

## Week 10 Word Embedding Quiz

---

1. Why is word embedding important?
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 coach.

3. What is the TF-IDF formula used in the assignment?
4. (T/F) If there is a review containing 100 words, then the vector representation of the review must have 100 dimension in order to capture all the information.
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 coach.

6. What is the main difference between Bag of Words and TF-IDF?
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.

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

Hope the pandemic will end soon

9. Briefly state the difference between Continuous Bag of Words (CBOW) and Skip Gram as two methods for Word2Vec.
10. Write the softmax function  $\sigma(z)_i$  and take the derivative w.r.t  $z_i$  and  $z_j$ .
11. Derive the loss function for CBOW and Skip Gram respectively.

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?