CS 224n Assignment 2: word2vec

Zhong Yinmin

May 13, 2020

1 Written: Understanding word2vec

(a) obviously

(b) $\partial \boldsymbol{J}_{naive-softmax}(v_c, o, \mathbf{U})/\partial v_c = \mathbf{U}(\hat{y} - y)$

(c) $\partial \boldsymbol{J}_{naive-softmax}(v_c, o, \mathbf{U})/\partial \mathbf{U} = v_c(\hat{y} - y)^T$

(d) $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

(e) $\partial \boldsymbol{J}_{neg-sample}(v_c, o, \mathbf{U})/\partial v_c = (\sigma(u_o^T v_c) - 1)u_o + \sum_{k=1}^{K} \sigma(u_k^T v_c)u_k$

$\partial \boldsymbol{J}_{neg-sample}(v_c, o, \mathbf{U})/\partial u_o = (\sigma(u_o^T v_c) - 1)v_c$

$\partial \boldsymbol{J}_{neg-sample}(v_c, o, \mathbf{U})/\partial u_k = \sigma(u_k^k v_c)v_c$ *for all k from 1 to K*

(f)

$$\partial \boldsymbol{J}_{skip-gram}(v_c, w_{t-m}, \ldots w_{t+m}, \mathbf{U})/\partial \mathbf{U} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \partial \boldsymbol{J}_{skip-gram}(v_c, w_{t+j}, \mathbf{U})/\partial \mathbf{U}$$

$$\partial \boldsymbol{J}_{skip-gram}(v_c, w_{t-m}, \ldots w_{t+m}, \mathbf{U})/\partial v_c = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \partial \boldsymbol{J}_{skip-gram}(v_c, w_{t+j}, \mathbf{U})/\partial v_c$$

$$\partial \boldsymbol{J}_{skip-gram}(v_c, w_{t-m}, \ldots w_{t+m}, \mathbf{U})/\partial v_w = 0 \qquad when \quad w \ne c$$

here $\partial \boldsymbol{J}_{skip-gram}(v_c, w_{t+j}, \mathbf{U})/\partial \mathbf{U}$ can be either *naive-softmax* or *neg-sample*,depending on your implementation.