Assignment2

April 13, 2020

1 Question(a)

We know that:

$$y_w = \begin{cases} 1 & w \text{ shown in context of } o \\ 0 & w \text{ not shown in context of } o \end{cases} \tag{1}$$

And since y is a one-hot vector, so only for y_o, it is 1.

$$\text{cross\_entropy}(y, \hat{y}) = \sum_{w \in V} y_w log(\hat{y_w}) = y_o log(\hat{y_o}) \tag{2}$$

2 Question(b)

U is a matrix where each column is u_i for context word i. c is the center word's index, o is the outside word's index.

$$\begin{aligned} J(o, c, U) &= -log \frac{exp(u_o^T v_c)}{\sum_{w \in V} exp(u_w^T v_c)} \\ &= -u_o^T v_c + log(\sum_{w \in V} exp(u_w^T v_c)) \end{aligned} \tag{3}$$

$$\begin{aligned} \frac{\partial J(o, c, U)}{\partial v_c} &= -u_o + \frac{\partial log \sum_{w \in V} exp(u_w^T v_c)}{\partial v_c} \\ &= -u_o + \frac{1}{\sum_{w \in V} exp(u_w^T v_c)} \frac{\partial \sum_{w \in V} exp(u_w^T v_c)}{\partial v_c} \\ &= -u_o + \frac{\sum_{w \in V} u_w exp(u_w^T v_c)}{\sum_{w \in V} exp(u_w^T v_c)} \\ &= -Uy + \sum_{w \in V} u_w \hat{y} \\ &= U(\hat{y} - y) \end{aligned} \tag{4}$$

3 Question(c)

If $w = o$(this word is the outside word), then:

$$\begin{aligned}
\frac{\partial J(o,c,U)}{\partial u_w} &= -v_c + \frac{\partial log \sum_{w \in V} exp(u_w^T v_c)}{\partial u_o} \\
&= -v_c + \frac{1}{\sum_{w \in V} exp(u_w^T v_c)} \frac{\partial \sum_{w \in V} exp(u_w^T v_c)}{\partial u_o} \\
&= -v_c + \frac{v_c exp(u_w^T v_c)}{\sum_{w \in V} exp(u_w^T v_c)} \\
&= -v_c + v_c \hat{y} \\
&= v_c(\hat{y} - y)
\end{aligned} \tag{5}$$