

START OF QUIZ

Student ID:

90660002,Benipal,Jaskirat

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 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 3

Topic: Lecture 8

Source: Lecture 8

What are some assumptions that we make when we are interpolating between a document and a corpus? When should we trust the corpus more, and when should we trust the document more? (2)

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

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 6

Topic: Lecture 7

Source: Lecture 7

Why do we generally care more about precision than recall in IR? (1)

Question 7

Topic: Lecture 7

Source: Lecture 7

Define $\mathbb{P} @ \mathbb{R}$. (1)

Question 8

Topic: Lecture 6

Source: Lecture 6

Why don't we just use k-means to cluster document-vectors (sparse or dense)? (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