

**START OF QUIZ**

**Student ID:**

**95174918,Maurin-  
Jones,Kai**

## Question 1

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 2

Topic: Topic2

Source: Lecture 2

Why is the Forgy initialization sub-optimal?

### Question 3

Topic: Topic4

Source: Lecture 4

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

## Question 4

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 5

Topic: Topic1

Source: Lecture 1

When is Manhattan distance more appropriate than Euclidean distance?

## Question 6

Topic: Topic2

Source: Lecture 2

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

## Question 7

Topic: Topic3

Source: Lecture 3

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



## Question 8

Topic: Topic3

Source: Lecture 3

If our vocabulary consists of just symbols A and B, and our corpus consists of the sequence: A B B A A B, 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 mean linkage criterion.

**END OF QUIZ**