Assignment 2 (Written)

a) $-\sum_{w \in Vocab} y_w \log(\hat{y}_w) = -\sum_{w \neq o} 0 . \log(\hat{y}_w) - 1 . \log(y_o)$

$= -\log(\hat{y}_o)$

b) $J_{naive-softmax}(v_c, o, U) = -\log(P(O=o \mid C=c))$

$= -\log\left[\frac{\exp(u_o^T v_c)}{\sum_{w \in Vocab} \exp(u_w^T v_c)}\right]$

$= -u_o^T v_c + \log \sum_{w \in Vocab} \exp(u_w^T v_c)$

b) $\frac{\partial J}{\partial v_c} = -u_o + \sum_{k \in Vocab} \frac{\exp(u_k^T v_c) u_k}{\sum_{w \in Vocab} \exp(u_w^T v_c)}$

$= -u_o + \sum_{k \in Vocab} y_k u_k$

c) $\frac{\partial J}{\partial u_o} = \cancel{-v_c} - (1 - y_c) v_c$

$\left(\frac{\partial J}{\partial u_w}\right)_{w \neq o} = y_w v_c$

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

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

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

$$\frac{\partial J_{neg}}{\partial v_c} = -[1-\sigma(u_o^T v_c)]\, u_o + \sum_{k=1}^{K} [1-\sigma(-u_k^T v_c)]\, u_k$$

$$\frac{\partial J_{neg}}{\partial u_o} = -[1-\sigma(u_o^T v_c)]\, v_c + \cancel{\sum_{k=1}^{K} [1-\sigma(-u_k^T v_c)]\, v_c}$$

$$\frac{\partial J_{neg}}{\partial u_k} = [1-\sigma(-u_k^T v_c)]\, v_c$$

In part (b) & (c) computing the summation over entire vocabulary is very computationally expensive. Here, summation is over just K negative samples.

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

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

(iii) $$\left(\frac{\partial J_{skip\text{-}gram}}{\partial v_w}\right)_{w \ne c} = 0$$