

**START OF QUIZ**

**Student ID:**

**70094511, Ma, Muyuan**

## 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 8

Source: Lecture 8

What do we mean by interpolation? (1)

### Question 3

Topic: Lecture 5

Source: Lecture 5

What advantages do sparse vectors have over dense ones? (1)

## Question 4

Topic: Lecture 5

Source: Lecture 5

The Frobenius norm looks very similar to a distance metric we've already observed. Explain which one. (1)

## Question 5

Topic: Lecture 7

Source: Lecture 7

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

## Question 6

Topic: Lecture 7

Source: Lecture 7

Explain why boolean filtering is usually insufficient for retrieval, and why we normally need some way of scoring the documents. (2)

## Question 7

Topic: Lecture 8

Source: Lecture 8

Why don't we use a higher-order language model to perform IR? (1)



## Question 8

Topic: Lecture 6

Source: Lecture 6

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

## 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**