Quiz 5: Latent Dirichlet Allocation Results for Sachin Sudheer

Score for this quiz: 9 out of 10
Submitted Jun 24 at 7:59pm
This attempt took 13 minutes.

Question 1
1 / 1 pts
Latent Dirichlet Allocation assumes a certain:
Correct!

- (selected) Data generating process
- Subset of dictionary words
- Distribution of structures in the entire English language
- Subset of business topics within a major topic

Question 2
1 / 1 pts
As an output of topicmodels package, we can find the per-topic-per-word probabilities in a matrix called β ("beta").
Correct!

- (selected) True
- False

Question 3
1 / 1 pts
The only output of Latent Dirichlet Allocation algorithm are topics with their membership probabilities in each document.

- True

Correct!

- (selected) False

This is one of two outputs.

Question 4
1 / 1 pts
The Latent Dirichlet Allocation output of topics is such that each topic is clearly and appropriately labelled to explain the content words.

- True

Correct!

- (selected) False

Question 5
1 / 1 pts
The core idea of LDA is to find a generative probabilistic model with latent topics that can generate the data we observe in document-term matrix.
Correct!

- (selected) True
- False

Question 6
1 / 1 pts

In general, perplexity is used to compare models with different number of topics. The model with the lowest perplexity is generally considered the “best”.

Correct!

- (selected) True
- False

Question 7

1 / 1 pts

Latent Semantic Analysis discovers hidden semantic content whereas topic models focus on the underlying subjects or themes that are present in the documents.

Correct!

- (selected) True
- False

Question 8

1 / 1 pts

Latent Dirichlet Allocation uses Singular Value Decomposition to break down the Term Document Matrix into three smaller matrices for smaller dimensional representation.

- True

Correct!

- (selected) False

LSA uses SVD.

Question 9

0 / 1 pts

One can run the LDAtuning package without having run the topicmodels package first.

Correct Answer

- True

You Answered

- (selected) False

Question 10

1 / 1 pts

One approach to find appropriate number of topics in LDA is by using LDAtuning package

Correct!

- (selected) True
- False

Quiz Score: 9 out of 10

Quiz 4: Latent Semantic Analysis Results for Sachin Sudheer

Score for this quiz: 8 out of 10
Submitted Jun 17 at 7:48pm
This attempt took 11 minutes.

Question 1
0 / 1 pts
In the Vector Space Model, we compute distance between words using Euclidean Distance.
You Answered

- (●) True

Correct Answer

- ( ) False

Question 2
1 / 1 pts
Basically, cosine similarity is nothing but dot product divided by the norm (magnitude) of the vectors.
Correct!

- (●) True
- ( ) False

Question 3
1 / 1 pts
Two fundamental issues of NLP are synonymy and polygloty.

- ( ) True

Correct!

- (●) False

Question 4
1 / 1 pts
LSA attempts to find a lower dimensional approximation of Term Document Matrix by using a matrix factorization algorithm called Singular Value Decomposition (SVD)
Correct!

- (●) True
- ( ) False

Question 5
1 / 1 pts

Singular Value Decomposition factorizes the original matrix into the following matrices:

- ( ) A left singular matrix of words and right singular matrix of documents

Correct!

- (●) A left singular matrix of words, singular diagonal matrix and right singular matrix of documents
- ( ) None of these
- ( ) A left singular matrix of documents, singular diagonal matrix and right singular matrix of words

Question 6
1 / 1 pts

Singular value matrix helps us determine how to reduce size of the three component matrices by restricting the matrices U, Σ, V^T to their first k<n rows.

Correct!

- (●) True
- ( ) False

Question 7

1 / 1 pts

Look at the R code below for creating a latent semantic space:

```
library(lsa)
```

```
library(LSAfun)
```

```
lsa(my_dfm, dims=dimcalc_share())
```

What is the purpose of dimcalc_share argument?

- ( ) By specifying 'dims=dimcalc_share()', we specify the number of dimensions we do not want to truncate

Correct!

- (●) By specifying 'dims=dimcalc_share()', R chooses the number of dimensions to retain by default
- ( ) By specifying 'dims=dimcalc_share()', R allows the number of dimensions to vary indefinitely
- ( ) None of these

Question 8

1 / 1 pts

Term Co-occurrence Matrix is a square matrix where all the terms occur in rows as well as columns, so that when some terms appear together, they may define some context.

Correct!

- (●) True
- ( ) False

Question 9

1 / 1 pts

If we choose to retain k=3 singular vectors or three latent dimensions after our SVD computation, we are saying that the 3 values of the Sigma matrix capture the important dimensions.

Correct!

- (●) True
- ( ) False

Question 10

0 / 1 pts

If we multiply the three truncated matrices, $U_{t\times k}$, $\Sigma_{k\times k}$, $(V_{d\times k})^T$, we get back our original matrix.

You Answered

- (●) True

Correct Answer

- ( ) False

Quiz Score: 8 out of 10

Quiz 3: Sentiment Analysis Results for Sachin Sudheer

Correct answers are hidden.

Score for this quiz: 10 out of 10
Submitted Jun 10 at 10:35pm
This attempt took 10 minutes.

Question 1
1 / 1 pts
Sentiment Analysis is same as opinion mining.

- True
- False

Question 2
1 / 1 pts
Sentiment Analysis can only be done on document level and sentence level.

