

START OF QUIZ

Student ID:

88179403,Du,Yiyang

Question 1

Topic: Lecture 3

Source: Lecture 3

Explain the purpose of padding in language modeling. (1)

Question 2

Topic: Lecture 2

Source: Lecture 2

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

Question 3

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 4

Topic: Lecture 4

Source: Lecture 4

How is it that EM can arrive at a good solution, even if we have a random initialization of parameters? (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 2, 2, and 2, respectively. What is the value we will put in cell (x, y) , given that the letters are NOT equal? (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 Levenshtein distance inappropriate?

(2)

Question 7

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 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 4

Please see the long question from lecture 4 in the quiz bank on Github. (3)

END OF QUIZ