START OF QUIZ
Student ID:
71936512,Jimenez,Daniel

Question 1

What do we mean by interpolation? (1)

Question 2

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

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

Question 4

Topic: Lecture 6
Source: Lecture 6

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

Question 5

What is the reasonining behind substituting TF-IDF with Okapi BM25? (1)

Question 6

Topic: Lecture 5
Source: Lecture 5

Why do we need methods like t-SNE? (1)

Question 7

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

Question 8

Topic: Lecture 7
Source: Lecture 7

Explain why the cosine similarity between a 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

Imagine that we have a Beta distribution for each document, and a Theta distribution for each document. 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