

START OF QUIZ
Student ID: 12582557,li
Chen,Mariana

Question 1

Topic: Lecture 8

Source: Lecture 8

In class, I mentioned that TF-IDF unfairly punishes words that appear in a lot, but not all, of the documents in our corpus. Explain how Okapi BM25 attempts to fix this. (1)

Question 2

Topic: Lecture 7

Source: Lecture 7

From a processing perspective, what is one benefit structured data has over unstructured data, and vice versa. (1)

Question 3

Topic: Lecture 5

Source: Lecture 5

What advantages do sparse vectors have over dense ones, and vice versa? (1)

Question 4

Topic: Lecture 5

Source: Lecture 5

Explain the logic behind the IDF part of TF-IDF (ie, why does it give higher weights to more "interesting" words?). (1)

Question 5

Topic: Lecture 6

Source: Lecture 6

In class, we saw a few topics that we were unable to identify. What could be a cause for such pointless topics (ie, how might we ensure that our topics are better? (2 reasons). (1)

Question 6

Topic: Lecture 7

Source: Lecture 7

When doing information retrieval, bag-of-words (and even just indicator functions) typically work very well. Explain why context is less important if we have a well-designed query. You may also want to explain your assumptions about a “well-designed” query. (2)

Question 7

Topic: Lecture 6

Source: Lecture 6

Imagine we performed LDA on the classes in this block. What might their Theta distributions look like? (2)

Question 8

Topic: Lecture 8

Source: Lecture 8

What are some assumptions that we make when we are interpolating between a document and a corpus? When should we trust the corpus more, and when should we trust the document more? (2)

Question 9

Topic: Long

Source: Lecture 8

In class, I mentioned that we might want to provide a weighted ranking of documents in an IR system (for example, we might have a list of relevant documents already sorted for relevancy, and we want our IR system to not only return those documents high in the returned document list, but in the same order). How might we modify the MAP metric to reward the ordering of the documents, as well? Explain (use an example, if you have to). (3)

END OF QUIZ