

Unit 3 - Week 2

Course outline

How to access the portal?

Week 1

Week 2

- Vector Space Models for NLP
- Document Similarity - Demo, Inverted index, Exercise
- Vector Representation of words
- Contextual understanding of text
- Co-occurrence matrix, n-grams
- Collocations, Dense word Vectors
- SVD, Dimensionality reduction, Demo
- Query Processing
- Topic Modeling
- Week 2 Lecture Materials
- Quiz : Assignment 2
- Assignment 2 : Programming Exercise for Self Assessment
- Week 2 Feedback : Applied Natural Language Processing

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

Week 12

DOWNLOAD VIDEOS

Assignment 2

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-08-21, 23:59 IST.

1) The raw Co-occurrence matrices are usually sparse 1 point

- True
 False

No, the answer is incorrect.
Score: 0

Accepted Answers:
True

2) Is one hot-vector a dense vector? 1 point

- True
 False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

3) Check all that are correct - Word embedding is 1 point

- A row vector of the left singular matrix (obtained using the decomposition of a Term-Document matrix)
 A vector that represents a row of a term-document matrix and each element of the vector represents TF
 A vector that represents a row of a term-document matrix and each element of the vector represents TF*IDF
 A vector that represents a row of a term-document matrix and each element of the vector represents co-occurrence value of its neighbors
 A One hot vector
 A vector that contains a distributed representation of words

No, the answer is incorrect.
Score: 0

Accepted Answers:
A row vector of the left singular matrix (obtained using the decomposition of a Term-Document matrix)
A vector that contains a distributed representation of words

4) Rank the documents with respect to the query $q = [0.1 \ 0.001 \ 0.0 \ 0.2]^T$ using cosine similarity using cosine similarity 1 point

	D1	D2	D3
car	0.1	0.0	0.1
comprehensive	0.1	0.1	0.1
third-party	0.0	0.9	0.0
insurance	0.1	0.9	0.02

- [D1,D2,D3]
 [D2,D1,D3]
 [D3,D1,D2]
 [D3,D2,D1]

No, the answer is incorrect.
Score: 0

Accepted Answers:
[D1,D2,D3]

A scientist extracted a table from a research paper that listed three molecular fingerprints in the binary form as given below.

$$\begin{bmatrix} 1.[0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0] \\ 2.[0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1] \\ 3.[0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0] \end{bmatrix}$$

He wants to find the similarity of the molecules using some similarity measure. He found that he could use Tanimoto coefficient to find the similarity. The formula for Tanimoto coefficient is

$$T_c(A, B) = \frac{c}{a + b - c}$$

where c is the length of $A \cap B$, and a and b are sizes of A and B , respectively. Use this description to answer question 5-7

5) The high ranking document will have a very small coefficient. Is this statement true? 1 point

- Yes
 No

No, the answer is incorrect.
Score: 0

Accepted Answers:
No

6) Compute the ranks the fingerprints with respect to the first one. Use the data given in question 5 to compute the ranks. Is rank order given below correct? 1 point

Document	rank
3	1
2	2

- Yes
 No

No, the answer is incorrect.
Score: 0

Accepted Answers:
Yes

7) What would be the range of Tanimoto coefficient? 1 point

- [0,2]
 [0,1]
 [0,3]
 unbounded

No, the answer is incorrect.
Score: 0

Accepted Answers:
[0,1]

8) Consider the 4 documents given below to construct a term-document matrix. Use the query "speech systems" to compute the rank (descending order :) using cosine similarity. Note: (1) Use normalized TF using document length (2) Each element in the matrix should represent TF*IDF 1 point

1. information extraction systems
2. natural language processing
3. speech signal systems
4. speech processing

Select the correct ranking order

- (1,2,3,4)
 (1,3,4,2)
 (2,3,4,1)
 (3,4,1,2)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(3,4,1,2)

9) IS-A vector could be used for (Check all that are correct) 1 point

- Creating Inverted Index
 Information extraction
 NER
 Document ranking

No, the answer is incorrect.
Score: 0

Accepted Answers:
Information extraction
NER

10) What is $AA^T U$ if $A = U \sum V^T$? 1 point

- $U \sum U^T$
 $U \sum^2 U^2$
 $U \sum^2$
 $U \sum^2 U^2$

No, the answer is incorrect.
Score: 0

Accepted Answers:
 $U \sum^2$