

START OF QUIZ

Student ID:

68022458,Chan,Douglas

Question 1

Topic: Lecture 4

Source: Lecture 4

What is the main purpose of semi-supervised learning in EM? That is, how does it affect the overall model, and where is the effect the largest? (1)

Question 2

Topic: Lecture 4

Source: Lecture 4

Iterative algorithms often require a stopping condition. Briefly explain why this is necessary, and why perplexity is a metric to use for stopping HMMs. (2)

Question 3

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 1, 3, and 5, respectively. What is the value we will put in cell (x, y) , given that the letters are equal? (1)

Question 4

Topic: Lecture 2

Source: Lecture 2

Why do outliers cause problems for clustering algorithms like k-means? How can we deal with them? (1)

Question 5

Topic: Lecture 2

Source: Lecture 2

Imagine we were using k-means to cluster misspellings around their correct spellings. How many clusters would we need, and what would be a good distance function? Explain. (2)

Question 6

Topic: Lecture 3

Source: Lecture 3

Imagine that we are doing machine translation instead of POS-tagging. What would be the equivalent of emission probabilities and transition probabilities? Explain. (2)

Question 7

Topic: Lecture 1

Source: Lecture 1

Explain why edit distance (given our formulation) will always choose a substitution, if it can. (1)

Question 8

Topic: Lecture 3

Source: Lecture 3

Explain the purpose of Laplace smoothing, and how it accomplishes its goal. (1)

Question 9

Topic: Long

Source: Lecture 1

We've all had an instance of autocorrect suggesting a bizarre correction for something. Given what you know about word similarity for error correction, explain why autocorrect doesn't always pick the word with the lowest edit distance. (3)

END OF QUIZ