Problem Set: Artificial Neural Networks

1. In the process of finding $\nabla E(\boldsymbol{w})$, we saw that $\frac{\partial E_n(\boldsymbol{w})}{\partial w_{ji}} = \delta_j z_i$, where $\delta_j \equiv \frac{\partial E_n(\boldsymbol{w})}{\partial a_j}$.

   Claim: If $j$ is an output unit $k$, then $\delta_k = f_k(x_n) - t_{nk}$.

   a) Show the above claim holds for $E_n(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{K}(f_k(x_n) - t_{nk})^2$, corresponding to the sum-of-squares error function, where $f_k(x_n) = b_k$ (the activation for unit $k$).

   b) Show that the claim holds for $E_n(\boldsymbol{w}) = -\sum_{k=1}^{K}[t_{nk}\log f_k(x_n) + (1 - t_{nk})\log(1 - f_k(x_n))]$, corresponding to the cross-entropy error function, where $f_k(x_n) = \sigma(b_k)$ with $\sigma(u) = \frac{1}{1+e^{-u}}$.

   c) Show that the claim holds for $E_n(\boldsymbol{w}) = -\sum_{k=1}^{K} t_{nk}\log f_k(x_n)$, corresponding to the multiclass cross-entropy error function, where $f_k(x_n) = \frac{e^{b_k}}{\sum_{i=1}^{K} e^{b_i}}$.

2. Consider a neural network with a single hidden layer used to solve a regression problem. Suppose the hidden unit activation function $h$ is the logistic sigmoid function $h(u) = \frac{1}{1+e^{-u}}$ and the output unit activation function $g_k$ is given by $g_k(b_1, \ldots, b_K) = b_k$ so that $Y_k = b_k$.

   Let the error function $E(\boldsymbol{w})$ be the sum-of-squares error function $E(\boldsymbol{w}) = \sum_{n=1}^{N} E_n(\boldsymbol{w})$, where $E_n(\boldsymbol{w}) = \sum_{k=1}^{K}(f_k(x_n) - t_{nk})^2$.

   a) Calculate $\delta_j \equiv \frac{\partial E_n(\boldsymbol{w})}{\partial a_j}$ for the case when $j$ is an output unit and for the case when $j$ is a hidden unit.

   b) Then calculate $\frac{\partial E_n(\boldsymbol{w})}{\partial w_{ji}}$ where $w_{ji}$ is a weight for a connection going from the input layer to the hidden layer.

   c) Calculate $\frac{\partial E_n(\boldsymbol{w})}{\partial w_{kj}}$ where $w_{kj}$ is a weight for a connection going from the hidden layer to the output layer.