

START OF QUIZ

Student ID:

74832403,Kumar,Rakesh

Question 1

Topic: Topic2

Source: Lecture 2

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

Question 2

Topic: Topic3

Source: Lecture 3

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

Question 3

Topic: Topic2

Source: Lecture 2

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

Question 4

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 equal?

Question 5

Topic: Topic3

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.

Question 6

Topic: Topic1

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?)

Question 7

Topic: Topic4

Source: Lecture 4

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

Question 8

Topic: Topic4

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?

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 max linkage criterion.

END OF QUIZ