START OF QUIZ
Student ID:
37469715,Sharma,Prakul

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 7
Source: Lecture 7

Define P@R. (1)

Question 3

Topic: Lecture 6
Source: Lecture 6

Why do we need a ”human in the loop” for topic modeling? (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 8
Source: Lecture 8

Why don't we use a higher-order language model to perform IR? (1)

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 5
Source: Lecture 5

Why do we need methods like t-SNE? (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: Long
Source: Lecture 6

Imagine that we have a Beta distribution for each document, and a Theta distribution for each document. We are at the Maximization state of EM write a short function that calculates the probability of a document, given these distributions. Pay special attention to edge cases and special considerations... (3)

END OF QUIZ