

START OF QUIZ

Student ID: 97132088, Bal-
akrishnan, Aswin
Chidambaram

Question 1

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 2

Topic: Topic2

Source: Lecture 2

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

Question 3

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 4

Topic: Topic1

Source: Lecture 1

Discuss why one might do unsupervised learning instead of supervised learning.

Question 5

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, 1, and 2, respectively. What is the value we will put in cell (x, y) , given that the letters are NOT equal?

Question 6

Topic: Topic4

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?

Question 7

Topic: Topic3

Source: Lecture 3

Explain why HMMs are a generative model, and how that differs from a discriminative model.

Question 8

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 9

Topic: Coding

Source: Lecture 1

Write a function that, given 2 vectors, calculates the cosine distance between them (don't just use the built-in cosine similarity function).

END OF QUIZ