(a)

$$y = [0, 0...1...0]$$

Since y only has 1 at the true outside value, and zeroes at all the rest. The sum will only consist of one element.

$$-\log \hat{y}_o$$

(b)

$$J = -\sum_{w \in V} y_w \log \hat{y}_w$$
$$a_i = u_i^T v_c$$
$$\partial J / \partial a_i = \hat{y}_i - y_i$$
$$\partial a_i / \partial v_c = u_i^T$$
$$\partial J / \partial v_c = \frac{\partial J}{\partial a_i} \frac{\partial a_i}{\partial v_c} = (\hat{y}_i - y_i) u_i^T$$
$$\partial J / \partial v_c = U^T (\hat{y} - y)$$

(c)

if $w \neq o$: only need to consider $-\log \hat{y}_0$

$$J = -\sum_{w \in V} y_w \log \hat{y}_w$$
$$a_i = u_i^T v_c$$
$$\partial J / \partial a_i = \hat{y}_i - y_i$$
$$\partial a_i / \partial u_w = u_i^T$$
$$\partial J / \partial u_w = \frac{\partial J}{\partial a_i} \frac{\partial a_i}{\partial u_w} = (\hat{y}_i - y_i) v_c^T$$
$$\partial J / \partial u_w = v_c^T (\hat{y} - y)$$

(d)

quotient rule

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

(e)

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

This may be more effecient to compute because you don't have to compute the softmax, and don't have to add all the $u_w^T v_c$.

(f)

(i) if w is part of $w_{t-m}...w_{t+m}$ $\partial J/\partial u_w = -(1 - \sigma(u_w^T v_c))v_c^T$
else if w is a negative sample $\partial J/\partial u_w = -(1 - \sigma(-u_k^T v_c)) - v_c^T$
else $\partial J/\partial u_w = 0$
(ii) $\partial J/\partial v_c = -\frac{1}{\sigma(u_o^T v_c)}(\sigma(u_o^T v_c))(1 - \sigma(u_o^T v_c))u_o^T - \sum_{k=1}^{K} \frac{1}{\sigma(-u_k^T v_c)}(\sigma(-u_k^T v_c))(1 - \sigma(-u_k^T v_c)) - u_k^T$
(iii) $\partial J/\partial v_w = 0$