

START OF QUIZ

Student ID:

43887546,Kumar,Abhi

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 6

Source: Lecture 6

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

Question 3

Topic: Lecture 7

Source: Lecture 7

What is the benefit of evaluating boolean queries using set operations instead of loops? (1)

Question 4

Topic: Lecture 8

Source: Lecture 8

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

Question 5

Topic: Lecture 6

Source: Lecture 6

In some ways, we could consider Beta distributions themselves to be an embedding of a topic. Explain, and explain how we might be able to leverage that. (2)

Question 6

Topic: Lecture 8

Source: Lecture 8

Why do we not simply take the probability of a word given its document (maybe with smoothing added in)? (1)

Question 7

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 8

Topic: Lecture 5

Source: Lecture 5

What impact do sparse matrices have on similarity metrics like cosine similarity? (1)

Question 9

Topic: Long

Source: Lecture 5

Imagine that we are working with a language other than English, such as Indonesian, with significant agglutinative morphology (words are inflected through the concatenation of affixes to a lemma). How do you think that this would impact our various vector space models? Which of them would be most affected, and which would be least affected? Explain. (3)

END OF QUIZ