START OF QUIZ
Student ID:
96733092,Zhou,Zhiyang

Question 1

Topic: Lecture 1
Source: Lecture 1

When is dynamic programming more efficient than brute force programming? (ie, what assumptions do we make about a problem when we use dynamic programming?) (1)

Question 2

Topic: Lecture 3
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 maximum likelihood estimate from the corpus, what is the probability P(B|A)? Please show your work. (2)

Question 3

Topic: Lecture 2
Source: Lecture 2

When is it more appropriate to use hierarchical clustering than k-means? (1)

Question 4

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 5

Topic: Lecture 3
Source: Lecture 3

Why do we use log-probability intstead of linear probability? (1)

Question 6

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 7

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 8

Topic: Lecture 1
Source: Lecture 1

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

Question 9

Topic: Long
Source: Lecture 2

Imagine that we are creating a bilingual dictionary, and we want to cluster words that are likely translations of each other (this task is known as "Bilingual Lexicon Induction", or BLI). What kind of features might be good features for this task, and how would we convert them to numerical representations? You can assume that we have a large bilingual corpus that is sentence aligned, but no further information. Do you think we could use K-Means for this task? If not, why not? If so, what kind of special considerations would we need to make, if any?

END OF QUIZ