Q1.1

$$\frac{\partial l}{\partial u_i} = \frac{\partial l}{\partial h_i}\frac{\partial h_i}{\partial u_i}$$

So

$$\frac{\partial l}{\partial \boldsymbol{u}} = \frac{\partial l}{\partial \boldsymbol{h}}.*\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}}$$

$$= \frac{\partial l}{\partial \boldsymbol{a}} * \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{u}}.*\boldsymbol{H(u)}^T$$

$$= \frac{\partial l}{\partial \boldsymbol{a}} * \frac{\partial(\boldsymbol{W^{(2)}h + b^{(2)}})}{\partial \boldsymbol{h}}.*\boldsymbol{H(u)}^T$$

$$= \frac{\partial l}{\partial \boldsymbol{a}} * \boldsymbol{W^{(2)}.*H(u)}^T$$

$$\frac{\partial l}{\partial ai} = \frac{\partial l}{\partial z_i} * \frac{\partial z_i}{\partial a_i}$$

$$= -(y_i == k) * \frac{1}{z_i} * (z_i * (1 - z_i))$$

$$= -(y_i == k) * (1 - z_i)$$

$$\frac{\partial l}{\partial \boldsymbol{a}} = \boldsymbol{y^T z - y^T}$$

$$\frac{\partial l}{w_{ji}^{(1)}} = \frac{\partial l}{\partial uj}\frac{\partial uj}{\partial w_{ji}^{(1)}} = \frac{\partial l}{\partial uj} * x_i$$

$$\frac{\partial l}{\partial \boldsymbol{W^{(1)}}} = (\frac{\partial l}{\partial \boldsymbol{u}})^T * \boldsymbol{x^T}$$

$$\frac{\partial l}{\partial \boldsymbol{b^{(1)}}} = \frac{\partial l}{\partial \boldsymbol{u}}\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{b}} = \frac{\partial l}{\partial \boldsymbol{u}}$$

$$\frac{\partial l}{w_{ji}^{(2)}} = \frac{\partial l}{\partial aj}\frac{\partial aj}{\partial w_{ji}^{(2)}} = \frac{\partial l}{\partial uj} * h_i$$

$$\frac{\partial l}{\partial \boldsymbol{W^{(2)}}} = (\frac{\partial l}{\partial \boldsymbol{a}})^T * \boldsymbol{h}^T$$

Q1.2 Because $\boldsymbol{W^{(1)}, W^{(2)}, b^{(1)}}$are all initialed to zero and from the above derivation, the gradient $\frac{\partial l}{\partial \boldsymbol{W^{(1)}}}, \frac{\partial l}{\partial \boldsymbol{W^{(2)}}}, \frac{\partial l}{\partial \boldsymbol{b^{(1)}}}$ are all 0.

Q 1.3

$$\boldsymbol{a = W^{(2)}u + b^{(2)}}$$
$$\boldsymbol{= W^{(2)}(W^{(1)}x + b^{(1)}) + b^{(2)}}$$
$$\boldsymbol{= W^{(2)}W^{(1)}x + W^{(2)}b^{(1)} + b^{(2)}}$$
$$\boldsymbol{U = W^{(2)}W^{(1)}}$$
$$\boldsymbol{v = W^{(2)}b^{(1)} + b^{(2)}}$$

Q2.1

$$\frac{\partial l}{\partial \boldsymbol{w}} = \sum_n \frac{\partial l(s, y_n)}{\partial s}\frac{\partial s}{\partial \boldsymbol{w}} + \lambda \boldsymbol{w}$$

$$\sum_n \frac{\partial l(s, y_n)}{\partial s}\boldsymbol{\phi(x_n)} + \lambda \boldsymbol{w}^* = 0$$

$$\boldsymbol{w}^* = -\frac{\sum_n \frac{\partial l(s, y_n)}{\partial s}\boldsymbol{\phi(x_n)}}{\lambda}$$

Q2.2

$$\boldsymbol{w^* = \Phi^T\alpha}$$

$$\sum_n l(\boldsymbol{w^T\phi(x_n)}, y_n) + \frac{\lambda}{2}||\boldsymbol{w}||_2^2$$

$$= \sum_n l(\boldsymbol{\alpha^T\Phi(x)\phi(x_n)}, y_n) + \frac{\lambda}{2}||\boldsymbol{w}||_2^2$$

$$= \sum_n l(\boldsymbol{\alpha}^T[K_{1n}, K_{2n}, \cdots, K_{nn}]^T, y_n) + \frac{\lambda}{2}||\boldsymbol{w}||_2^2$$