Part 1

(a)

As konwn $y_w = \begin{cases} 1 & w = o \\ 0 & w != o \end{cases}$

Then， $-\sum_{w \in Vocab} y_w \log\left(\hat{y}_w\right) = -y_o \log\left(\hat{y}_o\right) = -\log\left(\hat{y}_o\right)$

(b)

$$\frac{\partial J(v_c, o, U)}{\partial v_c} = \frac{\partial(-\log P(O = o \mid C = c)}{\partial v_c}$$

$$= \frac{\partial(-\log \frac{\exp\left(u_o^\top v_c\right)}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)})}{\partial v_c}$$

$$= -\frac{\partial(\log \exp\left(u_o^\top v_c\right))}{\partial v_c} + \frac{\partial(\log \sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right))}{\partial v_c}$$

The first part,

$$\frac{\partial(\log \exp\left(u_o^\top v_c\right))}{\partial v_c} = \frac{\partial(u_o^\top v_c)}{\partial v_c} = u_o$$

The second part:

$$\frac{\partial(\log \sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right))}{\partial v_c} = \frac{1}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)} \frac{\partial(\sum_{x \in \text{Vocab}} \exp\left(u_x^\top v_c\right))}{\partial v_c}$$

$$= \frac{1}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)} \sum_{x \in \text{Vocab}} \exp\left(u_x^\top v_c\right) \frac{\partial(u_x^\top v_c)}{\partial v_c}$$

$$= \frac{1}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)} \sum_{x \in \text{Vocab}} \exp\left(u_x^\top v_c\right) u_x$$

Brought into the original:

$$\frac{\partial J(v_c, o, U)}{\partial v_c} = -u_o + \frac{1}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)} \sum_{x \in \text{Vocab}} \exp\left(u_x^\top v_c\right) u_x$$

$$= -u_o + \sum_{x=1}^{V} \frac{\exp\left(u_x^\top v_c\right)}{\sum_{w \in \text{Vocab}} \exp\left(u_w^\top v_c\right)} u_x$$

$$= -u_o + \sum_{x=1}^{V} P(x \mid c) u_x$$

$$= U^T (\hat{y} - y)$$

(c)

$$\frac{\partial J(v_c,o,U)}{\partial u_w}=\frac{\partial(-\log P(O=o\mid C=c)}{\partial u_w}$$

$$=\frac{\partial(-\log\frac{\exp\left(u_o^\top v_c\right)}{\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right)})}{\partial u_w}$$

$$=-\frac{\partial(\log\exp\left(u_o^\top v_c\right))}{\partial u_w}+\frac{\partial(\log\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right))}{\partial u_w}$$

当w=o时，

$$\frac{\partial J(v_c,o,U)}{\partial u_w}=-v_c+\frac{1}{\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right)}\frac{\partial(\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right))}{\partial u_o}$$

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

$$=-v_c+\frac{\exp\left(u_o^\top v_c\right)}{\sum_{x\in\text{Vocab}}\exp\left(u_x^\top v_c\right)}v_c$$

$$=-v_c+P(O=o\mid C=c)v_c$$

$$=(P(O=o\mid C=c)-1)v_c$$

$$=(\hat{y}-y)v_c$$

当w$\neq$o时，

$$\frac{\partial J(v_c,o,U)}{\partial u_w}=\frac{1}{\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right)}\frac{\partial(\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right))}{\partial u_w}$$

$$=\frac{1}{\sum_{w\in\text{Vocab}}\exp\left(u_w^\top v_c\right)}\exp\left(u_w^\top v_c\right)\frac{\partial\left(u_w^\top v_c\right)}{\partial u_w}$$

$$=\frac{\exp\left(u_w^\top v_c\right)}{\sum_{x\in\text{Vocab}}\exp\left(u_x^\top v_c\right)}v_c$$

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

$$=\hat{y}v_c$$

$$=(\hat{y}-y)v_c(y=0)$$

Summary,

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

(d)

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

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

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

$$= \sigma(x)(1-\sigma(x))$$