Assignment 2 writeup

Dima Lipin

January 18, 2019

1

Figure 1.1: European swallow.

2 Theory

a)

$$-\sum_{w} y_w log(\hat{y}_w) = -log(\hat{y}_o) \tag{2.1}$$

as only $y_o = 1$ and other elements are zeros.
b)

$$\frac{\partial J(v_c, o, U)}{\partial v_c} = \tag{2.2}$$

$$-\frac{\partial(u_o^T v_c)}{\partial v_c} + \frac{\log(\sum_w exp(u_w^T v_c))}{\partial v_c} = \tag{2.3}$$

$$-u_o + \sum_w p(O = w|C = c)u_w = U(\hat{y} - y) \tag{2.4}$$

c)
Let $w \neq o$

$$\frac{\partial J(v_c, o, U)}{\partial u_w} = \tag{2.5}$$

$$-\frac{\partial(u_o^T v_c)}{\partial u_w} + \frac{\log(\sum_w exp(u_w^T v_c))}{\partial u_w} = \tag{2.6}$$

$$p(O = w|C = c)v_c = \hat{y}_w v_c \tag{2.7}$$

Now let $w = o$
We get: $\frac{\partial J(v_c,o,U)}{\partial u_w} = (p(O = o|C = c) - 1)v_c = (\hat{y}_c - y_c)v_c$
d)

$$\frac{\partial \sigma(x)}{\partial x} = \frac{1}{(1 + \exp(-x))^2} \exp(-x) = \sigma(x)(1 - \sigma(x)) \tag{2.8}$$

e)

$$J_{ns}(v_c, o, U) = -\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} \log(\sigma(-u_k^T v_c)) \tag{2.9}$$

$$\frac{\partial J_{ns}}{v_c} = \tag{2.10}$$

$$-(1 - \sigma(u_o^T v_c))u_o + \sum_{k=1}^{K} (1 - \sigma(-u_k^T v_c))u_k = \tag{2.11}$$

$$-\sigma(-u_o^T v_c)u_o + \sum_{k=1}^{K} \sigma(u_k^T v_c)u_k \tag{2.12}$$

$$\frac{\partial J_{ns}}{u_0} = -\sigma(-u_o^T v_c)v_c \tag{2.13}$$

$$\frac{\partial J_{ns}}{u_k} = \sigma(u_k^T v_c)v_c \tag{2.14}$$

$$\tag{2.15}$$

This loss function is more efficient as it requires less computation (about $\frac{K}{V}$ less operations for computing the loss) , also there is no expensive matrix multiplication to compute grad of v_c.
(f)

$$J_{sg}(v_c, w_{t-m} \ldots w_{t+m}, U) = \sum_{-m \leq j \leq m, j \neq 0} J(v_c, w_{t+j}, U) \tag{2.16}$$

(i) $\frac{\partial J_{sg}}{\partial U} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial U}$

(ii) $\frac{\partial J_{sg}}{\partial v_c} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c}$

(iii) $\frac{\partial J_{sg}}{\partial v_w} = 0$ (when $w \neq c$)

3 Practical part

Below is obtained visualization for trained word vectors. We can notice some patterns there: synonyms are grouped together, like amazing and wonderful. But as antonyms could belong to the same context frequently, they also occur with synonyms together (female and male, enjoyable and annoying).

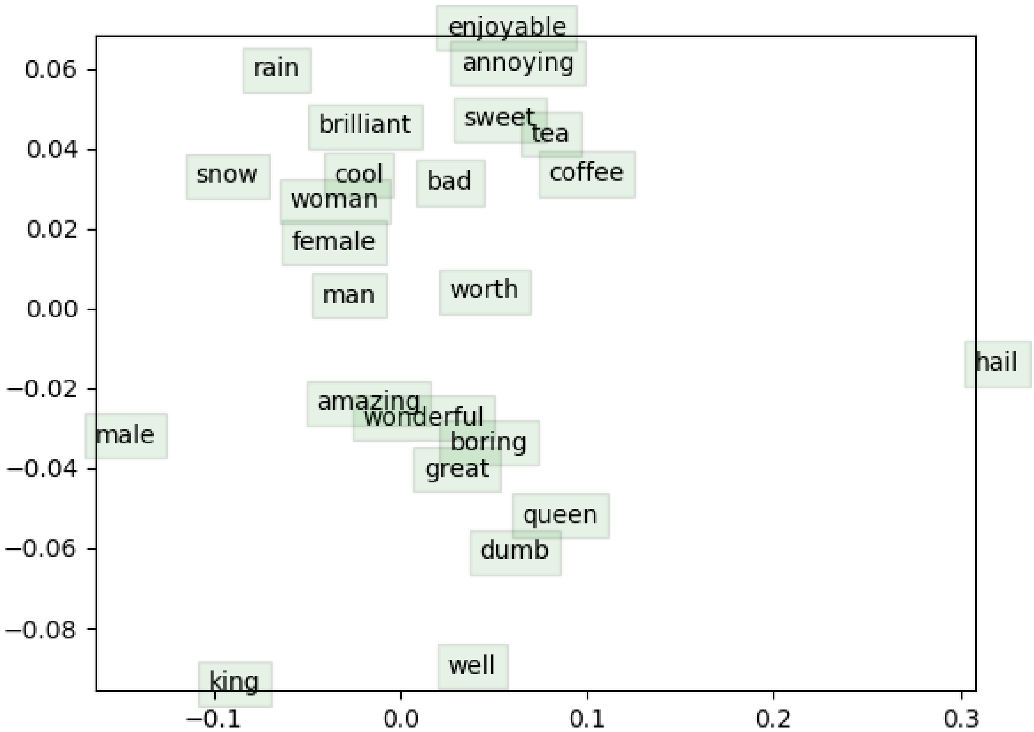

Figure 3.1: A boat.