CS 224n Assignment 2.

Abhishek Goswami.

January 20, 2019

1. Written : Understanding word2vec

   (a)

$$\sum_{w \in Vocab} y_w \log(\hat{y}_w) = -\log(\hat{y}_o). \tag{1}$$

because

$$y_i = \begin{cases} 1, & \text{if } i = o \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

   (b)

$$\frac{\partial \boldsymbol{J}_{\text{naive-softmax}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{v}_c} = -\boldsymbol{u}_o + \sum_{w=1}^{V} \hat{y}_w \boldsymbol{u}_w. \tag{3}$$

also equivalent to

$$\frac{\partial \boldsymbol{J}_{\text{naive-softmax}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{v}_c} = \boldsymbol{U}(\hat{\boldsymbol{y}} - \boldsymbol{y}). \tag{4}$$

   (c)

$$\frac{\partial \boldsymbol{J}_{\text{naive-softmax}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{U}} = \boldsymbol{v}_c(\hat{\boldsymbol{y}} - \boldsymbol{y})^\top. \tag{5}$$

also equivalent to

$$\frac{\partial \boldsymbol{J}_{\text{naive-softmax}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{U}} = \begin{cases} (\hat{y}_w - 1)\boldsymbol{v}_c, & \text{if } x \geq 1 \\ \hat{y}_w \boldsymbol{v}_c, & \text{otherwise} \end{cases}. \tag{6}$$

   (d)

$$\sigma'(\boldsymbol{x}) = \sigma(\boldsymbol{x})(1 - \sigma(\boldsymbol{x})) \tag{7}$$

   (e)

$$\frac{\partial \boldsymbol{J}_{\text{neg-sample}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{v}_c} = (\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c) - 1)\boldsymbol{u}_o - \sum_{k=1}^{k} (\sigma(\boldsymbol{u}_k^\top \boldsymbol{v}_c) - 1)\boldsymbol{u}_k \tag{8}$$

$$\frac{\partial \boldsymbol{J}_{\text{neg-sample}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{u}_c} = (\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c) - 1)\boldsymbol{v}_c. \tag{9}$$

$$\frac{\partial \boldsymbol{J}_{\text{neg-sample}}(\boldsymbol{v}_c, o, \boldsymbol{U})}{\partial \boldsymbol{u}_k} = -(\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c) - 1)\boldsymbol{v}_c. \tag{10}$$

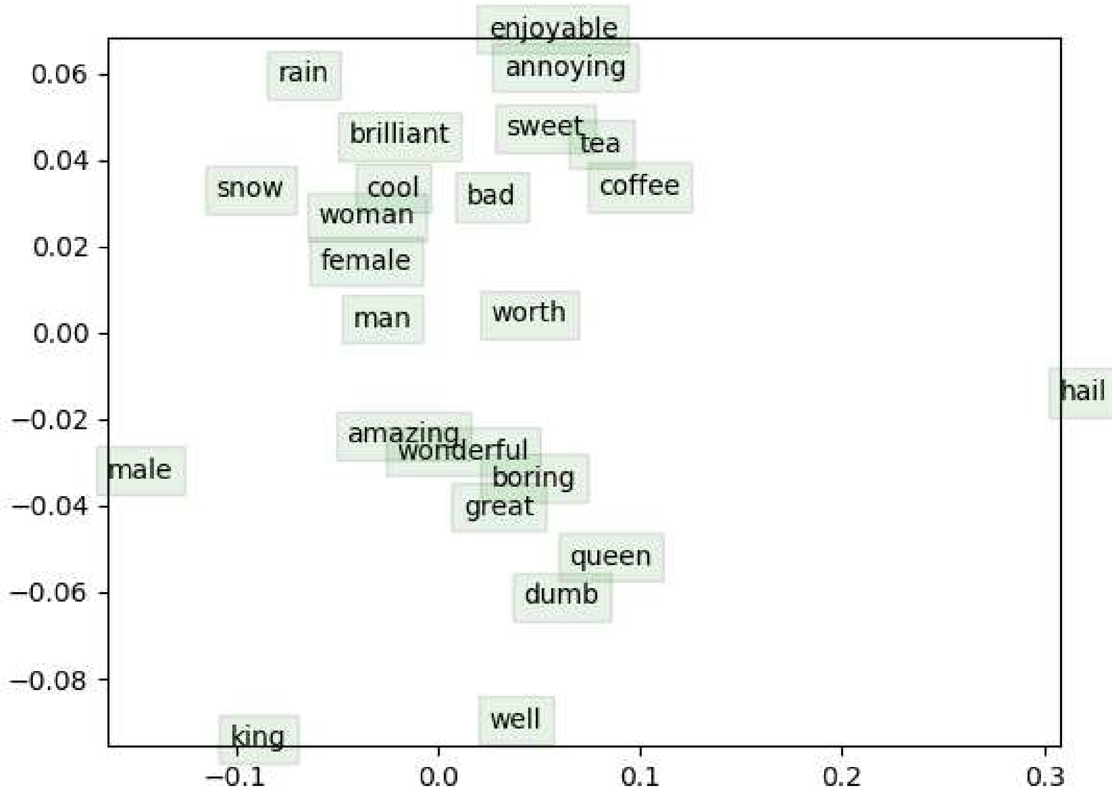

Figure 1: Word Vectors

(f) i.

$$\frac{\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, ...w_{t+m}, \boldsymbol{U})}{\partial \boldsymbol{U}} = \sum_{-m \le j \le m, j \ne 0} \frac{\partial \boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U})}{\partial \boldsymbol{U}}. \tag{11}$$

ii.

$$\frac{\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, ...w_{t+m}, \boldsymbol{U})}{\partial \boldsymbol{v}_c} = \sum_{-m \le j \le m, j \ne 0} \frac{\partial \boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U})}{\partial \boldsymbol{v}_c} \tag{12}$$

iii.

$$\frac{\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, ...w_{t+m}, \boldsymbol{U})}{\partial \boldsymbol{v}_w} = 0, \text{when} w \ne c \tag{13}$$

2. Coding