Q 2)

(a) $\sigma(x) = \dfrac{1}{1+e^{-x}} \rightarrow$ 1

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

$$= \frac{-(0-e^{-x})}{(1+e^{-x})^2}$$

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

$\longrightarrow$ add and subtract 1 to numerator

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

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

$$= (1+e^{-x})^{-1}\left(1-(1+e^{-x})^{-1}\right)$$

$$= \sigma(x)\left(1-\sigma(x)\right)$$

// Ans.

2.(b)

$$\frac{\partial \mathcal{L}}{\partial v_c} = \frac{\partial}{\partial v_c}\left[-\ln \frac{e^{u_o^T v_c}}{\sum_{w=1}^{W} u_w^T v_c}\right]$$

$$= \frac{\partial}{\partial v_c}\left[u_o^T v_c - \ln\left(\sum_{w=1}^{W} e^{u_w^T v_c}\right)\right]$$

$$= -\left[u_o^T - \frac{1}{\sum e^{u_w^T v_c}} \frac{\partial \sum e^{u_w^T v_c}}{\partial v_c}\right]$$

$$= -\left[u_o^T - \sum u_w^T \hat{y}_w\right]$$

$$= \left(\sum_{w=1}^{W} u_w^T \hat{y}_w\right) - u_o^T$$

2(c). $$\frac{\partial \mathcal{L}}{\partial u_w} = \begin{cases} \text{(i)}, & w = o \\ \text{(ii)} & w \neq o \end{cases}$$

(i) $$\frac{\partial \mathcal{L}}{\partial u_o} = -\frac{\partial}{\partial u_o}\left[u_o^T v_c - \ln\left(\sum e^{u_w^T v_c}\right)\right]$$

$$= -\left[v_c - \frac{1}{\sum e^{u_w^T v_c}} \cdot \frac{\partial \sum e^{u_w^T v_c}}{\partial u_o}\right]$$

$$= -\left[v_c - v_c \hat{y}_o\right] = v_c\left[\hat{y}_o - 1\right]$$

(ii) $$\frac{\partial \mathcal{L}}{\partial u_w} = -\frac{\partial}{\partial u_w}\left[u_o^T v_c - \ln\left(\sum e^{u_w^T v_c}\right)\right]$$

$$= -\left[0 - \frac{1}{\sum e^{u_w^T v_c}} \cdot \frac{\partial \sum e^{u_w^T v_c}}{\partial u_w}\right]$$

$$= -\left[-\hat{y}_w\right] = \hat{y}_w$$

Scanned with CamScanner

2(d).

$$\frac{\partial \mathcal{L}}{\partial v_c} = \frac{\partial}{\partial v_c}\left[-\ln\left(\sigma(u_o^T v_c)\right) - \sum_{k=1}^{K} \ln\left(\sigma(-u_k^T v_c)\right)\right]$$

$$= \left[\frac{1}{\cancel{\sigma(u_o^T v_c)}} \cdot (\cancel{\sigma(u_o^T v_c)}(1-\sigma(u_o^T v_c)))u_o^T - \sum_{k=1}^{K} (1-\sigma(-u_k^T v_c))u_k^T\right]$$

$$= \sum_{k=1}^{K} u_k^T(1-\sigma(-u_k^T v_c)) - (1-\sigma(u_o^T v_c))u_o^T$$

$$\frac{\partial \mathcal{L}}{\partial u_w} = \begin{cases} (i), & w=o \\ (ii), & w\neq o \end{cases} = \begin{cases} -1(1-\sigma(u_o^T v_c))v_c, & w=o \\ 1-\sigma(-u_w^T v_c))v_c, & w\neq o \end{cases}$$

$$2(e).\quad J = \sum_{-m \le j \le m,\, j \ne 0} F(w_{c+j}, v_c)$$

where $F(w_{c+j}, v_c)$ can be J_{CE} or $J_{neg\text{-}sample}$

(i) $$\frac{\partial J}{\partial U = [u_1, \dots u_W]} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial u} = \frac{\partial J}{\partial u_w} \quad (CE \text{ or neg-sam})$$

↙ from 2(c) & (d)

(ii) $$\frac{\partial J}{\partial v_c} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial v_c} = \frac{\partial J}{\partial v_c} \quad (CE \text{ or neg-samp})$$

↙ from 2(b) & (d)

$$\therefore \nabla J = \begin{cases} v_c[\hat{y}_o - 1], & w = o, \text{ CE w.r.t } u \\ \hat{y}_w, & w \ne o, \text{ CE w.r.t } u \\ (\sigma(u_o^T v_c) - 1) v_c, & w = o, \text{ neg-Samp w.r.t } u \\ (1 - \sigma(u_o^T v_c)) v_c, & w \ne o, \text{ neg Samp w.r.t. } u \\ \sum_{w=1}^{W} u_w^T \hat{y}_w - u_o^T, & \text{CE, w.r.t. } v_c \\ \sum u_k^T (1 - \sigma(-u_k^T v_c)) - (1 - \sigma(u_k^T v_c)) u_o^T, & \text{neg-Samp, w.r.t } v_c \end{cases}$$