Week 10 Word Embedding Quiz Solution

1. Why is word embedding important?
   answer: word embedding allows words to be represented by a series of real numbers that contain semantic meaning. Therefore, words can be efficient training data for various machine learning tasks.

2. With Bag of Words approach, vectorize the following two documents.

   Sentence 1 : The cat is running on the road.

   Sentence 2: The dog is sleeping on the couch.

   answer:

   [the, cat, is, running, on, the, road, dog, sleeping, couch]

   [1, 1, 1, 1, 1, 1, 1, 0, 0, 0]

   [1, 0, 1, 0, 1, 1, 0, 1, 1, 1]

3. What is the TF-IDF formula used in the assignment?

   answer:

$$tfidf(t, d, D) = \frac{n_{t,d}}{\sum_k n_{k,d}} * \ln(\frac{N}{df_t}) \tag{1}$$

4. (T/F) If there is a review containing 100 words, then the vector representation of the review must have at least 100 dimension in order to capture all the information.

   answer: F

5. Calculate the TF-IDF score of word "The" and "cat" for the following two documents.

   Sentence 1 : The cat is running on the road.
   Sentence 2: The dog is sleeping on the couch.

   answer:

   tfidf("The") = 1/7 * log(2/2) = 0
   tfidf("cat") = 1/7 * log(2/1)

6. What is the main difference between Bag of Words and TF-IDF?
   answer: Bag of Words creates a vector containing the count of word occurrences in the review, while the TF-IDF model contains the weights on the importance of the words to the review, in addition to term frequency.

7. Construct the co-occurrence matrix, window size = 3, with the following documents.

   Sentence 1 : I want to finish that show.

   Sentence 2: Did you watch that show on Netflix?

   Sentence 3: I will watch Netflix show.

   answer:

	I	want	to	finish	that	show	did	you	watch	on	Netflix	will
I	0	1	1	1	0	0	0	0	1	0	1	1
want	1	0	1	1	1	0	0	0	0	0	0	0
to	1	1	0	1	1	1	0	0	0	0	0	0
finish	1	1	1	0	1	1	0	0	0	0	0	0
that	0	1	1	1	0	2	1	1	1	1	1	0
show	0	0	1	1	2	0	0	1	2	1	2	1
did	0	0	0	0	1	0	0	1	1	0	0	0
you	0	0	0	0	1	1	1	0	1	0	0	0
watch	1	0	0	0	1	2	1	1	0	1	1	1
on	0	0	0	0	1	1	0	0	1	0	1	0
Netflix	1	0	0	0	1	2	0	0	1	1	0	1
will	1	0	0	0	0	1	0	0	1	0	1	0

8. Make training samples for Word2Vec for the following sentence with window size = 2.

   Hope the pandemic will end soon

   answer: (**Hope** the pandemic), (Hope **the** pandemic will), (Hope the **pandemic** will end), (the pandemic **will** end soon), (pandemic will **end** soon), (will end **soon**)

9. Briefly state the difference between Continuous Bag of Words (CBOW) and Skip Gram as two methods for Word2Vec.
   answer: CBOW takes in the context words and predicts the center word, while Skip Gram takes in the center word and predicts the context words.

10. Write the softmax function $\sigma(z)_i$ and take the derivative w.r.t z_i and z_j.
    answer:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{V} e^{z_j}} \tag{2}$$

$$\frac{\partial \sigma(z)_i}{\partial z_i} = \frac{e^{z_i} \sum_j^V e^{z_j} - e^{2z_i}}{(\sum_{j=1}^{V} e^{z_j})^2} \tag{3}$$

$$\frac{\partial \sigma(z)_i}{\partial z_j} = \frac{-e^{z_i} e^{z_j}}{(\sum_{j=1}^{V} e^{z_j})^2} \tag{4}$$

11. Derive the loss function for CBOW and Skip Gram respectively.
    answer:
    For CBOW:

$$\begin{aligned} E &= -\log p(W_O | W_{I,1}, ..., W_{I,C}) \\ &= -\log \frac{e^{u_{j*}}}{\sum_{j'=1}^{V} e^{u_{j'}}} \\ &= -u_{j*} + \log \sum_{j'=1}^{V} e^{u_{j'}} \end{aligned} \tag{5}$$

    For Skip Gram:

$$\begin{aligned} E &= -\log p(W_{O,1}, ..., W_{O,C} | W_I) \\ &= -\log \prod_{c=1}^{C} \frac{e^{u_{j_c^*}}}{\sum_{j'=1}^{V} e^{u_{j'}}} \\ &= -\sum_{c=1}^{C} u_{j_c^*} + C * \log \sum_{j'=1}^{V} e^{u_{j'}} \end{aligned} \tag{6}$$

12. For each question, choose your answers from these options: BoW, TFIDF, Co-occurrence, GloVe, Word2vec.

    - Which of the word embedding methods uses count-based/frequency-based strategy?

- Which of the word embedding methods incorporate predictive models?
- Which of the word embedding methods aim to learn geometric encoding of terms/tokens?

answer:
BoW, TFIDF, Co-occurrence, GloVe
GloVe, Word2vec
Co-occurrence, GloVe, Word2vec