

Part 1

(a)

$$\text{As known } y_w = \begin{cases} 1 & w = o \\ 0 & w \neq o \end{cases}$$

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

(b)

$$\begin{aligned} \frac{\partial J(v_c, o, U)}{\partial v_c} &= \frac{\partial(-\log P(O=o | C=c))}{\partial v_c} \\ &= \frac{\partial(-\log \frac{\exp(u_o^\top v_c)}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)})}{\partial v_c} \\ &= -\frac{\partial(\log \exp(u_o^\top v_c))}{\partial v_c} + \frac{\partial(\log \sum_{w \in \text{Vocab}} \exp(u_w^\top v_c))}{\partial v_c} \end{aligned}$$

The first part,

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

The second part:

$$\begin{aligned} \frac{\partial(\log \sum_{w \in \text{Vocab}} \exp(u_w^\top v_c))}{\partial v_c} &= \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \frac{\partial(\sum_{x \in \text{Vocab}} \exp(u_x^\top v_c))}{\partial v_c} \\ &= \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \sum_{x \in \text{Vocab}} \exp(u_x^\top v_c) \frac{\partial(u_x^\top v_c)}{\partial v_c} \\ &= \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \sum_{x \in \text{Vocab}} \exp(u_x^\top v_c) * u_x \end{aligned}$$

Brought into the original:

$$\begin{aligned} \frac{\partial J(v_c, o, U)}{\partial v_c} &= -u_o + \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \sum_{x \in \text{Vocab}} \exp(u_x^\top v_c) * u_x \\ &= -u_o + \sum_{x=1}^V \frac{\exp(u_x^\top v_c)}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} u_x \\ &= -u_o + \sum_{x=1}^V P(x | c) u_x \\ &= U^T (\hat{y} - y) \end{aligned}$$

(c)

$$\begin{aligned}
\frac{\partial J(v_c, o, U)}{\partial u_w} &= \frac{\partial(-\log P(O=o|C=c))}{\partial u_w} \\
&= \frac{\partial(-\log \frac{\exp(u_o^\top v_c)}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)})}{\partial u_w} \\
&= -\frac{\partial(\log \exp(u_o^\top v_c))}{\partial u_w} + \frac{\partial(\log \sum_{w \in \text{Vocab}} \exp(u_w^\top v_c))}{\partial u_w}
\end{aligned}$$

当 $w=o$ 时,

$$\begin{aligned}
\frac{\partial J(v_c, o, U)}{\partial u_w} &= -v_c + \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \frac{\partial(\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c))}{\partial u_o} \\
&= -v_c + \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \exp(u_o^\top v_c) \frac{\partial \exp(u_o^\top v_c)}{\partial u_o} \\
&= -v_c + \frac{\exp(u_o^\top v_c)}{\sum_{x \in \text{Vocab}} \exp(u_x^\top v_c)} v_c \\
&= -v_c + P(O=o|C=c) v_c \\
&= (P(O=o|C=c) - 1) v_c \\
&= (\hat{y} - y) v_c
\end{aligned}$$

当 $w \neq o$ 时,

$$\begin{aligned}
\frac{\partial J(v_c, o, U)}{\partial u_w} &= \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \frac{\partial(\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c))}{\partial u_w} \\
&= \frac{1}{\sum_{w \in \text{Vocab}} \exp(u_w^\top v_c)} \exp(u_w^\top v_c) \frac{\partial \exp(u_w^\top v_c)}{\partial u_w} \\
&= \frac{\exp(u_w^\top v_c)}{\sum_{x \in \text{Vocab}} \exp(u_x^\top v_c)} v_c \\
&= P(O=w|C=c) v_c \\
&= \hat{y} v_c \\
&= (\hat{y} - y) v_c (y=0)
\end{aligned}$$

*Summary,*

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

(d)

$$\begin{aligned}\frac{\partial \sigma(x)}{\partial x} &= \frac{\partial \left( \frac{1}{1+e^{-x}} \right)}{\partial x} = -\frac{1}{(1+e^{-x})^2} \frac{\partial (1+e^{-x})}{\partial x} \\ &= \frac{e^{-x}}{(1+e^{-x})^2} \\ &= \frac{1}{1+e^{-x}} \frac{e^{-x}}{1+e^{-x}} \\ &= \sigma(x)(1-\sigma(x))\end{aligned}$$