COURSE 02457

Non-Linear Signal Processing: Answers to Exercise 12

Checkpoint 12.1

This checkpoint is designed to illustrate the concept of HMM learning. We have a transition and emission matrix that have generated some data, and we want to estimate them using EM for different numbers of hidden state variables.

The dimensions of the teacher transition matrix is 4×4 and 4×5 for the emission matrix. By inspecting the transition matrix, we expect that the Markov chain that it produces seldomly changes state, since the probability for staying in the same state is high, and the probability for emission i given hidden state i is also high.

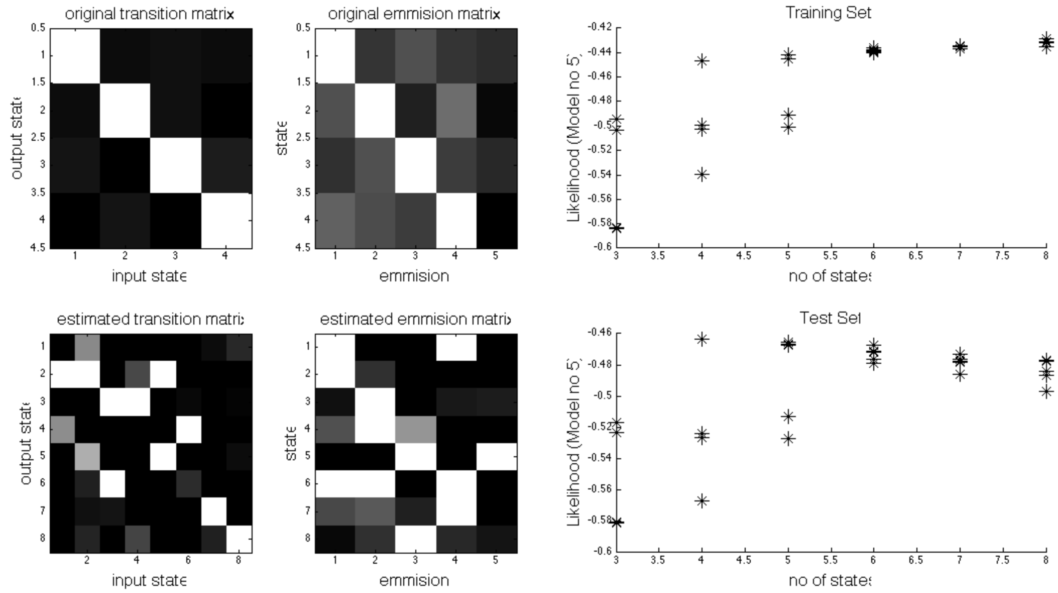
We also see that when the algorithm is run several times, it is not always the same number of hidden states that results in the maximum likelihood, so it is not consistent in predicting the optimal number of states.

For S number of hidden states and K number of emission states, the HMM algorithm will have to estimate $S(S-1) + K(S-1)$.

