Q1 Ans Given $J(\omega) = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2$

$$\hat{y}_k = \frac{\exp z_k}{\sum_{k'=1}^{c} \exp z_{k'}}$$

$$z_k = x^T \omega^{(k)} + b_k$$

Q1 Ans given $J(\omega) = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2$

$$\hat{y} = f_\omega(x) = \omega^T x$$

we have to prove,

$$\omega^* = (XX^T)^{-1} X y$$

Given that,

$$x\omega^* = \omega^{(k)} - H(f)(\omega^{(k)})^{-1} \nabla_\omega f(\omega^{(k)}) \quad \text{— (1)}$$

$$\nabla_\omega J(\omega) = \nabla_\omega \left(\frac{1}{2n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2\right.$$

$$= \frac{2}{2n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)}) \cdot x$$

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

$$H(\omega) = \nabla_\omega (\nabla_\omega J(\omega))$$

$$= \nabla_\omega \left(\frac{1}{n}\sum_{i=1}^{n} x(\hat{y}^{(i)} - y^{(i)})\right)$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} x x^T$$

$\therefore$ Substituting in equation ①

$$\cancel{w^{(k)} = w^{(k)} - \frac{1}{n}\sum_{i=1}^{n} x(\hat{y}^{(i)} - y^{(i)})}$$

$$w^* = w^{(k)} - \frac{\frac{1}{n}\sum_{i=1}^{n} x(\hat{y}^{(i)} - y^{(i)})}{\frac{1}{n}\sum_{i=1}^{n} x x^T}$$

$$= w^k - \frac{X(X^T w^{(k)} - Y)}{X X^T}$$

$\because$ Substituting $\hat{y} = X^T w$

and

$$\frac{1}{n}\sum_{i=1}^{n} x(\hat{Y}^{(i)} - Y^{(i)}) = X(X^T w^{(k)} - Y)$$

$$\frac{1}{n}\sum_{i=1}^{n} x x^T = X X^T$$

$$\therefore \quad w^* = w^k - w^{(k)} + \frac{XY}{XX^T}$$

$$\boxed{w^* = (XX^T)^{-1} X y}$$

Q2 $\nabla_{w^{(l)}} f_{CE}(W,b) = \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \nabla_{w^{(l)}} \log \hat{y}_k^{(i)}$

$$= \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \left(\frac{\nabla_{w^{(l)}} \hat{y}_k^{(i)}}{\hat{y}_k^{(i)}} \right.$$

Case I :

$\boxed{l = k}$

$$\nabla_{w^{(l)}} \hat{y}_k^{(i)} = \nabla_{w^{(l)}} \frac{\exp(z^{i})}{\sum_{k'=1}^{c} \exp z_{k'}^{(i)}}$$

$$= \nabla_{w^{(l)}} \left(\frac{\exp(x^T w^{(l)} + b^{(l)})^{(i)}}{\sum_{k'=1}^{c} \exp(x^T w^{(k')} + b_{k'})^{(i)}} \right.$$

for $f(x) = g(x)/h(x)$,

$$f'(x) = \frac{g'(x)\, h(x) - g(x)\, h'(x)}{[h(x)]^2}$$

$\therefore \nabla_{w^{(l)}} \hat{y}_k^{(i)} = \nabla_{w^{(l)}} \left(\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})^{(i)} \right)$

II ...

$$= \frac{\left(\nabla_{w^{(l)}} (\exp(x^T w^{(l)} + b^{(l)})^{(i)}\right) \cdot \sum_{k'=1}^{c} \exp(x^T w^{(k')} + b_{k'}) - \exp(x^T w^{(l)} + b^{(l)})^{(i)} \cdot \nabla_{w^{(l)}} \sum \exp(x^T w_{k'} + b_{k'})}{\left(\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'}) \right)^2}$$

Scanned by CamScanner

$$= \frac{\left[x^{(i)} \exp\left(x^T w^{(\ell)} + b^{(\ell)}\right)^{(i)} \sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right)^{(i)}\right] - \left[\exp\left(x^T w^{(\ell)} + b^{(\ell)}\right)^{(i)} \cdot \left(x^{i} \exp\left(x^T w^{(\ell)} + b^{(\ell)}\right)\right)^{(i)}\right]}{\left[\sum_{k'=1}^{c} \exp\left(x^T w^{k'} + b_{k'}\right)^{(i)}\right]^2}$$

$$= \frac{x^{(i)} \exp\left(x^T w^{(\ell)} + b^{(\ell)}\right)^{(i)} \left[\sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right) - \exp^{(i)}\left(x^T w_{\ell} + b_{\ell}\right)\right]}{\sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right) \sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right)}$$

