Name: Serena Yang
CSC578 HW2

Question 1:

Traning Set		
a	b	c
1	0	1

η
0.3

$\sigma(y) = \frac{1}{1+e^{-y}}$

Two neurons:

$o_c = \sigma(0.1 \cdot 1 + 0.1 \cdot 1 + 0.1 \cdot 0) = \sigma(0.2) = 0.5498$

$o_d = \sigma(0.1 \cdot 1 + 0.1 \cdot 0.5498) = \sigma(0.15498) = 0.53867$

Therefore, error terms for the two neurons:

$\delta_d = 0.53867 \times (1 - 0.53867) \times (1 - 0.53867) = 0.1146$

$\delta_c = 0.5498 \times (1 - 0.5498) \times 0.1 \times 0.1146 = 0.002837$

Therefore, correction terms:

$\Delta w_{d0} = 0.3 * 0.1146 * 1 = 0.03438$

$\Delta w_{dc} = 0.3 * 0.1146 * 0.5498 = 0.0189$

$\Delta w_{c0} = 0.3 * 0.002837 * 1 = 0.00085$

$\Delta w_{ca} = 0.3 * 0.002837 * 1 = 0.00085$

$\Delta w_{cb} = 0.3 * 0.002837 * 0 = 0$

New Weights:

$w_{d0} = 0.1 + 0.03438 = 0.13438$

$w_{dc} = 0.1 + 0.0189 = 0.1189$

$w_{c0} = 0.1 + 0.00085 = 0.10085$

$w_{ca} = 0.1 + 0.00085 = 0.10085$

$w_{cb} = 0.1 + 0 = 0.1$

Traning Set		
a	b	c
0	1	0

η
0.3

$\sigma(y) = \frac{1}{1+e^{-y}}$

Two neurons:

$o_c = \sigma(0.10085 * 0 + 0.1 * 1 + 0.10085 * 1) = \sigma(0.20085) = 0.55$
$o_d = \sigma(0.1189 * 0.55 + 0.13438) = \sigma(0.1198) = 0.5498$

Therefore, error terms for the two neurons:
$\delta_d = 0.5498 \times (1 - 0.5498) \times (0 - 0.5498) = -0.1361$
$\delta_c = 0.55 \times (1 - 0.55) \times 0.1189 \times -0.1361 = -0.004$

Therefore, correction terms:
$\Delta w_{d0} = 0.3 * -0.1361 * 1 + 0.9 * 0.03438 = -0.0099$
$\Delta w_{dc} = 0.3 * -0.1361 * 0.55 + 0.9 * 0.0189 = -0.0055$
$\Delta w_{c0} = 0.3 * -0.004 * 1 + 0.9 * 0.00085 = -0.00043$
$\Delta w_{ca} = 0.3 * -0.004 * 0 + 0.9 * 0.00085 = 0.00077$
$\Delta w_{cb} = 0.3 * -0.004 * 1 + 0.9 * 0 = -0.0012$

New Weights:
$w_{d0} = 0.13438 - 0.0099 = 0.12448$
$w_{dc} = 0.1189 - 0.0055 = 0.1134$
$w_{c0} = 0.10085 - 0.00043 = 0.10042$
$w_{ca} = 0.10085 + 0.00077 = 0.10162$
$w_{cb} = 0.1 - 0.0012 = 0.0988$

Question 2:

$$\begin{aligned}
\frac{\partial E}{\partial w_i} &= \frac{\partial}{\partial w_i} \cdot \frac{1}{2} \sum_{d \in \mathrm{D}} (t_d - o_d)^2 \\
&= \frac{1}{2} \sum_{d \in \mathrm{D}} \frac{\partial}{\partial w_i} (t_d - o_d)^2 \\
&= \frac{1}{2} \sum_{d \in D} 2 \cdot (t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d) \\
&= -\sum_{d \in D} (t_d - o_d) \frac{\partial}{\partial w_i} (\tanh(\vec{w} \cdot \vec{x}_d)) \\
&= -\sum_{d \in D} (t_d - o_d)\, tanh'(\vec{w} \cdot \vec{x}_d) \frac{\partial}{\partial w_i} (\vec{w} \cdot \vec{x}_d) \\
&= -\sum_{d \in D} (t_d - o_d)(1 - tanh^2(\vec{w} \cdot \vec{x}_d)) x_{i,d} \\
&= -\sum_{d \in D} (t_d - o_d)(1 - o_d^2) x_{i,d}
\end{aligned}$$

Therefore,
$\Delta w_i = \eta \sum_{d \in D} (t_d - o_d)(1 - o_d^2) x_{i,d}$

For each $\{\vec{x}, \vec{t}\}$ in traning examples, Do

Propagate the input forward through the newtwork:

1. Input the instance $\vec{x}$ to the network and compute the output o_u of every unit u in the network.

Propagate the errors backward through the network:

2. For each network output unit k, calculate its error term δ_k

$$\delta_k \leftarrow (1 - o_k^2)(t_k - o_k)$$

3. For each hidden unit h, calculate its error term δ_h
$$\delta_h \leftarrow (1 - o_h^2) \sum_{k \in outputs} w_{kh} \delta_k$$
4. Update each network weight w_{ji}
$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji}$$
Where
$$\Delta w_{ji} = \eta \delta_j \Delta x_{ji}$$