'outside' vectors u_w center vecs v_w

$$U = \begin{pmatrix} \; \end{pmatrix} \qquad V = \begin{pmatrix} \; \end{pmatrix}$$

loss: $J(v_c, o, U) = -\log P(O=o \mid C=c)$

(a)

$$-\sum_{w \in Vocab} y_w \log(\hat{y}_w) = -\log(\hat{y}_o)$$

$y = (0, \cdots 1, \cdots 0)$ one-hot vector.

(b)

$$J(v_c, o, U) = -\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)} = -u_o^T v_c + \log \sum_w \exp(u_w^T v_c)$$

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

$$= -u_o + \sum_k u_k \cdot \frac{\exp(u_k^T v_c)}{\sum_w \exp(u_w^T v_c)} \rightarrow \text{constant.}$$

$$= -u_o + \sum_k \hat{y}_k u_k$$

....

(c) when $w = o$

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

$$= \left(\frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)} - 1 \right) v_c = (\hat{y}_o - 1) v_c$$

when $w \neq o$

$$\frac{\partial J(v_c, o, U)}{\partial u_w} = \frac{v_c \exp(u_w^T v_c)}{\sum_k \exp(u_k^T v_c)} = v_c \cdot \hat{y}_{w \neq o}$$

(d)

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

(e)

$$J_{neg}(v_c, o, U) = -\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} \log(\sigma(-u_k^T v_c))$$

$$\frac{\partial J_{neg}}{\partial v_c} = -\frac{1}{\cancel{\sigma(u_o^T v_c)}} \cdot \cancel{\sigma(u_o^T v_c)} (1 - \sigma(u_o^T v_c)) \cdot u_o + \sum_{k=1}^{K} (1 - \sigma(-u_k^T v_c)) u_k$$

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

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

$$\frac{\partial J_{neg}}{\partial u_k} = +\left(1-\sigma(-u_k^T v_c)\right) v_c$$

$O(k)$ vs $O(|V|)$

The Negtive Sampling method is computationally simpler.

(f) (i) $\displaystyle\sum_{\substack{-m\le j\le m \\ j\ne 0}} \frac{\partial J(v_c, w_{t+j}, U)}{\partial U}$

(ii) $\displaystyle\sum_{\substack{-m\le j\le m \\ j\ne 0}} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c}$

(iii) 0.