Feedback — Week 3 Quiz

You submitted this quiz on **Sat 18 Apr 2015 2:49 AM PDT**. You got a score of **10.00** out of **10.00**.

Question 1

Assume you are using a unigram language model to calculate the probabilities of phrases. Then, the probabilities of generating the phrases "study text mining" and "text mining study" are **not** equal, i.e., P("study text mining") $\neq$ P("text mining study").

Your Answer		Score	Explanation
○ True			
◉ False	✔	1.00	
Total		1.00 / 1.00	

Question 2

You are given a vocabulary composed of only four words: "the", "computer", "science", and "technology". Below are the probabilities of three of these four words given by a unigram language model.

Word	Probability
the	0.4
computer	0.2
science	0.3

What is the probability of generating the phrase "the technology" using this unigram language model?

Your Answer		Score	Explanation
◉ 0.04	✔	1.00	

○ 0.1

○ 0.0024

○ 0.5

| Total | 1.00 / 1.00 |

Question 3

You are given the query Q= "online courses" and two documents:

D1 = "online courses search engine"

D2 = "online education is affordable"

Assume you are using the maximum likelihood estimator **without** smoothing to calculate the probabilities of words in documents (i.e., the estimated $p(w|D)$ is the relative frequency of word w in the document D). Based on the unigram query likelihood model, which of the following choices is correct?

Your Answer		Score	Explanation
◉ P(Q|D1) = 1/16 P(Q|D2) = 0		✔ 1.00	
○ P(Q|D1) = 0 P(Q|D2) = 1/4			
○ P(Q|D1) = 1/16 P(Q|D2) = 1/4			
○ P(Q|D1) = 1/2 P(Q|D2) = 1/2			
Total		1.00 / 1.00	

Question 4

Assume the same scenario as in Question 3, but using linear interpolation (Jelinek-Mercer) smoothing with $\lambda = 0.5$. Furthermore, you are given the following probabilities of **some** of the words in the collection language model:

Word	P(w|C)
online	1/4
courses	1/4

education	1/8

Based on the unigram query likelihood model, which of the following choices is correct?

Your Answer		Score	Explanation
○ P(Q|D1) = 1/16	P(Q|D2) = 1/16		
○ P(Q|D1) = 1/32	P(Q|D2) = 1/32		
○ P(Q|D1) = 1/16	P(Q|D2) = 0		
◉ P(Q|D1) = 1/16	P(Q|D2) = 1/32	✔ 1.00	
Total		1.00 / 1.00	

Question 5

The BM25 has more free parameters to tune than the ranking function of the Dirichlet Prior smoothing.

Your Answer	Score	Explanation
◉ True	✔ 1.00	
○ False		
Total	1.00 / 1.00	

Question 6

Assume you are using Dirichlet Prior smoothing to estimate the probabilities of words in a certain document. What happens to the smoothed probability of the word when the parameter μ is **increased**?

Your Answer	Score	Explanation
◉ It becomes closer to the probability of the word in the collection language model	✔ 1.00	
○ It becomes closer to the maximum likelihood estimate of the probability derived from the document		

○ It does not change

○ It tends to 1

Total	1.00 / 1.00

Question 7

It is possible that pseudo feedback decreases the precision and recall of a certain retrieval system.

Your Answer		Score	Explanation
● True	✔	1.00	
○ False			
Total		1.00 / 1.00	

Question 8

Refer to the Rocchio feedback formula in the slides. If you want to eliminate the effect of **non-relevant** documents when doing feedback, which of the following parameters must be set to zero?

Your Answer		Score	Explanation
● γ	✔	1.00	
○ γ and β			
○ β			
○ α			
Total		1.00 / 1.00	

Question 9

Let q be the original query vector, $D_R = \{P_1, \ldots, P_n\}$ be the set of positive document vectors, and $D_N = \{N_1, \ldots, N_m\}$ be the set of negative document vectors. Let q_1 be the expanded query vector after applying Rocchio on D_R and D_N with positive parameter values α, β, and γ. Let q_2 be the expanded query vector after applying Rocchio on D_R and D_N with the same values for α, β, but γ being set to zero. Which of the following is correct?

Your Answer	Score	Explanation
○ q_1 has strictly greater weights than q_2		
○ q_2 has strictly greater weights than q_1		
○ q_1 can have greater or equal weights to q_2		
◉ q_2 can have greater or equal weights to q_1	✔ 1.00	
Total	1.00 / 1.00	

Question 10

Which of the following is **not** true about the KL-divergence retrieval model?

Your Answer	Score	Explanation
◉ It cannot be computed as efficiently as the query likelihood model.	✔ 1.00	
○ It represents both queries and documents as language models.		
○ It supports relevance feedback.		
Total	1.00 / 1.00	