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

1)b.

$$\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial v_c}$$

$$\frac{\partial J}{\partial v_c} = \frac{\partial -\log\left(\frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}\right)}{\partial v_c}$$

$$= -\underbrace{\frac{\partial \log(\exp(u_o^T v_c))}{\partial v_c}}_{\text{chain rule} \; = \; \frac{1}{\cancel{\exp(u_o^T v_c)}} \times u_o \cancel{\exp(u_o^T v_c)}} + \underbrace{\frac{\partial \log\left(\sum_{w \in V} \exp(u_w^T v_c)\right)}{\partial v_c}}_{= \frac{1}{\sum_{w \in V} \exp(u_w^T v_c)} \times \sum u_w \exp(u_w^T v_c) \; = \; \sum u_w \times \hat{y}_w}$$

So: $\frac{\partial J}{\partial v_c} = -u_o + U\hat{y} = -U \times y + U\hat{y}$

one-hot encoded w/ 1 on o row.

$$\boxed{\frac{\partial J}{\partial v_c} = U(\hat{y} - y)}$$

1)c.

$$\frac{\partial J}{\partial u_w} = \frac{\partial \log(\sum_{w\in V} \exp(u_w v_c))}{\partial u_w} - \frac{\partial \log(\exp(\ldots))}{\partial u_w}$$

if $w = 0$: $= \frac{1}{u_o^T v_c}$

$$\frac{\partial J}{\partial u_{w_i}} = \underbrace{\frac{\partial \log\left(\sum_{w\in V} \exp(u_w^T v_o)\right)}{\partial u_{w_i}}}_{(1)} - \underbrace{\frac{\partial \log\left(\exp(u_o^T v_c)\right)}{\partial u_{w_i}}}_{(2)}$$

$$(2) \underset{\log\circ\exp = Id}{=} \frac{\partial\, u_o^T v_c}{\partial u_{w_i}} = \begin{cases} 0 & \text{if } w_i \neq 0 \\ v_c & \text{else} \end{cases}$$

$$(1) \underset{\text{chain rule}}{=} \frac{1}{\sum_{w\in V} \exp(u_w^T v_c)} \times \frac{\partial \sum_{w\in V} \exp(u_w^T v_c)}{\partial u_{w_i}}$$

$$= \frac{1}{\sum_{w\in V} \exp(u_w^T v_c)} \times v_c \exp(u_{w_i}^T v_c)$$

SHOT ON REDMI 7
AI DUAL CAMERA

$$(1) = v_c \times \hat{y}_{w_i}$$

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

1)d.

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

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

$$= \sigma(x) - \sigma(x)^2$$

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

1)e. u0 :

$$\frac{\partial J}{\partial u_o} = -\frac{\partial \log(\sigma(u_o^T v_c))}{\partial u_o} - \underbrace{\sum_{k=1}^{K} \frac{\partial \log(\sigma(-u_k^T v_c))}{\partial u_o}}_{\to 0}$$

$$\underset{\substack{\uparrow \\ \text{chain rule}}}{=} -v_c \times \underbrace{\frac{1}{\cancel{\sigma(u_o^T v_c)}}}_{\frac{\partial \log(x)}{\partial x}} \times \underbrace{\cancel{\sigma(u_o^T v_c)}(1-\sigma(u_o^T v_c))}_{\frac{\partial \sigma(x)}{\partial x}}$$

$$\boxed{\frac{\partial J}{\partial u_o} = v_c\left(\sigma(u_o^T v_c) - 1\right)}$$

Vc :

$$\frac{\partial J}{\partial v_c} = \underbrace{\frac{\partial(-\log(\sigma(u_o^T v_c)))}{\partial v_c}}_{(1)} - \underbrace{\sum_{k=1}^{K} \frac{\partial \log(\sigma(-u_k^T v_c))}{\partial v_c}}_{(2)}$$

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

$$= u_o\left(\sigma(u_o^T v_c) - 1\right).$$

$$(2) = -u_k \frac{1}{\cancel{\sigma(-u_k^T v_c)}} \cancel{\sigma(-u_k^T v_c)}\left(1-\sigma(-u_k^T v_c)\right)$$

$$= u_k\left(\sigma(-u_k^T v_c) - 1\right).$$

Soit $\boxed{\dfrac{\partial J}{\partial v_c} = u_o\left(\sigma(u_o^T v_c)-1\right) - \sum_{k=1}^{K} u_k\left(\sigma(-u_k^T v_c)-1\right)}$

uK :

$$\frac{\partial J}{\partial u_{k_i}} = -\underbrace{\frac{\partial \log(\sigma(u_o^T v_c))}{\partial u_{k_i}}}_{\to 0.} - \sum_{k=1}^{K} \frac{\partial \log(\sigma(-u_k^T v_c))}{\partial u_{k_i}}$$

$$= -\frac{\partial \log(\sigma(-u_{k_i}^T v_c))}{\partial u_{k_i}}$$

$$= +v_c \times \underbrace{\frac{1}{\sigma(-u_{k_i}^T v_c)}}_{\frac{\partial \log(x)}{\partial x}} \times \underbrace{\sigma(-u_{k_i}^T v_c)\left(1-\sigma(-u_{k_i}^T v_c)\right)}_{\frac{\partial \sigma(x)}{\partial x}}$$

$$\boxed{\frac{\partial J}{\partial u_{k_i}} = v_c\left(1-\sigma(-u_{k_i}^T v_c)\right)}$$

No need to sum over all vocab

1)f. Easy no calculus :