CS244 Exercise 9

Task 1: HMM Parameters

Suppose we have a hidden Markov model with N states, each of which can output one of k observation symbols, and we have an observation sequence of length T. Our HMM is represented by $\lambda = (A, B, \pi)$, where A represents the transition parameters, B represents the emission parameters, and π represents the initial starting state parameters.

For π, how many initial starting state parameters does our model have?

For A, how many transition parameters does our model have?

For B, how many emission parameters does our model have?

Download the Jupyter Notebook for Exercise 9 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

Task 2: HMM Execution

What is the optimal state sequence for the coin example?

What is the probability (not the natural log of the probability) of the optimal state sequence?

Task 3: Class Colors on Wellesley Campus

What is the most likely sequence of buildings that Wendy walked through?

Task 4: Auto-Correction of Typos in Text

How frequently is "o" followed by "u" in the training data?

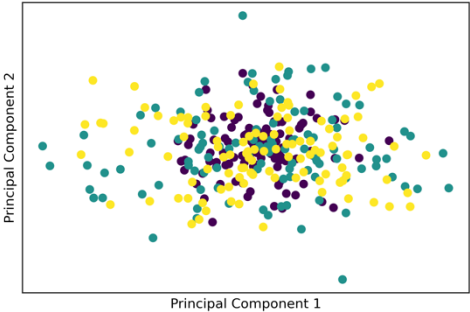What character is "h" most commonly followed by in the training data?

What percentage of characters in `typos_observation.txt` are typos?
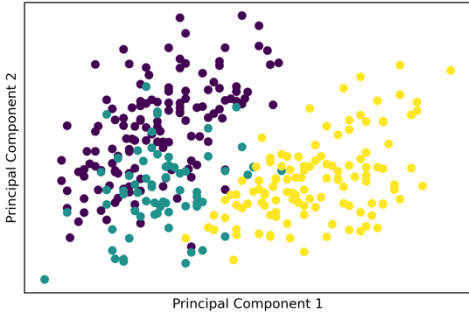
What percentage of characters in `path.txt` are typos?

Task 5: Dimensionality Reduction and Principal Component Analysis

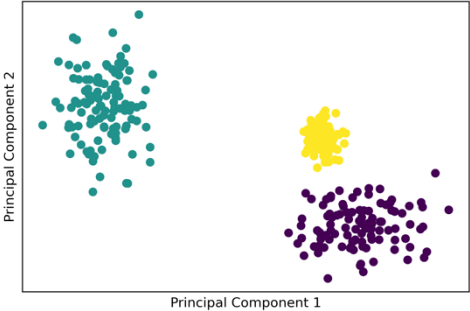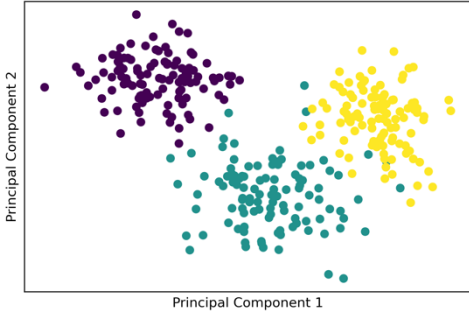Which of the four figures below is your plot most similar to?

(A)

(B)

(C)

(D)

What percentage of the variance in the data is explained by the first two principal components?