Assignment 2: word2vec solutions

Shubham Gupta

January 1, 2020

1 Understanding word2vec

1.1 Derivation

From the question, we know the following:

$$J(\nu_c, o, \mathbf{U}) = -logP(O = o|C = c).$$

We know cross entropy loss is given as:

$$-\sum_{w \epsilon Vocab} y_w log(\hat{y}_w) = -log(\hat{y}_o).$$

Since $\mathbf{y}$ is a one hot vector, it implies that y will be equal to 1 only when $i == o$. Hence, the LHS can be written as:

$$\begin{aligned}
&-\sum_{w \epsilon Vocab} y_w log(\hat{y}_w) \\
&= -[y_0 log(y_0) + y_1 log(y_1) + ... + y_o log(y_o) + ... + y_w log(y_w)] \\
&= -y_o log(y_o) \\
&= -log(y_o) \\
&= -log(P(O = o|C = c))
\end{aligned} \tag{1}$$

1.2 Derivation of $J_{\textbf{naive softmax}}$ wrt ν_c

$$\begin{aligned}
&CE(y, \hat{y}) \\
&\hat{y} = softmax(\theta)\frac{\partial J}{\partial \theta} = (\hat{y} - y)^T
\end{aligned} \tag{2}$$

Using chain rule, we can solve this as follows:

$$\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \theta}\frac{\partial \theta}{\partial v_c} \quad = (\hat{y} - y)\frac{\partial U^T v_c}{\partial v_c} \quad = U^T(\hat{y} - y)^T \tag{3}$$

1.3 Derivation of $J_{\textbf{naive softmax}}$ wrt outside vectors u_w