Jay Yen
Rakesh Veetekat

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

$$= \frac{\sum_{k'=1}^{c} \exp z_{k'} \cdot \frac{\partial}{\partial w}\exp z_l - \exp z_l \cdot \frac{\partial}{\partial w}\sum_{k'=1}^{c} \exp z_{k'}}{\left(\sum_{k'=1}^{c} \exp z_{k'}\right)^2}$$

$$= \frac{\sum_{k'=1}^{c} \exp(x^T w^{(k')} + b_{k'}) \cdot \exp(x^T w^{(l)} + b_k) \cdot x^T - \exp(x^T w^{(l)} + b_k) \cdot \exp(z_l) x^T}{\left(\sum_{k'=1}^{c} \exp(x^T w^{(k')} + b_{k'})\right)^2}$$

$$= \frac{\exp(z_l) x^T}{\sum_{k'=1}^{c} \exp(z_{k'})} - \frac{x^T (\exp(z_l))^2}{\left(\sum_{k'=1}^{c} \exp(z_{k'})\right)^2}$$

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

$$\boxed{= x^{(i)} \hat{y}_l^{(i)} (1 - \hat{y}_l^{(i)})}$$

$\nabla_{w^{(l)}} = \nabla_{w^{(l)}} \frac{\exp(z_k)}{\sum_{k'=1}^{c} \exp z_{k'}}$ $\quad l \neq k$

$$= \frac{\sum_{k'=1}^{c} \exp z_{k'} \cdot \frac{\partial}{\partial w_l}\exp z_k - \exp z_k \cdot \frac{\partial}{\partial w}\sum_{k'=1}^{c} \exp z_{k'}}{\left(\sum_{k'=1}^{c} \exp z_{k'}\right)^2}$$

$$= \frac{\sum_{k'=1}^{c} \exp z_{k'} \cdot 0 - \exp z_k \cdot \exp z_l \cdot x^T}{\left(\sum_{k'=1}^{c} \exp z_{k'}\right)^2}$$

$$= -\hat{y}_k \cdot \hat{y}_l \cdot x^T$$

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

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

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

$$= -\frac{1}{n}\sum_{i=1}^{n} \frac{y_l^{(i)}}{\hat{y}_k^{(i)}} \cdot \left(x^{(i)}\hat{y}_l^{(i)}(1-\hat{y}_l^{(i)}) - \sum_{k\neq l}^{c} x^{(i)} \hat{y}_k^{(i)} \hat{y}_l^{(i)}\right)$$

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

$$= -\frac{1}{n}\sum_{i=1}^{n} y_l^{(i)} x^{(i)}(1-\hat{y}_l^{(i)}) - \sum_{k\neq l}^{c} y_k^{(i)} x^{(i)} \hat{y}_l^{(i)}$$

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

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

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

2. $-\log P(D|W,b) = -\log \sum_{i=1}^{n} P(y|x,W,b)$, where n is the number of x,y pairs in D

$$= -\log \sum_{i=1}^{n} \left(\prod_{k=1}^{c} \hat{y}_k^{y_k}\right)$$

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

$$\boxed{= -\sum_{i=1}^{n}\sum_{k=1}^{c} y_k \log \hat{y}_k}$$

4. a. $f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \left[x_1 - 2x_2 + \frac{x_3}{4}\right]$

$$x_1 - 2x_2 + \frac{x_3}{4} = \begin{bmatrix} w_{21} & w_{22} & w_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b_2$$

$$\boxed{\begin{aligned} W_2 &= \begin{bmatrix} 1 & -2 & \frac{1}{4} \end{bmatrix} \\ b_2 &= 0 \end{aligned}}$$

$$g\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ x_2 \\ -x_1 + 3 \end{bmatrix}$$

$$\begin{matrix} x_1 + 2x_2 \\ x_2 \\ -x_1 + 3 \end{matrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

$$\boxed{\begin{aligned} W_1 &= \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \\ b_1 &= \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \end{aligned}}$$

e. $\boxed{\frac{\partial z_2}{\partial \text{vec}[W_2]} \cdot \frac{\partial W_2}{\partial z_1} \qquad \frac{\partial z_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_1}}$

f. $f(g(x)) = W_2 g(x)$, so $z_1 = g(x)$

$$\frac{\partial (f \circ g)}{\partial z_1}(x) = \boxed{W_2}$$

g. $\frac{\partial (f \circ g)}{\partial b_1} = \boxed{\frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial b_1}}$

$$\nabla_{W_1} (f \circ g)(x) = \boxed{\frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial W_1}}$$