CS224n a2 written

(a) ∵ if w is a outside word, $y_w = 1$

and if w is not a outside word, $y_w = 0$

∴ left $= -\sum_{w=o} \log(\hat{y}_w) = -\log(\hat{y}_o)$

(b)
$$\frac{\partial J}{\partial v_c} = \frac{\partial}{\partial v_c}\left(-\log \frac{\exp(u_o^T v_c)}{\sum_{w \in W} \exp(u_w^T v_c)}\right)$$

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

term 1 $= -u_o^T$

$$\text{term 2} = \frac{1}{\sum_{w \in W} \exp(u_w^T v_c)} \cdot \frac{\partial}{\partial v_c}\left(\sum_{w \in W} \exp(u_w^T v_c)\right)$$

$$= \frac{1}{\sum_{w \in W} \exp(u_w^T v_c)} \sum_{x \in W} \exp(u_x^T v_c) \cdot u_x^T$$

$p(O = x \mid C = c)$

$$= \sum_{x \in W} \boxed{\frac{\exp(u_x^T v_c)}{\sum_{w \in W} \exp(u_w^T v_c)}} u_x^T$$

$$= \sum_{x \in W} p(x \mid c)\, u_x^T$$

$$= \sum_{x \in W} \hat{y}_x u_x^T$$

$$\frac{\partial J}{\partial v_c} = -u_o^T + \sum_{x \in W} \hat{y}_x u_x^T = -u_o^T + \hat{y} U^T$$

(c) $\frac{\partial J}{\partial u_w} = \frac{\partial}{\partial u_w}(-u_o^T v_c) + \frac{\partial}{\partial u_w}\left(\log \sum_{w \in W} \exp(u_w^T v_c)\right)$

① $w = 0$:

$$\frac{\partial J}{\partial u_o} = -v_c^T + \hat{y}_o v_c^T \qquad \hat{y}_o = p(o|c)$$

② $w \neq 0$

$$\frac{\partial J}{\partial u_w} = \hat{y}_w v_c^T \qquad \hat{y}_w = p(w|c)$$

(d) $\frac{\partial J}{\partial U} = \frac{\partial J}{\partial u_1}, \frac{\partial J}{\partial u_2}, \cdots \frac{\partial J}{\partial u_{|vocab|}}$

$= \hat{y}_1 v_c^T, \hat{y}_2 v_c^T, \cdots -v_c^T + y_{o_1} v_c^T \cdots, -v_c^T + y_{o_2} v_c^T$

$\cdots\cdots$

(e) $f'(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$

(f) $f'(x) = \frac{(e^x+1)e^x - e^x \cdot e^x}{(e^x+1)^2} = \frac{e^x}{(e^x+1)^2} = \frac{e^x}{e^x+1}\left(1 - \frac{e^x}{e^x+1}\right)$

$= \sigma(x)(1 - \sigma(x))$

(g) (i)

$$\frac{\partial J}{\partial V_c} = \frac{\partial}{\partial V_c}\Big[\underbrace{-\log\big(\sigma(u_o^T V_c)\big)}_{\text{term 1}}\Big] \underbrace{- \frac{\partial}{\partial V_c}\sum_{s=1}^{K}\log(\sigma(-u_{w_s}^T V_c))}_{\text{term 2}}$$

$$\text{term 1} = -\frac{1}{\sigma(u_o^T V_c)} \cdot \sigma(u_o^T V_c)\,[1-\sigma(u_o^T V_c)] \cdot u_o^T$$

$$= [\sigma(u_o^T V_c) - 1] \cdot u_o^T$$

$\sigma(x) + \sigma(-x) = 1$

$$\text{term 2} = -\frac{\partial}{\partial V_c}\sum_{s=1}^{K}\log(\sigma(-u_{w_s}^T V_c))$$

$$= -\sum_{s=1}^{k}\frac{1}{\sigma(-u_{w_s}^T V_c)} \cdot \sigma(-u_{w_s}^T V_c)\,\underbrace{[1-\sigma(-u_{w_s}^T V_c)]}\,(-u_{w_s}^T)$$

$$= +\sum_{s=1}^{k}\sigma(\underbrace{u_{w_s}^T V_c})\,(+u_{w_s}^T)$$

(num,)

$$\frac{\partial J}{\partial V_c} = \underbrace{[\sigma(u_o^T V_c) - 1]\,u_o^T}_{(1,\ \text{vec})} + \sum_{s=1}^{K}\underbrace{\sigma(u_{w_s}^T V_c)}_{(\text{idx},\ )} \cdot u_{w_s}^T$$

(idx, vec) (vec,)

(idx,) (idx, vec)

$$\frac{\partial J}{\partial u_o} = [\sigma(u_o^T V_c) - 1]\,V_c^T$$

$$\frac{\partial J}{\partial u_{w_s}} = \sigma(u_{w_s}^T V_c)\,V_c^T$$

(ii) $\frac{dJ}{d\log} \cdot \frac{d\log}{d(u_k^T v_c)}$ ← reuse part.

(iii) The derivative of sigmoid can be written in terms of sigmoid (?).

(h) $\frac{\partial J}{\partial U_{ws}} = i \cdot \sum_{\substack{1 \le j \le k \\ w_j = w_s}} \sigma(U_w \cdot V_c) V_c^T$ (i = # of $w_j = w_s$ in K words)

(i) (i) $\frac{\partial J}{\partial U} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial J(V_c, W_{t+j}, U)}{\partial U}$

$\frac{\partial J}{\partial V_c} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial J(V_c, W_{t+j}, U)}{\partial V_c}$

$\frac{\partial J}{\partial V_w} = 0$.

Observation:

man and woman are close > strange.

male and female are not

= "king - queen" is parallel to "male - female"