

START OF QUIZ

Student ID:

47323894,Liao,Jingyi

Question 1

Topic: Lecture 5

Source: Lecture 5

The Frobenius norm looks very similar to a distance metric we've already observed. Explain which one. (1)

Question 2

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 3

Topic: Lecture 7

Source: Lecture 7

What is the benefit (in terms of efficiency) of placing the most discriminative search terms first in a boolean search? (1)

Question 4

Topic: Lecture 8

Source: Lecture 8

In class, I mentioned that high k value for BM25 TF weighting rewards documents with many, many instances of a term in them. Explain why that's the case. (2)

Question 5

Topic: Lecture 6

Source: Lecture 6

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

Question 6

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 7

Topic: Lecture 5

Source: Lecture 5

Why can we represent a rank- m matrix as the sum of m rank-1 matrices *or* the product of an $n \times m$ matrix and an $m \times n$ matrix (ie, what is matrix multiplication doing that we can take advantage of)? Explain. (2)

Question 8

Topic: Lecture 7

Source: Lecture 7

Explain why the cosine similarity between a 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 7

Imagine that we have 2 information retrieval systems, and we are evaluating on the same test set, which has 10 relevant documents. The first system returns them in positions [1, 5, 7, 15, 25, 50, 60, 70, 71, 90]. The second returns the documents at positions [2, 3, 6, 8, 10, 62, 80, 83, 91, 95]. Make an argument for each system being better, and provide support for both. Explain which system you would rather use, and why. If there are any other considerations, list them. (3)

END OF QUIZ