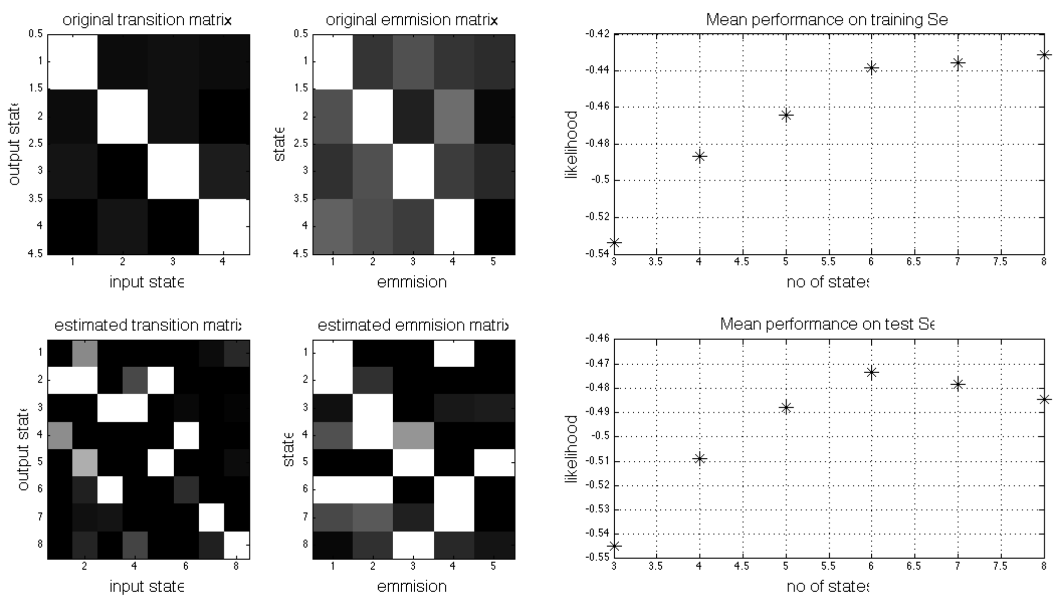
By inspecting the likelihood for the test set in the example in Fig. 1b, the optimal number of hidden states is 6. If we include more hidden states than this, the test likelihood decreases, which correspond to a higher test error. This seems reasonable, since including more hidden states is the same as increasing the model complexity, allowing the model to not only fit the true underlying model, but also the noise, leading to large variance in the generalization error.

Conversely, if the number of hidden states is limited, and may not be able to produce a good estimate, leading to a large bias in the generalization error.

It is also interesting to note that when the number of state variables in the student transition matrix is equal to that of the teacher matrix, the transition matrix always has the same behavior, but the emission matrix does not. This is because if the states are permuted (ordered differently), it has no effect on the transition matrix, but the mapping from the hidden states to the observables get mixed up, and the emission matrix is no longer the same.



(a) Top matrices: Teacher transition and emission matrices. Bottom Matrices: Student transmission and emission matrices. Right: Likelihood of the model for the test and training data for 5 runs.



(b) Top matrices: Teacher transition and emission matrices. Bottom Matrices: Student transmission and emission matrices. Right: Average of the likelihood of the model for the test and training data over 5 runs.

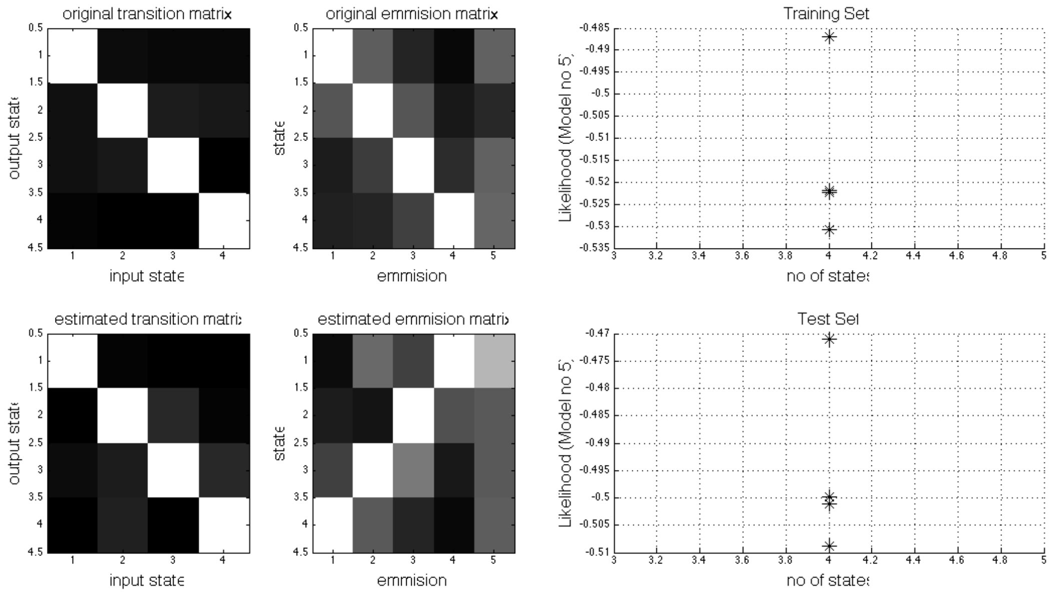


Figure 1: Illustration of that the transition and emission can not be directly compared.