$$= x^{(i)} \cdot \frac{\exp\left(x^T w^{(\ell)} + b^{(\ell)}\right)^{(i)}}{\sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right)} \left[1 - \frac{\exp\left(x^T w_{(\ell)} + b^{(\ell)}\right)}{\sum_{k'=1}^{c} \exp\left(x^T w_{k'} + b_{k'}\right)}\right]$$

$$= x^{(i)} \cdot \hat{y}_k^{(i)} \cdot \left(1 - \hat{y}_\ell^{(i)}\right)$$

$$\therefore \boxed{\nabla w^{(\ell)} \hat{y}_k^{(\ell)} = x^{(i)} \hat{y}_\ell^{(i)} \left(1 - \hat{y}_\ell^{(i)}\right)} \quad \ldots\ldots \text{(II)}$$

case II

if $l \neq k$,

$$\nabla_{w^{(l)}} \hat{y}_k^{(i)} = \nabla_{w(l)} \frac{\exp(z_k^{(i)})}{\sum_{k'=1}^{C} \exp(z_{k'}^{(i)})}$$

$$= \nabla_{w^{(l)}} \frac{\exp(x^T w_k + b_k)}{\sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})}$$

As, $f(x) = g(x)/h(x)$

$$f'(x) = \frac{g'(x)h(x) - h'(x)g(x)}{[h(x)]^2}$$

$$\therefore \nabla_{w^{(l)}} \hat{y}_k^{(i)} = \frac{\left(\nabla_{w^{(l)}} \exp(x^T w_k^{(i)} + b_k^{(i)}) \cdot \sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})\right)}{\cancel{[h(x)]^2} \left[\sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})\right]^2} - \frac{\nabla_{w^{(l)}} \sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'}) \cdot \exp(x^T w_k + b_k)}{\cancel{[h(x)]^2} \sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})}$$

$$= \frac{0 \cdot \sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'}) - x^{(i)} \cdot \exp(x^T w_l + b_{(l)}) \cdot \exp(x^T w_k + b_k)}{\left[\sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})\right]^2}$$

$$= -x^{(i)} \cdot \frac{\exp(x^T w_k + b_k)}{\sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})} \cdot \frac{\exp(x^T w_l + b_l)}{\sum_{k'=1}^{C} \exp(x^T w_{k'} + b_{k'})}$$

$$\boxed{\nabla_{w^{(l)}} \hat{y}_k^{(i)} = -x^{(i)} \hat{y}_k^{(i)} \hat{y}_l^{(i)}} \quad \text{--- III}$$

$$f_{CE}(W, b) = \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \log \hat{y}_k^{(i)}$$

$$\nabla_{w^{(\ell)}} f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \log \hat{y}_k^{(i)}$$

$$\nabla f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} \left[\sum_{k=\ell} \frac{y_k^{(i)}}{\hat{y}_k(\ell)} \nabla_{w(\ell)} (\hat{y}_k(i)) + \sum_{(k \neq \ell)} \frac{y_k^{(i)}}{\hat{y}_k^{(i)}} \nabla_{w(k)} \hat{y}_k^{(i)} \right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[\sum_{k=\ell} \frac{y_\ell^{(i)}}{\hat{y}_\ell^{(i)}} x^{(i)} \hat{y}_\ell^{(i)} (1 - \hat{y}_\ell^{(i)}) + \sum_{k \neq \ell} \frac{y_k^{(i)}}{\hat{y}_k^{(i)}} (-x^{(i)} \hat{y}_k^{(i)} \hat{y}_\ell^{(i)}) \right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[\sum_{k=\ell} y_\ell^{(i)} x^{(i)} (1 - \hat{y}_\ell^{(i)}) + \sum_{k \neq \ell} y_k^{(i)} (-x^{(i)}) \hat{y}_\ell^{(i)} \right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[y_\ell^{(i)} x^{(i)} - y_\ell^{(i)} x^{(i)} \hat{y}_\ell^{(i)} - \sum_{k \neq \ell} x^{(i)} y_k^{(i)} \hat{y}_\ell^{(i)} \right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[\sum x^{(i)} y_\ell^{(i)} - \sum_{k=1}^{c} (x^{i} y_k^{(i)} \hat{y}_\ell^{(i)} \right]$$

Scanned by CamScanner

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[x^{(i)} y_l^{(i)} - x^{(i)} \hat{y}_l^{(i)} \sum_{k=1}^{c} y_k^{(i)} \right]$$

$$\sum_{k=1}^{c} y_k^{(i)} = 1$$

$$\therefore \nabla_w f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} \left[x^{(i)} y_l^{(i)} - x^{(i)} \hat{y}_l^{(i)} \right]$$

$$\boxed{\nabla_w f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} x^{(i)} \left(y_l^{(i)} - \hat{y}_l^{(i)} \right)}$$

