1. $-\sum_{w \in Vocab} y_m \log(\hat{y}_m) = -(y_1 \log(\hat{y}_1) + \ldots + y_o \log(\hat{y}_o) + \ldots$
$+ y_w \log(\hat{y}_w)) = -y_o \log(\hat{y}_o) = -\log(\hat{y}_o)$

2. $J(v_c, o, U)'_{v_c} = \frac{\partial}{\partial v_c}(-\log P(O=o \mid C=c))'_{v_c} =$

$= \frac{\partial}{\partial v_c}\left(-\log\left(\frac{\exp(u_o^T v_c)}{\sum \exp(u_w^T v_c)}\right)\right)'_{v_c} = (-\log \exp(u_o^T v_c) +$

$+ \log(\sum \exp(u_w^T v_c)))'_{v_c} = (-u_o^T v_c + \log(\sum_w \exp(u_w^T v_c)))'_{v_c}$

$= -u_o + \frac{\sum \exp(u_w^T v_c)\, u_w}{\sum \exp(u_w^T v_c)} = -u_o + \sum P(O=w \mid C=c)\, u_w = -u_o + \sum \hat{y}_w u_w$

3. $\frac{\partial}{\partial u_w}(-\log P(O=o \mid C=c))'_{u_w}$

$w = o : -v_c + P(O=o \mid C=c) v_c = \hat{y}_w v_c - v_c$

$w \neq o : 0 + P(O=w \mid C=c) v_c = \hat{y}_w v_c$

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

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

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

5. $J(v_c, o, U) = -\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} \log \sigma(-u_k^T v_c)$

$$J'_{v_c} = -\frac{\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))}{\sigma(u_o^T v_c)} u_o - \sum^{K} \frac{\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))}{\sigma(-u_k^T v_c)} \cdot (-u_k)$$

$$= (\sigma(u_o^T v_c) - 1) u_o + \sum^{K} (1-\sigma(-u_k^T v_c)) u_k$$

$$J'_{u_o} = -(1-\sigma(u_o^T v_c)) v_c = (\sigma(u_o^T v_c) - 1) v_c$$

$$J'_{u_k} = +(1-\sigma(-u_k^T v_c)) \cdot v_c = \sigma(-u_k^T v_c) v_c \quad \text{for } 1 \text{ to } k$$

6. $J_{sk}(v_c, w_{t-m}, \dots, w_{t+m}, U) = \sum_{\substack{-m \le j \le m \\ j \ne 0}} J(v_c, w_{t+j}, U)$

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

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

c) $$\frac{\partial J_{sk}}{\partial v_w} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_w} = 0$$