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

$$\frac{\partial}{\partial x}\sigma(x) = \frac{\partial}{\partial x}\left[\frac{1}{1+e^{x}}\right]$$

$$\frac{\partial}{\partial x}\left[(1+e^{-x})^{-1}\right]$$

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

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

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

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

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

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

2.) $\mathcal{L} = -\sum_i y_i \log(\hat{y}_i)$

$$= -\sum_i y_i \log(p(o|c))$$

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

$$= -\frac{\partial}{\partial v_c}\left[u_o^T v_c - \log\left(\sum_1^W e^{u_w^T v_c}\right)\right]$$

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

$$= -\left[u_o^T - \frac{1}{\sum_{w=1}^{W} \exp(u_w^T v_c)} \cdot \sum \exp(u_w^T v_c) \cdot u_w^T\right]$$

$$= -\left[u_o^T - \frac{\sum_1^W \exp(u_w^T \cdot v_c)}{\sum_{w=1}^{W} \exp(u_w^T v_c)} \cdot u_w^T\right]$$

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

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

$$= v \cdot \hat{y} - v \cdot y$$

$$= u(\hat{y} - y)$$

3) $$\frac{\partial L}{\partial v_w} = \frac{\partial L}{\partial v} = \begin{cases} v_c[\hat{y_o} - 1], & w = o \\ \hat{y}_w & \text{otherwise} \end{cases}$$

$$\frac{\partial L}{\partial v_o} = \frac{-\partial}{\partial v_o}\left[u_o^T \cancel{v_c} - \log\left(\sum \exp(u_w^T \cdot v_c)\right) \right]$$

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

$$= -\left[v_c - \frac{1}{\sum exp(u_w^T v_c)} \cdot \sum exp(u_w^T v_c) v_c \right]$$

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

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

$$\frac{\partial L}{\partial u_w} = \frac{-\partial}{\partial u_w} \left[u_o^T v_c - \log\left(\sum exp(u_w^T \cdot v_c) \right) \right]$$

$$= -\left[0 - \frac{1}{\sum exp(u_w^T \cdot v_c)} \cdot \sum exp(u_w^T \cdot w) v_c \right]$$

$$= -\left[0 - \hat{y} w \right]$$

$$= \hat{y} w.$$

4) $J_{neg\text{-}sample}(o, v_c, U) =$

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

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

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

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

$$= \begin{cases} -(1-\sigma(u_o^T v_c))\, v_c, & w=0 \\ (1-\sigma(-u_w^T v_c))\, v_c, & O/W \end{cases}$$

ives

giv

e) Let U be the collection of all output vectors for all words in the vocabulary.

The derivatives for a given cost function F are given as follows,

$$\frac{\partial F(w_i, \hat{v})}{\partial U}, \quad \frac{\partial F(w_i, \hat{v})}{\partial \hat{v}}$$

∴ the gradients for the cost function of one context window are,

$$\frac{\partial J_{skip\text{-}gram}(word_{c-m, \dots c+m})}{\partial U}$$

$$= \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial F(w_{c+j}, v_c)}{\partial U},$$

$$\frac{\partial J_{skip\text{-}gram}(word_{c-m,\dots c+m})}{\partial v_c}$$

$$= \sum_{-m \le j \le m,\, j \ne 0} \frac{\partial F(w_{c+j}, v_c)}{\partial v_c},$$

$$\frac{\partial J_{skip\text{-}gram}(word_{c-m \dots\dots c+m})}{\partial v_j} = 0$$

for all $j \neq c$