START OF QUIZ
Student ID:
50656347,Zhang,Lisa

Question 1

Topic: Lecture 5
Source: Lecture 5

Why can we be confident that a low-rank approximation of a matrix contains the most important information in a document? (1)

Question 2

Topic: Lecture 7
Source: Lecture 7

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

Question 3

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

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

Question 6

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

Question 7

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

Question 8

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 9

Topic: Long
Source: Lecture 8

In class, I mentioned that we might want to provide a weighted ranking of documents in an IR system (for example, we might have a list of relevant documents already sorted for relevancy, and we want our IR system to not only return those documents high in the returned document list, but in the same order). How might we modify the MAP metric to reward the ordering of the documents, as well? Explain (use an example, if you have to). (3)

END OF QUIZ