

START OF QUIZ

Student ID:

22039382,Amal,Cenith

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 7

Source: Lecture 7

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

Question 3

Topic: Lecture 6

Source: Lecture 6

Why don't we just use k-means to cluster document-vectors (sparse or dense)? (1)

Question 4

Topic: Lecture 8

Source: Lecture 8

What is the intuition behind MAP? (1)

Question 5

Topic: Lecture 7

Source: Lecture 7

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

Question 6

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 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 8

Source: Lecture 8

In class (and in the lab) you saw some examples of using a language model for IR. How do you think we could incorporate an LLM into the IR pipeline? In what ways do you think an n-gram lm might be more appropriate? (2)

Question 9

Topic: Long

Source: Lecture 5

Imagine that we are working with a language other than English, such as Indonesian, with significant agglutinative morphology (words are inflected through the concatenation of affixes to a lemma). How do you think that this would impact our various vector space models? Which of them would be most affected, and which would be least affected? How might we go about adapting our model to solve these problems? Explain. (3)

END OF QUIZ