START OF QUIZ
Student ID:
92642370,Chen,Tiffany

Question 1

Topic: Lecture 5
Source: Lecture 5

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

Question 2

Topic: Lecture 8
Source: Lecture 8

What is the intuition behind MAP? (1)

Question 3

Topic: Lecture 6
Source: Lecture 6

Why do we need a "human in the loop" for topic modeling? (1)

Question 4

What is the reasonining behind substituting TF-IDF with Okapi BM25? (1)

Question 5

Topic: Lecture 5
Source: Lecture 5

Why can we be confident that a low-rank approximation of a matrix contains the most important information in a document? (1)

Question 6

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 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 8

Explain why the cosine similarity between a (TF-IDF-weighted) document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query. (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