2.

(a)

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

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

(b)

$$J = -\sum_{z=1} y_z \log(y_z) = -\sum_{z=1} y_z \log(p(x_z|c)) = -\log\left(\frac{\exp(u_o^T v_c)}{\sum_{w=1}^{W}\exp(u_w^T v_c)}\right)$$

$$= -(u_w^T v_c) - \log\left(\sum_{w=1}^{W}\exp(u_w^T v_c)\right) \quad ①$$

$$\frac{\partial \hat{y}_o}{\partial v_c} = \frac{u_o \exp(u_o^T v_c)\sum_{w}^{w=1}\exp(u_w^T v_c)}{\left(\sum_{w=1}^{W}\exp(u_w^T v_c)\right)^2} - \frac{\exp(u_o^T v_c)\cdot\sum_{w=1}^{W} u_w\cdot\exp(u_w^T\cdot v_c)}{\left(\sum_{w=1}^{W}\exp(u_w^T v_c)\right)^2}$$

$$= \underset{\exp(u_o^T v_c)}{\overset{y_o}{\parallel}} \left[\frac{u_o\sum_{w}^{w=1}\exp(u_w^T v_c) - \sum_{w=1}^{W} u_w\cdot\exp(u_w^T\cdot v_c)}{\sum_{w}^{W}\exp(u_w^T\cdot v_c)}\right]$$

$$\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \hat{y}_o}\cdot\frac{\partial \hat{y}_o}{\partial v_c} = -\sum_{w=1}^{W} y_o\cdot\frac{1}{\hat{y}_o}\cdot\frac{\partial \hat{y}_o}{\partial v_c} = y_o\left(u_o - \sum_{w=1}^{W}\hat{y}_w u_w\right) = U^T(\hat{y}-y)$$

(c)

case 1: $U_w = U_o$

from ①

$$\frac{\partial}{\partial u_w} = -(u_o^T v_c) = -v_c \qquad \frac{\partial}{\partial w}\log\left(\sum_{w=1}^{W}\exp(u_w^T v_c)\right) = \sum_{w=1}^{W}\frac{\exp(u_x^T v_c)v_c}{\sum_{w=1}^{W}\exp(u_w^T v_c)}$$

$$= \sum_{w=1}^{W}\hat{y}_o v_c$$

$$\frac{\partial J}{\partial w} = -v_c + \sum_{w=1}^{W}\hat{y}_o v_c = v_c(1-\hat{y}_o)$$

case 2: $u_w \neq u_o$

$$\frac{\partial}{\partial u_w} = -(u_o^T v_c) = 0 \qquad \frac{\partial}{\partial w} \log\left(\sum_{w=1}^{W} \exp(u_w^T v_c)\right) = \sum_{w=1}^{W} \frac{\exp(u_x^T v_c) v_c}{\sum_{w=1}^{W} \exp(u_w^T v_c)}$$

$$= \sum_{w=1}^{W} \hat{y}_o v_c$$

$$\frac{\partial J}{\partial w} = 0 + \sum_{w=1}^{W} \hat{y}_o v_c = \hat{y}_o v_c$$

So, combine two case $\underline{v_c(\hat{y}-y)^T}$

(d)

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

$$= -\log\left(\frac{1}{1+e^{-u_o^T v_c}}\right) - \sum_{k=1}^{K} \log\left(\frac{1}{1+e^{u_k^T v_c}}\right) = \log(1+e^{-u_o^T v_c}) + \sum_{k=1}^{K} \log(1+e^{u_k^T v_c})$$

$$\frac{\partial J_{neg}}{\partial v_c} = \frac{-u_o e^{-u_o^T v_c}}{1+e^{-u_o^T v_c}} + \sum_{k=1}^{K} \frac{u_k \cdot e^{u_k^T v_c}}{1+e^{-u_o^T v_c}}$$

case 1: $u_w = u_o$

$$\frac{\partial J_{neg}}{\partial u_w} = \frac{-v_c \cdot e^{-u_o^T v_c}}{1+e^{-u_o^T v_c}} + 0 = \underline{\frac{-v_c \cdot e^{-u_o^T v_c}}{1+e^{u_o^T v_c}}}$$

because $\frac{\partial J_{neg}}{\partial w_o}(-u_k^T v_c) = 0$

case 2: $u_w \neq u_o$ (, $u_w = u_k$)

$$\frac{\partial J_{neg}}{\partial u_w} = 0 + \sum_{k=1}^{K} \frac{v_c \cdot e^{u_k^T v_c}}{1+e^{-u_o^T v_c}} = \underline{\sum_{k=1}^{K} \frac{v_c \cdot e^{u_k^T v_c}}{1+e^{-u_o^T v_c}}}$$

(e) Skip-gram $= [\text{word}_{c-m} \cdots \text{word}_c \cdots \text{word}_{c+m}] = \sum_{-m \le j \le m, j \ne 0} F(w_{c+j}, v_c)$

$$\frac{\partial J_{\text{skip-gram}}}{\partial v_c} = \sum_{-m \le j \le m, j \ne 0} \frac{\partial F(o, v_c)}{\partial v_k} \qquad \frac{\partial J_{\text{skip-gram}}}{\partial v_c} = \sum_{-m \le j \le m, j \ne 0} \frac{\partial F(o, v_c)}{\partial u_k}$$