

X


<https://swayam.gov.in>

https://swayam.gov.in/nc_details/NPTEL

2111cs010024@mallareddyuniversity.ac.in ✓

NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)

Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

☐ One-hot representations of words

Week 9 : Assignment 9

The due date for submitting this assignment has passed.

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

As per our records you have not submitted this assignment.

1) Let X be the co-occurrence matrix such that the (i, j) -th entry of X captures the PMI between the i -th and j -th word in the corpus. Every row of X corresponds to the representation of the i -th word in the corpus. Suppose each row of X is normalized (i.e., the L_2 norm of each row is 1) then the (i, j) -th entry of XX^T captures the: **1 point**

☐
PMI between word i and word j
☐
Euclidean distance between word i and word j
☐
Probability that word i
☐
Cosine similarity between word i

No, the answer is incorrect.

Score: 0

Accepted Answers:

Cosine similarity between word i

2) Consider the following corpus: "human machine interface for computer applications. user opinion of computer system response time. user interface management system. system engineering for improved response time". What is the size of the vocabulary of the above corpus? **1 point**

☐ 13

☐ 14

☐ 15

☐ 16


(unit?
unit=115&less
on=116)

☐ Distributed
Representatio
ns of words
(unit?
unit=115&less
on=117)

☐ SVD for
learning word
representation
s (unit?
unit=115&less
on=118)

☐ SVD for
learning word
representation
s (Contd.)
(unit?
unit=115&less
on=119)

☐ Continuous
bag of words
model (unit?
unit=115&less
on=120)

☐ Skip-gram
model (unit?
unit=115&less
on=121)

☐ Skip-gram
model (Contd.)
(unit?
unit=115&less
on=122)

☐ Contrastive
estimation
(unit?
unit=115&less
on=123)

☐ Hierarchical
softmax (unit?
unit=115&less
on=124)

☐ GloVe
representation
s (unit?
unit=115&less
on=125)

No, the answer is incorrect.

Score: 0

Accepted Answers:

15

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

☐

the representation of the i -th word in the vocabulary

☐

the i -th eigen vector of the co-occurrence matrix

No, the answer is incorrect.

Score: 0

Accepted Answers:

the representation of the i -th word in the vocabulary

4) 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.

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Numeric) 1000

1 point

5) Consider a skip-gram model trained using hierarchical softmax for analyzing **1 point**
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?

☐

'Neuron' and 'Brain' frequently appear in similar contexts

☐

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

No, the answer is incorrect.

Score: 0

Accepted Answers:

'Neuron' and 'Brain' frequently appear in similar contexts

The model's learned representations will indicate a high similarity between 'Neuron' and

'Synapse'



- ☐ Evaluating word representation s (unit? unit=115&lesson=126)
- ☐ Relation between SVD and Word2Vec (unit? unit=115&lesson=127)
- ☐ Lecture Material for Week 9 (unit? unit=115&lesson=128)
- ☐ Week 9 Feedback Form: Deep Learning - IIT Ropar (unit? unit=115&lesson=192)
- ☐ **Quiz: Week 9 : Assignment 9 (assessment? name=297)**

week 10 ()

Week 11 ()

Week 12 ()

Download Videos ()

Books ()

Text Transcripts ()

Problem Solving Session - July 2024 ()

6) Which of the following is an advantage of the CBOW model compared to the Skip-gram model? **1 point**

- ☐ It is faster to train
- ☐ It requires less memory
- ☐ It performs better on rare words
- ☐ All of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

It is faster to train

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

No, the answer is incorrect.

Score: 0

Accepted Answers:

The skip-gram method performs better on rare words

8) What is the role of the softmax function in the skip-gram method? **1 point**

- ☐ To calculate the dot product between the target word and the context words
- ☐ To transform the dot product into a probability distribution
- ☐ To calculate the distance between the target word and the context words
- ☐ To adjust the weights of the neural network during training

No, the answer is incorrect.

Score: 0

Accepted Answers:

To transform the dot product into a probability distribution

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

- ☐ $O(n)$
- ☐ $O(n^2)$
- ☐ $O(n \log n)$
- ☐ $O(\log n)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$O(n)$



10) How does Hierarchical Softmax reduce the computational complexity of computing the softmax function? **1 point**

- ☐ It replaces the softmax function with a linear function
- ☐ It uses a binary tree to approximate the softmax function
- ☐ It uses a heuristic to compute the softmax function faster
- ☐ It does not reduce the computational complexity of computing the softmax function

No, the answer is incorrect.

Score: 0

Accepted Answers:

It uses a binary tree to approximate the softmax function

