CS224N
Assignment 2

Name: Oscar
SUID: 05735451

Problem 1: Understanding word2vecl

(i) A The proof that these two are equal should be

$$-\sum y_w log(\hat{y}_w) = -log(\hat{y}_o)$$

This is true because y_w will be equal to zero whenever w is not o. This means that the summation will, in practical terms, disappear, making the above expression true.

(ii) B: $\frac{dj}{dv_c}$

$$J = -log(\frac{exp(u_o^T v_c)}{\sum exp(u_o^T v_c)}) = -log(exp(u_o^T v_c)) + log(\sum exp(u_o^T v_c))$$

$$\frac{dj}{dv_c} = -u_0^T + \frac{1}{\sum exp(u_o^T v_c)} \sum exp(u_o^T v_c) u_o^T$$

$$\frac{dj}{dv_c} = -u_0^T + \sum \hat{y} u_w = \sum_w u_w(\hat{y}_w - y_w)$$

(iii) C: $\frac{dj}{du}$

o = w

$$J = -log(\frac{exp(u_o^T v_c)}{\sum exp(u_o^T v_c)}) = -log(exp(u_o^T v_c)) + log(\sum exp(u_o^T v_c))$$

$$\frac{dj}{du} = -v_c^T + \frac{1}{\sum exp(u_o^T v_c)} \sum exp(u_o^T v_c) v_c^T = -v_c^T + \hat{y} v_c$$

o ≠ w

$$\frac{dj}{du} = \frac{1}{\sum exp(u_o^T v_c)} \sum exp(u_o^T v_c) v_c^T = \hat{y}_w v_c$$

(iv) D

$$\frac{d\sigma}{dx} = -(1 + e^{-z})^{-2} e^{-z} = \frac{= e^{-z} + 1 - 1}{(1 + e^{-z})^2}$$

$$\frac{d\sigma}{dx} = \frac{1}{1 + e^{-z}} - \frac{1}{1 + e^{-z}}^2 = \sigma(1 - \sigma)$$

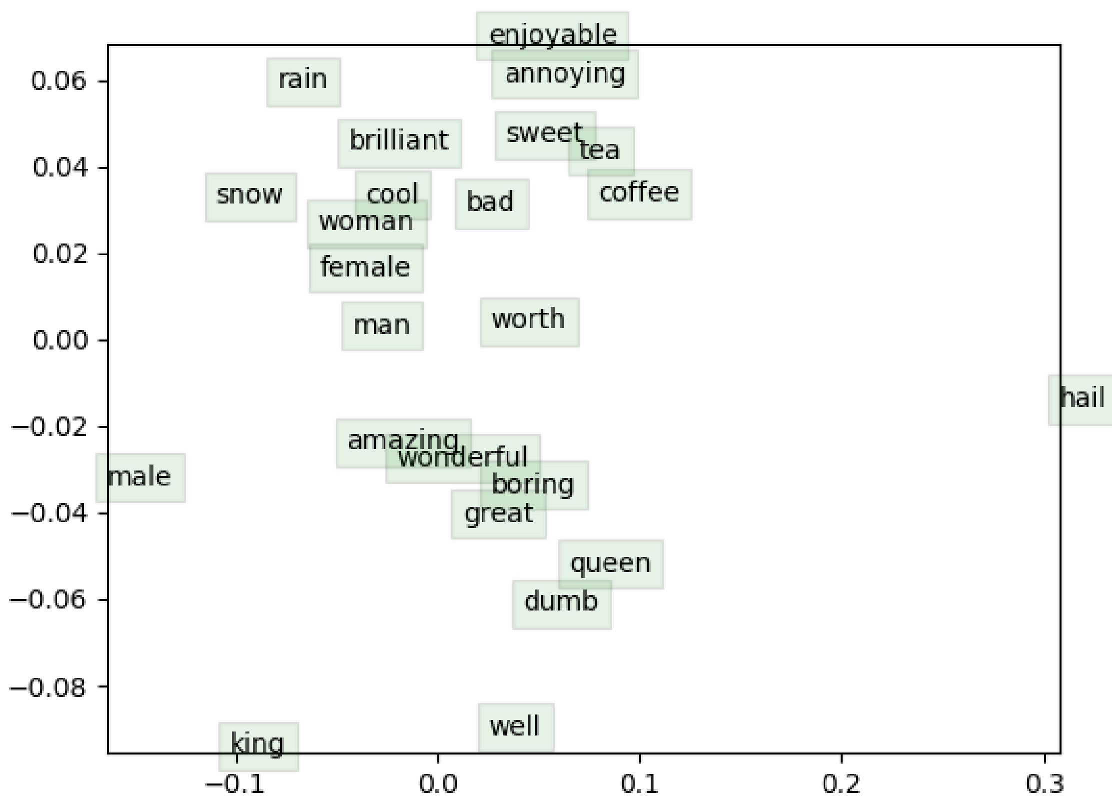
(v) E

$$J = -log(\sigma(u_o^T v_c)) - \sum^K log(\sigma(-u_k^T v_c))$$

$$\frac{dJ}{v_c} = -\frac{1}{\sigma(u_o^T v_c)}\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))u_o - \sum^K \frac{1}{\sigma(-u_k^T v_c)}\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))(-u_k)$$
$$\frac{dJ}{v_c} = -(1-\sigma(u_o^T v_c))u_o - \sum^K (1-\sigma(-u_k^T v_c))(-u_k)$$
$$\frac{dJ}{u_o} = -\frac{1}{\sigma(u_o^T v_c)}\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))v_c^T$$
$$\frac{dJ}{u_o} = -(1-\sigma(u_o^T v_c))v_c^T$$
$$\frac{dJ}{u_k} = \frac{1}{\sigma(-u_k^T v_c)}\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c)) - (v_c)$$
$$\frac{dJ}{u_k} = (1-\sigma(-u_k^T v_c)) - v_c$$

(vi) F

$$\frac{dJ}{dU} = \sum^m \frac{dJ}{dU}$$
$$\frac{dJ}{dv_c} = \sum^m \frac{dJ}{dv_c}$$
$$\frac{dJ}{dv_w} = \sum^m \frac{dJ}{dv_w}$$



This plot is interesting because we can see certain associations between word vectors. There's a group of adjectives grouped in the center. It's also interesting seeing the associations between man and king and queen and women. t