START OF QUIZ
Student ID:
38595138,Christilaw,Tim

Question 1

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 2

Topic: Lecture 3
Source: Lecture 3

Describe the noisy channel model, and how it can be used to represent Machine Translation. (1)

Question 3

Topic: Lecture 3
Source: Lecture 3

If we have the sentence “You keep using that word - I do not think it means what you think it means”, what is the probability of the bigram “you think”, assuming that the sentence is the entire corpus? (1)

Question 4

Topic: Lecture 1
Source: Lecture 1

When is dynamic programming more efficient than brute force programming? (ie, what assumptions do we make about a problem when we use dynamic programming?) (1)

Question 5

Topic: Lecture 1
Source: Lecture 1

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

Question 6

Topic: Lecture 4
Source: Lecture 4

Imagine that we are doing OCR (optical character recognition; ie, the translation of handwritten text into digital text) instead of POS tagging. Do you think we could use an HMM? If so, what would the states, transitions, and emissions be? If not, describe why it's an inappropriate tool for the task. (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 2
Source: Lecture 2

Imagine we were using k-means to cluster misspellings around their correct spellings. How many clusters would we need, and what would be a good distance function? 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