START OF QUIZ
Student ID:
56633415,Cady,Amandisa

Question 1

Topic: Lecture 3
Source: Lecture 3

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

Question 2

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 3

Topic: Lecture 3
Source: Lecture 3

When you were using Naive Bayes, a bag of words model was sufficient for classification. Why is it too simplistic for language modeling? (1)

Question 4

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 5

Topic: Lecture 4
Source: Lecture 4

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

Question 6

Topic: Lecture 1
Source: Lecture 1

What is the primary concern of a semantic vector space (ie, a vector space representing meaning), and how does it relate to our use of cosine similarity to measure word similarity? Can you think of any sorts of words for which it might be very difficult to satisfy this concern? (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 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 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