CO–495: Advanced Statistical Machine Learning & Pattern Recognition

Coursework #2: Hidden Markov Models

Exercise I

For this coding exercise, you should be implementing the EM algorithm to estimate the parameters for a discrete and continued valued Hidden Markov Model (HMM). You may work either on Matlab or on Python. In the `zip` file you will find a directory with the name `matlab` and one with the name `python`. All files are commented and you should be following the instructions provided there.

You **must** report your results on the data provided with the provided initialisation. **Courseworks that do not comply with the provided data and initialisations will not be assessed**. You are, however, more than welcome to report additional results on other data[1] and with other initialisations (e.g. more than two mixing distributions, random initialisation, etc.).

Important: The Jupyter notebook contains additional information (in the form of comments). Matlab users are encouraged to read it.

i) Given the provided data, initialisations and skeletons, fill in the functions that perform the EM algorithm to estimate the parameters for the discrete and continued valued HMM.

(70 marks)

ii) *Bonus*[2]: Given the provided data, initialisations and skeletons, fill in the function that performs Viterbi decoding.

(15 marks)

[1] Data generation function can be given on demand.

[2] Your maximum score will be 100 in total. However, if you attempt Viterbi decoding you could gain extra marks and thus still get perfect score even if you have minor mistakes elsewhere in your coursework.

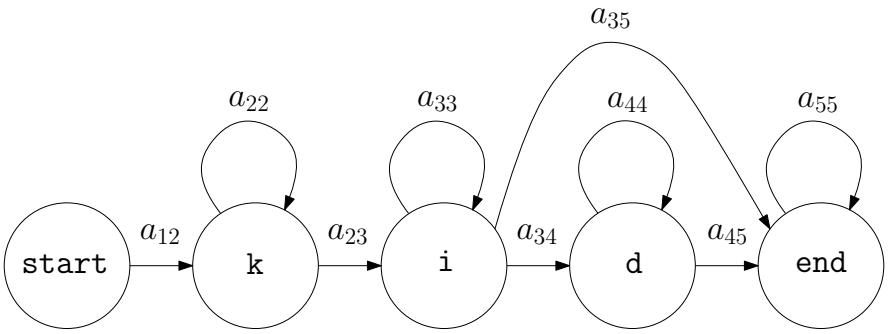

Figure 1: Stochastic automaton for the word "kid"

Exercise II

i) Assume the stochastic automaton given in Fig. 1, which represents a Markov Chain for the word "kid", consisting of phonemes `k`,`I`,`d`, as well as a starting and an ending state `start` and `end`, respectively. Assume that N strings of length T are given. Apply the Maximum Likelihood algorithm to estimate $\pi_1, \ldots, \pi_5$, as well as the transition probabilities a_{ij}. Add your derivations and results in a file named `answers.pdf`.

(15 marks)

ii) Assume an HMM with emission probabilities $p(\mathbf{x}|\mathbf{z})$ with discrete observations taking 5 values. Given a set of N sequences of length T of observations, devise the EM algorithm to find the parameters. Add your derivations and results in a file named `answers.pdf`.

(15 marks)