Word and sentence embeddings [Solution by Karthikeyan.S]

LATEST SUBMISSION GRADE

100%

1.Question 1

Compute a second-order co-occurrence between the words *'These'* and *'So'* (the cosine similarity between their first-order co-occurrence vectors). Use the toy corpus:

These are the wrong sort of bees. Quite the wrong sort. So I should think they would make the wrong sort of honey.

- Let's define a context of a word as three words to the left and three words to the right from the target word, **occurred within the same sentence** (if there are any).
- For the first-order co-occurrence, let's consider pPMI values (the formula was given on slide 5 of the first video).

Hint: in this question you actually do not need to *compute* anything... And the answer would be the same for any type of first-order co-occurrence.

- [x] 0
- [ ] 1
- [ ] $\frac{\sqrt{3}}{2}$
- [ ] 2
- [ ] $\frac{1}{2}$

Correct

2.Question 2

Choose correct statements about Singular Value Decomposition (SVD), an important notion from the linear algebra. Feel free to consult any additional resource like wiki if needed.

- [x] Squares of singular values of a matrix X are eigenvalues of $X^T X$ (or $X X^T$).
- [x] Singular values decomposition is not unique (for example, the zero matrix can be decomposed in infinitely many ways).

☑

Truncated SVD is the best rank k approximation of the original matrix in terms of Frobenius norm.

☑

Any rectangular matrix with real entries has a singular value decomposition.

☐

Singular values can be negative.

☐

Singular values of a rectangular matrix are its eigenvalues.

Correct

3.Question 3

Find the objective function of the skip-gram negative sampling (SGNS) model.

◉

\sum\limits_{u \in W} \sum\limits_{v \in C} \left(n_{uv} \log \sigma(\langle \phi_u, \theta_v\rangle) + k \mathbb{E}_{\bar{v}} \log \sigma(-\langle \phi_u, \theta_{\bar{v}} \rangle) \right)$\sum_{u\in W}\sum_{v\in C}(n_{uv}\log\sigma(\langle\phi_u,\theta_v\rangle)+k\mathrm{E}_{\bar{v}}\log\sigma(-\langle\phi_u,\theta_{\bar{v}}\rangle))$

○

\sum\limits_{u \in W} \sum\limits_{v \in C} n_{uv} \frac{\exp\langle \phi_u, \theta_v\rangle}{\sum_{\bar{u} \in W} \exp \langle \phi_{\bar{u}}, \theta_v \rangle}$\sum_{u\in W}\sum_{v\in C}n_{uv}\frac{\exp\langle\phi_u,\theta_v\rangle}{\sum_{\bar{u}\in W}\exp\langle\phi_{\bar{u}},\theta_v\rangle}$

○

$\sum_{u\in W}\sum_{v\in C}(n_{uv}\langle\phi_u,\theta_v\rangle-k\mathrm{E}_{\bar{v}}\langle\phi_u,\theta_{\bar{v}}\rangle)$

○

\sum\limits_{u \in W} \sum\limits_{v \in C} f(n_{uv}) \left(\langle \phi_u, \theta_v\rangle + b_{u} + b_{v}' - \log n_{uv} \right)^2$\sum_{u\in W}\sum_{v\in C}f(n_{uv})(\langle\phi_u,\theta_v\rangle+b_u+b_{v'}-\log n_{uv})^2$

Correct

4.Question 4

How are word embeddings usually evaluated (qualitatively or quantitively)?

☑

By the accuracy of analogy prediction (using some pre-defined dataset of 4-word analogies).

☑

By Spearman's correlation (or similar rank correlation measure) with human judgements on word similarity task.

☐

By the amount of positive components of word vectors.

☑

By the interpretability of the components of the vectors.

☐

By comparing maximal lengths of word vectors (the more is the length, the better is the model).

Correct

5.Question 5

Choose the correct statements.

☐

Word2vec works fine for word analogies, but there are many concerns with word similarities.

☑

For word similarity tasks, count-based methods perform on par with predictive methods.

☐

Skip-gram negative sampling (SGNS) model is too hard to train, and it is often approximated with softmax.

☑

Representations of word or character n-grams may improve the quality of the model.

Correct