DEEP LEARNING WEEK 9

1. We have 100 words in a vocabulary. What will be the size of the one hot encoded vector for a given word of this vocabulary?
   a) 100
   b) 2^{10}
   c)2^{100}
   d)10000

   Answer:a)

   Solution Each one-hot vector contains a 1 at a unique position and 0 at others. The size of the vector needed to accumulate 100 such words is 100.

2. In CBOW we multiply the one hot encoded-word with the a matrix. What does the ith column of this matrix represent?
   a) Eigenvector of W
   b) Contextual representation of ith word in vocabulary
   c) Word representation of the ith word in the vocabulary
   d) Insufficient information

   Answer: b)

   Solution: In CBOW we predict the word given the context. Hence the first matrix gives the contextual representation of the ith word and the second matrix gives the word representation of the ith word.

3. Which of the following is a possible mathematical representation of one hot encoding for a word w in a vocabulary V? (Assume all the nonvisible values in the vector to be 0)
   a) [0, 1, 0, ..., 0]
   b) [1, 0, 1, ..., 0]
   c) [0, 0, 0, ..., 0]
   d) [1, 1, 1, ..., 1]

   Answer: a) [0, 1, 0, ..., 0]

   Explanation: In one hot encoding, a word w in a vocabulary V is represented as a binary vector where only one element is 1 and the rest are 0.

4. Which of the following is a major limitation of the CBOW model?
   a. It is not suitable for learning word embeddings on large datasets
   b. It does not capture the order of words in the context
   c. It is prone to overfitting
   d. It is slower to train than skip-gram

   Answer: b. It does not capture the order of words in the context

   Solution: The CBOW model does not capture the order of words in the context, since the context is represented as a bag of words. This can be a limitation in tasks where the order of words is important, such as natural language generation.

5. What is the objective of the skip-gram method?
   a) To predict the probability distribution of target words given context words
   b) To predict the probability distribution of context words given a target word
   c) To identify the most frequent words in a corpus

d) To cluster similar words together

Answer: b) To predict the probability distribution of context words given a target word

6. Suppose we are learning the representations of words using Glove representations. If we observe that the cosine similarity between two representations v_i and v_j for words 'i' and 'j' is very high. which of the following statements is true?(parameter $b_i = 0.02$ and $b_j = 0.05$

a)log(Xij)= 0.03.
b)log(Xij) = 0.8.
c)log(Xij) = 0.35.
d)log(X_{ij})= 0.

Answer: b)
Solution: Since the word representations are similar we know $v_i^T v_j$ is high but $v_i^T v_j = X_{ij} - b_i - b_j$. Hence X_{ij} is high but the only high value for X_{ij} is 0.8

7. What is the computational complexity of computing the softmax function in the output layer of a neural network?
a) O(n)
b) $O(n^2)$
c) $O(nlogn)$
d) $O(logn)$

Answer: a)

Explanation: The computational complexity of computing the softmax function in the output layer of a neural network is O(n), where n is the number of output classes

We are given the following corpus. Answer the following questions based on the given corpus. (Treat uppercase and lowercase as same, ignore the punctuation marks like ',')
Given a sound clip of a person or people speaking, separate it into words
Given a text, transform those units and produce a spoken representation
Separate a chunk of continuous text into separate words
Given a sentence, determine the part of speech for each word

8. What is the size of the co-occurrence matrix created using this vocabulary? (Assume stride=1)
a)50×50
b)44×44
c)25×25
d)31×31

Answer: d)

Solution: The size of the vocabulary is 31, hence the dimension of the co-occurrence matrix is 31x31

9. What is the PMI of (Given, a) in the text?
a)2.9
b)3.9
c)1.8
d)2.5

Answer: a)

Solution: PMI=log(c(Given,a)*N/c(Given)*c(a)). Putting the respective values and calculating will give the correct answer

10. Suppose we get a co-occurrence matrix of dimension $R^{m \times n}$ and we run SVD on this matrix. We do a k-rank approximation of this matrix using SVD. What is the dimension of $W_{word} = U \sum$?
a)$m \times n$
b)$n \times k$
c)$k \times n$
d)$m \times k$

Answer: d) Solution: Refer to lectures from week 5 on SVD.