

START OF QUIZ
Student ID:
32029357,Oh,Jacob

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 5

Source: Lecture 5

What impact do sparse matrices have on similarity metrics like cosine similarity? (1)

Question 3

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 4

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 5

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 6

Topic: Lecture 6

Source: Lecture 6

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

Question 7

Topic: Lecture 8

Source: Lecture 8

What do we mean by interpolation? (1)

Question 8

Topic: Lecture 7

Source: Lecture 7

What is the benefit (in terms of efficiency) of placing the most discriminative search terms first in a boolean search? (1)

Question 9

Topic: Coding

Source: Coding

Imagine that our corpus contains 1M documents. We have 3 queries that we are looking at. The first query has 5 relevant documents, returned in positions 1, 5, 10, 20, and 50. The second query has 3 relevant documents, returned at 10, 11, and 12. The third query has only one relevant document, and it is returned in position 7. What is the MAP of our system? (3)

END OF QUIZ