Checkpoint 12.2

We are now going to combine a number of techniques that we have learned throughout the course to perform speech recognition on audio samples of the pronunciation of the numbers 0 through 9.

We have L versions of the pronunciation of each word, and we intend to extract the cepstral features and train a HMM model for each word, see flowchart.

In Figs. 3-4 we see that the number of hidden states is 5 and that the size of the **codebook** is 16.

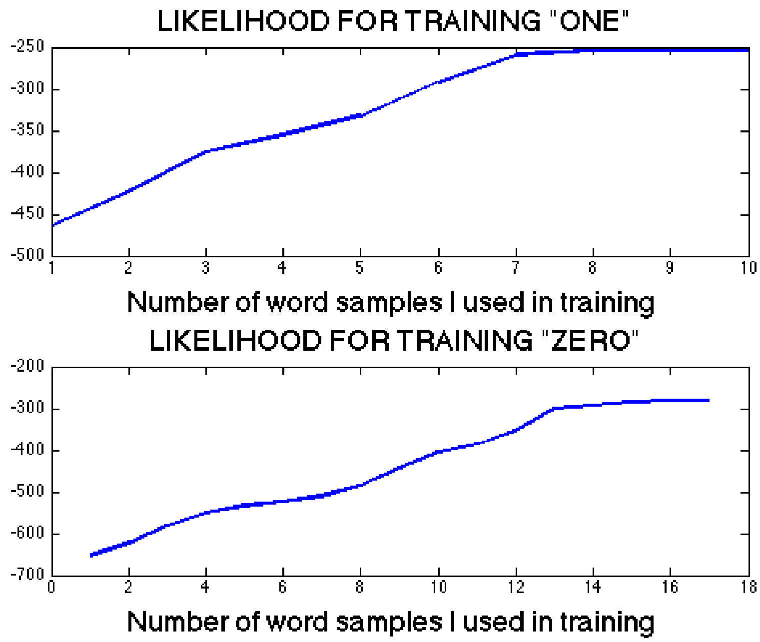


Figure 2: Example of the evolution of the likelihood during training of a HMM to the word ONE. It increases as expected.

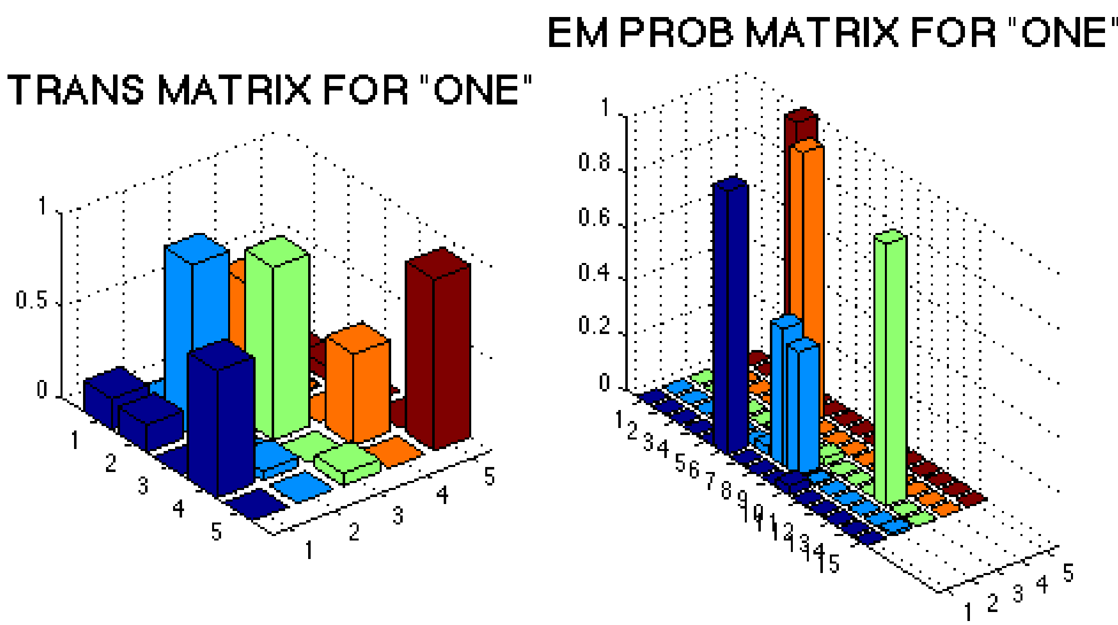


Figure 3: The trained HMM for the word $ZERO$.

