Assignment 2

Fudi(Fred) Wang

November 2019

Part 1

(a) Since **y** is a one-hot vector with a 1 for the true outside word o, and 0 everywhere else, then when $w = o$, $y_w = 1$; when $w \neq o$, $y_w = 0$, hence,

$$-\sum_{w \in Vocab} y_w log(\hat{y}_w) = 0 - 1 \times log(\hat{y}_0) = -log(\hat{y}_0)$$

(b)

$$\begin{aligned}
-\frac{\partial P(O = o \mid C = c)}{\partial \boldsymbol{v}_c} &= -\frac{\partial [\boldsymbol{u}_o^T \boldsymbol{v}_c - log \sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)]}{\partial \boldsymbol{v}_c} \\
&= \frac{\sum_{j \in Vocab} exp(\boldsymbol{u}_j^T \boldsymbol{v}_c) \boldsymbol{u}_j}{\sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)} - \boldsymbol{u}_o \\
&= \sum_{j \in Vocab} \Big(\frac{exp(\boldsymbol{u}_j^T \boldsymbol{v}_c)}{\sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)} \Big) \boldsymbol{u}_j - \boldsymbol{u}_o \\
&= \sum_{j \in Vocab} P(w_j \mid w_c) \boldsymbol{u}_j - \boldsymbol{u}_o \\
&= \sum_{j \in Vocab} \hat{y}_j \boldsymbol{u}_j - \boldsymbol{u}_o \\
&= \boldsymbol{U}^T (\hat{\boldsymbol{y}} - \boldsymbol{y})
\end{aligned}$$

(c)
When $w = o$, we have

$$\begin{aligned}
-\frac{\partial P(O = o \mid C = c)}{\partial \boldsymbol{u}_w} &= -\frac{\partial [\boldsymbol{u}_o^T \boldsymbol{v}_c - log \sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)]}{\partial \boldsymbol{u}_o} \\
&= \frac{exp(\boldsymbol{u}_o^T \boldsymbol{v}_c) \boldsymbol{v}_c}{\sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)} - \boldsymbol{v}_c \\
&= (P(w_o \mid w_c) - 1) \boldsymbol{v}_c \\
&= (\hat{y}_o - y_o) \boldsymbol{v}_c \\
&= (\hat{\boldsymbol{y}} - \boldsymbol{y}) \boldsymbol{v}_c
\end{aligned}$$

When $w \neq o$, we have

$$\begin{aligned}
-\frac{\partial P(O=o \mid C=c)}{\partial \boldsymbol{u}_w} &= -\frac{\partial[\boldsymbol{u}_o^T \boldsymbol{v}_c - log \sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)]}{\partial \boldsymbol{u}_w} \\
&= \frac{\sum_{j \in Vocab, j \neq o} exp(\boldsymbol{u}_j^T \boldsymbol{v}_c) \boldsymbol{v}_c}{\sum_{w \in Vocab} exp(\boldsymbol{u}_w^T \boldsymbol{v}_c)} \\
&= \sum_{j \in Vocab, j \neq o} P(w_j \mid w_c) \boldsymbol{v}_c \\
&= \sum_{j \in Vocab, j \neq o} \hat{y}_j \boldsymbol{v}_c \\
&= \hat{\boldsymbol{y}} \boldsymbol{v}_c
\end{aligned}$$

(d)

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