Middle East Technical University Department of Computer Engineering

CENG 499

Introduction to Machine Learning

Spring 2018-2019

Homework 3 - Hidden Markov Models

version 1

Due date: 22 05 2019, Wednesday, 23:59

1 Introduction

In this assignment, you are going to solve **evaluation** and **decoding** tasks of Hidden Markov Models(HMM). To solve them, you are going to implement **forward** and **viterbi** algorithms. The template of the functions are given in "hmm.py". No report is expected.

2 Data

Only filling the **forward** and **viterbi** functions is enough for this homework. Arguments will be directly fed into those functions. Therefore, there is no need to explicitly parse the input.

Although the outputs of the tasks are different their inputs are the same. The inputs are as mentioned in Table 1.

Input Symbol	**Input Name**	**Input Size**
A	State Transition Matrix	NxN
B	Observation Probability Matrix	NxM
pi	Initial State Probabilities	N
O	Observation Sequence	T

Table 1: Explanation of the input symbols in tasks. N, M, T represent number of states, number of possible observations, length of the observation sequence, respectively. ($2 \leq N \leq 10$, $2 \leq M \leq 10$, $1 \leq T \leq 30$)

To test your implementation, you can use "tester.py" which uses the example inputs and outputs in "data" folder. Final grading will not be done using only those inputs; therefore, passing the given examples may not mean you will get 100 points.

The data will be given in numpy arrays that consist np.float64 items. **In your implementations, please continue to use np.float64 precision.** A[i, j] is the state transition probability from state i to state j. B[i, j] is the probability of observing observation j in state i. pi[i] is the probability of initial state being i. O[t] is the t^{th} observation, which is an index between 0 and M-1(inclusive).

3 Evaluation Task

For this task, you are going to implement **forward** algorithm by filling the forward function in "hmm.py". The output should be the probability of an observation sequence given A(state transition matrix), B(observation probability matrix), pi(initial state probabilities). This can be done by calculating this probability for every possible state sequence and summing them up; however, this takes exponential time. Therefore, in this homework, **you are asked to implement the forward algorithm**, which runs in N^2T time. If your algorithm runs in exponential time, you won't be able to get full points.

4 Decoding Task

Similar to the evaluation task, you are going to implement **viterbi** algorithm by filling the viterbi function in "hmm.py". The output should be the **numpy array** of most likely observation sequence given A(state transition matrix), B(observation probability matrix), pi(initial state probabilities). This can be done by calculating this probability for every possible state sequence then selecting the maximum; however, this takes exponential time. Therefore, in this homework, **you are asked to implement the Viterbi algorithm**, which runs in N^2T time. If your algorithm runs in exponential time, you won't be able to get full point.

5 Specifications

- The codes must be in Python and use only numpy. You may also use python lists to implement the algorithms. Any other programming language or library will not be accepted. Python 3 is preferable but you are allowed to use Python 2 as well.
- Falsifying results is strictly forbidden and you will receive 0 if this is the case. Your programs will be examined to see if you have actually reached the results and if it is working correctly.
- You have total of 3 late days for **all** your homeworks. For each day you have submitted late, you will lose 10 points. The homeworks you submit late after your total late days have exceeded 3 will not be graded.
- Using any piece of code that is not your own is strictly forbidden and constitutes as cheating. This includes friends, previous homeworks, or the internet. The violators will be punished according to the department regulations.
- Follow the course page on ODTUCLASS or COW for any updates and clarifications. Please ask your questions on discussion section of ODTUCLASS or COW instead of e-mailing if the question does not contain code or solution.

6 Submission

Submission will be done via ODTUCLASS. If you do not have access to ODTUCLASS send your homework to this address "artun@ceng.metu.edu.tr" before the deadline. You will submit a zip file called "hw3.zip" that contains "hmm.py" and other source files(if you have used any).