START OF QUIZ
Student ID:
22039382,Amal,Cenith

Question 1

Topic: Lecture 2
Source: Lecture 2

What is the impact of choosing a poor value for k in k-means clustering? How can we determine a more appropriate k? (1)

Question 2

Topic: Lecture 3
Source: Lecture 3

In your own words, explain the Markov assumption, and how it is used in language modeling. (1)

Question 3

Topic: Lecture 2
Source: Lecture 2

Desribe the concept of cluster homogeneity, and how it relates to precision. (1)

Question 4

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 5

Topic: Lecture 4
Source: Lecture 4

How does semi-supervised learning differ from unsupervised and fully-supervised learning? (1)

Question 6

Topic: Lecture 1
Source: Lecture 1

Let's consider a variant of the string alignment problem where instead of aligning characters, we're aligning sequences of characters (maybe we're doing machine translation...). What would need to be modified to handle a situation where we likely have a much higher vocabulary, and there's a lot less copying going on? What assumptions would we be making about the data? Would any of these assumptions make Levensthein distance inappropriate? (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 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 9

Topic: Long
Source: Lecture 1

Do you think that auto-correct has a bias for where in a word an error occurs (ie, the index of the mistake)? If so, how might you approach fixing this problem? If not, explain why the position doesn't matter. As always, list any assumptions you're making. (3)

END OF QUIZ