START OF QUIZ
Student ID:
42135814,Lopez Gonzalez,Nico

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 4
Source: Lecture 4

What makes dynamic programming methods, such as the Viterbi algorithm, more efficient for sequence prediction tasks compared to brute-force approaches? (1)

Question 3

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 4

Topic: Lecture 4
Source: Lecture 4

Why can we use logarithms for the Viterbi algorithm, but not the forward algorithm? (1)

Question 5

Topic: Lecture 1
Source: Lecture 1

Why is cosine distance typically a more suitable distance metric for semantic spaces than Euclidean distance? (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 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 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 3

In class, we built a collocation matrix for a bigram language model. Modify the function so that it can handle a trigram language model and implements "add-alpha" smoothing, instead of "add-one" smoothing. (3)

END OF QUIZ