**Congratulations! You passed!**

**TO PASS** 70% or higher

Keep Learning

**GRADE**
80%

# Week 3 Practice Quiz

**TOTAL POINTS 10**

1. You are given two unigram language models, θ1 and θ2, as defined in the table below: **1 / 1 point**

| w | P(w\|θ1) | P(w\|θ2) |
|---|---|---|
| the | 0.4 | 0.05 |
| of | 0.4 | 0.05 |
| technology | 0.1 | 0.5 |
| machine | 0.1 | 0.4 |

Suppose we are using a mixture model for document clustering based on the two given unigram language models, θ1 and θ2, such that P(θ1)=0.3 and P(θ2)=0.7. To generate a document, first, one of the two language models is chosen according to P(θi), and then all the words in the document are generated based on the chosen language model.

The probability of generating a document composed only of the one word “technology” using the given mixture model is P(“technology”)=

- 0.58
- 0.38
- 0.3
- 0.7

✓ **Correct**

0.1 * 0.3 + 0.5 * 0.7

2. Assume the same given as in Question 1. What is the probability of generating a document composed only of the phrase “the technology”, i.e., P(“the technology”)? **1 / 1 point**

- 0.0589
- 0.0295
- 0.3
- 0.1444

**Correct**

0.4 * 0.1 * 0.3 + 0.05 * 0.5 * 0.7

3. In mixture model, why do different components tend to assign high probability on different words? **1 / 1 point**

- Because during training, when different components assign high probability to the same model, the training restarts
- Because it gives a higher overall likelihood
- Because the model was initialized with components with high probability assigned to different words

**Correct**

If they all assign the high probability to the same words, then it generally indicates that the chosen number of components is too many.

4. Why it is good to have the "background" component? Check all that apply. **0 / 1 point**

- [x] To prevent overfitting

**Correct**

By applying human knowledge that words in background components does not form any useful topic, adding such background components improves the robustness of the model

- [x] To improve model likehood

**This should not be selected**

- [x] To better filter topic words into other components

**Correct**

As all non-topic words such as stop words goes into the background component

5. What type of words are usually assigned with high probability in the background component? **1 / 1 point**

- (•) "the", "he", "she", "is"
- ( ) "computer", "information", "data"
- ( ) "car", "cat", "catch"

**Correct**

those are words with very high frequency (stop words)

6. Which of the following about the EM algorithm is false? **1 / 1 point**

- (•) The result of the EM algorithm does not depend on the initialization.
- ( ) It always increase the likelihood.
- ( ) It is generally considered a fast algorithm for optimizing likelihood.
- ( ) It can be trapped into a local optimal solution.

**Correct**

EM algorithm has different results for different initialization

7. In EM, what does the E-step do? **1 / 1 point**

- ○ Given the predicted values of unseen data, maximizes the joint likelihood
- ◉ Predicts values of unseen (hidden) variables

✓ **Correct**

8. Which of the following generative descriptions is not TRUE about PLSA? **1 / 1 point**

- ○ To generate a word, a topic is drawn from the document's topic weight distribution, and a word is drawn according to the topic's word distribution.
- ○ To generate a document, a distribution of topic weights (multinomial distribution) is assumed, which is considered part of the model.
- ◉ To generate a topic assignment for a word, a coin is tossed to decide if the topic is from the background topic or not, and the probability of the background is a constant specified by the user.

✓ **Correct**
the probability of background component weight is learnt not fixed.

9. In PLSA, which of the following variables are part of the model? Check all that apply. **0 / 1 point**

- ☑ $\lambda$, background percentage

✓ **Correct**

- ☑ $z$, topic assignment

! **This should not be selected**

- ☑ $\theta$, topic's word distribution

**Correct**

10. True or false? Let $\theta_1,...,\theta_k$ be the k unigram language model's output by PLSA. Then, for a specific word w, the following relation always holds: $\sum_{i=1}^{k} P(w|\theta_i) = 1.$ **1 / 1 point**

- (●) False
- ( ) True

**Correct**

$\sum_w P(w|\theta_i) = 1$