CSCI567 Homework 2

Question 1.1

$$\frac{\partial l}{\partial u} = \frac{\partial l}{\partial a}\left(\frac{\partial a}{\partial u}\right) = \frac{\partial l}{\partial a}\left(\frac{\partial a}{\partial h}\right)\left(\frac{\partial h}{\partial u}\right)$$

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

$$\frac{\partial l}{\partial a} = \frac{\partial l}{\partial z}\left(\frac{\partial z}{\partial a}\right)$$

$$\frac{\partial l}{\partial a} = \left(-\sum_k \frac{y_k}{z_k}\right)\begin{pmatrix} \frac{(e^{a_1}\sum_k e^{a_k} - e^{a_1} e^{a_1})}{(\sum_k e^{a_k})^2} \\ \vdots \\ \frac{(e^{a_K}\sum_k e^{a_k} - e^{a_K} e^{a_K})}{(\sum_k e^{a_k})^2} \end{pmatrix}$$

$$\frac{\partial l}{\partial W^{(1)}} = \frac{\partial l}{\partial u}\left(\frac{\partial u}{\partial W^{(1)}}\right)$$

$$\frac{\partial l}{\partial W^{(1)}} = \frac{\partial l}{\partial u} X$$

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

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

$$\frac{\partial l}{\partial W^{(2)}} = \frac{\partial l}{\partial a}\left(\frac{\partial a}{\partial W^{(2)}}\right)$$

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

Question 1.2

Upon each iteration, each of the listed variables are updated with a value that is directly proportional to their respective gradient, so if their respective gradient is equal to zero, there will be no change to that value regardless of how many iterations.

Question 1.3

$$a = W^{(2)}\left(W^{(1)}x + b^{(1)}\right) + b^{(2)}$$

$$= W^{(2)}W^{(1)}x + W^{(2)}b^{(1)} + b^{(2)}$$

$$U = W^{(2)}W^{(1)}$$

$$v = W^{(2)}b^{(1)} + b^{(2)}$$

Question 2.1

$$\frac{\partial}{\partial w}\sum_n l\left(w^T\varphi(x_n), y_n\right) + \frac{\lambda}{2}\|w\|_2^2 = 0$$

$$\sum_n \frac{\partial l\left(w^T\varphi(x_n), y_n\right)}{\partial w^T\varphi(x_n)}\left(\frac{\partial w^T\varphi(x_n)}{\partial w}\right) + \lambda\|w\|_2 = 0$$

$$\|w\|_2 = -\frac{1}{\lambda}\sum_n \frac{\partial l\left(w^T\varphi(x_n), y_n\right)}{\partial w^T\varphi(x_n)}\varphi(x_n)$$

Question 2.2

$$\|w\|_2 = -\frac{1}{\lambda}\sum_n \frac{\partial l\left(w^T\varphi(x_n), y_n\right)}{\partial w^T\varphi(x_n)}\varphi(x_n)$$

$$\alpha_n = -\frac{1}{\lambda}\left(\frac{\partial l\left(w^T\varphi(x_n), y_n\right)}{\partial w^T\varphi(x_n)}\right)$$

$$\|w\|_2 = \sum_n \alpha_n\varphi(x_n) = \varphi^T\alpha$$

$$\min_w \sum_n l\left(w^T\varphi(x_n), y_n\right) + \frac{\lambda}{2}\|w\|_2^2 = \min_\alpha \sum_{m,n} l\left(\alpha\varphi(x_m)\varphi(x_n), y_n\right) + \frac{\lambda}{2}\left\|\varphi^T\alpha\right\|_2^2$$

$$= \min_\alpha \sum_{m,n} l\left(\alpha K_{mn}, y_n\right) + \frac{\lambda}{2}\left\|\varphi^T\alpha\right\|_2^2$$