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

$$\left.\begin{array}{lll} \text{when} & w \neq o & y_{w_i} = 0 \\ \text{when} & w = o & y_o = 1 \end{array}\right\} \Rightarrow -\overset{1}{y_o} \log(\hat{y}_o) = -\log(\hat{y}_o)$$

(b) $$\left(J'_{naive\text{-}softmax}\right)_{v_c} = \left(-\ln\left(\frac{e^{u_o^T v_c}}{\sum_{w \in Vocab} e^{u_w v_c}}\right)\right)'_{v_c} =$$

$$= -\left(\ln e^{u_o v_c} - \ln \sum_{w \in Vocab} e^{u_w v_c}\right)'_{v_c} =$$

$$= -\left(+u_o - \frac{\sum e^{u_w v_c} \cdot u_w}{\sum_{w \in Vocab} e^{u_w v_c}}\right.$$

(c) $w = o$

$$\left(J'_{naive\text{-}softmax}\right)_{u_o} = -v_c + \frac{e^{u_o v_c} \cdot v_c}{\sum_{w \in Vocab} e^{u_w v_c}}$$

$w \neq o$

$$\left(J_{naive\text{-}softmax}\right)'_{u_w} = \frac{\left(\left(\sum e^{u_w v_c}\right) - e^{u_o v_c}\right) \cdot v_c}{\sum_{w \in Vocab} e^{u_w v_c}}$$

$$(d)\begin{cases} J'_{u_{w_1}} = \dfrac{e^{u_{w_1} \cdot v_c}}{\sum\limits_{w \in vocab} e^{u_w v_c}} \\ J'_{u_{w_2}} = \dfrac{e^{u_2 v_c}}{\sum\limits_{w \in vocab} e^{u_w v_c}} \end{cases} \quad w \in vocab \setminus \{0\}$$

$$w = 0 \quad J'_{u_0} = -v_c + \frac{e^{u_0 v_c} \cdot v_c}{\sum\limits_{w \in vocab} e^{u_w v_c}}$$

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

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

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

$$(f) \quad (J'_{neg\text{-}sample})'_{v_c} = -\frac{1 \cdot e^{u_0 v_c} \cdot u_0}{\sigma(u_0 v_c)(1+e^{u_0 v_c})(1+e^{u_0 v_c})} - \sum_{k=1}^{K} \frac{1 \cdot e^{-u_k v_c} \cdot (-u_k)}{\sigma(-u_k v_c)(1+e^{-u_k v_c})(1+e^{-u_k v_c})}$$

$$(J_{neg})'_{u_0} = - \frac{1}{\sigma(u_0 v_c)} \frac{e^{u_0 v_c} \cdot v_c}{(1+e^{u_0 v_c})(1+e^{u_0 v_c})}$$

$$(J_{neg})'_{u_K} = - \sum_{k=1}^{K} \frac{1}{\sigma(-u_k v_c)} \frac{e^{-u_k v_c} \cdot (-v_c)}{(1+e^{-u_k v_c})(1+e^{-u_k v_c})}$$

(g) $w \in K$ not distinct; k-number of not disti. word u_k

$$(J_{neg})'_{u_k} = -\left(\sum_{n=1}^{N} \log(\sigma(-u_m v_c)) \cdot k + \sum_{k=1}^{K-N} \ln(\sigma(-u_k v_c)) \right)'_{u_k}$$

(h)

i) $$\frac{\partial J_{skip-Gram}}{\partial u} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \left(-\log(\sigma(u_{o} \cdot v_c, \cancel{\ldots})) - \right.$$

$$- \sum_{k=1}^{K} \ln(\sigma(-u_k v_c))'_u = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \left(- \frac{1}{\sigma(u_{j+t} \cdot v_c)} \frac{e^{u_{j+t} \cdot v_c} \cdot v_c}{(1+e^{u_{j+t} v_c})(1+e^{u_{j+t} v_c})} - \text{—}\,\text{''}\,\text{—} \right)$$

ii) $\dfrac{\partial J}{\partial v_c} = \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \left(J'_{neg\ sample}\right)_{v_c}$

iii) 0