1

(a) $$-\sum_{w\in V} y_w \log(\hat{y}_w) = -y_o\cdot\log(\hat{y}_o) - \sum_{w\in V\setminus\{o\}} y_w \log(\hat{y}_w) = -1\cdot\log(\hat{y}_o) - \sum_{w\in V\setminus\{o\}} 0\cdot\log(\hat{y}_w) = -\log(\hat{y}_o)$$

(b) $$J = -\log(P(o|c)) = -\log\left(\frac{\exp(u_o^T v_c)}{\sum_{w\in V}\exp(u_w^T v_c)}\right) = -\log(\exp(u_o^T v_c)) + \log\left(\sum_{w\in V}\exp(u_w^T v_c)\right) = -u_o^T v_c + \log\left(\sum_{w\in V}\exp(u_w^T v_c)\right)$$

$$\frac{\partial J}{\partial v_c} = -u_o + \frac{\partial}{\partial v_c}\left(\log\left(\sum_{w\in V}\exp(u_w^T v_c)\right)\right) = -u_o + \frac{\sum_{w\in V}\frac{\partial}{\partial v_c}\left(\exp(u_w^T v_c)\right)}{\sum_{w\in V}\exp(u_w^T v_c)} = -u_o + \frac{\sum_{w\in V}\exp(u_w^T v_c)\, u_w}{\sum_{w\in V}\exp(u_w^T v_c)}$$

(c) $$\frac{\partial J}{\partial u_o} = -v_c + \frac{\partial}{\partial u_o}\left(\log\left(\sum_{w\in V}\exp(u_w^T v_c)\right)\right) = -v_c + \frac{\frac{\partial}{\partial u_o}\left(\sum_{w\in V}\exp(u_w^T v_c)\right)}{\sum_{w\in V}\exp(u_w^T v_c)} = -v_c + \frac{\frac{\partial}{\partial u_o}\left(\exp(u_o^T v_c)\right) + \frac{\partial}{\partial u_o}\left(\sum_{w\in V\setminus\{o\}}\exp(u_w^T v_c)\right)}{\sum_{w\in V}\exp(u_w^T v_c)} = -v_c + \frac{\exp(u_o^T v_c)\cdot v_c}{\sum_{w\in V}\exp(u_w^T v_c)}$$

(c) $$\frac{\partial J}{\partial u_{w\neq o}} = -0 + \frac{\partial}{\partial u_{w\neq o}}\left(\log\left(\sum_{x\in V}\exp(u_x^T v_c)\right)\right) = \frac{\frac{\partial}{\partial u_{w\neq o}}\left(\sum_{x\in V}\exp(u_x^T v_c)\right)}{\sum_{x\in V}\exp(u_x^T v_c)} = \frac{\frac{\partial}{\partial u_{w\neq o}}\exp(u_{w\neq o}^T v_c) + \left(\sum_{x\in V\setminus\{w\neq o\}}\exp(u_x^T v_c)\right)'_{u_{w\neq o}}}{\sum_{x\in V}\exp(u_x^T v_c)} = \frac{\exp(u_{w\neq o}^T v_c)\, v_c}{\sum_{x\in V}\exp(u_x^T v_c)}$$

(d) $$\overrightarrow{\frac{\partial J}{\partial U}} = \begin{bmatrix} \frac{\partial J}{\partial u_1} \\ \frac{\partial J}{\partial u_2} \\ \vdots \\ \frac{\partial J}{\partial u_o} \\ \vdots \\ \frac{\partial J}{\partial u_{|V|}} \end{bmatrix}$$

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

(f) $$J = -\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K}\log(\sigma(-u_k^T v_c))$$

$$\frac{\partial J}{\partial v_c} = \frac{\sigma^2(u_o^T v_c)\exp(-u_o^T v_c)\cdot u_o}{\sigma(u_o^T v_c)} - \sum_{k=1}^{K}\frac{\sigma^2(-u_k^T v_c)\exp(u_k^T v_c)\cdot u_k}{\sigma(-u_k^T v_c)} = \sigma(u_o^T v_c)\exp(-u_o^T v_c)\, u_o - \sum_{k=1}^{K}\sigma(-u_k^T v_c)\exp(u_k^T v_c)\, u_k$$

$$\frac{\partial J}{\partial u_o} = \frac{\sigma^2(u_o^T v_c)\exp(-u_o^T v_c)\cdot v_c}{\sigma(u_o^T v_c)} - 0 = \sigma(u_o^T v_c)\exp(-u_o^T v_c)\, v_c$$

$$\frac{\partial J}{\partial u_k} = -0 - \frac{\sigma^2(-u_k^T v_c)\exp(u_k^T v_c)\, v_c}{\sigma(-u_k^T v_c)} - 0 = -\sigma(-u_k^T v_c)\exp(u_k^T v_c)\, v_c$$

(g) $$\frac{\partial J}{\partial u_k} = -0 - \sum_{\{i=\overline{1,K}:\ u_i = u_k\}}\sigma(-u_i^T v_c)\exp(u_i^T v_c)\, v_c - 0 = -\sum_{\{i=\overline{1,K}:\ u_i = u_k\}}\sigma(-u_i^T v_c)\exp(u_i^T v_c)\, v_c$$

(h) (i) $$\frac{\partial J_{sg}}{\partial U} = \sum_{\substack{-m\le i\le m\\ i\neq 0}}\frac{\partial}{\partial U}J(v_c, w_{t+i}, U)$$

(ii) $$\frac{\partial J_{sg}}{\partial v_c} = \sum_{\substack{-m\le i\le m\\ i\neq 0}}\frac{\partial}{\partial v_c}J(v_c, w_{t+i}, U)$$

(iii) $$\frac{\partial J_{sg}}{\partial v_w} = \frac{\partial J(v_c, v_w, U)}{\partial u_w} = \frac{\partial J_{sg}}{\partial U}[w]$$