X

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

shweta_s@cs.iitr.ac.in

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

Click to register for Certification exam (https://examform.nptel.ac.in/2025_10/exam_form/dashboard)

If already registered, click to check your payment status

Course outline

- **About NPTEL ()**
- **How does an NPTEL online course work? ()**
- **Week 0 ()**
- **Week 1 ()**
- **Week 2 ()**
- **Week 3 ()**
- **week 4 ()**

Week 9 : Assignment 9

The due date for submitting this assignment has passed.

Due on 2025-03-26, 23:59 IST.

Assignment submitted on 2025-03-26, 09:31 IST

1) 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

No, the answer is incorrect.
Score: 0

Accepted Answers:
It is less accurate than computing the softmax function directly

2) Consider the following corpus: ``AI driven user experience optimization. Perception of AI decision making speed. Intelligent interface adaptation system. AI system engineering for enhanced processing efficiency". What is the size of the vocabulary of the above corpus? ***1 point***

- 18
- (●) 20
- 22
- 19

No, the answer is incorrect.
Score: 0

Accepted Answers:
19

3) We add incorrect pairs into our corpus to maximize the probability of words that occur in the same context and minimize the probability of words that occur in different contexts. ***1 point***

This technique is called:

- (•) Negative sampling
- ( ) Hierarchical softmax
- ( ) Contrastive estimation
- ( ) Glove representations

Yes, the answer is correct.
Score: 1

Accepted Answers:
Negative sampling

4) 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

5) 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 20,000 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) 2000

1 point

6) You are given the one hot representation of two words below:
GEMINI=[1, 0, 0, 0, 1], CLAUDE=[0, 0, 0, 1, 0]
What is the Euclidean distance between GEMINI and CLAUDE?

[1.732]

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Range) 1.7,1.74

1 point

7) Let $count(w, c)$ be the number of times the words w and c appear together in the 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: ***1 point***

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

Yes, the answer is correct.
Score: 1
Accepted Answers:
$\log \frac{count(w,c)*N}{count(w)*count(c)}$

8) 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'

9) 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 = 0.07) ***1 point***

- $X_{ij} = 0.04$
- $X_{ij} = 0.17$
- $X_{ij} = 0$
- $X_{ij} = 0.95$

Yes, the answer is correct.

Download Videos ()

Books ()

Text Transcripts ()

Problem Solving Session - Jan 2025 ()

Score: 1

Accepted Answers:
$X_{ij} = 0.95$

10) 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