- True
- False

Question 3
1 / 1 pts
The outcome of sentiment analysis always includes information about the target, source and different types of attitudes.

- True
- False

Question 4
2 / 2 pts
Negation of superlative words is problematic because the outcome of such negation does not result in a word opposite in meaning.

- True
- False

Question 5
1 / 1 pts
One approach to dealing with negatives could be to append a negative between every negation and clause level punctuation.

- True
- False

Question 6
1 / 1 pts
Part of speech tagging can be very useful for finding out about product features.

- True
- False

Question 7
1 / 1 pts
The reason why we use Binarized Multinomial Naïve Bayes for classifying different sentiments is because word occurrence matters more than word frequency

- True

False

Question 8

2 / 2 pts

Using library sentimentr we can calculate text sentiment at the sentence level and can also aggregate by rows or grouping variable(s).

True

False

Quiz Score: 10 out of 10

Quiz 2: Webscraping and APIs Results for Sachin Sudheer

Score for this quiz: 10 out of 10
Submitted Jun 3 at 8:46pm
This attempt took 7 minutes.

Question 1
1 / 1 pts
HTML code is organized hierarchically, where the highest level there is a < $html$ > tag. The next level has < $head$ > and < $body$ > tags, and within the < $body$ > of the webpage there are various elements separated by different tags.
Correct!

- (selected) True
- False

Question 2
1 / 1 pts
One can get the xpath code for scraping an element from Selector Gadget extension.
Correct!

- (selected) True
- False

Question 3
2 / 2 pts
What is the primary benefit of an API?

- It provides a convenient data structure
- It provides a matrix representation of unstructured text data

Correct!

- (selected) It allows one program to interact with the features or elements of another program or service
- It is available for any website

Question 4
1 / 1 pts
To fetch data from an existing resource (website), we will use this command from library httr.

- POST
- PUT

Correct!

- (selected) GET
- FETCH

Question 5
2 / 2 pts
The text output from fetching data using HTTR is generally in JSON format. JSON stands for:

- JavaScript Onward Notation

Correct!

- JavaScript Object Notation
- Java Source Object Notation
- None of these

Question 6

1 / 1 pts

The general procedure for web scraping with package rvest is to first download the webpage into html file and then convert to text.

Correct!

- True
- False

Question 7

1 / 1 pts

Whether web scraping is legal or not, can be found by accessing donotscrape.txt file.

- True

Correct!

- False

Question 8

1 / 1 pts

Ralger is an R package to access the APIs of different websites.

- True

Correct!

- False

Quiz Score: 10 out of 10

Quiz 1: Text Preprocessing Results for Sachin Sudheer

ⓘ Correct answers are hidden.

Score for this quiz: 6 out of 10
Submitted May 27 at 5:19pm
This attempt took 6 minutes.

IncorrectQuestion 1
0 / 1 pts

Without any kind of pre-processing, please calculate the number of tokens in the statement below:

It was clean, good service and suite style rooms; however, it's a little older than some Residence Marriotts's I've stayed at.

- (●) 26
- ( ) 22
- ( ) 25
- ( ) 21

IncorrectQuestion 2
0 / 1 pts

In a document term matrix, we have documents in rows and tokens/words in columns.

- (●) True for Term Document Matrix
- ( ) None of these
- ( ) True for Tidy Text Data
- ( ) True

Question 3
1 / 1 pts

Which of the following is not the way to address the curse of dimensionality in text data?

- ( ) Stemming
- ( ) Removing punctuations
- ( ) Lemmatization
- (●) Adding n-Grams to the dataset

IncorrectQuestion 4
0 / 1 pts

Lemmatization uses vocabulary and morphological analysis to come up with the correct lemma. For instance, words, 'less', 'lesser', 'lessen', 'lessor' may be reduced to a common lemma "less".

- (●) Concept and example, both true
- ( ) Concept and example, both false
- ( ) Concept false, example true
- ( ) Concept true, example false

Question 5
1 / 1 pts

In the Bag-of-word models, the order of terms does not matter.

- None of these
- True
- False
- True only for TFIDF

Question 6
1 / 1 pts
In a wordcloud, the word size indicates the frequency of occurrence of the documents.

- False, word size indicates similarity of terms
- True, word size indicates frequency of documents
- False, word size indicates frequency of occurrence of words
- False, word size indicates importance of documents

IncorrectQuestion 7
0 / 1 pts
N grams significantly reduces the number of columns in the document feature matrix and consequently addresses the problem of dimensionality and sparsity.

- Depends on the document term matrix
- None of these
- True
- False

Question 8
1 / 1 pts

What is the purpose of stemming in text preprocessing?

- To check the spelling of words in the text data
- To convert all text to lowercase
- To increase the variety of words in the text data
- To reduce words to their root forms, thereby consolidating similar forms of a word

Question 9
1 / 1 pts
The purpose of TFIDF weighting is to address the problem of:

- length of document only
- high frequency of terms in document only
- length of document and high frequency of terms across documents
- None of these

Question 10
1 / 1 pts

Which of the following best describes the function of part-of-speech tagging in text preprocessing?

- To increase the frequency of certain important words in a text
- To remove punctuation and special characters from text
- To convert all text data into a single standard font

- [x] To identify and categorize each word in a text according to its grammatical role

Quiz Score: 6 out of 10