CSCI 567 - Home Work - 2

Sai Swarna Latha Vempaty - 5824899937

1.1)

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

$\partial l/\partial a =$ -$\frac{y^T}{Z}.Z'(a)$; where each element of (kxk) matrix $Z'(a)$is $\mathbf{Z}'(a) = \frac{e^{a_m}}{\sum_k e^{a_k}}[I(m == n) - \frac{e^{a_n}}{\sum_k e^{a_k}}]$

$\partial l/\partial W^{(1)} = (\partial l/\partial u).x^T$

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

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

1.2)

When $W^{(1)}, W^{(2)}, b^{(1)}$ is initialized to zero vectors/matrices,loss derivatives w.r.t these also become zero., so at every step of gradient descent ($v = \alpha V - \eta g$ and $\omega = \omega + V$), will not change from initial values.

1.3)

$u = W^{(1)}.x + b^{(1)}$ => a $= W^{(2)}$. ($W^{(1)}.x + b^{(1)}$) + $b^{(2)}$

comparing this to $a = Ux + v$

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

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

2.1)

$J(\omega) = min_\omega \sum_n l(\omega^T\phi(x_n), y_n) + (\lambda/2)||\omega||_2{}^2$

$\partial J(\omega)/\partial\omega = \sum_n \partial l(s,y)/\partial s.(\phi(x_n)) + \lambda\omega = 0$

$\lambda\omega^* = -\sum_n \partial(s,y)/\partial s.\phi(x_n)$

$\omega^* = \phi^T\alpha$

where $\phi^T = (\phi(x_1)\phi(x_2)......\phi(x_n))$ => R^{MxN}

and $\alpha_n = \frac{-1}{\lambda}.\frac{\partial l(s,y)}{\partial s}$

$\omega^* = \sum_n \alpha_n.\phi(x_n)$

which proves that optimal parameter is a linear combination of features.

2.2)

$J(\omega) = min_\omega \sum_n l(\omega^T\phi(x_n), y_n) + (\lambda/2)||\omega||_2{}^2$

$J(\omega) = l(\phi\phi^T\alpha, y) + \frac{\lambda}{2}||\phi^2\alpha||_2{}^2$

objective function in terms of k and α

$J(\alpha) = l(K\alpha, y) + \frac{\lambda}{2}\alpha^T K\alpha$