CS 224N assignment2

kugwzk

February 9, 2019

1. Question (a):

   The cross entropy loss is $-\sum_{w \in Vocab} y_w \log(\hat{y_w})$. Because the vector of true distribution y is s a one-hot vector with a 1 for the true outside word o, and 0 everywhere else.
   So the answer is

$$-\sum_{w \in Vocab} y_w \log(\hat{y_w}) = -\sum_{w \in Vocab} [w \; is \; the \; true \; word \; o] \times \log(\hat{y_w}) = -\log \hat{y_o} \tag{1}$$

2. Question (b):
   $U \in R^{d*n}$, the $J = CE(y, \hat{y}) = CE(y, softmax(z))$, the derivative is that is $\frac{\partial y}{\partial z} = \hat{y} - y$.
   And $z = U^T v_c$, so $\frac{\partial J}{\partial v_c} = U(\hat{y} - y)$

3. Question (c):

$$\frac{\partial J}{\partial U_{ij}} = \sum_k \frac{\partial J}{\partial z_k} = \sum_k (\hat{y} - y)|_k \frac{\partial z_k}{\partial U_{ij}} \tag{2}$$

   and if $j = k$,the derivative is v_i, so the total derivative is:

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

4. Question (d):
   $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

5. Question (e):

$$\frac{\partial J}{\partial v_c} = -u_o(1 - \sigma(u_o^T v_c)) - \sum_{k=1}^{K} (\sigma(-u_k^T v_c) - 1)u_k \tag{4}$$

$$\frac{\partial J}{\partial u_o} = -(1 - \sigma(u_o^T v_c)v_c \tag{5}$$

$$\frac{\partial J}{\partial u_k} = (1 - \sigma(-u_k^T v_c))v_c \tag{6}$$

   We only need to compute the gradient of the neg samples not whole.

6. Question (f):

$$\frac{\partial J_{skip-gram}(v_c, w_{t-m}, ..., w_{t+m}, U)}{\partial U} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial U} \tag{7}$$

$$\frac{\partial J_{skip-gram}(v_c, w_{t-m}, ..., w_{t+m}, U)}{\partial v_c} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c} \tag{8}$$

$$\frac{\partial J_{skip-gram}(v_c, w_{t-m}, ..., w_{t+m}, U)}{\partial v_w} = 0, w \neq c \tag{9}$$