

START OF QUIZ

Student ID:

**70152863,Khan,Muhammad
Mujtaba**

Question 1

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 2

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 3

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 4

Topic: Lecture 6

Source: Lecture 6

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

Question 5

Topic: Lecture 8

Source: Lecture 8

In class, I mentioned that TF-IDF unfairly punishes words that appear in a lot, but not all, of the documents in our corpus. Explain how Okapi BM25 attempts to fix this. (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 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 7

Source: Lecture 7

Explain why the cosine similarity between a (TF-IDF-weighted) document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query. (2)

Question 9

Topic: Long

Source: Lecture 6

Imagine that we have a Beta distribution for each document, and a Theta distribution for each topic. We are at the Maximization state of EM write a short function that calculates the probability of a document, given these distributions. Pay special attention to edge cases and special considerations... (3)

END OF QUIZ