(a)

$$-\sum_{w\in Vocab} y_w \log(\hat{y}_w) = -y_o \log(\hat{y}_o) - \sum_{\substack{w\in Vocab \\ w\neq o}} y_w \log(\hat{y}_w).$$

(here y_i expresses as the i-th element in the $\vec{y}$.

and $y_o = 1$ $y_w = 0$ $(w \neq o)$).

$$= -\log(\hat{y}_o)$$

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

$$= -u_o + \frac{1}{\sum_{w\in Vocab} \exp(u_w^T v_c)} \cdot \left(\sum_{w\in Vocab} u_w \cdot \exp(u_w^T v_c)\right)$$

$$= -u_o + \sum_w \frac{u_w \cdot \exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)}$$

$$= -u_o + \sum_w u_w \cdot P(O=w \mid C=c)$$

$$= -u_o + \sum_w u_w y_w$$

$$= U(\hat{y} - y).$$

(c) if $w \neq o$:

$$\frac{\partial J(v_c, o, U)}{\partial u_w} = -\frac{\partial (u_o^T v_c)}{\partial u_w} + \frac{\partial \log \sum_w \exp(u_w^T v_c)}{\partial u_w}$$

$$= 0 + v_c \frac{\exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)}$$

$$= 0 + v_c P(O = w \mid C = c)$$

$$= \hat{y}_w v_c.$$

if $w = o$:

$$\frac{\partial J(v_c, o, U)}{\partial u_o} = -v_c + v_c \hat{y}_o$$

(d). $$\frac{\partial J(v_c, o, U)}{\partial U} = \left(\frac{\partial J(v_c, o, U)}{\partial u_1}, \frac{\partial J(v_c, o, U)}{\partial u_2}, \dots, \frac{\partial J(v_c, o, U)}{\partial u_{|Vocab|}} \right)^T$$

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

$$= \sigma(x) \left(1 - \sigma(x)\right).$$

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$$