } p \text{ being pattern in training set } P$$

$${}^pE = \frac{1}{2} \sum_{j \in M} \left(\hat{y}_m - y_m\right)^2 \quad \text{with } \hat{y}_m \text{ being the teacher of } m\text{-th output } y_m$$

Weight function:

$${}^p\triangle\, w_{hm} \sim -\nabla_w \cdot {}^pE \quad \text{with } w_{hm} \text{ being the weight from neuron } h \text{ to neuron } m$$

$$\triangle\, w_{hm} = -\eta \frac{\partial E\left(w_{hm}\right)}{\partial w_{hm}} \quad \text{with } \eta \text{ being the learning rate}$$

For output neurons:

$net_m = \sum_{i=0}^{H} w_{im}\tilde{o}_i$

$o_m = y_m = f_m(net_m)$

$\frac{\partial E(w_{hm})}{\partial w_{hm}} = \frac{\partial E}{\partial net_m} \cdot \frac{\partial net_m}{\partial w_{hm}}$

$$\begin{aligned}
\frac{\partial net_m}{\partial w_{hm}} &= \frac{\partial}{\partial w_{hm}} \cdot net_m \\
&= \frac{\partial}{\partial w_{hm}} \sum_{i=0}^{H} w_{im}\tilde{o}_i \\
&= \sum_{i=0}^{H} \frac{\partial}{\partial w_{hm}} w_{im}\tilde{o}_i \\
&= \frac{\partial}{\partial w_{hm}} \tilde{o}_h w_{hm} \\
&= \boxed{\tilde{o}_h}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E}{\partial net_m} &= \frac{\partial E}{\partial y_m} \cdot \frac{\partial y_m}{\partial net_m} \quad (= -\delta_m) \\
&= \frac{\partial E}{\partial y_m} \cdot \frac{\partial}{\partial net_m} f_m(net_m) \\
&= \underbrace{\frac{\partial E}{\partial y_m} \cdot \boxed{f'_m(net_m)}}_{=:-\delta_m}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E}{\partial y_m} &= \frac{\partial}{\partial y_m} \cdot \frac{1}{2} \sum_{j=1}^{M} \left(\hat{y}_j - y_j\right)^2 \\
&= \boxed{-(\hat{y}_m - y_m)}
\end{aligned}$$

$$\begin{aligned}
\triangle\, w_{hm} &= -\eta \frac{\partial E}{\partial w_{hm}} \\
&= -\eta \frac{\partial E}{\partial y_m} \frac{\partial y_m}{\partial net_m} \frac{\partial net_m}{\partial w_{hm}} \\
&= \eta(\hat{y}_m - y_m) f'_m(net_m)\tilde{o}_h
\end{aligned}$$

$\boxed{\delta_m = (\hat{y}_m - y_m) \cdot f'_m(net_m)}$

$\boxed{\triangle\, w_{hm} = \eta \cdot \delta_m \cdot \tilde{o}_h}$ Widrow-Hoff-Rule / δ-Rule

For hidden neurons:

$net_h = \sum_{i=0}^{H} w_{ih}\tilde{o}_i$

$\frac{\partial E}{\partial w_{kh}} = \frac{\partial E}{\partial net_h} \cdot \frac{\partial net_h}{\partial w_{kh}}$

$$\begin{aligned} \delta_h &= -\frac{\partial E}{\partial net_h} \\ &= -\frac{\partial E}{\partial o_h} \cdot \frac{\partial o_h}{\partial net_h} \end{aligned}$$

$$\begin{aligned} -\frac{\partial E}{\partial o_h} &= -\frac{\partial E\left(\underline{net}_{l=1}, \underline{net}_{l=2}, \ldots, \underline{net}_{l=L}\right)}{\partial o_h} \\ &= \sum_{l=1}^{L}\left(-\frac{\partial E}{\partial \underline{net}_l}\right) \cdot \frac{\partial \underline{net}_l}{\partial o_h} \\ &= \sum_{l=1}^{L} \underline{\delta}_l \cdot \frac{\partial}{\partial o_h} \sum_{j=0}^{H} \underline{w}_{jl} \cdot o_j \\ &= \sum_{l=1}^{L} \underline{\delta}_l \cdot \underline{w}_{hl} \end{aligned}$$

$\boxed{\delta_h = \sum_{l=1}^{L} \left(\underline{\delta}_l \cdot \underline{w}_{hl}\right) \cdot f'\left(net_h\right)}$

$\boxed{\triangle\ w_{kh} = \eta \cdot \delta_h \cdot \tilde{o}_k}$