1a)

$$\frac{\partial C^n(w)}{\partial w^i} = \frac{\partial C^n(w)}{\partial \hat{y}^n} \times \frac{\partial \hat{y}^n}{\partial f(x^n)} \times \frac{\partial f(x^n)}{\partial w^i}$$

As $\hat{y}^n = f(x^n)$, $\frac{\partial \hat{y}^n}{\partial f(x^n)} = 1$

$$\therefore \frac{\partial C^n(w)}{\partial w^i} = \frac{\partial C^n(w)}{\partial \hat{y}^n} \times \frac{\partial f(x^n)}{\partial w^i}$$

$$\frac{\partial C^n(w)}{\partial \hat{y}^n} = -\left(y^n \times \frac{1}{\hat{y}^n} + (1-y^n)\left(\frac{1}{1-\hat{y}^n}\right)(-1)\right)$$

$$= \frac{1-y^n}{1-\hat{y}^n} - \frac{y^n}{\hat{y}^n}$$

$$= \frac{\hat{y}^n(1-y^n) - y^n(1-\hat{y}^n)}{(\hat{y}^n)(1-\hat{y}^n)}$$

$$= \frac{\hat{y}^n - \hat{y}^n y^n - y^n + y^n\hat{y}^n}{(\hat{y}^n)(1-\hat{y}^n)}$$

$$= \frac{\hat{y}^n - y^n}{(\hat{y}^n)(1-\hat{y}^n)}$$

We know $\frac{\partial f(x^n)}{\partial w}$ from the question, Hence,

$$\frac{\partial C^n(w)}{\partial w_i} = \frac{\hat{y}^n - y^n}{(\hat{y}^n)(1-\hat{y}^n)} \times (x_i^n)\, f(x^n)(1-f(x^n))$$

As $\hat{y}^n = f(x^n)$,

$$\frac{\partial C^n(w)}{\partial w_i} = x_i^n \frac{\hat{y}^n - y^n}{(\hat{y}^n)(1-\hat{y}^n)} \times (\hat{y}^n)(1-\hat{y}^n)$$

$$= -(y^n - \hat{y}^n)\, x_i^n \quad \text{(Shown)}$$

(b) $C^n(w) = -\sum_{k=1}^{K} y_k^n \ln(\hat{y}_k^n)$ $\quad \hat{y}_k^n = \frac{e^{z_k}}{\sum_{k'}^{K} e^{z_{k'}}}$ $\quad z_k = \sum_{i}^{I} w_{k,i} \cdot x_i$

$$\frac{\partial C^n(w)}{\partial w_{kj}} = \frac{\partial C^n(w)}{\partial \hat{y}_k^n} \times \frac{\partial \hat{y}_k^n}{\partial z_k} \times \frac{\partial z_k}{\partial w_{kj}}$$

$\frac{\partial z_{k'}}{\partial w_{k'j}} = x_j$ as all other w in matrix treated as const.

$$\frac{\partial \hat{y}_k^n}{\partial z_{k'}} = \frac{\partial}{\partial z_{k'}} \frac{e^{z_k}}{\sum_{k'}^{K} e^{z_{k'}}}$$

$$\frac{\partial \hat{y}_k^n}{\partial z_{k'}} = \frac{\left(\sum e^{z_{k'}}\right)\frac{\partial}{\partial z_{k'}} e^{z_k} - e^{z_k} \frac{\partial}{\partial z_{k'}} \sum_{k'}^{K} e^{z_{k'}}}{\left(\sum e^{z_{k'}}\right)^2}$$

when $k = k'$,

$$\frac{\partial \hat{y}_k^n}{\partial z_{k'}} = \frac{\left(\sum e^{z_{k'}}\right) e^{z_{k'}} - e^{z_{k'}} \cdot e^{z_{k'}}}{\left(\sum e^{z_{k'}}\right)^2}$$

$$\frac{e^{z_{k'}}}{\quad} - \left(\frac{e^{z_{k'}}}{\sum e^{z_{k'}}}\right)^2$$

$$= \quad \sum_{k'} e^{z_{k'}} \qquad \left(\sum_{k'} \right)$$

$$= \hat{y}^n_{k'} - \left(\hat{y}^n_{k'}\right)^2 = \hat{y}^n_{k'}\left(1 - \hat{y}^n_{k'}\right)$$

when $k \neq k'$

$$\frac{\partial \hat{y}^n_k}{\partial z_{k'}} = \frac{\sum_{k'} e^{z_{k'}} \times 0 - e^{z_k} \cdot e^{z_{k'}}}{\left[\sum_{k'}^{K} e^{z_{k'}}\right]^2}$$

