1) a) در معادله cross entropy loss، y_w زمانی یک می‌شود که w کلمه پیش بینی شده باشد

$$\Rightarrow -\sum_w y_w \log(\hat{y}_w) = -[y_1 \log(\hat{y}_1) + \dots + y_o \log(\hat{y}_o) + \dots + y_w \log(\hat{y}_w)]$$

$$= -y_o \log(\hat{y}_o) = -\log(\hat{y}_o)$$

1b)

$$\frac{\partial J}{\partial v_c} = \frac{\partial(-\log P(O=o|C=c))}{\partial v_c} = \frac{\partial\left(-\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)}\right)}{\partial v_c}$$

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

می‌دانیم که اگر $y = \log_a u$ باشد آنگاه $y' = \frac{u'}{u} \ln a$

$$= -\frac{1}{\exp(u_o^T v_c)} \cdot \frac{\partial \exp(u_o^T v_c)}{\partial v_c} + \frac{1}{\sum_w \exp(u_w^T v_c)} \cdot \frac{\partial \sum_w \exp(u_w^T v_c)}{\partial v_c}$$

$$= \frac{-1}{\exp(u_o^T v_c)} \exp(u_o^T v_c)\, u_o + \sum_w \underbrace{\frac{\exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)}}_{P(O=o|C=c)} u_w$$

$$= -u_o + \sum_w P(O=o|C=c)\, u_w = u^T(\hat{y}-y)$$

1c)

$$\frac{\partial J}{\partial v_m} = \frac{\partial\left(-\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)}\right)}{\partial v_m} = -\left(\frac{\partial \log \exp(u_o^T v_c)}{\partial v_m} - \frac{\partial \log \sum_w \exp(v_m^T v_c)}{\partial v_m}\right)$$

$$\xrightarrow{w \neq o} = 0 + \frac{\exp(u_m^T v_c)}{\sum_w \exp(u_w^T v_c)} v_c = P(O=o|C=c)\, v_c$$

$$\xrightarrow{w = o} = \frac{\exp(u_m^T v_c)}{\sum_w \exp(u_m^T v_c)} v_c = v_c(P(O=o|C=c)-1)$$

subject: ۹۴۴۷۲۱۰۴ - محمد یارمقدم date:

$$\Rightarrow \frac{\partial J}{\partial v_m} = (\hat{y}-y)^T v_c$$

d) $$\frac{\partial J}{\partial U} = \left[\frac{\partial J(v_c,o,U)}{\partial u_1}, \dots, \frac{\partial J(v_c,o,U)}{\partial u_{|w\in vocab|}} \right]$$

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

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

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

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

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

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

چون مستقل از u_o است پس ۰ میشود

$$\frac{\partial J}{\partial u_k} = \underbrace{\frac{\partial(-\log(\sigma(u_o^T v_c)))}{\partial u_k}}_{0} - \sum_{k=1}^{K} \frac{\sigma \log(\sigma(-u_k^T v_c))}{\partial u_k} = -\sum_{k=1}^{K} \frac{\sigma' \log(\sigma(-u_k^T v_c))}{\partial u_k}$$

مانند مورد بالا صفر میشود

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

(h

$$J = J_{skip-gram}(v_c, w_{t-m}, \dots, w_{t+m}, U).$$

$$\frac{\partial J}{\partial U} = \sum_{-m<j<m} \frac{\partial J}{\partial U} = \sum_{-m<j<m} v_c (\hat{y}-y)^T$$

$$\frac{\partial J}{\partial v_c} = \sum_{-m<j<m} \frac{\partial J}{\partial v_c} = \sum_{-m<j<m} U^T (\hat{y}-y)^T$$

$$\frac{\partial J}{\partial v_m} = \sum_{\substack{-m<j<m \\ w \neq c}} \frac{\partial J}{\partial v_m} = 0$$

(g

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

$$= 0 - \frac{\partial(\log(\sigma(u_1^T v_c)) + \dots + \log(\sigma(u_k^T v_c)))}{\partial u_k}$$

$$= \frac{-k(-v_c)}{\sigma(-u_k^T v_c)} \sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c)) = k(v_c)(1-\sigma(-u_k^T v_c))$$