

# START OF QUIZ

Student ID: 75576249, Tse,  
Timothy

## Question 1

Topic: Lecture 2

Source: Lecture 2

Describe the concept of cluster homogeneity, and how it relates to precision. (1)

## Question 2

Topic: Lecture 3

Source: Lecture 3

Explain the purpose of padding in language modeling. (1)

### Question 3

Topic: Lecture 1  
Source: Lecture 1

Unsupervised Learning typically tries to find structure in unlabeled data. Give two reasons why we might want to still annotate a small dataset to use with our algorithm. (1)

## Question 4

Topic: Lecture 2

Source: Lecture 2

Describe the purpose of linkage in hierarchical clustering. (1)

## Question 5

Topic: Lecture 4

Source: Lecture 4

How is it that EM can arrive at a good solution, even if we have a random initialization of parameters? (1)

## Question 6

Topic: Lecture 3

Source: Lecture 3

Imagine that we have a trigram model that encounters a trigram where none of the tokens are in the vocabulary. How do you think that might impact our probability calculation for the sentence? How might we go about finding a solution? (2)

## Question 7

Topic: Lecture 1

Source: Lecture 1

Let's consider a variant of the string alignment problem where instead of aligning characters, we're aligning sequences of characters (maybe we're doing machine translation...). What would need to be modified to handle a situation where we likely have a much higher vocabulary, and there's a lot less copying going on? What assumptions would we be making about the data? Would any of these assumptions make Levenshtein distance inappropriate?

(2)



## Question 8

Topic: Lecture 4

Source: Lecture 4

Imagine that we are doing OCR (optical character recognition; ie, the translation of hand-written text into digital text) instead of POS tagging. Do you think we could use an HMM? If so, what would the states, transitions, and emissions be? If not, describe why it's an inappropriate tool for the task. (2)

## Question 9

Topic: Long

Source: Lecture 4

Please see the long question from lecture 4 in the quiz bank on Github. (3)

**END OF QUIZ**