

START OF QUIZ

Student ID:

21741764,Hou,Bingyang

Question 1

Topic: Lecture 6

Source: Lecture 6

In class, we talked about bookstores and streaming algorithms classifying books / movies. How can we tell that they don't use a topic modeling algorithm (or, if you think they do, what would be some clues)? (1)

Question 2

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 3

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 4

Topic: Lecture 8

Source: Lecture 8

Why don't we use a higher-order language model to perform IR? (1)

Question 5

Topic: Lecture 7

Source: Lecture 7

What is the purpose of an inverted index? (1)

Question 6

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 7

Topic: Lecture 7

Source: Lecture 7

Explain why boolean filtering is usually insufficient for retrieval, and why we normally need some way of scoring the documents. (2)

Question 8

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

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