Taking gradient w.r.t b

$$\nabla_b f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \nabla_b \log \hat{y}_k^{(i)}$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \sum_{k=1}^{c} y_k^{(i)} \left[\frac{\nabla_b \hat{y}_k^{(i)}}{y_k^{(i)}} \right]$$

But, $\nabla_b \hat{y}_k^{(i)} = \nabla_b \left[\frac{\exp(x^T w_k + b_k)}{\sum_{k'=1} \exp(x^T w_{k'} + b_{k'})} \right]$

when, $k = l$

$$\nabla_b \hat{y}_k^{(i)} = \frac{\exp(x^T w_k + b_k) \sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'}) - \exp(x^T w_k + b_k) \exp(x^T w_l + b_l)}{\left[\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'}) \right]^2}$$

as $f(x) = \frac{g(x)}{h(x)}$ $\quad f'(x) = \frac{g'(x) h(x) - h'(x) g(x)}{[h(x)]^2}$

as $k = l$,

$$= \frac{\exp(x^T w_l + b_l) \sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'}) - \exp(x^T w_l + b_l)\exp(x^T w_l + b_l)}{\left[\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})\right]^2}$$

$$= \frac{\exp[x^T w_l + b_l]}{\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})} \left[1 - \frac{\exp(x^T w_l + b_l)}{\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})}\right]$$

$$\therefore \boxed{\nabla b\, \hat{y}_k^{(i)} = \hat{y}_l^{(i)} \left(1 - \hat{y}_l^{(i)}\right)}$$

and when $k \neq l$,

$$\nabla b\, \hat{y}_k^{(i)} = \nabla b \left[\frac{\exp(x^T w^{(k)} + b^{(k)})}{\sum_{k'=1}^{c} \exp(x^T w^{k'} + b^{k'})}\right]$$

$f(x) = g(x) / h(x)$ then, $f'(x) = \dfrac{g'(x)h(x) - h'(x)g(x)}{[h(x)]^2}$

$$\therefore \nabla b\, \hat{y}_k^{(i)} = \frac{\nabla b\left(\exp(x^T w_k^k + b_k) \cdot \sum_{k'=1}^{c} \exp(x^T w^{k'} + b^{k'})\right) - \nabla b\left(\sum_{k'=1}^{c} \exp(x^T w^{k'} + b^{k'}) \cdot \exp(x^T w_k^k + b_k)\right)}{\left[\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})\right]^2}$$

$$= \frac{0 - \exp(x^T w_k + b_k)\exp(x^T w_\ell + b_\ell)}{\left[\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})\right]^2}$$

$$= \frac{-\exp(x^T w_k + b_k)}{\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})} \cdot \frac{\exp(x^T w_\ell + b_\ell)}{\sum_{k'=1}^{c} \exp(x^T w_{k'} + b_{k'})}$$

$$\boxed{\underset{(k \neq \ell)}{\nabla_b \hat{y}_k^{(i)}} = -\hat{y}_k^{(i)} \hat{y}_\ell^{(i)}}$$

As $\sum_k a_k = a_\ell + \sum_{k \neq \ell} a_k$

Substituting ∇b values for $k = \ell$ & $k \neq \ell$,

$$\nabla_{b \hat{y}^{(i)}} f_{CE}(w,b) = \frac{-1}{n}\sum_{i=1}^{n}\left[\frac{y_\ell^{(i)}}{\hat{y}_\ell^{(i)}} \cdot \hat{y}_\ell^{(i)}(1 - \hat{y}_\ell^{(i)}) + \sum \frac{y_k^{(i)}}{\hat{y}_k^{(i)}}(-\hat{y}_k^{(i)}\hat{y}_\ell^{(i)})\right]$$

$$= \frac{-1}{n}\sum_{i=1}^{n}\left[\frac{y_\ell^{(i)}}{\hat{y}_\ell^{(i)}} \hat{y}_\ell^{(i)}(1 - \hat{y}_\ell^{(i)}) + \sum_{k \neq \ell}(-y_k^{(i)})\hat{y}_\ell^{(i)}\right]$$

$$= \frac{-1}{n}\sum_{i=1}^{n}\left[y_\ell^{(i)}(1 - \hat{y}_\ell^{(i)}) + \sum_{k \neq \ell}(-y_k^{(i)})\hat{y}_\ell^{(i)}\right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[y_\ell^{(i)} - \hat{y}_\ell^{(i)} y_\ell^{(i)} + \sum_{k \neq \ell} (-y_k)^{(i)} \hat{y}_\ell^{(i)} \right]$$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[y_\ell^{(i)} - \hat{y}_\ell^{(i)} \sum_{k=1}^{c} y_k^{(i)} \right]$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \left[y_\ell^{(i)} - \hat{y}_\ell^{(i)} \right]$$

$$\boxed{\nabla_b f_{CE}(w,b) = -\frac{1}{n} \sum_{i=1}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]}$$

Scanned by CamScanner