Extra Credit #2 ①

$$J(v_c, o, U) = -\log(\hat{y}_o)$$

$$= -\log\left(\frac{\exp(u_o^T v_c)}{\sum_{w=1}^{V} \exp(u_w^T v_c)}\right)$$

$$= -\left(\log(\exp(u_o^T v_c)) - \log\left(\sum_{w=1}^{V} \exp(u_w^T v_c)\right)\right)$$

$$= -u_o^T v_c + \log\left(\sum_{w=1}^{V} \exp(u_w^T v_c)\right)$$

Taking derivative w.r.t v_c

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

② Let $$\frac{\partial}{\partial v_c}\left(\log\left(\sum_{w=1}^{V} \exp(u_w^T v_c)\right)\right) = \frac{\partial I}{\partial v_c}$$

Applying chain rule:

Let $F = u_w^T v_c \rightarrow \frac{\partial F}{\partial v_c} = u_w$

$G = \exp(F) \rightarrow \frac{\partial G}{\partial F} = \exp(F)$

$H = \log\left(\sum_{w=1}^{V} G\right) \rightarrow \frac{\partial H}{\partial G} = \sum_{w=1}^{V} \frac{1}{G}$

$$\Rightarrow \frac{\partial I}{\partial v_c} = \frac{\partial H}{\partial G}\frac{\partial G}{\partial F}\frac{\partial F}{\partial v_c}$$

$$\frac{1}{\sum_{w_2=0}^{V} \exp(u_{w_2}^T v_c)}\left[\sum_{w_1=0}^{V} \exp(u_{w_1}^T v_c)\, u_{w_1}\right]$$

(2)

Note the nested summation with the inner $\sum_{w_1=0}^{V}$ going from w_1 & outer $\sum_{w_2=0}^{V}$ going from w_2.

Only the denominator is summed over w_2

This reduces to Eq (1) from the handout

where $P(O=o \mid C=c) = \dfrac{\exp(u_o^T v_c)}{\sum_{w=1}^{V} \exp(u_w^T v_c)} = \hat{y}_o$

for all words w_1

$$\therefore \frac{\partial J}{\partial v_c} = \sum_{w=0}^{V} \hat{y}_w u_w$$

Putting everything back, we get:

$$\frac{\partial J}{\partial v_c} = -u_o + \sum_{w=0}^{V} \hat{y}_w u_w$$

which is Eq (4) from handout

Extra Credit #3

$$J(v_c, o, U) = -u_o^T v_c + \log\left(\sum_{w=0}^{V} \exp(u_w^T v_c)\right)$$

Let's break the 2nd expression as follows:

$F = u_w^T v_c \qquad \frac{\partial F}{\partial u_w} = v_c$

$G = \sum_{w=0}^{V} \exp(F_w) \qquad \frac{\partial G}{\partial F_w} = \sum_{w=0}^{V} \exp(F_w)$

$H = \log(G) \qquad \frac{\partial H}{\partial G} = \frac{1}{G}$

Applying the chain rule we get

$$\frac{\partial H}{\partial U} = \frac{1}{\sum_{w=0}^{V} \exp(u_w^T v_c)} \left[\sum_{w=0}^{V} \exp(u_w^T v_c)\, v_c \right]$$

Reducing to $\hat{y}_w$ similar to problem #2.

we end up with

$\frac{\partial H}{U} = \sum_{w=0}^{V} \hat{y}_w v_c \quad$ or $\quad \frac{\partial H}{\partial u_w} = \hat{y}_w v_c$

Considering the 1st expression of $J(v_c, o, U)$

$$\frac{\partial}{\partial u_w}(-u_o^T v_c) = \begin{cases} -v_c & \text{if } w = 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is because u_w is not in the expression when $w \neq 0$.

Putting it together.

when $w=0$ we get : $-v_c + \hat{y}_w v_c = (\hat{y}_w - 1) v_c$

when $w \neq 0$, we get : $\hat{y}_w v_c$

Hence Proved.