START OF QUIZ
Student ID:
64243512,ZENG,Min

Question 1

Topic: Topic2
Source: Lecture 2

Are both K-means and agglomerative clustering iterative? Explain, and for each that is, explain when the algorithm ends.

Question 2

Topic: Topic2
Source: Lecture 2

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

Question 3

Topic: Topic3
Source: Lecture 3

Describe the noisy channel model, and how it can be used to represent POS-Tagging.

Question 4

Topic: Topic1
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 4, respectively. What is the value we will put in cell (x, y), given that the letters are equal?

Question 5

Topic: Topic4
Source: Lecture 4

What is the main difference between the Viterbi algorithm and the Forward algorithm, and why does it allow us to find the optimal path through a sequence?

Question 6

Topic: Topic4
Source: Lecture 4

Briefly describe why soft EM might provide more accurate tagging results than hard EM.

Question 7

Topic: Topic1
Source: Lecture 1

When is cosine similarity appropriate as a similarity measure?

Question 8

Topic: Topic3
Source: Lecture 3

If our vocabulary consists of just symbols A and B, and our corpus consists of the sequence: B A A B B A, and we build a bigram language model by applying add-one smoothing to the MLE from the corpus, what is the probability of P(B||A)? Please show your work.

Question 9

Topic: Coding
Source: Lecture 2

Imagine we have three clusters [[X, Y], [M, N, P], [A, B, C, D]], and a point [R]. Write a function that determines which cluster to add R to, given the min linkage criterion.

END OF QUIZ