

X


<https://swayam.gov.in>

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

2111cs010077@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 representation s of words

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, 20:17 IST

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

Yes, the answer is correct.

Score: 1

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)

Yes, the answer is correct.

Score: 1

Accepted Answers:

15

3) Let $count(w, c)$ be the number of times the words w and c appear together in the **1 point** corpus (i.e., occur within a window of few words around each other). Further, let $count(w)$ and $count(c)$ be the total number of times the word w and c appear in the corpus respectively and let N be the total number of words in the corpus. The PMI between w and c is then given by:

☐ $\log \frac{count(w,c)*count(w)}{N*count(c)}$

☒ $\log \frac{count(w,c)*count(c)}{N*count(w)}$

☐ $\log \frac{count(w,c)*N}{count(w)*count(c)}$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$\log \frac{count(w,c)*N}{count(w)*count(c)}$

4) 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

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) 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

Yes, the answer is correct.

☐ 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 ()

Score: 1

Accepted Answers:

It is faster to train

6) Which of the following is true about the input representation in the CBOW model? **1 point**

- ☒ Each word is represented as a one-hot vector
- ☐ Each word is represented as a continuous vector
- ☐ Each word is represented as a sequence of one-hot vectors
- ☐ Each word is represented as a sequence of continuous vectors

Yes, the answer is correct.

Score: 1

Accepted Answers:

Each word is represented as a one-hot vector

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

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

Yes, the answer is correct.

Score: 1

Accepted Answers:

$O(n)$

9) 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

Yes, the answer is correct.

Score: 1

Accepted Answers:

It uses a binary tree to approximate the softmax function

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