

START OF QUIZ

Student ID:

34552166,Kuang,Jessie

Question 1

Topic: Lecture 5

Source: Lecture 5

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

Question 2

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 3

Topic: Lecture 6

Source: Lecture 6

Why can't we just run an HMM over documents to discover the latent states like we do for POS-tagging? (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 7

Source: Lecture 7

Define $P @ R$. (1)

Question 6

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 7

Topic: Lecture 5

Source: Lecture 5

We often weight our matrices using something like PMI or TF-IDF. Do you think it would make sense to do this after applying SVD? Why or why not? (2)

Question 8

Topic: Lecture 6

Source: Lecture 6

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

Question 9

Topic: Coding

Source: Coding

Imagine that our corpus contains 1M documents. We have 3 queries that we are looking at. The first query has 5 relevant documents, returned in positions 1, 5, 10, 20, and 50. The second query has 3 relevant documents, returned at 10, 11, and 12. The third query has only one relevant document, and it is returned in position 7. What is the MAP of our system? (3)

END OF QUIZ