START OF QUIZ
Student ID:
27841444,Gu,Martin

Question 1

Topic: Lecture 1
Source: Lecture 1

Why is cosine distance typically a more suitable distance metric for semantic spaces than Euclidean distance? (1)

Question 2

Topic: Lecture 2
Source: Lecture 2

Describe the intuition behind K-means++ (ie, why do we use it, and what is it trying to accomplish?) (1)

Question 3

Topic: Lecture 3
Source: Lecture 3

If we have the sentence “I am what I am, and that’s all that I am”, what is the probability of the bigram “I am”, assuming that the sentence is the entire corpus? (1)

Question 4

Topic: Lecture 1
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 4, 3, and 3, respectively. What is the value we will put in cell (x, y), given that the letters are NOT equal? (1)

Question 5

Topic: Lecture 4
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? (1)

Question 6

Topic: Lecture 2
Source: Lecture 2

What kinds of data might be difficult to cluster using k-means? Is it a shortcoming of the algorithm, or does it just need very careful feature engineering and distance calculations? (2)

Question 7

Topic: Lecture 4
Source: Lecture 4

Imagine that we are doing OCR (optical character recognition; ie, the translation of handwritten 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 8

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 9

Topic: Long
Source: Lecture 4

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

END OF QUIZ