Harsh Agrawal Homework 2

Q1) $E(w) = \sum_{i=1}^{m} (w^T . x^{(i)} - y^{(i)})^2 + \lambda \sum_{i=1}^{m} w_i^2$

$$= ((xw)^T - y^T)(xw - y)$$

$$= (xw)^T (xw) - (xw)^T (y)$$
$$- (xw)(y)^T + y^T y + \lambda w^T w$$

$$= w^T x^T x w - 2(xw)^T y = y^T y + \lambda w^T w$$

Also, $\frac{\partial E}{\partial w} = 0$

∴ Differentiating prev. equation, we get

$$(x^T x + \lambda I) w = x^T y$$

$$\Rightarrow w = \frac{y}{(x^T x + \lambda I)}$$

$$= (x^T x + \lambda I)^{-1} x^T y$$

$$w = (\lambda I + x^T . x)^{-1} . x^T . y$$

Q2) $\hat{P}_k = S(s_k(x))_k = \dfrac{\exp(s_k(x))}{\sum_{j=1}^{k} \exp(s_j(x))}$, $s_k(x) = \theta_k^T . x$

Shape of $x \rightarrow n$ (column vector)
and
k is the no. of classes.

To learn this softmax, we estimate $(n+1)$ parameters which are from $\theta_0 \ldots \theta_n$.

Now, $J(\theta) = \dfrac{-1}{m} \sum_{i=1}^{m} \sum_{k=1}^{k} y_k^{(i)} \log(\hat{p}_k^{(i)})$

Here, $y_k^{(i)} = 1$ if the i^{th} instance belongs to the k^{th} class, 0 otherwise.

$\therefore J(\theta) = \dfrac{-1}{m} \sum_{i=1}^{m} \sum_{k=1}^{k} y_k^{(i)} \log\left(\dfrac{\exp(s_k(x)^i)}{\sum_{j=1}^{k} \exp(s_j(x)^i)}\right)$

$= \dfrac{-1}{m} \sum_{i=1}^{m} \left(\sum_{x=1}^{k} y_k(i) \log(\exp(s_k(x^{(i)}))\right)$

$= \sum_{k=1}^{k} y_k^{(i)} \log(\sum_{j=1} \exp(s_j(x)^i))$

Now from $J(\theta)$ eqn, we differentiate wrt. θ,

$$\frac{\partial J(\theta)}{\partial(\theta)} = \frac{-1}{m} \sum_{i=1}^{m} \left(\hat{p}_k^{(i)} - y_k^{(i)}\right) x^{(i)}$$

$$= \frac{-1}{m} \sum_{i=1}^{m} x^{(i)} - \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T x^{(i)})} \left(\exp(\theta_k^T x^{(i)}) x^{(i)}\right)$$

Now integrating wrt θ, we get,

$$J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} \sum_{k=1}^{m} y_k^{(i)} \log\left(\frac{\exp(s_k(x))}{\sum_{j=1}^{k} \exp(s_j(x))}\right)$$

$$\therefore J(\theta) = \frac{-1}{m} \sum_{i=1} \sum_{k=1} y_k^{(i)} \log(\hat{p}_k^{(i)})$$

$\therefore$ We proved that, the gradient of $J(\theta)$ ~~expre~~ wrt. θ is.

$$\nabla_\theta J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} x^{(i)} - \frac{\exp(\theta^T x(x)^{(i)})x^{(i)}}{\sum_{j=1}^{k} \exp(\theta_j^T x(i))}$$