



Modeling text topics with Latent Dirichlet Allocation

12 questions

1
point

1.

Identify the top 3 most probable words for the first topic.

- ☐ institute
- ☒ university
- ☒ professor
- ☒ research
- ☐ studies
- ☐ game
- ☐ coach

1
point

2.

What is the sum of the probabilities assigned to the top 50 words in the 3rd topic? Round your answer to 3 decimal places.

0.210

1
point

3.

What is the topic most closely associated with the article about former US President George W. Bush? Use the average results from 100 topic predictions.

general politics

1
point

4.

What are the top 3 topics corresponding to the article about English football (soccer) player Steven Gerrard? Use the average results from 100 topic predictions.

- ☐ science and research
- ☐ music, TV, and film
- ☒ Great Britain and Australia
- ☒ international athletics
- ☒ team sports

1
point

5.

Using the LDA representation, compute the 5000 nearest neighbors for American baseball player Alex Rodriguez. For what value of k is Mariano Rivera the k -th nearest neighbor to Alex Rodriguez?

194

1
point

6.

Using the TF-IDF representation, compute the 5000 nearest neighbors for American baseball player Alex Rodriguez. For what value of k is Mariano Rivera the k -th nearest neighbor to Alex Rodriguez?

1
point

7. What was the value of α used to fit our original topic model?

1
point

8.

What was the value of γ used to fit our original topic model? Remember that GraphLab Create uses "beta" instead of "gamma" to refer to the hyperparameter that influences topic distributions over words.

1
point

9.

How many topics are assigned a weight greater than 0.3 or less than 0.05 for the article on Paul Krugman in the **low α** model? Use the average results from 100 topic predictions.

1
point

10.

How many topics are assigned a weight greater than 0.3 or less than 0.05 for the article on Paul Krugman in the **high alpha** model? Use the average results from 100 topic predictions.

1
point

11.

For each topic of the **low gamma model**, compute the number of words required to make a list with total probability 0.5. What is the average number of words required across all topics? (HINT: use the `get_topics()` function from GraphLab Create with the `cdf_cutoff` argument.)

1
point

12.

For each topic of the **high gamma model**, compute the number of words required to make a list with total probability 0.5. What is the average number of words required across all topics? (HINT: use the `get_topics()` function from GraphLab Create with the `cdf_cutoff` argument).



I, **Devendra Singh Sachan**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera's Honor Code

Submit Quiz

