Written Solution

Ahmed Bahaa

09/12/19

Question 1:

a- $\mathcal{Y}$ is a one hot vector ([0, 0, 0,...... 1, 0, 0,.....] where 1 corresponds to the target word. So this means that all the other terms from -log($\hat{\mathcal{Y}}$) will be neglected except the term corresponding to the 1.
b- $\frac{\partial(J)}{\partial(v_c)} = u_o(\hat{\mathcal{Y}} - \mathcal{Y})$
c- $\frac{\partial(J)}{\partial(u_c)} = v_c(\hat{\mathcal{Y}} - \mathcal{Y})$
d- $\frac{\partial(\sigma)}{\partial(x)} = \sigma(x)(1 - \sigma(x))$
$\sigma(x) = \frac{1}{1+e^{-x}}$
e- $J_{negativesampling}(vc, o, U) = -log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} log(\sigma(-u_o^T v_c))$
Derivative is composed of two parts and we differentiate according to $\frac{\partial(J)}{\partial(v_c)} = \frac{\partial(J)}{\partial(logterm)} \frac{\partial(logterm)}{\partial(\sigma)} \frac{\partial(\sigma)}{\partial(v_c)}$
First part we will see that we differentiate using chain rule as above: $\frac{1}{\sigma(u_o^T v_c)} \sigma(u_o^T v_c)(1 - \sigma(u_o^T v_c))u_o^T$
The $\sigma(u_o^T v_c)$ terms can be cancelled together. So final term is $-(1 - \sigma(u_o^T v_c))u_o^T = (\sigma(u_o^T v_c) - 1)u_o^T$

and Second part will be $\sum_{k=1}^{K}(1 - \sigma(-u_k^T v_c)) - u_k^T$

All Together $(\sigma(u_o^T v_c) - 1)u_o^T + \sum_{k=1}^{K}(1 - \sigma(-u_k^T v_c))u_k^T$

f- i) $\frac{\partial(J_s kipgram)}{\partial(U)} = \sum_{-m<j<m, j\neq 0} \frac{\partial(J_{skipgram}(v_c, w_{j+1}, U)}{\partial(U)}$
ii) $\frac{\partial(J_s kipgram)}{\partial(v_c)} = \sum_{-m<j<m, j\neq 0} \frac{\partial(J_{skipgram}(v_c, w_{j+1}, U)}{\partial(v_c)}$
iii) w != c $\frac{\partial(J_s kipgram)}{\partial(v_w)} = 0$