Modeling text topics with Latent Dirichlet Allocation

12 试题

1 point

1.
Identify the top 3 most probable words for the first topic.

- [ ] institute
- [x] university
- [x] professor
- [x] 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.

- [x] Great Britain and Australia
- [ ] music, TV, and film
- [x] team sports
- [x] international athletics
- [ ] science and research

---

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?

116

---

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?

52

---

1 point

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

5.0

---

1 point

8.
What was the value of gamma 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.

0.1

---

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 alpha** model? Use the average results from 100 topic predictions.

8

---

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.

2

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.)

252.4

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).

576.2

沈伟臣

提交测试