CS598: CXZ	**Fall 2016**
Problem Set 6	
chen zhang, NetId: czhang49	*Handed In: November 3, 2016*

1 Word Embeddings via Singular Value Decomposition

1.1 a

As attached code file.

1.2 b

1.2.1 i

We can use the following equation:

$$\vec{q} = \vec{e}_x - \vec{e}_y + \vec{e}_z$$
$$\hat{q} = \frac{\vec{q}}{||q||}$$

The scoring factor is then

$$\text{score} = \vec{v} \cdot \hat{q}$$

in which $\vec{v}$ is the vector of words and $\vec{v} \in W$

1.2.2 ii

What I got is :
king (0.991194367409)
husband (0.991003453732)
mother (0.990785479546)
father (0.990462422371)
miss (0.990008413792)
daughter (0.990000665188)
brother (0.989433646202)
club (0.989403545856)
friends (0.989162325859)
hall (0.988909602165)

These are the top 10 words I got, and queen is not among them. This may have to do with insufficient word pair of (woman, queen).

Analogy : student to school is as teacher to XX
teacher (0.988462507725)
college (0.988054871559)
brown (0.98684155941)
miss (0.986574411392)
parents (0.986404180527)
child (0.985893189907)
couple (0.985881924629)
mary (0.985849678516)
martin (0.985593140125)
dr (0.985047757626)

Qualitatively the results is acceptable, it falls in the range of words that make sense.

2 Evaluating Word Embeddings via an Analogy Task

2.1 a

MRR considers the rank first relevant document. Accuracy has the expression of $\frac{\text{true positive} + \text{true negati}}{\text{positive} + \text{negative}}$

If we want one good response and that response needs to be on top of the rank, we prefer MRR. In the scenario where the user would want one good response which appears on top of the rank, MRR is a good measure.
If we want several response and we are willing to look over all the response, we can use accuracy. In the scenario where the user would want several response and would want to browse through, accuracy is a good measure.

2.2 b

The modified code is as in the attachment.

2.3 c

2.3.1 i

SVD p=1.0: 4.9e-40
SVD p=0.5: 3.2e-39
SVD p=0.0: 5.3e-41
GloVe: 3.58e-41

We see that as p decrease, the MRR has a trend to first increase and then decrease. Maybe this is due to effect of the position and value of the singular values.

2.3.2 ii

In order to run the t-test, the process needs to be done again and calculate the p value.
SVD p=1.0: 4.72e-40
SVD p=0.5: 5.65e-39
SVD p=0.0: 4.8e-41
GloVe: 7.24e-40

The p value calculated is 0.0544, which is considered to be almost statistically significant.

3 Evaluating Word Embeddings via a Similarity Task

3.1 a

SVD p=1.0: 0.396551724138
SVD p=0.5: 0.379310344828
SVD p=0.0: 0.5
GloVe: 0.362068965517

From the results it turns out that when $p = 0$, it gives the best result. This is sort of expected because we are giving 0 weights to the singular values.

3.2 b

In order to do the t-test, the simulator needs to be run again, so the result for the second time is :
SVD p=1.0: 0.392625469443
SVD p=0.5: 0.385455786958
SVD p=0.0: 0.50495
GloVe: 0.35848412427

The p value calculated is 0.0009, which is considered to be extremely statistically significant.

3.3 c

yes I meet this problem. I suggest allowing the user to input more than one answer, so that we can get this information.

3.4 d

yes I meet this problem. I suggest putting another choice : None of the above. So that we can capture this information.