< Word2Vec Homework 1 > - NLP

Computer Software Engineering
KIM SUNG HYUK (김성혁)
2020081421

$\boxed{(a)}$ let. [$y_k = 1$ (o is k^{th} entry of vector y), then $y_{\neq k} = 0$.
except

$\because$ y is one-hot vector.

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

$(\because y_k = y_o)$

$\boxed{(b)}$ $J(v_c, o, U) = -\log P(O=o \mid C=c) = -\log \left\{ \dfrac{\exp(u_o^T v_c)}{\sum_{w \in Vocab} \exp(u_w^T v_c)} \right\}$

$$= \underbrace{-u_o^T v_c}_{\text{①}} + \underbrace{\log \left\{ \sum_{w \in Vocab} \exp(u_w^T v_c) \right\}}_{\text{②}}$$

i) $\dfrac{\partial ①}{\partial v_c} = \dfrac{\partial u_o^T v_c}{\partial v_c} = u_o$

ii) $\dfrac{\partial ②}{\partial v_c} = \dfrac{\partial}{\partial v_c} \left\{ \log \sum_{w \in Vocab} \exp(u_w^T v_c) \right\} = \dfrac{\sum_{x=1}^{V} \exp(u_x^T v_c) \cdot u_x}{\sum_{w \in Vocab} \exp(u_w^T v_c)}$

$$= \sum_{x=1}^{V} P(x \mid c) \cdot u_x$$

$$\therefore \frac{\partial J}{\partial v_c} = \sum_{x=1}^{V} P(x \mid c) \cdot u_x - u_o$$

$$= \sum_{x=1}^{V} \hat{y}_x \cdot u_x - u_o$$

(where u_x is x^{th} column of U)

$$= U \hat{y} - u_o$$

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

$$U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_{|V|} \\ | & & | \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_{|V|} \end{bmatrix} \qquad y = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

(c)

i) $w = o$ (actual outside word)

$$\frac{\partial ㉠}{\partial u_o} = v_c$$

$$\frac{\partial ㉡}{\partial u_o} = \frac{\exp(u_o^T v_c) \cdot v_c}{\sum_{w \in Vocab} \exp(u_w^T v_c)}$$

$$\therefore \frac{\partial J}{\partial u_o} = \frac{\exp(u_o^T v_c)}{\sum_{w \in Vocab} \exp(u_w^T v_c)} \cdot v_c - v_c = P(O|C) \cdot v_c - v_c$$

$$= \boxed{\hat{y}_o v_c - v_c}$$

ii) $w \neq o$

$$\frac{\partial ㉠}{\partial u_w} = 0$$

$$\frac{\partial ㉡}{\partial u_w} = \frac{\exp(u_w^T v) \cdot v_c}{\sum_{x \in Vocab} \exp(u_x^T v_c)} = P(w|c) \cdot v_c$$

$$\therefore \boxed{\frac{\partial J}{\partial u_w} = \hat{y}_w v_c}$$

(d) $$\frac{\partial J(v_c, o, U)}{\partial U} = \left[\underline{\hat{y}_1 v_c} \quad \underline{\hat{y}_2 v_c} \quad \cdots \quad \underline{\hat{y}_o v_c - v_c} \quad \cdots \quad \underline{\hat{y}_{|V|} v_c} \right]$$

only different term

(e)

i) $$J_{neg}(v_c, o, U) = \underbrace{-\log(\sigma(u_o^T v_c))}_{㉠} - \underbrace{\sum_{s=1}^{K} \log(\sigma(-u_{w_s}^T v_c))}_{㉡}$$

$$\frac{\partial ㉠}{\partial v_c} = \frac{\cancel{\sigma(u_o^T v_c)}}{\cancel{\sigma(u_o^T v_c)}} (1 - \sigma(u_o^T v_c)) \cdot u_o \quad \left(\because \frac{d\sigma(x)}{dx} = \sigma(x)(1-\sigma(x))\right)$$

$$\frac{\partial ㉡}{\partial v_c} = \sum_{s=1}^{K} \frac{\cancel{\sigma(-u_{w_s}^T v_c)}}{\cancel{\sigma(-u_{w_s}^T v_c)}} (1 - \sigma(-u_{w_s}^T v_c))(-u_{w_s})$$

1) $$\therefore \frac{\partial J_{neg}}{\partial v_c} = u_o \sigma(u_o^T v_c) - u_o + \sum_{s=1}^{K} (u_{w_s} - u_{w_s} \sigma(-u_{w_s}^T v_c))$$

$$= u_o(\sigma(u_o^T v_c) - 1) + \sum_{s=1}^{K} \sigma(u_{w_s}^T v_c) \cdot u_{w_s}$$

— ② —

case ①
$w = o$
outside word.

$$\frac{\partial ①}{\partial u_o} = (1 - \sigma(u_o^T v_c)) \cdot v_c$$

$$\frac{\partial ②}{\partial u_o} = 0$$

2) $$\boxed{\frac{\partial J_{neg}}{\partial u_o} = -v_c(1 - \sigma(u_o^T v_c))}$$ //

case ②
$w \neq o$

$$\frac{\partial ①}{\partial u_w} = 0$$

$$\frac{\partial ②}{\partial u_w} = (1 - \sigma(-u_w^T v_c)) \cdot (-v_c)$$

3) $$\boxed{\frac{\partial J_{neg}}{\partial u_w} = v_c - v_c\,\sigma(-u_w^T v_c)}$$

where $w \in \{w_1, \cdots, w_k\}$

(e) -ⅰ)

Compute only k number of $u_{w_s}^T v_c$ is much more efficient than compute all $|V|$ of $u_w^T v_c$ //

(f)

(i) $\dfrac{\partial J_{\text{skip-gram}}(v_c, w_{t-m}, \cdots, w_{t+m}, U)}{\partial U}$

$$= \cancel{\left[2m\hat{y}_1 v_c \quad 2m\hat{y}_2 v_c \quad \cdots \quad 2m\hat{y}_{w_{t-m}} v_c - v_c \quad \cdots \quad 2m\hat{y}_{w_{t+m}} v_c - v_c \quad \cdots \quad 2m\hat{y}_{|V|} v_c \right]}$$

$$\boxed{= \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial J_{\text{skip}}(v_c, w_{t+j}, U)}{\partial U}}$$ (∵ using ~~result of~~ (c))

(ii) $\dfrac{\partial J(v_c, w_{t-m}, \cdots, w_{t+m}, U)}{\partial v_c}$ $\boxed{= \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c}}$

$$= \cancel{\left(\sum_{x=1}^{V} \hat{y}_x u_x - u_{w_{t-m}} \right) + \left(\sum_{x=1}^{V} \hat{y}_x u_x - u_{w_{t-m+1}} \right) + \cdots + \left(\sum_{x=1}^{V} \hat{y}_x u_x - u_{w_{t+m}} \right)}$$

$$= \cancel{2m \sum_{x=1}^{V} \hat{y}_x u_x - \sum_{-m \le j \le m,\, j \neq 0} u_{w_{t+j}}}$$ (∵ using result of (b))

(iii) $\boxed{0}$ (when $w \neq c$)