Pattern and Speech Recognition WS2015-16
Exercise 10

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

January 22, 2016

Hidden Markov Models

Simulation Algorithm

1 In MATLAB, it is a 'structure' with four fields, where
First field is **alphabet** which is a character array of size m.
Second field is **initial distribution** which is a vector of size n which is the number of states.
Third field is the **transition matrix** of size n x n
Last field is the **emission matrix** of size n x m
This structure allows for the easy access of the parameters.

2 See *random_select.m*

3 See *simulate.m*

4 Simulation-1:

> '1a' '2a' '3c' '1a' '2b' '3b' '3c' '1a' '2c' '1a' '2b' '3c' '2a' '3c' '3c' '1a' '1c' '2b' '1a' '1c'
> '2b' '1a' '2b' '3c' '3b' '3c' '1a' '2b' '1a' '1a' '3c' '1a' '2b' '3a' '1c' '2a' '3c' '3c' '3c' '3c'
> '3c' '3c' '1a' '2a' '3c' '1a' '2b' '3c' '3c' '3c' '2b' '2b' '1a' '3c' '1a' '1a' '3b' '1a' '2b' '1a'
> '2a' '3c' '2a' '3c' '2a' '3b' '1a' '1c' '1a' '1c' '2b' '1a' '1c' '2b' '1a' '1a' '2c' '3c' '3c' '1a'
> '2a' '1a' '1a' '2c' '3a' '3c' '1a' '1c' '1a' '2b' '2a' '1a' '2c' '1c' '2a' '3c' '3c' '3c' '3c' '3c'

Simulation-2:

> '1b' '1a' '1a' '1a' '2b' '3c' '1a' '3c' '3c' '3c' '3c' '1a' '3c' '3c' '3c' '3c' '3c' '2a' '1a' '1c'
> '1c' '1a' '1a' '1a' '1b' '2b' '1c' '1a' '2a' '3c' '2a' '1c' '2a' '1c' '2b' '1a' '3c' '3a' '3a' '1c'
> '2a' '1c' '2a' '1c' '2c' '3c' '3c' '3c' '1a' '1a' '2b' '3c' '3a' '3c' '3a' '3c' '2a' '3b' '1a' '2b'
> '3c' '1c' '2b' '1a' '1b' '1a' '1a' '1a' '2a' '1b' '2c' '1c' '2a' '1a' '1a' '2b' '1a' '2b' '1a' '1a'
> '2b' '3c' '3c' '2a' '3c' '2a' '1c' '2a' '2b' '1b' '2a' '1c' '2b' '1a' '2b' '3c' '3c' '3c' '1a' '2a'

Viterbi Algorithm

1 See *viterbi.m*

2 Done. The most likely sequence is,

> 1 2 3 3 2 3 3

And the probability of this state sequence is, *0.0015*

4 Done. See *simulate_viterbi.m*

We run simulate hmm for 100,000 times, then we consider only the sequences which are only the same as the observation sequence. And then, we take the sequence of states which are the most frequent for the given observation.

For larger hmms and bigger sequence of observations this would become too inefficient.