(a): $-\sum_{w\in vocab} y_w \log(\hat{y}_w) = -(0+0+...+y_o\log(\hat{y}_o)+0+...+0) = -\log(\hat{y}_o)$

(b):

$$\frac{\partial J}{\partial v_c} = u_o - \sum_{x\in v} p(x|c)u_x = u_o - \sum_{x\in v} p(x|c)u_x = U(y-\hat{y})$$

(c):

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

Then we first calculate the later part, we give the annotation l as the value of the later part, so we can deduct that:

$$l = -\frac{1}{\sum_{w\in V}\exp(u_w^T v_c)}\cdot\frac{\partial}{\partial u_w}\exp(u_w^T v_c) = -\frac{\exp(u_w^T v_c)}{\sum_{w\in V}\exp(u_w^T v_c)}\cdot v_c = -\hat{y}\cdot v_c$$

So we can get the final solution:

$$\frac{\partial J}{\partial u_w} = \begin{cases} (1-\hat{y})\cdot v_c & \text{if } w = o \\ -\hat{y}\cdot v_c & \text{if } w \neq o \end{cases}$$

(d):

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

(e):

1):

$$\frac{\partial J}{\partial u_o} = -\frac{\partial}{\partial u_o}\log(\sigma(u_o^T v_c)) = -\frac{1}{\sigma(u_o^T v_c)}\cdot\frac{\partial}{\partial u_o}\sigma(u_o^T v_c) = (\sigma(u_o^T v_c)-1)v_c$$

2):

$$\frac{\partial J}{\partial u_k} = -\frac{\partial}{\partial u_k}\log(\sigma(-u_k^T v_c)) = -(1-\sigma(-u_k^T v_c))\cdot -v_c = (1-\sigma(-u_k^T v_c))\cdot v_c$$

3):

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

we give the annotations l as the left part, r as the right part, then we get following:

$$l = -\frac{1}{\sigma(u_o^T v_c)} \cdot \frac{\partial}{\partial v_c}\sigma(u_o^T v_c) = (\sigma(u_o^T v_c) - 1)u_o$$

$$r = \sum_{k=1}^{K}(1 - \sigma(-u_k^T v_c)) \cdot u_k$$

So we can get the final derivative:

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

(f):

$$\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right) / \partial \boldsymbol{U} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial \boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U})}{\partial \boldsymbol{U}}$$

$$\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right) / \partial \boldsymbol{v}_c = \sum_{-m \leq j \leq m, j \neq 0} \partial \boldsymbol{J}\left(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}\right) / \partial \boldsymbol{v}_c$$

$$\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right) / \partial \boldsymbol{v}_w = 0 \text{ when } w \neq c$$