


Feedback — Week 4 Practice Quiz

[Help Center](#)

You submitted this quiz on **Sat 4 Jul 2015 1:29 AM PDT**. You got a score of **10.00** out of **10.00**.


Question 1

Which one of the following statements is **not** an opinion?

Your Answer	Score	Explanation
<input type="radio"/> PLSA always performs similarly to LDA.		
<input type="radio"/> PLSA is the best method for a topic mining task.		
<input checked="" type="radio"/> PLSA is a mixture model.	 1.00	
Total	1.00 / 1.00	

Question 2

Word unigrams are the best performing features for sentiment classification.

Your Answer	Score	Explanation
<input type="radio"/> True		
<input checked="" type="radio"/> False	 1.00	
Total	1.00 / 1.00	

Question Explanation

The best feature will depend at the task and dataset at hand. In general, word bigrams tend to perform better than unigrams for sentiment analysis.

Question 3

Suppose we are using logistic regression for binary classification (i.e., $k = 2$) where the number

of features is M . Then, the number of parameters to be estimated is $M+1$.

Your Answer	Score	Explanation
<input checked="" type="radio"/> True	✓ 1.00	
<input type="radio"/> False		
Total	1.00 / 1.00	

Question Explanation

In addition to the M weights associated with the M features, an additional threshold weight has to be estimated.

Question 4

Which of the following is not true about Latent Aspect Rating Analysis (LARA)?

Your Answer	Score	Explanation
<input type="radio"/> LARA is composed of two stages: Aspect Segmentation and Latent Rating Regression.		
<input checked="" type="radio"/> LARA assumes that all latent aspects have equal weights, and the overall rating is the average of the aspect rating.	✓ 1.00	
<input type="radio"/> LARA is a generative model for inferring ratings of latent aspects.		
Total	1.00 / 1.00	

Question Explanation

The latent aspect weights are not necessarily equal; they are inferred using maximum likelihood.

Question 5

NetPLSA leverages the power of both the text and the network structure to mine topics.

Your Answer	Score	Explanation
<input type="radio"/> False		

☒ True  1.00

Total 1.00 / 1.00

Question 6

NetPLSA tries to smooth the topic transitions by forcing neighbor nodes in the network to have different topic coverage.

Your Answer	Score	Explanation
-------------	-------	-------------

☐ True

☒ False  1.00

Total 1.00 / 1.00

Question Explanation

NetPLSA has an additional term in its objective function which penalizes cases where neighbor nodes are assigned different topic coverage.

Question 7

Contextual Probabilistic Latent Semantic Analysis (CPLSA) can be applied to which of the following tasks?

Your Answer	Score	Explanation
-------------	-------	-------------

☐ Discovering temporal trends of topics in text

☐ Revealing how the coverage of topics in different locations evolves over time

☒ All of the above  1.00

Total 1.00 / 1.00

Question Explanation

Both time and location can be treated as context when performing topic modeling, which makes CPLSA an appropriate choice.

Question 8

Suppose we are interested in discovering topics whose coverage in Twitter has strong correlations with airline prices. Which method would be best suited for this task?

Your Answer	Score	Explanation
<input type="radio"/> LDA		
<input type="radio"/> PLSA		
<input checked="" type="radio"/> Iterative Topic Modeling with Time Series Feedback	✓ 1.00	
<input type="radio"/> Contextual PLSA (CPLSA)		
Total	1.00 / 1.00	

Question Explanation

Airline prices can be viewed as a time series, so Iterative Topic Modeling with Time Series Feedback is a suitable choice in this case.

Question 9

Assume we are using word n-grams as features to perform sentiment classification. Then, higher values of n will usually be **less** prone to overfitting (i.e., for higher values of n, the difference between training and testing accuracies will be smaller).


Your Answer	Score	Explanation
<input checked="" type="radio"/> False	✓ 1.00	
<input type="radio"/> True		
Total	1.00 / 1.00	

Question Explanation

Large values of n will usually allow the classifier to achieve a high accuracy on the training data, while that on the testing data will be considerably lower.

Question 10

Deep learning is a new topic emerging in machine learning. Suppose we are interested in knowing whether US researchers and those outside the US have different focuses when working on this topic. For this purpose, we can collect research publications with metadata such as the author names, their affiliations, and locations. Which of the following text mining techniques is most suitable for this task?

Your Answer	Score	Explanation
<input checked="" type="radio"/> Contextual PLSA (CPLSA)	 1.00	
<input type="radio"/> Iterative Topic Modeling with Time Series Supervision		
<input type="radio"/> Text clustering		
Total	1.00 / 1.00	

Question Explanation

Location can be viewed as the context.