NAME: BARZA NISAR

COLLABORATORS: YASH KANT, PAVEL GOLIKOV

Q1) a) $h = \sigma(z_1)$

Since $h \in \mathbb{R}^d$ and σ does not change the dimension of its input vector, hence $\sigma : \mathbb{R}^d \to \mathbb{R}^d$ which means $\underline{z_1 \in \mathbb{R}^d}$ and $\underline{W^{(1)} \in \mathbb{R}^{d \times d}}$

$$\underset{d\times 1}{z_1} = \underset{d\times d}{W^{(1)}} \underset{d\times 1}{x}$$

$$\underset{d\times 1}{h} = \sigma(\underset{d\times 1}{z_1})$$

$$\underset{d\times 1}{z_2} = \underset{d\times 1}{h} + \underset{d\times 1}{x} \implies \underline{z_2 \in \mathbb{R}^d}$$

$$\underset{1\times 1}{y} = \underset{1\times d}{W^{(2)}} \underset{d\times 1}{z_2} \implies \underline{W^{(2)} \in \mathbb{R}^{1\times d}}$$

$$\underset{1\times 1}{L} = \frac{1}{2}(\underset{1\times 1}{y} - \underset{1\times 1}{t})^2$$

b) Number of parameters = Total number of elements in $W^{(1)}$ and $W^{(2)}$ $= \underline{d^2 + d}$

c) $$\bar{y} = \frac{\partial L}{\partial y} = (y-t)$$

$$\underset{1\times d}{\bar{W}^{(2)}} = \frac{\partial L}{\partial W^{(2)}} = \bar{y}\,\frac{\partial y}{\partial W^{(2)}} = \underset{1\times 1}{(y-t)}\;\underset{1\times d}{z_2^{T}}$$

$$\underset{d\times 1}{\bar{z}_2} = \frac{\partial L}{\partial z_2} = \bar{y}\,\frac{\partial y}{\partial z_2} = \underset{d\times 1}{W^{(2)T}}\;\underset{1\times 1}{(y-t)}$$

$$\underset{d\times 1}{\bar{h}} = \bar{z}_2\,\underset{\text{identity } d\times d}{\frac{\partial z_2}{\partial h}} = \underset{d\times 1}{W^{(2)T}}\;\underset{1\times 1}{(y-t)}$$

$$\underset{d\times 1}{\bar{z}_1} = \bar{h}\,\frac{\partial h}{\partial z_1} = \underset{d\times 1}{\bar{h}} \circ \underset{d\times 1}{\sigma'(z_1)}$$

element wise product

$$\underset{d\times d}{\bar{W}^{(1)}} = \bar{z}_1\,\frac{\partial z_1}{\partial W^{(1)}} = \underset{d\times 1}{\bar{z}_1}\;\underset{1\times d}{x^{T}}$$

$$\underset{d\times 1}{\bar{x}} = \bar{z}_1\,\frac{\partial z_1}{\partial x} = \underset{d\times d}{W^{(1)T}}\;\underset{d\times 1}{\bar{z}_1}$$

Q2) $y_k = \dfrac{e^{z_k}}{e^{z_1} + \dots + e^{z_k} + \dots + e^{z_N}} = \dfrac{e^{z_k}}{\sum_{j=1}^{N} e^{z_j}}$

Changing notation from question for clarity:-
Let $K = N$.

Let $g = e^{z_k}$ and $h = \sum_{j=1}^{N} e^{z_j}$

Then $\dfrac{\partial y_k}{\partial z_{k'}} = \dfrac{g'h - h'g}{h^2}$ (quotient rule)

$g' = \dfrac{\partial g}{\partial z_{k'}} = \begin{cases} e^{z_k} & \text{if } k = k' \\ 0 & \text{otherwise} \end{cases}$

$h' = \dfrac{\partial h}{\partial z_{k'}} = \dfrac{\partial (e^{z_1} + \dots + e^{z_N})}{\partial z_{k'}} = e^{z_{k'}}$

Case ① :- If $k = k'$, using quotient rule:-

$$\frac{\partial y_k}{\partial z_j} = \frac{e^{z_k} \sum_{j=1}^{N} e^{z_j} - e^{z_{k'}} e^{z_k}}{\left(\sum_{j=1}^{N} e^{z_j}\right)^2}$$

$$= \frac{e^{z_k} \left(\sum_j e^{z_j} - e^{z_{k'}} \right)}{\left(\sum_j e^{z_j} \right)^2}$$

$$= \frac{e^{z_k}}{\sum_j e^{z_j}} \left(\frac{\sum_j e^{z_j} - e^{z_{k'}}}{\sum_j e^{z_j}} \right)$$

$$= \frac{e^{z_k}}{\sum_j e^{z_j}} \left(1 - \frac{e^{z_{k'}}}{\sum_j e^{z_j}} \right)$$

$$= y_k \left(1 - y_{k'} \right)$$

$$\frac{\partial y_k}{\partial z_{k'}} = \underline{y_k \left(1 - y_k \right)} \quad \text{when } k = k'$$

Case (2) :- If $k \neq k'$

Using quotient rule again :-

$$\frac{\partial y_k}{\partial z_{k'}} = \frac{0 - e^{z_{k'}} e^{z_k}}{\left(\sum_{j=1}^{N} e^{z_j}\right)^2}$$

$$= - \frac{e^{z_{k'}}}{\sum_j e^{z_j}} \frac{e^{z_k}}{\sum_j e^{z_j}}$$

$$= - y_{k'} y_k$$

b) $$\mathcal{L}_{CE} = - \sum_{j=1}^{N} t_j \log\left(\frac{e^{z_j}}{e^{z_1} + \dots + e^{z_N}}\right)$$

Separating $j=k$ term from the sum :-

$$\mathcal{L}_{CE} = - \sum_{\substack{j'=1, \\ j' \neq k}}^{N} t_{j'} \log\left(\frac{e^{z_{j'}}}{e^{z_1} + \dots + e^{z_N}}\right) - t_k \log\left(\frac{e^{z_k}}{e^{z_1} + \dots + e^{z_N}}\right)$$

$$\mathcal{L}_{CE} = -\sum_{j'} t_{j'} \left[\log\left(e^{z_{j'}}\right) - \log\left(e^{z_1} + \cdots + e^{z_N}\right) \right]$$

$$- t_k \left[\log e^{z_k} - \log\left(e^{z_1} + \cdots + e^{z_N}\right) \right]$$

$$\frac{\partial L}{\partial z_k} = -\sum_{j'} t_{j'} \left[0 - \frac{e^{z_k}}{e^{z_1} + \cdots + e^{z_N}} \right]$$

$$- t_k \left[1 - \underbrace{\frac{e^{z_k}}{e^{z_1} + \cdots + e^{z_N}}}_{= y_k} \right]$$

$$= \sum_{j'} t_{j'} y_k - t_k + t_k y_k$$

$$= y_k \underbrace{\left(\sum_{j'} t_{j'} + t_k \right)}_{= 1 \ (t \text{ is a one hot vector})} - t_k$$

$$\frac{\partial L}{\partial z_k} = y_k - t_k$$

$$\frac{\partial z_k}{\partial w_k} = x$$

$$\frac{\partial L}{\partial w_k} = \frac{\partial L}{\partial z_k}\,\frac{\partial z_k}{\partial w_k} = \underline{\underline{(y_k - t_k)}}\,x$$