

Assessment sub
X

(https://swayam.gov.in)



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

2021cs85ve@mitsgwl.ac.in ✓

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

Course
outlineAbout
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
representatio
ns of words
(unit?
unit=115&less
on=116)

Thank you for taking the Week 9 : Assignment 9.

Week 9 : Assignment 9

Your last recorded submission was on 2024-09-25, 22:15 Due date: 2024-09-25, 23:59 IST.

1) Let X be the co-occurrence matrix such that the (i, j) -th entry of X captures the **1 point** 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:

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

2) 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

Assessment submitted. X

- ☐ Distributed Representations of words (unit? unit=115&less on=117)
- ☐ SVD for learning word representations (unit? unit=115&less on=118)
- ☐ SVD for learning word representations (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 representations (unit? unit=115&less on=125)
- ☐ Evaluating word representations (unit?

- 3) 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
- 4) 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
- 5) 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
- 6) 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
- 7) 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. This technique is called- **1 point**
- ☐ Hierarchical softmax
 - ☐ Contrastive estimation
 - ☒ Negative sampling
 - ☐ Glove representations
- 8) What is the computational complexity of computing the softmax function in the output layer of a neural network? **1 point**
- ☒ $O(n)$
 - ☐ $O(n^2)$
 - ☐ $O(n \log n)$
 - ☐ $O(\log n)$

Assessment submitted.
X

☐ Relation between SVD and Word2Vec (unit? unit=115&less on=127)

☒ Lecture Material for Week 9 (unit? unit=115&less on=128)

☐ Week 9 Feedback Form: Deep Learning - IIT Ropar (unit? unit=115&less on=192)

☒ **Quiz: Week 9 : Assignment 9 (assessment? name=297)**

week 10 ()

Download Videos ()

Books ()

Text Transcripts ()

Problem Solving Session - July 2024 ()

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

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

You may submit any number of times before the due date. The final submission will be considered for grading.

Submit Answers