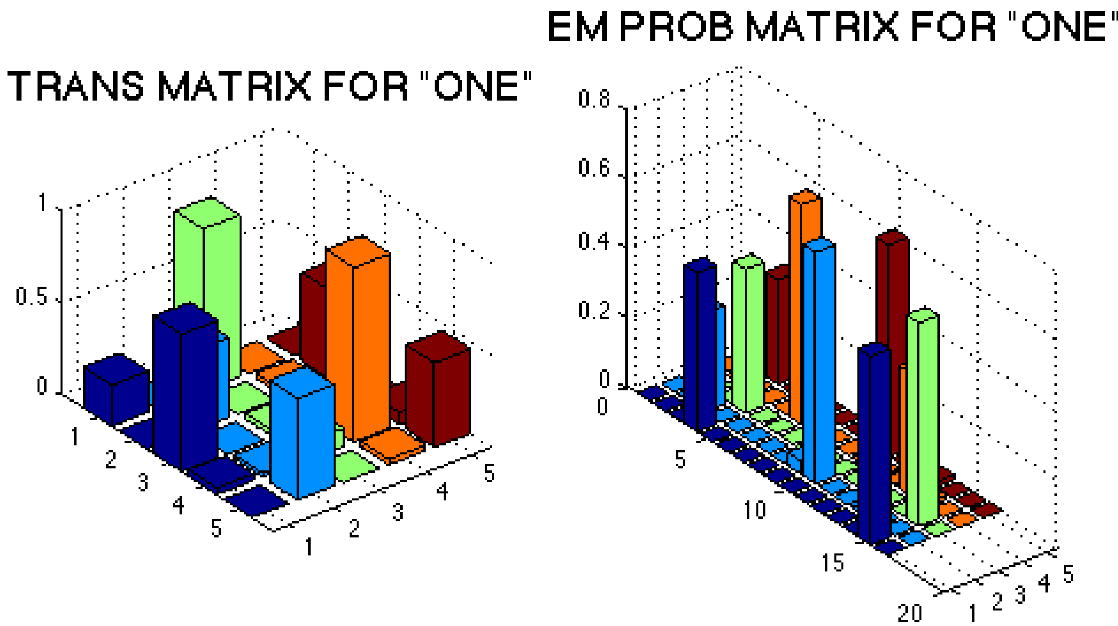


Figure 4: The trained HMM for the word ONE.

