START OF QUIZ
Student ID:
38522413,Wang,Yuxi

Question 1

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 2

Topic: Lecture 3
Source: Lecture 3

Why do we use log-probability intstead of linear probability? (1)

Question 3

Topic: Lecture 3
Source: Lecture 3

Describe the noisy channel model, and how it can be used to represent [Machine Translation, ASR, POS-tagging]. (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 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 6

Topic: Lecture 1
Source: Lecture 1

What is the main difference between Hamming Distance and Edit Distance? (1)

Question 7

Why do outliers cause problems for clustering algorithms like k-means? How can we deal with them? (1)

Question 8

Topic: Lecture 1
Source: Lecture 1

Explain what modifications would need to be made to our dynamic edit distance algorithm to incorporate weighted edit distance. (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