Week 4 Quiz

10 试题

1 point

1。
What is NOT the motivation for text clustering?

- [x] To remove spam documents based on a small collection of human annotated spam documents
- [ ] To create structure of text data
- [ ] To link similar documents and remove duplicated documents
- [ ] To quickly get an idea about a large collection of documents

1 point

2。
What is TRUE about the mixture model and topic modeling?

- [ ] Only topic modeling can learn topics, while the mixture model does not yield such information after learning.
- [ ] Topic modeling can also be used for document clustering directly.
- [x] In topic modeling, the topic of each word is independently sampled, while in the mixture model, only one topic is drawn for each document.

1 point

3.
In the mixture model, if we want to encourage the formation of a large cluster:

- (●) Add prior to $P(\theta)$ so that the distribution is skewed
- ( ) Try different initialization
- ( ) Use a smaller number of clusters for training

---

1 point

4.
In the EM algorithm, which step improves the model likelihood?

- ( ) E-step
- (●) M-step

---

1 point

5.
True or false? In the EM algorithm, the model likelihood monotonically increases.

- ( ) False
- (●) True

---

1 point

6.

What is the most difficult part of directly applying maximal likelihood to PLSA?

- ○ The objective function needs to sum over all words for each document.
- ○ The objective function needs to sum over all documents in the collection.
- ● The objective function needs to sum over all topics for each word.

---

1 point

7。
For the agglomerative clustering algorithm, which of the following is not TRUE?

- ○ The user needs to specify a similarity measurement.
- ○ It's a bottom-up algorithm to form a hierarchy.
- ● The depth of the hierarchy is always $log_2(N)$ where N is the number of items.

---

1 point

8。
Which evaluation method is best for clustering results of a large collection of documents?

- ○ Use the direct evaluation method and create human annotations for each document in the collection.
- ● Use the indirect evaluation method and test performance for an application with or without clustering.

---

1 point

9。
Which of the following is NOT sensitive to outliers?

- Complete-link
- Single-link
- Average-link

1 point

10。
Which of the following is a generative classification algorithm?

- Naive Bayes
- K-NN
- Logistic Regression
- SVM

我（**伟臣 沈**）了解提交不是我自己完成的作业 将永远不会通过此课程或导致我的 Coursera 帐号被关闭。
了解荣誉准则的更多信息

提交测试