# Week 2 Practice Quiz

**10/10** points earned (100%)

Excellent!

Retake

[Course Home](https://www.coursera.org/learn/text-mining/home/welcome)

Correct

1 / 1 points

1. You are given a unigram language model *θ* distributed over a vocabulary set *V* composed of **only** 4 words: “the”, “machine”, “learning”, and “data”. The distribution of *θ* is given in the table below:

|  |  |
| --- | --- |
| *w* | *P*(*w*|*θ*) |
| machine | 0.1 |
| learning | 0.2 |
| data | 0.3 |
| the | 0.4 |

*P*(“machine learning”|*θ*)=

1. **0.02**

**Correct Response**

Since in a unigram language model words are assumed to be generated independently, we have*P*(“machine learning”|*θ*)=*P*(“machine”|*θ*)*P*(“learning”|*θ*)=0.1∗0.2=0.02

1. 0.004
2. 0.3
3. 0.2

Correct

1 / 1 points

2. Assume the same unigram language model as in Question 1. Then, *P*(“learning machine”|*θ*)=

1. 0.3
2. 0.2
3. **0.02**

**Correct Response**

Due to the independence assumption, the word order does not matter when generating text based on a unigram language model. Thus, the answer is the same as that of Question 1.

1. 0.004

Correct

1 / 1 points

3. Assume the same unigram language model as in Question 1. Then, *P*(“learning machine learning”|*θ*)=

1. 0.2
2. **0.004**

**Correct Response**

0.2 \* 0.1 \* 0.2

1. 0.3
2. 0.02

Correct

1 / 1 points

4. Assume that words are being generated by a mixture of two unigram language models, *θ*1 and *θ*2, where *P*(*θ*1)=0.5 and *P*(*θ*2)=0.5. The distributions of the two models are given in the table below:

|  |  |  |
| --- | --- | --- |
| *w* | *P*(*w*|*θ*1) | *P*(*w*|*θ*2) |
| the | 0.4 | 0.15 |
| and | 0.4 | 0.15 |
| genes | 0.05 | 0.3 |
| biology | 0.15 | 0.4 |

Then *P*(“biology”)=

1. **0.275**

**Correct Response**

*P*(“biology”)=*P*(“biology”|*θ*1)*P*(*θ*1)+*P*(“biology”|*θ*2)*P*(*θ*2)=0.15∗0.5+0.4∗0.5=0.275

1. 0.55
2. 0.15
3. 0.175

Correct

1 / 1 points

5. Assume the same given as in Question 4. Then *P*(“the biology”)=

1. 0.275625
2. 1
3. 0.275
4. **0.075625**

**Correct Response**

*P*(“the biology”)=(*P*(“the”|*θ*1)*P*(*θ*1)+*P*(“the”|*θ*2)*P*(*θ*2))(*P*(“biology”|*θ*1)*P*(*θ*1)+*P*(“biology”|*θ*2)*P*(*θ*2))=0.075625

Correct

1 / 1 points

6. Suppose we have the following word counts for two documents *d*1 and *d*2.

Table 1: Counts for words in document set

|  |  |  |  |
| --- | --- | --- | --- |
| **Vocabulary Words** | *c*(*w*,*d*1) | *c*(*w*,*d*2) | *P*(*w*|*θB*) |
| text | 5 | 0 | 0.15 |
| mining | 4 | 0 | 0.05 |
| the | 4 | 4 | 0.50 |
| fifa | 0 | 4 | 0.10 |
| football | 0 | 3 | 0.20 |

We are interested in applying topic modeling to discover two topics, *θ*0 and *θ*1, in our corpus of two documents. Suppose that we run PLSA with the number of topics set to 2 (i.e. *k*=2) while using an additional known (fixed) background word distribution *θB* as shown in Table 1. Using the EM algorithm, and after *n* iterations, the E-step gives the following estimates:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Documents** | **Words** | *P*(*zw*,*d*=0) | *P*(*zw*,*d*=1) | *P*(*zw*, *d*=*B*) |
| *d*1 | text | 1.00 | 0.00 | 0.20 |
|  | mining | 1.00 | 0.00 | 0.10 |
|  | the | 0.60 | 0.40 | 0.90 |
| *d*2 | the | 0.40 | 0.60 | 0.90 |
|  | fifa | 0.00 | 1.00 | 0.20 |
|  | football | 0.00 | 1.00 | 0.20 |

Assume *λB*=*P*(*θB*)=0.20 and recall that *P*(*zw*,*d*=0)+*P*(*zw*,*d*=1)=1 as discussed in the lectures. After completing the M-step, *P*(football|*θ*1) =

1. 0.2
2. 0.3
3. **0.4**

**Correct Response**

We have that:

*P*(text|*θ*1)∝5(1−0.20)(0.00)=0.00

*P*(mining|*θ*1)∝4(1−0.10)(0.00)=0.00

*P*(the|*θ*1)∝4(1−0.90)(0.60)+4(1−0.90)(0.40)=0.40

*P*(fifa|*θ*1)∝4(1−0.20)(1.00)=3.20

*P*(football|*θ*1)∝3(1−0.20)(1.00)=2.40

Thus our normalizing term is 0.40+3.20+2.30=6.0, then

*P*(football|*θ*0)=2.40/6.0=0.40

1. 0.1

Correct

1 / 1 points

7. Assume the same given as in Question 6 and recall that *πd*2,0+*πd*2,1=1. Then, *πd*2,1, rounded to the nearest hundredth, is equal to:

1. 1.00
2. 0.98
3. **0.97**

**Correct Response**

Computing the distribution for document 2 we have

*πd*2,1∝3(1−0.20)(1.00)+4(1−0.20)(1.00)+4(1−0.90)(0.60)=5.84

*πd*2,0∝3(1−0.20)(0.00)+4(1−0.20)(0.00)+4(1−0.90)(0.40)=0.16

Setting the normalization term as 5.84+0.16=6.00, we can then compute the topic probabilities of document 2 as follows:

*πd*2,1=5.84/6.00≈0.97

*πd*2,0=0.16/6.00≈0.03

1. 0.99

Correct

1 / 1 points

8. True or false? A random variable X with P(X=1)=1 achieves the minimum possible entropy.

1. False
2. **True**

**Correct Response**

This is a deterministic variable with H(X) = 0, which is the lowest possible value for entropy.

Correct

1 / 1 points

9. True or false? The outcome of an unbiased coin is easier to predict than the outcome of a biased coin.

1. True
2. **False**

**Correct Response**

It's harder to predict. An extreme case is for a coin with only head, after observing it's first few outcome, one can easily determine the next outcomes given it is known that the coin has only one face.

Correct

1 / 1 points

10. Which of the following is not true?

1. **If H(X) = H(Y), then X and Y follow the same distribution**

**Correct Response**

Counterexample for "If H(X) = H(Y) then X and Y follow the same distribution": Let P(X=1) = 0.9 and P(Y=1) = 0.1. Clearly, H(X) = H(Y), however, the distributions of the two random variables are different. "If H(X|Y) = H(Y|X) then H(X) = H(Y)" is always true because H(X)=I(X;Y)+H(X|Y) and H(Y)=I(X;Y)+H(Y|X).

1. If H(X|Y) = H(Y|X), then H(X) = H(Y)
2. I(X;Y) = I(Y;X)