

Feedback — Week 3 Practice Quiz

[Help Center](#)

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Sun 5 Apr 2015 6:08 AM PDT**. You got a score of **8.00** out of **8.00**.

Question 1

You are given a vocabulary composed of only three words: “text”, “mining”, and “research”.

Below are the probabilities of two of these three words given by a unigram language model:

Word	Probability
text	0.4
mining	0.2

What is the probability of generating the phrase “text mining research” using this unigram language model?

Your Answer	Score	Explanation
<input type="radio"/> 0.4		
<input type="radio"/> 0.08		
<input type="radio"/> 0		
<input checked="" type="radio"/> 0.032	✓ 1.00	
Total	1.00 / 1.00	

Question Explanation

The probability of “research” is $P(\text{“research”}) = 1 - (P(\text{“text”}) + P(\text{“mining”})) = 1 - (0.4 + 0.2) = 0.4$. The probability of generating the given phrase $P(\text{“text mining research”}) = P(\text{“text”}) \times P(\text{“mining”}) \times P(\text{“research”}) = 0.4 \times 0.2 \times 0.4 = 0.032$.

Question 2

You are given the query $Q = \text{"food safety"}$ and two documents:

$D1 = \text{"food quality regulations"}$

$D2 = \text{"food safety measures"}$

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
<input checked="" type="radio"/> $P(Q D1) = 0$ $P(Q D2) = 1/9$	✓ 1.00	
<input type="radio"/> $P(Q D1) = 1/3$ $P(Q D2) = 1/9$		
<input type="radio"/> $P(Q D1) = 1/3$ $P(Q D2) = 0$		
<input type="radio"/> $P(Q D1) = 1/2$ $P(Q D2) = 1/2$		
Total	1.00 / 1.00	

Question Explanation

$P(Q|D1) = P(\text{"food"}|D1) \times P(\text{"safety"}|D1) = 1/3 \times 0 = 0$. $P(Q|D2) = P(\text{"food"}|D2) \times P(\text{"safety"}|D2) = 1/3 \times 1/3 = 1/9$.

Question 3

Probability smoothing avoids assigning zero probabilities to unseen words in documents.

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

Question 4

Assume you are given two scoring functions:

$$S_1(Q, D) = P(Q|D)$$

$$S_2(Q, D) = \log P(Q|D)$$

For the same query and corpus, S_1 and S_2 will give the same ranked list of documents.

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

Question Explanation

log is a monotonically increasing function so it will preserve the ranking of documents, but not the scores.

Question 5

Assume you are using linear interpolation (Jelinek-Mercer) smoothing to estimate the probabilities of words in a certain document. What happens to the smoothed probability of the word when the parameter λ is **decreased**?

Your Answer	Score	Explanation
<input type="radio"/> It does not change		
<input type="radio"/> It becomes closer to the probability of the word in the collection language model		
<input checked="" type="radio"/> It becomes closer to the maximum likelihood estimate of the probability derived from the document.	1.00	
Total	1.00 / 1.00	

Question Explanation

The smoothed probability can be thought of as a weighted average of the maximum likelihood estimate and the probability of the word in the collection. When λ decreases, the weight assigned to maximum likelihood estimate increases, thus having a higher effect on the smoothed probability value.

Question 6

Refer to the Rocchio feedback formula in the slides. If you want to reduce the effect of the **relevant** documents in the updated query, which of the following should be done?

Your Answer	Score	Explanation
<input type="radio"/> Increase γ		
<input type="radio"/> Increase β		
<input checked="" type="radio"/> Reduce β	✓ 1.00	
<input type="radio"/> Reduce γ		
Total	1.00 / 1.00	

Question Explanation

The weight assigned to the centroid of the relevant documents is directly proportional to β .

Question 7

Assume that $\beta = 1$ is a good choice when performing relevance feedback using Rocchio's method. What is a reasonable value of β to use when relying on pseudo feedback?

Your Answer	Score	Explanation
<input type="radio"/> More than 1		
<input checked="" type="radio"/> Less than 1	✓ 1.00	
<input type="radio"/> 1		
Total	1.00 / 1.00	

Question Explanation

When doing relevance feedback, the judgments are usually reliable since human assessors generate them after reading the queries and documents. However, in pseudo feedback, the top k documents retrieved by the system are “blindly” assumed to be relevant, which makes the judgments less reliable compared to relevance feedback. The reasonable choice is to lower the parameter β , which can be thought of as the degree of “confidence” in the documents being used as “positive” examples in feedback.

Question 8

Let q be the original query vector, $D_R = \{P_1, \dots, P_n\}$ be the set of positive document vectors, and $D_N = \{N_1, \dots, N_m\}$ 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.

In which updated query do you expect stopwords to have higher weights?

Your Answer		Score	Explanation
<input checked="" type="radio"/> q_2	✓	1.00	
<input type="radio"/> q_1			
Total		1.00 / 1.00	