ASSIGNMENT 2

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

$y_w = \begin{cases} 1, & \text{if } w = o \\ 0, & \text{otherwise,} \end{cases}$ so $-(\overset{0}{\cancel{y_1}} \log(\hat{y}_1) + \overset{0}{\cancel{y_2}} \log(\hat{y}_2) + \ldots \overset{1}{\cancel{y_o}} \log \hat{y}_o + \ldots) = -\log \hat{y}_o$

b) $\dfrac{\partial J_{naive-softmax}(v_c, o, U)}{\partial v_c}$ – ?

$$J_{naive-softmax}(v_c, o, U) = -\log P(o|c) = -\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)}$$

$$\frac{\partial}{\partial v_c} - \log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)} = -\Big(\frac{\partial}{\partial v_c} \overset{①}{\log \exp(u_o^T v_c)} - \frac{\partial}{\partial v_c} \overset{②}{\log \sum_w \exp(u_w^T v_c)}\Big) \circleddash$$

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

② $\dfrac{\partial}{\partial v_c} \log \sum_w \exp(u_w^T v_c) = \dfrac{1}{\sum_w \exp(u_w^T v_c)} \dfrac{\partial}{\partial v_c} \sum_x \exp(u_x^T v_c) = \dfrac{1}{\sum_w \exp(u_w^T v_c)} \cdot \sum_x \exp(u_x^T v_c) u_x$

$$\circleddash \; -\Big(u_o - \frac{1}{\sum_w \exp(u_w^T v_c)} \cdot \sum_x \exp(u_x^T v_c) \cdot u_x\Big) = \sum_x \underbrace{\frac{\exp(u_x^T v_c)}{\sum_w \exp(u_w^T v_c)}}_{P(x|c)} \cdot u_x - u_o =$$

$$= \sum_x P(x|c) \cdot u_x - u_o = U(\hat{y} - y)$$

c) $\dfrac{\partial J_{naive-softmax}(v_c, o, U)}{\partial v_u}$ – ?

$$J_{naive-softmax}(v_c, o, U) = -\log P(o|c) = -\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)}$$

$$-\frac{\partial}{\partial u_w} \log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)} = -\Big(\overset{①}{\frac{\partial}{\partial u_w} u_o^T v_c} - \overset{②}{\frac{\partial}{\partial u_w} \log \sum_w \exp(u_w^T v_c)}\Big) \circleddash$$

① $= v_c$, if $u_o = u_w$; 0, if $u_o \neq u_w$

② $\dfrac{1}{\sum_w \exp(u_w^T v_c)} \cdot \exp(u_w^T v_c) \cdot v_c$

$u_o = u_w$: $\circleddash \; -v_c + \dfrac{v_c \exp(u_w^T v_c)}{\sum \exp(u_w^T v_c)} = v_c\Big(\dfrac{\exp(u_w^T v_c)}{\sum \exp(u_w^T v_c)} - 1\Big) = v_c(P-1)$

$u_o \neq u_w$: $\circleddash \; -0 + \dfrac{v_c \exp(u_w^T v_c)}{\sum \exp(u_w^T v_c)} = v_c\Big(\dfrac{\exp(u_w^T v_c)}{\sum \exp(u_w^T v_c)} - 0\Big) = v_c(P-0)$

$$= v_c(\hat{y} - y)^T$$

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

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

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)\,(1-\sigma(x))$$

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

① $$\frac{\partial J_{neg\text{-}sample}}{\partial v_c} = \overbrace{\frac{\partial}{\partial v_c}\left[-\log(\sigma(u_o^T v_c))\right]}^{①} - \overbrace{\frac{\partial}{\partial v_c}\sum_{k=1}^{K}\log(\sigma(-u_k^T v_c))}^{②}$$

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

$$= -(1-\sigma(u_o^T v_c))\cdot u_o$$

② $$\sum_{k=1}^{K}\frac{\partial}{\partial v_c}\log(\sigma(-u_k^T v_c)) = \sum_{k=1}^{K}\frac{1}{\sigma(-u_k^T v_c)}\cdot\frac{\partial}{\partial v_c}\sigma(-u_k^T v_c) = -\sum_{k=1}^{K}\frac{\cancel{\sigma(-u_k^T v_c)}(1-\sigma(-u_k^T v_c))}{\cancel{\sigma(-u_k^T v_c)}}\cdot u_k$$

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

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

② $$\frac{\partial J_{neg\text{-}sample}}{\partial u_o} = \overbrace{\frac{\partial}{\partial u_o}\left[-\log(\sigma(u_o^T v_c))\right]}^{①} - \overbrace{\frac{\partial}{\partial u_o}\sum_{k=1}^{K}\log(\sigma(-u_k^T v_c))}^{=0,\ \text{since } o\notin\{w_1,\dots,w_K\}} = \boxed{-(1-\sigma(u_o^T v_c))\,v_c}$$

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

③ $$\frac{\partial J_{neg\text{-}sample}}{\partial u_k} = \overbrace{\frac{\partial}{\partial u_k}\left[\cancel{-\log(\sigma(u_o^T v_c))}\right]}^{0,\ \text{since } o\notin\{w_1,\dots,w_K\}} - \overbrace{\frac{\partial}{\partial u_k}\sum_{k=1}^{K}\log(\sigma(-u_k^T v_c))}^{①} \ominus$$

① $$\underbrace{\sum_{x=1}^{K}}_{=0\ k\neq x}\frac{\partial}{\partial u_k}\log(\sigma(-u_x^T v_c)) = \frac{1}{\cancel{\sigma(-u_k^T v_c)}}\cdot\cancel{\sigma(-u_k^T v_c)}\cdot(1-\sigma(-u_k^T v_c))\cdot(-v_c) = -(1-\sigma)\cdot v_c$$

$\ominus$ $$\boxed{(1-\sigma(-u_k^T v_c))\,v_c}$$

h) $$\frac{\partial J_{skip\text{-}gram}(v_c, w_{t-m}, \dots w_{t+m}, U)}{\partial U} = \sum_{\substack{-m\le j\le m\\ j\neq 0}}\frac{\partial J(v_c, w_{t+j}, U)}{\partial U}$$

$$\frac{\partial J_{skip\text{-}gram}(v_c, w_{t-m}, \dots w_{t+m}, U)}{\partial v_c} = \sum_{\substack{-m\le j\le m\\ j\neq 0}}\frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c}$$

$$\frac{\partial J_{skip\text{-}gram}(v_c, w_{t-m}, \dots w_{t+m}, U)}{\partial v_c} = \sum_{\substack{-m\le j\le m\\ j\neq 0}}\frac{\partial J(v_c, w_{t+j}, U)}{\partial v_w} = 0$$

(since $j\neq 0$)