START OF QUIZ
Student ID: 75287953,Apparaju,Ananya

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

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 NOT equal?

Question 3

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 4

Topic: Topic1
Source: Lecture 1

When is Manhattan distance more appropriate than Euclidean distance?

Question 5

Topic: Topic3
Source: Lecture 3

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 6

Topic: Topic2
Source: Lecture 2

Why is the Forgy initialization sub-optimal?

Question 7

Topic: Topic4
Source: Lecture 4

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

Question 8

Topic: Topic3
Source: Lecture 3

Explain why HMMs are a generative model, and how that differs from a discriminative model.

Question 9

Topic: Coding
Source: Lecture 1

Write a function that, given 2 vectors, calculates the cosine distance between them (don't just use the built-in cosine similarity function).

END OF QUIZ