In the name of GOD.

Stochastic Process

Fall 2023
Hamid R. Rabiee

Sharif University of Technology

Project	Markov Chains and HMM and Sampling	Deadline : 5 Bahman

1. **Text generation via Markov Chains**

   You are going to implement a simple text generator using Markov Chains. This will be a character based model that takes the previous k-character of the chain and generates the next letter in the sequence. You are going to use Hafez poems for train this text generator.

   (a) Explain how it can be modeled via Markov Chains. Bring an example for simplicity.

   (b) Complete the jupyter notebook

   (c) Generate text for k = {1,4,10, 20} and check the results. Use of the poems for the starting word. Analyze the results.

   (d) What are the disadvantages of this model? What can be done to improve the model?

2. **HMM**

   The diagram of the hidden states of an HMM is given in Figure 1. These hidden states represent the mood of the TA who is going to design the final exam questions. The difficulty of the questions is an integer number in the range of [0,5]. Therefore, there can be six observations for each hidden state.

   (a) Implement the Baum-Welch algorithm to find HMM parameters. Run the algorithm on the given training data available in the 'train.json' file. Store all parameters in a file with appropriate names. (Use the numbers in parentheses to indicate each state.)

   (b) Implement the forward-backward algorithm to calculate the log-probabilities of each sequence in the test data available in the 'test.json' file. Store the probabilities in a file.

   (c) Implement the Viterbi algorithm to find the most probable states in each test sample. Store the results in a file.

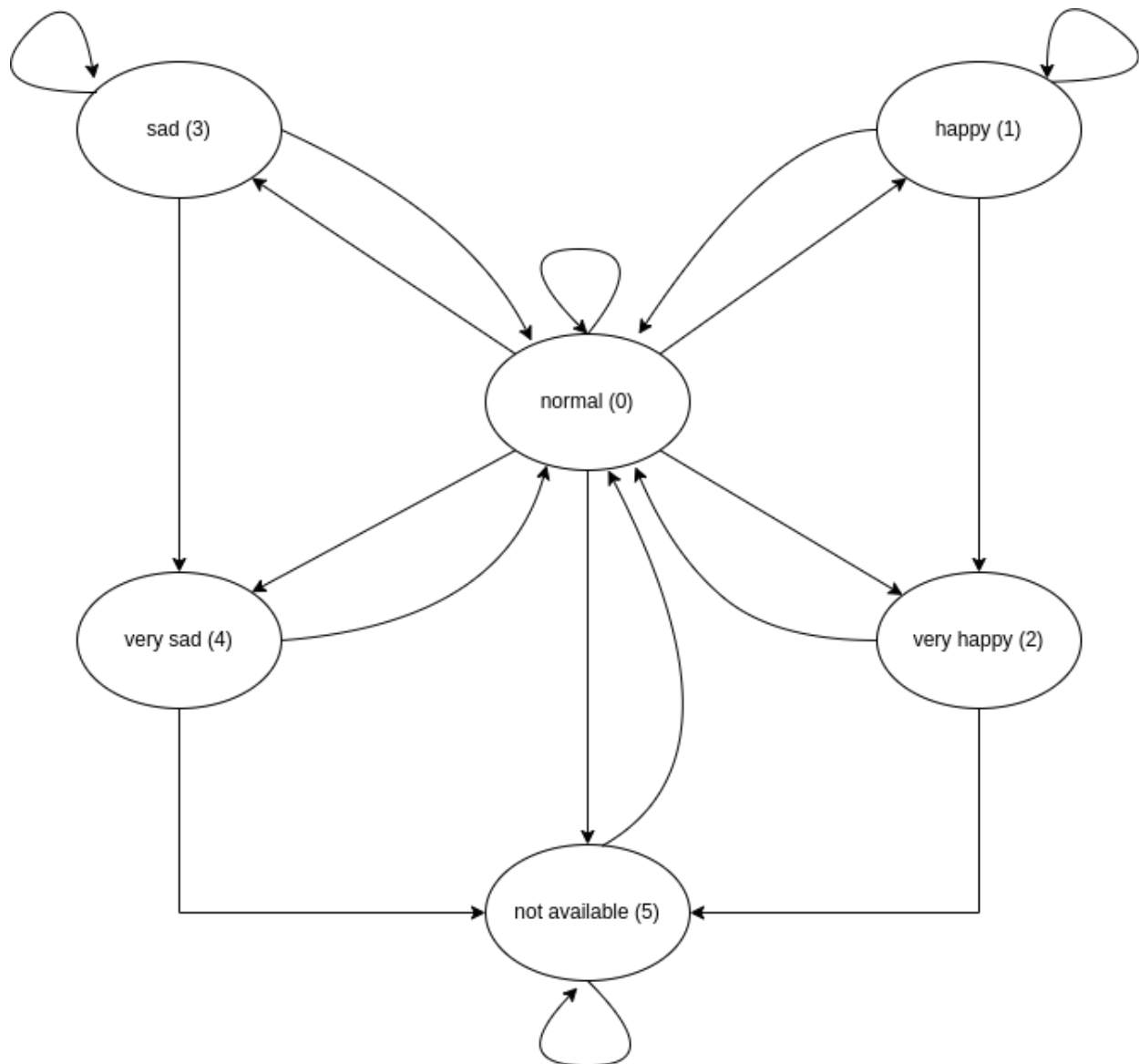   (d) Explain every challenge you encountered while implementing the algorithms. How did you solve them?

Figure 1: Hidden States of HMM