

1)a. Easy one. Just observe that $y_0 = 1$ and $y_w = 0$ for $w \neq 0$.

1)b.

Handwritten derivation of the derivative of the log-likelihood function with respect to the bias parameter v_c .

$$\frac{\partial \mathcal{L}}{\partial v_c} = \frac{\partial}{\partial v_c} \log \left(\frac{\exp(u_0^T v_c)}{\sum_w \exp(u_w^T v_c)} \right)$$

$$= \frac{\partial \log(\exp(u_0^T v_c))}{\partial v_c} + \frac{\partial \log(\sum_w \exp(u_w^T v_c))}{\partial v_c}$$

Chain rule:

$$= \frac{1}{\exp(u_0^T v_c)} \times u_0 \exp(u_0^T v_c) - \frac{1}{\sum_w \exp(u_w^T v_c)} \times \sum_w u_w \exp(u_w^T v_c)$$

$$= u_0 - \sum_w u_w \hat{y}_w$$

$$\frac{\partial \mathcal{L}}{\partial v_c} = -u_0 + U \hat{y} = -Ux + U\hat{y}$$

one-hot encoded, w/ non zero.

$$\frac{\partial \mathcal{L}}{\partial v_c} = U(\hat{y} - y)$$

SHOT ON REDMI 7
AI DUAL CAMERA

1)c.

$\delta \log(\sum_{w \in V} \exp(u_w^T v_c)) = \frac{\delta \log(\exp(u_w^T v_c))}{\delta u_w}$
 if $w = 0$:
 $= \frac{1}{u_w^T v_c}$

$\frac{\delta \tau}{\delta u_w} = \underbrace{\frac{\delta \log(\sum_{w \in V} \exp(u_w^T v_c))}{\delta u_w}}_{(1)} - \underbrace{\frac{\delta \log(\exp(u_w^T v_c))}{\delta u_w}}_{(2)}$

$(2) = \frac{\delta \log \exp(u_w^T v_c)}{\delta u_w} = \begin{cases} 0 & \text{if } w_i \neq 0 \\ v_c & \text{else} \end{cases}$

$(1) = \frac{1}{\sum_{w \in V} \exp(u_w^T v_c)} \times \frac{\delta \sum_{w \in V} \exp(u_w^T v_c)}{\delta u_w}$
 chain rule
 $= \frac{1}{\sum_{w \in V} \exp(u_w^T v_c)} \times v_c \exp(u_w^T v_c)$

SHOT ON REDMI 7
AI DUAL CAMERA

$$(\lambda) = v_c \times \hat{y}_{w_i}$$

$$\sum \frac{\partial \mathcal{L}}{\partial u_w} = \begin{cases} v_c \hat{y}_w, & w \neq 0 \\ (\hat{y}_w - 1) v_c, & w = 0 \end{cases}$$

1)d.

$$\sigma(x) = \frac{e^x}{e^x + 1}$$

$$\frac{\partial \sigma}{\partial x} = \frac{e^x(e^x + 1) - e^x e^x}{(e^x + 1)^2}$$

$$= \sigma(x) \cdot (1 - \sigma(x))$$

$$\boxed{\frac{\partial \sigma}{\partial x} = \sigma(x) (1 - \sigma(x))}$$

SHOT ON REDMI 7
AI DUAL CAMERA

1)e. u_0 :

$$\frac{\partial \mathcal{L}}{\partial u_0} = \frac{\partial \log(\sigma(u_0^T v_c))}{\partial u_0} - \sum_{i=1}^K \frac{\partial \log(\sigma(-u_i^T v_c))}{\partial u_0}$$

\searrow
 $\rightarrow 0$

$$= -v_c \times \frac{1}{\sigma(u_0^T v_c)} \times \underbrace{\sigma(u_0^T v_c)(1 - \sigma(u_0^T v_c))}_{\frac{\partial \sigma(x)}{\partial x}}$$

\uparrow
chain rule $\frac{\partial \log(x)}{\partial x}$

$$\frac{\partial \mathcal{L}}{\partial u_0} = v_c (\sigma(u_0^T v_c) - 1)$$

SHOT ON REDMI 7
AI DUAL CAMERA

v_c :

$$\frac{\partial \mathcal{L}}{\partial v_c} = \underbrace{\frac{\partial (-\log(\sigma(u_0^T v_c)))}{\partial v_c}}_{(1)} - \sum_{i=1}^K \underbrace{\frac{\partial \log(\sigma(-u_i^T v_c))}{\partial v_c}}_{(2)}$$

$$(1) = -u_0 \frac{1}{\sigma(u_0^T v_c)} \times \underbrace{\sigma(u_0^T v_c)(1 - \sigma(u_0^T v_c))}_{\frac{\partial \sigma(x)}{\partial x}}$$

\uparrow
 $\frac{\partial \log(x)}{\partial x}$

$$= u_0 (\sigma(u_0^T v_c) - 1)$$

REDMI 7
CAMERA

$$\begin{aligned}
 (2) &= -u_k \frac{1}{\cancel{\sigma(-u_k^T v_c)}} \cancel{\sigma(-u_k^T v_c)} (1 - \sigma(-u_k^T v_c)) \\
 &= u_k (\sigma(-u_k^T v_c) - 1) \\
 \text{So } \left[\frac{\partial \mathcal{L}}{\partial v_c} \right] &= u_0 (\sigma(u_0^T v_c) - 1) - \sum_{k=1}^K u_k (\sigma(-u_k^T v_c) - 1)
 \end{aligned}$$

uK :

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial u_k} &= - \frac{\cancel{\partial \log(\sigma(u_k^T v_c))}}{\cancel{\partial u_k}} - \sum_{i=1}^K \frac{\partial \log(\sigma(-u_i^T v_c))}{\partial u_k} \\
 &\quad \rightarrow 0 \\
 &= - \frac{\partial \log(\sigma(-u_k^T v_c))}{\partial u_k} \\
 &= + v_c \times \underbrace{\frac{1}{\sigma(-u_k^T v_c)}}_{\frac{\partial \log(1)}{\partial u_k}} \times \underbrace{\sigma(-u_k^T v_c) (1 - \sigma(-u_k^T v_c))}_{\frac{\partial \sigma(u)}{\partial u_k}} \\
 \left[\frac{\partial \mathcal{L}}{\partial v_c} \right] &= v_c (1 - \sigma(-u_k^T v_c))
 \end{aligned}$$

No need to sum over all words

1)f. Easy no calculus :

$$\begin{aligned}
 \text{a)} \quad \frac{\delta E_{\text{mean}}(u, u_{\text{in}}, u_{\text{out}}, U)}{\delta U} &= \sum_{\text{neuron } j \in \text{out}} \frac{\delta S(u_j, u_{\text{in}}, U)}{\delta U} \\
 \text{ii)} \quad \frac{\delta E_{\text{mean}}}{\delta \theta} &= \sum_{\text{neuron } j \in \text{out}} \frac{\delta E(u_j, u_{\text{in}}, U)}{\delta \theta_j} \\
 \text{iii)} \quad \underline{\hspace{2cm}} &= 0
 \end{aligned}$$

SHOTON REDMI 7
AI DUAL CAMERA