START OF QUIZ
Student ID:
95153730,Ihn,Jae

Question 1

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 2

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 3

Topic: Topic2
Source: Lecture 2

When is it more appropriate to use hierarchical clustering than k-means?

Question 4

Topic: Topic4
Source: Lecture 4

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

Question 5

In your own words, explain the Markov assumption, and how it is used for language modeling.

Question 6

If our vocabulary consists of just symbols A and B, and our corpus consists of the sequence: B A B A 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 7

Topic: Topic1
Source: Lecture 1

When is Manhattan distance more appropriate than Euclidean distance?

Question 8

Topic: Topic4
Source: Lecture 4

How is it that EM can arrive at a good solution, even if we have a random initialization of parameters?

Question 9

Topic: Coding
Source: Lecture 3

In class, we built a collocation matrix for a bigram language model. Modify the function so that it can handle trigram language model and implements "add-alpha" smoothing, instead of "add-one" smoothing.

END OF QUIZ