# Week 4 Quiz

Passed

**9/10** points earned (90%)

Quiz passed!

Incorrect

0/1 points

1. What is NOT the motivation for text clustering?

1. To remove spam documents based on a small collection of human annotated spam documents
2. To link similar documents and remove duplicated documents
3. To create structure of text data
4. To quickly get an idea about a large collection of documents

Correct  
1 / 1 points

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

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

Correct  
1 / 1 points

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

1. Use a smaller number of clusters for training
2. Try different initialization
3. **Add prior to *P*(*θ*) so that the distribution is skewed**

Correct  
1 / 1 points

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

1. **M-step**
2. E-step

Correct  
1 / 1 points

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

1. **True**
2. False

Correct  
1 / 1 points

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

1. The objective function needs to sum over all documents in the collection.
2. The objective function needs to sum over all words for each document.
3. **The objective function needs to sum over all topics for each word.**

Correct  
1 / 1 points

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

1. **The depth of the hierarchy is always *log*2(*N*) where N is the number of items.**
2. The user needs to specify a similarity measurement.
3. It's a bottom-up algorithm to form a hierarchy.

Correct  
1 / 1 points

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

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

Correct  
1 / 1 points

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

1. **Single-link**
2. Complete-link
3. Average-link

Correct  
1 / 1 points

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

1. K-NN
2. **Naive Bayes**
3. Logistic Regression
4. SVM