Solution Set: Artificial Neural Networks

1. a) If j is an output unit k, then $\delta_k = \frac{\partial E_n(\boldsymbol{w})}{\partial b_k}$ where b_k is the activation for unit k.

$$E_n(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{K}(f_k(x_n) - t_{nk})^2$$
$$= \frac{1}{2}\sum_{k=1}^{K}(b_k - t_{nk})^2$$

$\Longrightarrow \quad \frac{\partial E_n(\boldsymbol{w})}{\partial b_k} = \frac{1}{2}\frac{\partial (b_k - t_{nk})^2}{\partial b_k} = b_k - t_{nk} = f_k(x_n) - t_{nk}.$

$\Longrightarrow \quad \delta_k = f_k(x_n) - t_{nk}.$

b) If j is an output unit k, then $\delta_k = \frac{\partial E_n(\boldsymbol{w})}{\partial b_k}$ where b_k is the activation for unit k.

$$E_n(\boldsymbol{w}) = -\sum_{k=1}^{K}[t_{nk}\log f_k(x_n) + (1 - t_{nk})\log(1 - f_k(x_n))]$$
$$= -\sum_{k=1}^{K}[t_{nk}\log \sigma(b_k) + (1 - t_{nk})\log(1 - \sigma(b_k))]$$

$\Longrightarrow \quad \frac{\partial E_n(\boldsymbol{w})}{\partial b_k} = -\frac{\partial [t_{nk}\log \sigma(b_k) + (1 - t_{nk})\log(1 - \sigma(b_k))]}{\partial b_k}$

$$= -[t_{nk}\frac{1}{\sigma(b_k)} \cdot \sigma'(b_k) + (1 - t_{nk}) \cdot \frac{1}{1 - \sigma(b_k)} \cdot (-\sigma'(b_k))]$$

$$= -[t_{nk}\frac{1}{\sigma(b_k)} \cdot \sigma(b_k)(1 - \sigma(b_k)) + (1 - t_{nk}) \cdot \frac{1}{1 - \sigma(b_k)} \cdot (-\sigma(b_k) \cdot (1 - \sigma(b_k))]$$

because $\sigma' = \sigma(1 - \sigma)$

$= -[t_{nk}\big(1 - \sigma(b_k)\big) + (1 - t_{nk})(-\sigma(b_k))]$

$= -[t_{nk} - \sigma(b_k)]$
$= \sigma(b_k) - t_{nk}$
$= f_k(x_n) - t_{nk}$

$\Longrightarrow \quad \delta_k = f_k(x_n) - t_{nk}.$

c) $E_n(\boldsymbol{w}) = -\sum_{k=1}^{K} t_{nk} \log f_k(x_n)$

$$= -\sum_{k=1}^{K} t_{nk} \log\left(\frac{e^{b_k}}{\sum_{i=1}^{K} e^{b_i}}\right)$$

$$\Longrightarrow \qquad \frac{\partial E_n(\boldsymbol{w})}{\partial b_k} = \frac{\partial\left(-\sum_{j=1}^{K} t_{nj} \log\left(\frac{e^{b_j}}{\sum_{i=1}^{K} e^{b_i}}\right)\right)}{\partial b_k}$$

$$= -\sum_{j \neq k} t_{nj} \frac{\partial \log\left(\frac{e^{b_j}}{\sum_{i=1}^{K} e^{b_i}}\right)}{\partial b_k} - t_{nk} \frac{\partial \log\left(\frac{e^{b_k}}{\sum_{i=1}^{K} e^{b_i}}\right)}{\partial b_k}$$

$$= -\sum_{j \neq k} t_{nj} \frac{1}{f_j(x_n)} \cdot \left(-f_k(x_n) f_j(x_n)\right) - t_{nk} \cdot \frac{1}{f_k(x_n)} \cdot f_k(x_n)(1 - f_k(x_n))$$

$$= -\sum_{j \neq k} t_{nj} \left(-f_k(x_n)\right) - t_{nk}(1 - f_k(x_n))$$

$$= f_k(x_n) \sum_{j \neq k} t_{nj} + t_{nk} f_k(x_n) - t_{nk}$$

$$= f_k(x_n) \sum_{j} t_{nj} - t_{nk}$$

$= f_k(x_n) - t_{nk}$ because $\sum_j t_{nj} = 1$

$$\Longrightarrow \qquad \delta_k = f_k(x_n) - t_{nk}$$

2. a) If j is an output unit k, then $\delta_k = \frac{\partial E_n(\boldsymbol{w})}{\partial b_k}$ where b_k is the activation for unit k. From problem 1, we already know $\delta_k = f_k(x_n) - t_{nk}$
$\Longrightarrow \qquad \delta_k = b_k - t_{nk}$.

   If j is a hidden unit, then (by the backpropagation equations),

   $\delta_j = h'(a_j) \sum_k w_{kj} \delta_k$ where k runs over all non-bias units in the layer after the layer for unit j.

   $= h(a_j)(1 - h(a_j)) \sum_k w_{kj} \delta_k$ because $h' = h(1 - h)$.

   b) $\frac{\partial E_n(\boldsymbol{w})}{\partial w_{ji}} = \delta_j z_i$ where z_i is the value for unit i in the input layer.

Since j is a hidden unit, $\delta_j = h(a_j)(1 - h(a_j)) \sum_k w_{kj}\delta_k$.

c) $\frac{\partial E_n(\boldsymbol{w})}{\partial w_{kj}} = \delta_k z_j$ where z_j is the value for unit j in the hidden layer.

Since k is an output unit,

$\delta_k = f_k(x_n) - t_{nk}$

$= b_k - t_{nk}$ where b_k is the activation for unit k.