**Introduction**

For this homework assignment we were tasked with classifying several sequences and estimating the number of states in the processes, which were used to generate the sequences. I accomplished this by first training the HMM using the observations in data1 and data2 to solve for the parameters. I ran the training process several times using different state counts to find the state count which corresponded to the most optimal logarithmic likelihood values. Then I used these parameters to classify the datasets X1 – X6. I accomplished this by calculating the logarithmic likelihood of each observation sequence given the parameters from data1 and data2 and assigning the sequence to the observation that had the highest likelihood.

**Methods**

Below are functions that are critical to the implementation of the code:

1. [loglik, errors] = **log\_lik\_dhmm**(data, prior, transmat, obsmat)
   1. HMM toolbox function that computes the logarithmic likelihood of a dataset given the hmm parameters: initial state, transition, and observation probabilities.
2. [ll\_trace, prior, transmat, obsmat, iterNr] = **learn\_dhmm**(data, prior0, transmat0, obsmat0, ...)
   1. HMM toolbox function that finds the parameters of the Hidden Markov Model using Expectation-Maximization.
3. [T,Z] = **mk\_stochastic**(T)
   1. HMM toolbox function that ensures that the parameter is a stochastic matrix.
4. function [class, classProb] = **classify**(X, prior1, prior2, transmat1, transmat2, obsmat1, obsmat2)
   1. User function that classifies each of the observation sequences as either being generated using prior1, transmat1, and obsmat1 or prior2, transmat2, and obsmat2.
5. function [stateCount, prior, transmat, obsmat, llVal, LL, loglik] = **estParams**(data, obCount)
   1. User function that trains the parameters using the inputted model using Expectation-Maximation.
   2. Returns the number of states and the initial states, transition, and observation probabilities.
6. function **classResults**(class, classProb)
   1. User function that displays the classification results of the observation sequences.

**Results**



Figure 1. Parameters for data1



Figure 2. Parameters for data2

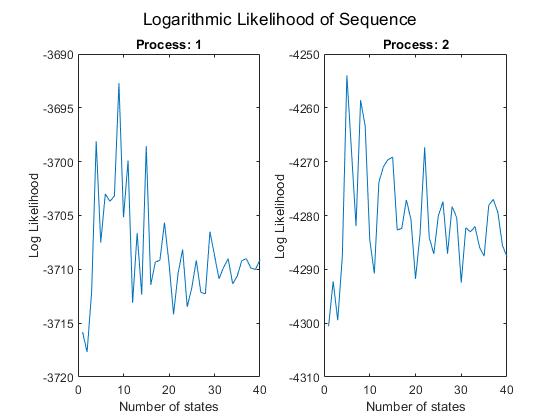


Figure 3. Log likelihood of sequence given number of states



Figure 4. Classification of sequences and log likelihood

**Discussion**

The first task in this homework assignment was to learn the parameters for each of the two Markov processes. This was done by iteratively running the dhmm\_em function, which uses Expectation-Maximization to estimate the parameters, with varying numbers of states. I then calculated the log likelihood for each of the parameters given the estimated parameters and chose the parameters which resulted in the highest likelihood. The parameters for the processes can be seen in figures 1 and 2, and the plot of the number of states versus the log likelihood of the sequence for each process can be seen in figure 3.

Then I used the calculated parameters for each process to classify the observation sequences in matrices X1-X6. This was done by calculating the log likelihood of each of the sequences given the two sets of parameters and classifying the sequence to have been generated using the parameters which resulted in the highest likelihood. This information can be seen in figure 4.

**Software listing and executable software**

Start the program by pressing “Run”.