

**START OF QUIZ**

**Student ID:**

**34447581,Ong,Claudia**

## Question 1

Topic: Lecture 6

Source: Lecture 6

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

## Question 2

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 3

Topic: Lecture 8

Source: Lecture 8

$P(d|q)$  is not what we are solving with the language model. Why is this not generally a problem? (1)

## Question 4

Topic: Lecture 7

Source: Lecture 7

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

## Question 5

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 6

Topic: Lecture 8

Source: Lecture 8

Why do we not simply take the probability of a word given its document (maybe with smoothing added in)? (1)

## Question 7

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 8

Topic: Lecture 7

Source: Lecture 7

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

## Question 9

Topic: Long

Source: Lecture 8

In class, we considered two different types of information retrieval systems - one that uses Boolean terms to find matches, and one that uses a language model to allow for "natural language" queries. Can you think of a way that we might be able to leverage the strengths of both, while minimizing the disadvantages? Briefly explain how that might work. (2)

**END OF QUIZ**