

Started on	Thursday, March 29, 2018, 9:39 PM
State	Finished
Completed on	Thursday, March 29, 2018, 9:58 PM
Time taken	18 mins 51 secs
Grade	60.00 out of 60.00 (100%)

Question 1

Complete

4.00 points out of
4.00

Which of the following can be used to quantify the likelihood of two words occurring together in a document?

Select one:

- ☐ a. Collocation
- ☐ b. Bigram
- ☐ c. Correlation
- ☒ d. Pointwise Mutual Information

Question 2

Complete

4.00 points out of
4.00

Which of the following labels is NOT used as a tag for a word in the Part of Speech Tagging?

Select one:

- ☐ a. NOUN
- ☒ b. NNN
- ☐ c. ADJ
- ☐ d. DET

Question 3

Complete

4.00 points out of
4.00

Can taggers be linked in NLTK?

Select one:

- ☒ a. Yes
- ☐ b. No.

Question 4

Complete

4.00 points out of
4.00

Which of the following sentences best describes a corpus in the context of natural language processing?

Select one:

- ☒ a. It is a collection of documents.
- ☐ b. It classifies (or recognizes) chunks of text that refer to pre-defined categories.
- ☐ c. It refers to the grammatical properties of a word.
- ☐ d. It is a collection of tags.

Question 5

Complete

4.00 points out of
4.00

A PMI close to ____ implies two words are nearly independent and rarely occur together, while A PMI close to ____ implies two words almost always occur together.

Select one:

- ☐ a. 1; 0
- ☐ b. 0; 0
- ☒ c. 0; 1
- ☐ d. 1; 1

Question 6

Complete

4.00 points out of
4.00

The ____ library creates _____ models that can be leveraged to construct topic models.

Select one:

- ☒ a. gensim; vector-space
- ☐ b. nltk; sparse-distributed
- ☐ c. sklearn; vector-space
- ☐ d. numpy; vector-space

Question 7

Complete

4.00 points out of
4.00

Suppose we have preprocessed our corpus by mapping it into a dictionary and then storing it as bag of words in the variable `bagOfWords`, for the purposes of topic modeling. Following this, we import the `gensim` library:

```
import gensim
```

From the options below, how do we transform our text data into a TFIDF model?

Select one:

☐

a.

```
tfidf = models.TfidfModel(crps)
```

☐

b.

```
tfidf_model = gensim.corpora.TfidfModel(bagOfWords)
```

☐

c.

```
tfidf_model = models.TfidfModel(bagOfWords)
```

☒

d.

```
tfidf_model = gensim.models.TfidfModel(bagOfWords)
```

Question 8

Complete

4.00 points out of
4.00

The simplest topic model is latent Dirichlet allocation. It is a type of ____ model with ____ variables.

Select one:

☐

a. probabilistic; visible

☐

b. deterministic; hidden

☒

c. probabilistic; hidden

☐

d. deterministic; visible

Question 9

Complete

4.00 points out of
4.00

Which module in sci-kit contains the implementation of the non-negative matrix factorization?

Select one:

- ☐ a. sklearn.feature_extraction.NMF
- ☒ b. sklearn.decomposition.NMF
- ☐ c. sklearn.preprocessing.NMF
- ☐ d. sklearn.ensemble.NMF

Question 10

Complete

4.00 points out of
4.00

When the gensim library creates the vector space, which data structure is used to map the indices to the words?

Select one:

- ☐ a. set
- ☒ b. dictionary
- ☐ c. list
- ☐ d. tuple

Question 11

Complete

4.00 points out of
4.00

Which function of word2vec in gensim library computes cosine similarity between two sets of words?

Select one:

- ☐ a. most_similar
- ☒ b. n_similarity
- ☐ c. doesnt_match
- ☐ d. most_similar_cosmul

Question 12

Complete

4.00 points out of
4.00

In the context of Word2Vec, what algorithm can be used to predict a word given a context?

Select one:

- ☐ a. n-gram
- ☐ b. NER
- ☒ c. Skip gram
- ☐ d. Continuous bag of words

Question 13

Complete

4.00 points out of
4.00

Which of the following is NOT a method to improve the accuracy of a word2vec model?

Select one:

- ☐ a. Choose a better model architecture from skip-gram and CBOW.
- ☐ b. Increase the number of vector dimensions.
- ☐ c. Increase the window size of words considered by the algorithm.
- ☒ d. Increase computational complexity and model generation time.
- ☐ e. Increase the training data set.

Question 14

Complete

4.00 points out of
4.00

Which of the following sentences best describes a wordnet?

Select one:

- ☒ a. An English lexical database that groups words into synonym sets.
- ☐ b. It is a neural network model that was developed to provide an efficient continuous bag of words and skip-gram algorithms for word-vector representations.
- ☐ c. It has same definitions for a word, associated synonyms, lemmas, and other information.
- ☐ d. A group of related models that are used to produce word embeddings.

Question 15

Complete

4.00 points out of
4.00

In the context of Word2Vec, what algorithm can be used to predict a context given a word?

Select one:

- ☐ a. Skip gram
- ☒ b. Continuous bag of words
- ☐ c. n-gram
- ☐ d. NER