

START OF QUIZ

Student ID:

74874876, Wang, Chenxin

Question 1

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 1, 3, and 5, respectively. What is the value we will put in cell (x, y) , given that the letters are equal? (1)

Question 2

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 3

Topic: Lecture 1

Source: Lecture 1

When is cosine similarity appropriate as a similarity measure? (1)

Question 4

Topic: Lecture 3

Source: Lecture 3

Imagine that we are doing machine translation instead of POS-tagging. What would be the equivalent of emission probabilities and transition probabilities? Explain. (2)

Question 5

Topic: Lecture 3

Source: Lecture 3

Explain the purpose of Laplace smoothing, and how it accomplishes its goal. (1)

Question 6

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 7

Topic: Lecture 2

Source: Lecture 2

Why is the Forgy initialization sub-optimal? (1)

Question 8

Topic: Lecture 4

Source: Lecture 4

Imagine that we are doing ASR instead of POS tagging. Briefly describe what the emissions and transitions would be. (2)

Question 9

Topic: Long

Source: Lecture 2

Imagine that we are creating a bilingual dictionary, and we want to cluster words that are likely translations of each other (this task is known as "Bilingual Lexicon Induction", or BLI). What kind of features might be good features for this task, and how would we convert them to numerical representations? You can assume that we have a large bilingual corpus that is sentence aligned, but no further information. Do you think we could use K-Means for this task? If not, why not? If so, what kind of special considerations would we need to make, if any?

END OF QUIZ