X

Week 9 : Assignment 9

The due date for submitting this assignment has passed.

Due on 2024-09-25, 23:59 IST.

Assignment submitted on 2024-09-25, 20:37 IST

1) At the input layer of a continuous bag of words model, we multiply a one-hot vector $x \in \mathbb{R}^{|V|}$ with the parameter matrix $W \in \mathbb{R}^{k \times |V|}$. What does each column of W correspond to? ***1 point***

- (●) the representation of the i-th word in the vocabulary
- ( ) the i-th eigen vector of the co-occurrence matrix

Yes, the answer is correct.
Score: 1

Accepted Answers:
the representation of the i-th word in the vocabulary

2) Suppose that we use the continuous bag of words (CBOW) model to find vector representations of words. Suppose further that we use a context window of size 3 (that is, given the 3 context words, predict the target word $P(w_t|(w_i, w_j, w_k))$). The size of word vectors (vector representation of words) is chosen to be 100 and the vocabulary contains 10,000 words. The input to the network is the one-hot encoding (also called 1-of-V encoding) of word(s). How many parameters (weights), excluding bias, are there in W_{word}? Enter the answer in thousands. For example, if your answer is 50,000, then just enter 50.

1000

Yes, the answer is correct.
Score: 1

Accepted Answers:
(Type: Numeric) 1000

1 point

3) You are given the one hot representation of two words below:
CAR$= [1, 0, 0, 0, 0]$, BUS$= [0, 0, 0, 1, 0]$

What is the Euclidean distance between CAR and BUS?

1.414

Yes, the answer is correct.
Score: 1
Accepted Answers:
(Type: Range) 1.40,1.42

1 point

4) Consider a skip-gram model trained using hierarchical softmax for analyzing scientific literature. We observe that the word embeddings for 'Neuron' and 'Brain' are highly similar. Similarly, the embeddings for 'Synapse' and 'Brain' also show high similarity. Which of the following statements can be inferred? ***1 point***

- [x] 'Neuron' and 'Brain' frequently appear in similar contexts
- [x] The model's learned representations will indicate a high similarity between 'Neuron' and 'Synapse'
- [ ] The model's learned representations will not show a high similarity between 'Neuron' and 'Synapse'
- [ ] According to the model's learned representations, 'Neuron' and 'Brain' have a low cosine similarity

Yes, the answer is correct.
Score: 1
Accepted Answers:
'Neuron' and 'Brain' frequently appear in similar contexts
The model's learned representations will indicate a high similarity between 'Neuron' and 'Synapse'

5) Suppose we are learning the representations of words using Glove representations. ***1 point***
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 = 0.07)

- $X_{ij} = 0.02$
- $X_{ij} = 0.2$
- $X_{ij} = 0.88$
- $X_{ij} = 0$

Yes, the answer is correct.
Score: 1
Accepted Answers:
$X_{ij} = 0.88$

6) Which of the following is a disadvantage of one hot encoding? ***1 point***

- It requires a large amount of memory to store the vectors

- It can result in a high-dimensional sparse representation
- It cannot capture the semantic similarity between words
- All of the above

Yes, the answer is correct.
Score: 1
Accepted Answers:
All of the above

7) Which of the following is an advantage of using the skip-gram method over the bag-of-words approach? ***1 point***

- The skip-gram method is faster to train
- The skip-gram method performs better on rare words
- The bag-of-words approach is more accurate
- The bag-of-words approach is better for short texts

Yes, the answer is correct.
Score: 1
Accepted Answers:
The skip-gram method performs better on rare words

8) Suppose we are learning the representations of words using Glove representations. ***1 point***
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

- $X_{ij} = 0.03.$
- $X_{ij} = 0.8.$
- $X_{ij} = 0.35.$
- $X_{ij} = 0.$

Yes, the answer is correct.
Score: 1
Accepted Answers:
$X_{ij} = 0.8.$

9) What is the computational complexity of computing the softmax function in the output layer of a neural network? ***1 point***

- $O(n)$
- $O(n^2)$
- $O(nlogn)$
- $O(logn)$

Yes, the answer is correct.
Score: 1
Accepted Answers:
$O(n)$

10) What is the disadvantage of using Hierarchical Softmax? ***1 point***

- (●) It requires more memory to store the binary tree
- ( ) It is slower than computing the softmax function directly
- ( ) It is less accurate than computing the softmax function directly
- ( ) It is more prone to overfitting than computing the softmax function directly

Yes, the answer is correct.
Score: 1

Accepted Answers:
It requires more memory to store the binary tree