INDIAN INSTITUTE OF INFORMATION TECHNOLOGY KALYANI

Kalyani, Nadia-741235, West Bengal
(An institute of National Importance)

Course Title: Machine Learning Lab Paper Code: CSC612 Spring 2025

Assignment 9

Date: 25/03/2025 Due date: 25/03/2025 1 PM 5 Marks

An HMM consists of two types of variables: hidden states and observations.

- The hidden states are the underlying variables that generate the observed data but are not directly observable. Here we consider NE tags as the hidden states.
- The observations are the output variables that are observed. Here we consider POS tags as the observations.

The relationship between the hidden states and the observations is modelled using a probability distribution. The Hidden Markov Model (HMM) is the relationship between the hidden states and the observations using three sets of probabilities: initial probability (Pi), transition probability (A) and emission probability (B).

- The initial probabilities describe the probability of observing a state such as P(B-org)
- The transition probabilities describe the probability of transitioning from one hidden state to another such as P(B-org/O).
- The emission probabilities describe the probability of observing an output given a hidden state such as P(NN/B-org).

Implement the Hidden Markov Model (HMM) algorithm for evaluation problem using the following steps. Use first 47,700 sentences of Assignment9.csv for training and last 259 sentences for testing. (Dataset is attached in the classroom).

1. Define the state space and observation space.
2. Calculate the initial state distribution (Pi) directly from the dataset.
3. Calculate the state transition probabilities (A) directly from the dataset.
4. Calculate the state observation probabilities (B) directly from the dataset.
5. Calculate the probabilities of the test sentences.