WORD2VEC

Huy Nguyen

November 23, 2019

Problem 1.

1. Apply Sofmax function

We have loss function is given by:

$$J_{navive-sofmax}(v_c, o, U) = -\log\big(\frac{\exp({u_0}^T v_c)}{\sum_{w\varepsilon Vocab}\exp({u_w}^T v_c)}\big)$$

From above, we can compute:

$$\frac{\partial J_{navive-sofmax}}{\partial v_c} = -u_0 + \sum_x \frac{\exp({u_x}^T v_c){u_x}}{\sum_{w\varepsilon Vocab}\exp({u_w}^T v_c)}$$

$$= U^T(\hat{y} - y)$$

$$\frac{\partial J_{avive-sofmax}}{\partial U} = (\hat{y} - y){v_c}^T$$

2. Negative Sampleing

We have loss function is given by:

$$J_{neg-sample}(v_c, o, U) = -\log\big(\sigma({u_0}^T v_c) - \sum_{k=1}^{K}\log(\sigma({u_k}^T v_c))\big)$$

From above, we can compute:

$$\frac{\partial J_{neg-sampling}}{\partial v_c} = \big(\sigma(u_0^T v_c) - 1\big)u_0 + u_k^T\big(1 - \sigma(-u_k v_c)\big)$$

$$\frac{\partial J_{neg-sampling}}{\partial u_0} = \big(\sigma(u_0^T v_c) - 1\big)v_c$$

$$\frac{\partial J_{neg-sampling}}{\partial u_k} = \big(1 - \sigma(-u_k^T v_c)\big)v_c$$