START OF QUIZ
Student ID:
74035403,Zheng,Meiyu

Question 1

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 2

Topic: Lecture 6
Source: Lecture 6

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

Question 3

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 4

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 5

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 6

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 7

Topic: Lecture 7
Source: Lecture 7

Define P@R. (1)

Question 8

Topic: Lecture 5
Source: Lecture 5

What advantages do sparse vectors have over dense ones. (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