START OF QUIZ
Student ID:
91877605,Ren,Justin

Question 1

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 2

Topic: Lecture 8
Source: Lecture 8

What are some assumptions that we make when we are interpolating between a document and a corpus? When should we trust the corpus more, and when should we trust the document more? (2)

Question 3

Topic: Lecture 7
Source: Lecture 7

What is the benefit of evaluating boolean queries using set operations instead of loops? (1)

Question 4

Topic: Lecture 6
Source: Lecture 6

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

Question 5

Topic: Lecture 8
Source: Lecture 8

In class, I mentioned that high k value for BM25 TF weigthing rewards documents with many, many instances of a term in them. Explain why that's the case. (2)

Question 6

Topic: Lecture 5
Source: Lecture 5

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

Question 7

Topic: Lecture 6
Source: Lecture 6

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

Question 8

Topic: Lecture 7
Source: Lecture 7

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

Question 9

Topic: Coding
Source: Coding

Write a short function that confirms that the sum of n rank-1 matrices is identical to the product of an nxk matrix and a kxn matrix. (3)

END OF QUIZ