START OF QUIZ
Student ID:
29014990,Mirjalili,Sara

Question 1

Topic: Lecture 6
Source: Lecture 6

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

Question 2

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 x m matrix and an m x n matrix (ie, what is matrix multiplication doing that we can take advantage of?)? Explain. (2)

Question 3

Topic: Lecture 7
Source: Lecture 7

Define P@R. (1)

Question 4

Topic: Lecture 7
Source: Lecture 7

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

Question 5

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 6

Topic: Lecture 6
Source: Lecture 6

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

Question 7

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 8

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 9

Topic: Coding
Source: Coding

Write a function that returns the most likely n documents given a term-document matrix, a smoothing parameter, and a query. (3)

END OF QUIZ