Quiz: Week 3 Quiz

TOTAL POINTS 10

1. You are given two unigram language models θ_1 and θ_2 as defined in the table below: 1 point

w	P(w|θ_1)	P(w|θ_2)
concert	0.1	0.4
music	0.1	0.4
data	0.4	0.1
software	0.4	0.1

Suppose we are using a mixture model for document clustering based on the two given unigram language models, θ_1 and θ_2, such that P(θ_1)=0.5 and P(θ_2)=0.5. To generate a document, first, one of the two language models is chosen according to P(θ_i), and then **all** the words in the document are generated based on the chosen language model. The probability of generating the document d: “music software” using the given mixture model is P(“music software”)=

- 0.05
- 0.6
- 0.04 (selected)
- 0.5

2. Assume the same unigram language models, θ1 and θ2, defined as in the table of Question 1 with P(θ1)=0.5 and P(θ2)=0.5. We now want to generate documents based on the mixture model used in topic modeling. To generate a document **for each word**, we first choose one of the two language models, θ1 and θ2, and then generate the word according to the chosen model. The probability of generating the document d: “music software” according to this mixture model is P(“music software”)= 1 point

- 0.125
- 0.0625 (selected)
- 0.0125
- 0.625

3. Let X_w be a random variable denoting whether word w occurs in a text document in a collection of English news articles. Which random variable do you expect to have a **lower**entropy? 1 point

- (●) $H(X_{the})$
- ( ) $H(X_{learning})$

4. We want to run PLSA on a collection of N documents with a fixed number of topics k where the vocabulary size is M. What is the number of parameters that PLSA tries to estimate? Consider each P(w|θj) or πd,j as a separate parameter. 1 point

- (●) Mk+Nk
- ( ) Nk
- ( ) Mk
- ( ) MNk

5. You are given a document d that contains only two words: “the” and “machine”. Assume that this document was generated from a mixture of two unigram language models: a known background language model θ_B and an unknown topic language model θ_d. Let $P(\theta_B) = \lambda$ and $P(\theta_d) = 1 - \lambda$ and assume that $P(\text{“the”}|\theta_B) = 0.9$ and $P(\text{“machine”}|\theta_B) = 0.1$. We want to estimate θ_d using maximum likelihood. Then, as λ increases, $P(\text{“machine”}|\theta_d)$ will: 1 point

Hint: First get the maximum likelihood estimates of the two words in θ_d (refer to the lecture on "Probabilistic Topic Models: Mixture Model Estimation"). Then, write $P(\text{“machine”}|\theta_d)$ as a function of λ and study the behavior of the function.

- ( ) Remain the same
- ( ) Decrease
- (●) Increase

6. True or false? In general, PLSA using the EM algorithm does not stop until it achieves the global maximum of the likelihood function. 1 point

- (●) False

- True

7. True or false? Let $\theta_1,...,\theta_k$ be the k unigram language model's output by PLSA and V be the vocabulary set. Then, for any i∈{1,...,k}, the following relation always holds: ∑w∈VP(w|θ_i)=1. 1 point

- False
- True

8. True or false? The EM algorithm **cannot** decrease the likelihood of the data. 1 point

- True
- False

9. True or false? Assume that the likelihood function of PLSA has multiple local maxima and one global maximum. There exists an initial set of parameters for which PLSA will converge to the global maximum of the likelihood function. 1 point

- False
- True

10. True or false? When using PLSA to mine topics from a text collection, the number of parameters of the PLSA model stays the same as we keep adding new documents into the text collection assuming that the new documents do not introduce new words that have not occurred in the current text collection. 1 point

- True
- False

I, **BAL KRISHNA NYAUPANE**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code