

START OF QUIZ

Student ID:

55343529,Nadal,Jacob

Question 1

Topic: Lecture 1

Source: Lecture 1

What intuition about substitutions allows the DP version of Levenshtein distance to work as it does? Briefly explain. (1)

Question 2

Topic: Lecture 2

Source: Lecture 2

Explain the purpose of a centroid in K-means clustering, and how we can think of it with respect to its cluster. (1)

Question 3

Topic: Lecture 2

Source: Lecture 2

Describe the intuition behind K-means++ (ie, why do we use it, and what is it trying to accomplish?) (1)

Question 4

Topic: Lecture 4

Source: Lecture 4

What is the main difference between the Viterbi algorithm and the Forward algorithm, and why does it allow us to find the optimal path through a sequence? (1)

Question 5

Topic: Lecture 1
Source: Lecture 1

Suppose we are filling the table for the Levenshtein distance algorithm. We are in cell (x, y) . The values of cell $(x-1, y-1)$, $(x-1, y)$, and $(x, y-1)$ are 3, 4, and 3, respectively. What is the value we will put in cell (x, y) , given that the letters are equal? (1)

Question 6

Topic: Lecture 3

Source: Lecture 3

Imagine you were trying to pitch a new version of Scrabble to Hasbro that included "digraphs" (ie, combinations of two consecutive letters, like "th"). Do you think that you could score them as a simple combination of the single letter scores (ie, "th" is worth "t" + "h"), or would you need to do some more complex scoring calculations? Explain. (2)

Question 7

Topic: Lecture 4

Source: Lecture 4

Let's imagine we're modifying our HMM to handle 2nd-order Markov operations (ie, consider the previous two states). Does anything in the model fundamentally change? Describe which aspects of the forward/Viterbi algorithm would need to be modified, if any. (2)

Question 8

Topic: Lecture 3

Source: Lecture 3

Imagine that we have a trigram model that encounters a trigram where none of the tokens are in the vocabulary. How do you think that might impact our probability calculation for the sentence? How might we go about finding a solution? (2)

Question 9

Topic: Long

Source: Lecture 2

Imagine you are tasked with clustering social media posts to identify trends or topics. You have access to a large amount of unstructured text data. What kind of features do you think would be helpful, how would you preprocess the data, and how would you verify that the clustering is a good one? (3)

END OF QUIZ