Sol^n 1 $\quad J(\omega) = \frac{1}{2n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2$

$\hat{y} = X^T\omega$

$\nabla_\omega J(\omega) = \cancel{\frac{1 \times 2}{2n}} \sum_{i=1}^{n} X(X^T\omega - y)$ — ①

Hessian for $J(\omega)$, (w.r.t. ω) — ②

$\frac{1}{n} XX^T$

According to Newton's method –

$\omega^* = \omega^{(0)} - H(f)(\omega^{(0)})^{-1} \nabla f(\omega^{(0)})$ — ③

Putting ① & ② in ③ $\quad$ ~~$f(\omega) = J(\omega)$~~

$\omega^* = \omega^{(0)} - \cancel{n}(XX^T)^{-1}\left[\frac{1}{\cancel{n}}(XX^T\omega^{(0)} - Xy)\right]$

$= \omega^{(0)} - (XX^T)^{-1}(XX^T)\omega^{(0)} + (XX^T)^{-1}Xy$

$\omega^* = \omega^{(0)} - \cancel{I} \omega^{(0)} + (XX^T)^{-1}Xy$

$(\because (XX^T)^{-1}(XX^T) = I)$

Hence, $\boxed{\omega^* = (XX^T)^{-1}Xy}$

Scanned by CamScanner

$sol^n 2$ Given that,

$$\hat{y}_K = \frac{e^{z_K}}{\sum_{K'=1}^{c} e^{z_{K'}}} \quad \& \quad z_K = x^T w^{(K)} + b$$

$$\nabla_{w^{(l)}} f_{CE}(W, b) = \frac{-1}{n} \sum_{i=1}^{n} \sum_{K=1}^{c} y_K^{(i)} \nabla_{w^{(l)}} \log \hat{y}_K^{(i)}$$

$\Rightarrow$ To find, $\nabla_{w^{(l)}} \hat{y}_K^{(i)}$ when $l = K$

$\hookrightarrow$ $$\nabla_{w^{(l)}} \hat{y}_K^{(i)} = \frac{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right) \nabla_w e^{z_K} - e^{z_K} \nabla_w \sum e^{z_{K'}}}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)^2}$$

$\because K = l$

$$\nabla_{w^{(l)}} \hat{y}_K^{(i)} = \frac{\left[\left(\sum_{K'=1}^{c} e^{z_{K'}}\right) e^{z_l} - \left(e^{z_K} e^{z_K}\right)\right]}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)^2} \nabla_w z^l$$

$$\boxed{\nabla_w z^l = x}$$

$$= \frac{x e^{z_l}}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)} - \frac{x e^{z_l^2}}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)^2}$$

$$= x \hat{y}_l - x \hat{y}_l^2$$

$$\boxed{\nabla_{w^{(l)}} \hat{y}_K^{(i)} = x^{(i)} \hat{y}_l^{(i)} \left(1 - \hat{y}_l^{(i)}\right)}$$

$\Rightarrow$ To find $\nabla_{w^{(l)}} \hat{y}_K^{(i)}$ when $l \neq K$.

$$= \frac{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right) \nabla_{w^{(l)}} e^{z_K} - \left[\left(e^{z_K} \nabla_{w^{(l)}} \sum e^{z_{K'}}\right)(\nabla z_{K'})\right.}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)^2}$$

$\nabla z^l = X$)

$\because K \neq l$ $\quad \nabla_{w^{(l)}} e^{z_K} = 0$ & $\left(\nabla_{w^{(l)}} \sum e^{z_K} = e^{z_l}\right)$

Hence

above equation becomes –

$$= \frac{0 - X e^{z_K} e^{z_l}}{\left(\sum_{K'=1}^{c} e^{z_{K'}}\right)^2}$$

$$\boxed{= - x^{(i)} \hat{y}_K^{(i)} \hat{y}_l^{(i)}}$$

$\Rightarrow$ $\nabla_{w^{(l)}} f_{CE}(W,b) = -\frac{1}{n} \sum_{i}^{n} \sum_{K}^{c} y_K^{(i)} \nabla_{w^{(l)}} \log \hat{y}_K^{(i)}$

From above 2 eq sol^{n}s we know $\nabla_{w^{(l)}} \hat{y}_K^{(i)}$ at $K = l$ & $K \neq l$.

Hence,

$$\frac{-1}{n} \sum_{i}^{n} \left[\sum_{K=l} \frac{y_K}{\hat{y}_K} \left(X \hat{y}_K - X \hat{y}_K \hat{y}_K \right) - \sum_{K \neq l} \frac{y_K x \hat{y}_K \hat{y}_l}{\hat{y}_K} \right]$$

$$= \frac{-1}{n} \sum_{i} \left(\sum_{K=l} y_K X - \sum_{K=l} X \hat{y}_K y_K \right) - \left(\sum_{K \neq l} y_K X \hat{y}_l \right)$$

$$= \frac{-1}{n} \left[\sum_{l=K} X y_l - \sum_{K=l} X \hat{y}_K y_l - \sum_{K \neq l} y_K X \hat{y}_l \right]$$

Scanned by CamScanner

On combining $k = l$ & $k \neq l$

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

$\because \sum_{k=1}^{c} y_k = 1$ because of softmax.

Hence,

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

$$\cancel{= -\frac{1}{n} x^{(i)}} \quad \boxed{= -\frac{1}{n} \sum_{i=1}^{n} x^{(i)} \left(y_l^{(i)} - \hat{y}_l^{(i)}\right)}$$

→ Using similar approach as previous parts, on differentiating ~~for~~ $\hat{y}_k$ with respect to b, $\nabla_{w^{(l)}} z^k$ becomes 1

Hence,

$\boxed{k = l}$

$$\nabla_{w^{(l)}} \hat{y}_k = \frac{\left[\sum_{k'=1}^{c} e^{z_{k'}} (e^{z_l}) - e^{z_l^2}\right] \nabla_w z^l}{\left(\sum_{k'=1}^{c} e^{z_{k'}}\right)^2}$$

$\boxed{\nabla_w z^l = 1}$

$$= \left(\hat{y}_l - \hat{y}_l^2\right)$$

For $\boxed{l \neq k}$

$$\nabla_{w^l} \hat{y}_k = \frac{0 - e^{z_k} e^{z_l}}{\left(\sum_{k'=1}^{c} e^{z_{k'}}\right)^2}$$

Scanned by CamScanner

For $l \neq k$, $= -\cancel{} \hat{y}_k \hat{y}_l$

Combining the 2 cases, we get–

$$\nabla_{w^{(l)}} f_{CE}(w, b) = \frac{-1}{n} \sum_{i=1}^{n} \left[\sum_{l=k} y_k \frac{(\hat{y}_l - \hat{y}_l^2)}{\hat{y}_k} - \sum_{l \neq k} y_k \frac{(\hat{y}_k \hat{y}_l)}{\hat{y}_k} \right]$$

Since $k=l$

$$= \frac{-1}{n} \sum_{i=1}^{n} \left[\sum_{l=k} y_l - \sum_{l \neq k} \hat{y}_k y_l - \sum_{l \neq k} y_k \hat{y}_l \right]$$

$\hat{y}_k = \hat{y}_l$

On combining $k=l$ & $k \neq l$,

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

$\sum_{k=1} y_k = 1$

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