Q1.1

a) $\frac{\partial \mathrm{l}}{\partial \mathbf{u}} = \frac{\partial \mathrm{l}}{\partial \mathbf{a}} . \frac{\partial \mathbf{a}}{\partial \mathbf{u}}$

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

b) $Z_i = \frac{e^{a_i}}{\sum_k e^{a_k}}$

$\frac{\partial Z_i}{\partial a_j} = \{Z_i(1 - Z_i), if\ i == j \middle| -Z_i Z_j, if\ i \neq j\}$

$\frac{\partial Z_i}{\partial a_j} = -\sum_k y_k \frac{1}{Z_k} \frac{\partial Z_k}{\partial a_i}$

$\frac{\partial \mathrm{l}}{\partial \mathbf{a}} = \frac{-\boldsymbol{y}^{\boldsymbol{T}}}{\boldsymbol{Z}} . (Z.* (1 - Z))$

c) $\frac{\partial \mathrm{l}}{\partial \boldsymbol{W}^{(1)}} = \frac{\partial \mathrm{l}}{\partial \mathbf{u}} . X$

d) $\frac{\partial \mathrm{l}}{\partial \boldsymbol{b}^{(1)}} = \frac{\partial \mathrm{l}}{\partial \mathbf{u}} . 1$

e) $\frac{\partial \mathrm{l}}{\partial \boldsymbol{W}^{(2)}} = \frac{\partial \mathrm{l}}{\partial \mathbf{a}} . \boldsymbol{h}$

Q1.2

If we initialize$\boldsymbol{W}^{(1)}$, $\boldsymbol{W}^{(2)}$ and $\boldsymbol{b}^{(1)}$ as Zero/Zero vectors we have$\frac{\partial \mathrm{l}}{\partial \boldsymbol{W}^{(1)}}, \frac{\partial \mathrm{l}}{\partial \boldsymbol{b}^{(1)}}, \frac{\partial \mathrm{l}}{\partial \boldsymbol{W}^{(2)}}$ to be zero vectors as well. When using stochastic gradient descent for learning we update values using their corresponding gradients. Now that the gradients are all Zero, immaterial of the number of iterations there would be no learning taking place.

Q1.3

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

$a = W^{(2)} + \ b^{(2)}$

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

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

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

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

Q2.1

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

$\Longrightarrow \sum_n l'(w^T \emptyset(x_n), y_n)\phi(x_n) + \lambda w = 0$

$\Longrightarrow \lambda w = \sum_n -l'\,(w^T \emptyset(x_n), y_n)\phi(x_n)$

$\Longrightarrow w = \sum_n -\frac{1}{\lambda} l'\,(w^T \emptyset(x_n), y_n)\phi(x_n)$

$where, \alpha_n = -\frac{1}{\lambda} l'(w^T \emptyset(x_n), y_n)$

$\Longrightarrow w = \sum_n \alpha_n \, \emptyset(x_n)$

Q2.2

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

We Know, $w^* = \Phi^T \alpha$; $\Phi\Phi^T = \boldsymbol{K}$

$\sum_n l\,(\Phi^T \alpha)\emptyset(x_n), y_n) + \frac{\lambda}{2}\alpha^T \Phi\Phi^T \alpha$

$l\big((\boldsymbol{K}\alpha), y_n\big) + \frac{\lambda}{2}\alpha^T \boldsymbol{K}\alpha$
