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

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

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

$$= \frac{1}{1+e^{-x}} \left[\frac{1 + e^{-x} - 1}{1 + e^{-x}} \right]$$

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

$$\frac{\partial \mathcal{L}}{\partial v_c} = \frac{\partial}{\partial v_c}\left[-\ln \frac{e^{u_o^T v_c}}{\sum_1^W e^{u_w^T v_c}}\right]$$

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

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

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

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

$$\frac{\partial \mathcal{L}}{\partial u_w} = \frac{\partial \mathcal{L}}{\partial U} = \begin{cases} \text{Ⓐ}, & w=o \\ \text{Ⓑ}, & o/w \end{cases} = \begin{cases} v_c[\hat{y}_o - 1], & w=o \\ \hat{y}_w, & o/w \end{cases}$$

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

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

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

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

$$= -\left[0 - \frac{1}{\text{''}} \frac{\partial}{\partial u_w} \text{''}\right]$$

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

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

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

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

$$\frac{\partial \mathcal{L}}{\partial u_w} = \begin{cases} Ⓐ, & w=o \\ Ⓑ, & o/w \end{cases} = \begin{cases} -\left(1-\sigma(u_o^T v_c)\right) v_c, & w=o \\ \left(1-\sigma(-u_w^T v_c)\right) v_c, & o/w \end{cases}$$

• Add ... conv ...

• matlab.engine

Given $J = \sum_{-m \le j \le m,\, j \neq 0} F(w_{c+j}, v_c)$

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

Then, $$\frac{\partial J}{\partial \xi} = \begin{cases} ⓐ, & \xi \in U = [u_1, \dots u_w] \\ ⓑ, & \xi = v_c \\ 0, & \text{o/w} \end{cases}$$

$$ⓐ = \frac{\partial}{\partial U} \sum F = \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial U} \cdot \frac{\partial J}{\partial U_w} \text{ (smx or neg-samp)}$$

↳ from part a) & d)

$$ⓑ = \frac{\partial}{\partial v_c} \sum F = \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial v_c} \cdot \frac{\partial J}{\partial v_c} \text{ (smx or neg-samp)}$$

↳ from part a) & d)

$$\frac{\partial J}{\partial \xi} = \begin{cases} v_c[\hat{y}_o - 1], & w = o, \text{ smx CE}, \xi \in U \\ \hat{y}_w, & \text{o/w}, \text{ smx CE}, \xi \in U \\ -(1 - \sigma(u_o^T v_c)) v_c, & w = o, \text{ neg-CE}, \xi \in U \\ (1 - \sigma(-u_o^T v_c)) v_c, & \text{o/w}, \text{ neg-CE}, \xi \in U \\ \sum_{w=1}^{W} u_w^T \hat{y}_w - u_o^T, & \text{smx CE}, \xi = v_c \\ \sum u_k^T (1 - \sigma(-u_k^T v_c)) - (1 - \sigma(u_o^T v_c)) u_o^T, & \text{neg-CE}, \xi = v_c \\ 0, & \text{o/w} \end{cases}$$