$$= -\frac{e^{z_k}}{\sum_{k'}^{K} e^{z_{k'}}} \times \frac{e^{z_{k'}}}{\sum_{k'}^{K} e^{z_{k'}}}$$

$$= -\hat{y}^n_k \times \hat{y}^n_{k'}$$

$$\frac{\partial C^n(w)}{\partial \hat{y}^n_k} = -\sum_{k'=1}^{K} \frac{\partial}{\partial \hat{y}^n_{k'}} \left(y^n_{k'}\right) \ln\left(\hat{y}^n_{k'}\right)$$

$$= -\sum_{k'=1}^{K} \frac{y^n_{k'}}{\hat{y}^n_{k'}}$$

$$\frac{\partial C^n(w)}{\partial w_{kj}} = -\sum_{k'=1}^{K} \frac{y^n_{k'}}{\hat{y}^n_{k'}} \times \frac{\partial \hat{y}^n_k}{\partial z_{k'}} \times x_j$$

$$= -x_j \sum_{k'=1}^{K} \frac{y^n_{k'}}{\hat{y}^n_{k'}} \times \frac{\partial \hat{y}^n_k}{\partial z_{k'}} \quad \Bigg\}$$ Only 1 instance when $k = k'$, rest $k \neq k'$

$$= -x_j \left[\frac{y^n_{k'}}{\hat{y}^n_{k'}} \times \hat{y}^n_{k'}\left(1 - \hat{y}^n_{k'}\right) + \sum_{k'=1,\, k' \neq k}^{K} \frac{y^n_{k'}}{\hat{y}^n_{k'}} \left(-\hat{y}^n_k \times \hat{y}^n_{k'}\right) \right]$$

$- y_{k'}$

$$= -x_j \left[y^n_{k'}\left(1-\hat{y}^n_{k'}\right) + \sum_{k'=1,\, k'\neq k}^{K} \left(- y^n_{k'}\, \hat{y}^n_{k}\right) \right]$$

$$= x_j \left[y^n_{k'}\, \hat{y}^n_{k'} - y^n_{k'} + \sum_{k'=1,\, k'\neq k}^{K} \hat{y}^n_{k} \times y^n_{k'} \right]$$

$$= x_j \left[-y^n_{k'} + y^n_{k'}\, \hat{y}^n_{k} + \sum_{k'=1,\, k'\neq k}^{K} \hat{y}^n_{k} \times y^n_{k'} \right],$$

as in $y^n_{k'}\hat{y}^n_{k'}$, $k=k'$, $\therefore \hat{y}^n_{k'} = \hat{y}^n_{k}$,

$$= x_j \left[-y^n_{k'} + \hat{y}^n_{k}\left[y^n_{k'} + \sum_{k'=1,\, k'\neq k}^{K} y^n_{k'} \right]\right]$$

$$= x_j \left[-y^n_{k'} + \hat{y}^n_{k}\left(\sum_{k'=1}^{K} y^n_{k'} \right)\right]$$

$$= -x_j \left[y^n_{k} - \hat{y}^n_{k}(1) \right]$$

as $k=k'$ in that instance and $\sum_{k'=1}^{K} \hat{y}^n_{k'} = 1$

$$= -x_j \left[y^n_{k} - \hat{y}^n_{k} \right] \quad \text{(Shown)}$$

4a)

$$J(w) = C(w) + \lambda R(w)$$

$$\frac{\partial J(w)}{\partial w} = \frac{\partial C(w)}{\partial w} + \lambda \frac{\partial R(w)}{\partial w}$$

$$\frac{\partial R(w)}{\partial w} = \frac{\partial}{\partial w} \frac{1}{2} \sum_{i,j} w_{i,j}^2$$

Derivative will only be non-zero when $i' = i$ and $j' = j$

Hence at each matrix cell, the derivative is:

$$\frac{1}{2} \times \frac{\partial}{\partial w} w_{i,j}^2 = w_{i,j}$$

Hence $\frac{\partial R(w)}{\partial w} = w$, where w is the weight matrix.

$\therefore \quad \lambda \cdot \frac{\partial R(w)}{\partial w} = \lambda w \longleftarrow$ Update due to L2

We know $C(w) = \frac{1}{N} \sum_{n=1}^{N} C^n(w)$

$$\therefore \quad \frac{\partial C(w)}{\partial w} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial C^n(w)}{\partial w}$$

As earlier computed, we know $\frac{\partial C^n(w)}{\partial w}$ for softmax regression.

hence we have our update term for softmax regression with L2 regularization, $\frac{\partial J(w)}{\partial w}$.

$$\frac{\partial J(w)}{\partial w} = \frac{1}{N}\sum_{n=1}^{N}\left[-x_j^n\left(y_k^n - \hat{y}_k^n\right)\right] + \lambda w$$