

START OF QUIZ

Student ID:

27841444,Gu,Martin

Question 1

Topic: Lecture 6

Source: Lecture 6

Why do we need a “human in the loop” for topic modeling? (1)

Question 2

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 3

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 4

Topic: Lecture 8

Source: Lecture 8

What is the intuition behind MAP? (1)

Question 5

Topic: Lecture 7

Source: Lecture 7

From a processing perspective, what is one benefit structured data has over unstructured data, and vice versa. (1)

Question 6

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 7

Topic: Lecture 6

Source: Lecture 6

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