A.1

For given example (x, y), the least square loss function is

$$E(w) = \frac{1}{2} \sum \left((x_j \cdot w) - y_j \right)^2$$

$$\frac{dE(w)}{dw} = \sum \left((x_j \cdot w) - y_j \right) \cdot x_j$$

Cross-Entropy:

$$E(w) = -\sum_{j=1}^{N} y_j \ln(p_j^+) + (1-y_j) \ln(1-p_j^+)$$

$$\frac{dE}{dw} = \frac{d}{dw} -\sum_{j=1}^{N} y_j \ln(p_j^+) + (1-y_j) \ln(1-p_j^+)$$

$$= -\sum_{j=1}^{N} y_j \frac{1}{p_j^+} (\sigma(1-\sigma)) \cdot x_j + (1-y_j) \left(\frac{1}{1-p_j^+} \right) \left(-(\sigma(1-\sigma) \cdot x_j) \right)$$

So, Basically we just apply input vector x_j and forward propagate to find all input and output, then evaluate the error signals Δ_k for all output nodes. Backpropagate the Δ_k to obtain error signals Δ_j for each hidden node, then perform the gradient decent updates for each weight vector w_{ij} to minimize the loss.

$$w_{ij} \leftarrow w_{ij} + \underbrace{\alpha}_{\text{learning rate}} \times \text{input} \times \Delta[j]$$

A.2

$$\begin{pmatrix} 0.5 & 1.5 \\ 3.5 & 2 \end{pmatrix} \begin{pmatrix} 0.5 & 1 \\ 3 & 2 \end{pmatrix} + \begin{pmatrix} 1.25 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} 4.75 & 3.5 \\ 7.75 & 7.5 \end{pmatrix} + \begin{pmatrix} 1.25 & 3 \end{pmatrix}$$

Apply broadcasting:

$$= \begin{pmatrix} 6 & 6.5 \\ 9 & 10.5 \end{pmatrix}$$

A.3 (1) $a_c = \sigma(W_{ac} a_a + W_{bc} a_b + W_{0c})$

$\Delta[c] = \sigma'(W_{ac} a_a + W_{bc} a_b + W_{0c}) \cdot W_{cd} \cdot \Delta[d]$

$W_{ac} \leftarrow W_{ac} + \alpha \times a_a \times \Delta[c]$

$W_{bc} \leftarrow W_{bc} + \alpha \times a_b \times \Delta[c]$

$W_{0c} \leftarrow W_{0c} + \alpha \times 1 \times \Delta[c]$

$a_d = \sigma(W_{cd} a_c + W_{0d})$

$\Delta[d] = \sigma'(W_{cd} a_c + W_{0d})(y - a_d)$

$W_{cd} \leftarrow W_{cd} + \alpha \times a_c \times \Delta[d]$

$W_{0d} \leftarrow W_{0d} + \alpha \times 1 \times \Delta[d]$

(2)

Data Point	a_c	$\Delta[c]$	a_d	$\Delta[d]$	w_{0c}	w_{ac}	w_{bc}	w_{cd}	w_{0d}
x_1	0.599	0.0026	0.565	0.107	0.20026	0.20026	0.1	0.106	0.211
x_2	0.5745	-0.0036	0.5676	-0.139	0.1999	0.20026	0.0996	0.098	0.197

$W_{ac} = 0.2 \quad W_{bc} = 0.1 \quad W_{0c} = 0.2$

$W_{cd} = 0.1 \quad W_{0d} = 0.2$

$$a_c = \sigma(0.2 \cdot 1 + 0.1 \cdot 0 + 0.2)$$
$$= \sigma(0.4)$$
$$= 0.599$$

$$a_d = \sigma(0.1 \cdot 0.599 + 0.2) = 0.565$$

$$\Delta[d] = 0.565 \cdot (1 - 0.565)(1 - 0.0.565) = 0.107$$

$$W_{cd} = (0.1 + 0.1 \times 0.599 \times 0.107) = 0.106$$

$$W_{0d} = (0.2 + 0.1 \times 0.107) = 0.211$$

$$\Delta[c] = 0.599 \cdot (1 - 0.599) \cdot 0.1 \cdot 0.107$$
$$= 0.0026$$

$$W_{ac} = 0.2 + 0.1 \times 0.0026 = 0.20026$$

$w_{bc} = 0.1$

$w_{0c} = 0.2 + 0.1 \times 0.0026 = 0.20026$

0 1 0

$a_c = \sigma(0.30026) = 0.5745$

$a_d = \sigma(0.106 \times 0.5745 + 0.211)$

$= 0.5676$

$\Delta[d] = -0.139$

$w_{cd} = 0.106 + 0.1 \times 0.5745 \times (-0.139)$

$= 0.098$

$w_{0d} = 0.211 + 0.1 \times 1 \times -0.139 = 0.197$

$\Delta[c] = -0.0036$

$$w_{ac} = 0.20026$$

$$w_{bc} = 0.0996$$

$$w_{ac} = 0.20026 + 0.1 \times (-0.0036)$$

$$= 0.1999$$