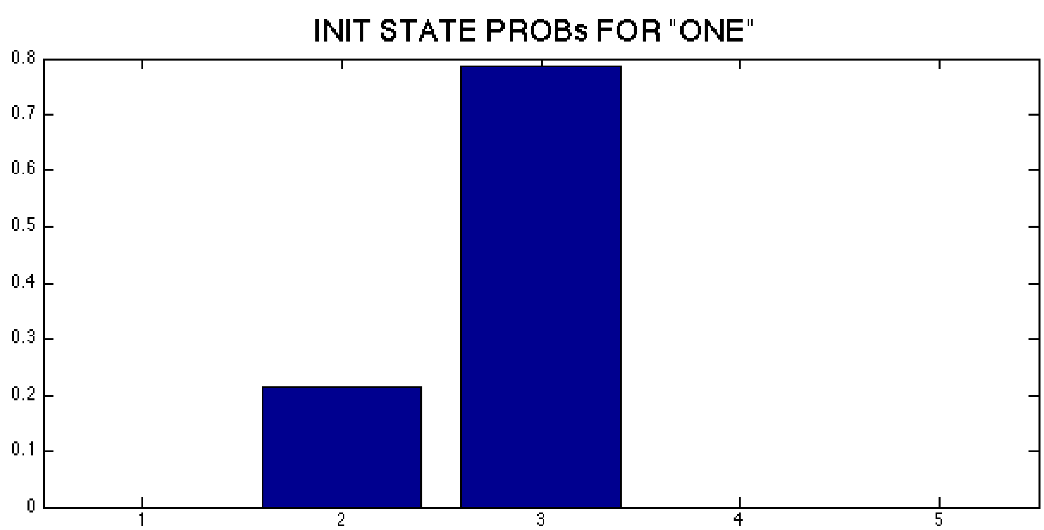
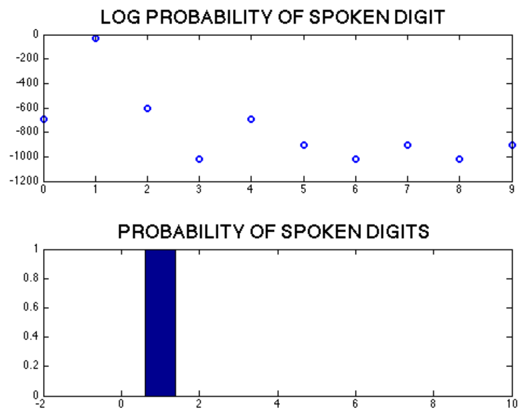
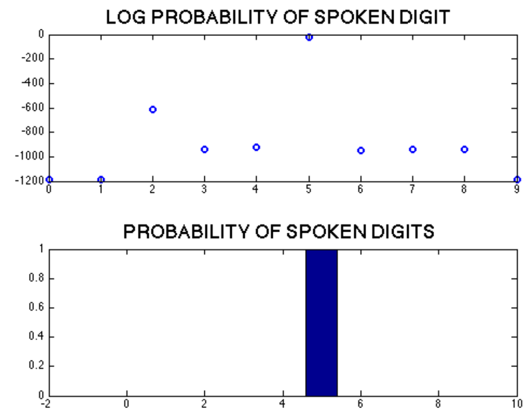


Figure 5: The estimated initial probability of the state vector $\pi_i(1)$ for the trained HMM in Fig. 4.

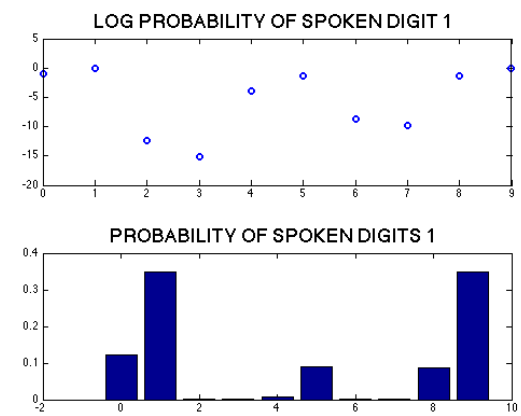


(a) Example of the posterior probability $p(y|I, M)$.

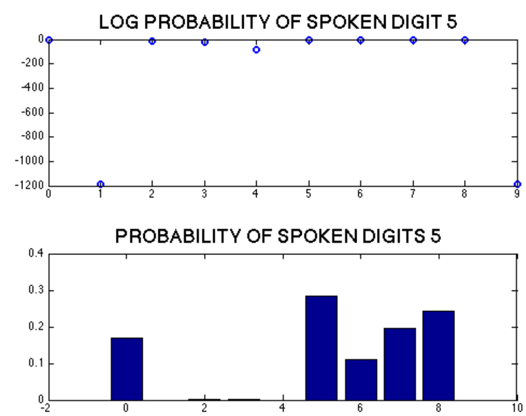


(b) Example of the posterior probability $p(y|I, M)$.

Figure 6



(a) Example of the posterior probability $p(y|I, M)$ with K=2 where the method almost fails.



(b) Example of the posterior probability $p(y|I, M)$ with K=2 where the method almost fails.

Figure 7