

START OF QUIZ

Student ID:

94325370,Qiao,Grace

Question 1

Topic: Lecture 6

Source: Lecture 6

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

Question 2

Topic: Lecture 8

Source: Lecture 8

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

Question 3

Topic: Lecture 7

Source: Lecture 7

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

Question 4

Topic: Lecture 6

Source: Lecture 6

Why can't we just run an HMM over documents to discover the latent states like we do for POS-tagging? (1)

Question 5

Topic: Lecture 8

Source: Lecture 8

$P(d|q)$ is not what we are solving with the language model. Why is this not generally a problem? (1)

Question 6

Topic: Lecture 5

Source: Lecture 5

Explain the logic behind the IDF part of TF-IDF (ie, why does it give higher weights to more "interesting" words?). (1)

Question 7

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 8

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 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 $n \times k$ matrix and a $k \times n$ matrix. (3)

END OF QUIZ