


<https://swayam.gov.in>

[https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL)

p22cs013@coed.svnit.ac.in ✓

 NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Natural Language Processing (course)

 Click to register  
for Certification  
exam

[https://examform.nptel.ac.in/2023-10/exam\\_form/dashboard](https://examform.nptel.ac.in/2023-10/exam_form/dashboard)

 If already  
registered, click  
to check your  
payment status

 Course  
outline

 How does an  
NPTEL online  
course work?  
( )

Week 0 ( )

Week 1 ( )

Week 2 ( )

Week 3 ( )

Week 4 ( )

Week 5 ( )

## Week 7 : Assignment

Your last recorded submission was on 2023-09-11, 17:02 IST Due date: 2023-09-13, 23:59 IST.

1)

1 point

Suppose you have a raw text corpus and you compute word co occurrence matrix from there. Which of the following algorithm(s) can you utilize to learn word representations? (Choose all that apply) [1 mark]

- a. CBOW
- b. SVD
- c. PCA
- d. GloVe

- ☒ a.  
☐ b.  
☐ c.  
☒ d.

2)

1 point

What is the method for solving word analogy questions like, given A, B and D, find C such that A:B::C:D, using word vectors? [1 mark]

- a.  $v_c = v_a + (v_b - v_d)$ , then use cosine similarity to find the closest word of  $v_c$ .
- b.  $v_c = v_a + (v_d - v_b)$  then do dictionary lookup for  $v_c$
- c.  $v_c = v_d + (v_b - v_a)$  then use cosine similarity to find the closest word of  $v_c$ .
- d.  $v_c = v_d + (v_a - v_b)$  then do dictionary lookup for  $v_c$ .
- e. None of the above

**Week 6 ()****Week 7 ()**

- Lecture 31 :  
Distributional  
Semantics -  
Introduction  
(unit?  
unit=66&lesson  
=67)

- Lecture 32 :  
Distributional  
Models of  
Semantics  
(unit?  
unit=66&lesson  
=68)

- Lecture 33 :  
Distributional  
Semantics :  
Applications,  
Structured  
Models (unit?  
unit=66&lesson  
=69)

- Lecture 34 :  
Word  
Embeddings -  
Part I (unit?  
unit=66&lesson  
=70)

- Lecture 35 :  
Word  
Embeddings -  
Part II (unit?  
unit=66&lesson  
=71)

- Week 7 -  
Lecture  
Materials (unit?  
unit=66&lesson  
=72)

- Quiz: Week 7 :  
Assignment  
(assessment?  
name=186)

- ☐ a.  
☐ b.  
☐ c.  
☒ d.  
☐ e.

3)

**1 point**

What is the value of  $PMI(w_1, w_2)$  for  $C(w_1) = 100$ ,  $C(w_2) = 2000$ ,  $C(w_1, w_2) = 64$ ,  $N = 100000$ ?  $N$ : Total number of documents.

$C(w_i)$ : Number of documents,  $w_i$  has appeared in.

$C(w_i, w_j)$ : Number of documents where both the words have appeared in. Use base 2 in logarithm.

Note:

**[1 mark]**

- a. 4  
b. 5  
c. 6  
d. 5.64

- ☐ a.  
☒ b.  
☐ c.  
☐ d.

4)

**2 points**

Given two binary word vectors  $w_1$  and  $w_2$  as follows:

$w_1 = [1010101010]$

$w_2 = [0011111100]$

Compute the Dice and Jaccard similarity between them. **[2 marks]**

- a. 6/11, 3/8  
b. 10/11, 5/6  
c. 4/9, 2/7  
d. 5/9, 5/8

- ☒ a.  
☐ b.  
☐ c.  
☐ d.

5)

**2 points**

● Feedback Form  
(unit?  
unit=66&lesson  
=187)

Week 8 ()

Download  
videos ()

Text  
Transcripts ()

Books ()

Consider two probability distributions for two words be  $p$  and  $q$ . Compute their similarity scores with KL-divergence. [2mark]

$p = [0.20, 0.75, 0.50]$

$q = [0.90, 0.10, 0.25]$

Note: Use base 2 in logarithm.

- a. 4.704, 1.720
- b. 1.692, 0.553
- c. 2.246, 1.412
- d. 3.213, 2.426

- ☐ a.
- ☐ b.
- ☒ c.
- ☐ d.

6)

2 points

Consider the following word co-occurrence matrix given below. Compute the cosine similarity between

(i)  $w_1$  and  $w_2$ , and (ii)  $w_1$  and  $w_3$ .

[2 mark]

	$w_4$	$w_5$	$w_6$
$w_1$	2	9	4
$w_2$	1	5	6
$w_3$	3	0	1

- a. 0.773, 0.412
- b. 0.881, 0.764
- c. 0.665, 0.601
- d. 0.897, 0.315

- ☐ a.
- ☐ b.
- ☐ c.
- ☒ d.

7)

1 point

Which of the following statement(s) is/are True? [1 mark]

- a. In structured distributional semantics, co-occurrence statistics are collected using parser extracted relations.
- b. Term mismatch occurs from the word independence assumption during document indexing.
- c. We can use distribution semantic models for query expansion.
- d. Attributional similarity depends on the degree of correspondence between attributes.

- ☒ a.

☒ b.☒ c.☐ d.

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

**Submit Answers**