Assignment 3: Hidden Markov Models
Machine Learning

Deadline: Sunday 25 Nov 2018, 21:00

Introduction

In this assignment, you will further deepen your understanding of Hidden Markov Models (HMMs). Please provide a latex based report in the PDF format. We provide you a sample latex project you might use for writing and generating your report. If latex is new to you, we recommend using overleaf.

Your report, code and all generated files must be archived in a file named firstname.lastname and uploaded to the iCorsi website before the deadline expires. Late submissions will result in 0 points.

Where to get help

We encourage you to use the tutorials to ask questions or to discuss exercises with other students. However, do not look at any report written by others or share your report with others. Violation of that rule will result in 0 points for all students involved. For further questions you can send email to *aleksandar@idsia.ch*.

Tasks

1. **(50 points)** In recent weeks the weather in Lugano was either sunny or rainy, 60% of the time rainy. The chance it being sunny tomorrow if today is sunny is 50%, and the chance of having a sunny day after a rainy day is 40%. On weekends Matteo chooses to do one of his favourite three activities: playing a game, chatting with friends or resting. His choice depends on the weather. If it is sunny, 40% of the time he plays a game, 30% chats with friends and 30% rests, otherwise, if it is rainy, 10% plays a game, 40% chats with friends, 50% rests. Sarah, Matteo's girlfriend who lives in Lausanne, knows about his habits. When they are not meeting on a weekend, she tries to guess the weather of Lugano based on what Matteo did from Friday to Sunday.
In this exercise, all calculations should be done manually in written form.

(a) **(10 points)** Show what Sarah's Hidden Markov Model of Matteo's weekend looks like, and provide all probabilities she needs to define the model. Note: you can draw the HMM by hand and include as graphics in your latex report.

(b) **(10 points)** What is (according to the HMM model) the probability of Matteo chatting with friends on Friday, playing a game on Saturday, and resting on Sunday? What is the "stupid way" of calculating this? What is the clever one?

(c) **(10 points)** What would be Sarah's guess about the weather in Lugano on Monday, if she knows that Matteo chatted with friends on Friday, played a game on Saturday, and rested on Sunday? Calculate the probabilities using the forward algorithm.

(d) **(20 points)** Use the Viterbi algorithm to manually calculate the most probable weather-sequence from Friday to Sunday given Matteo's weekend activities from (c).

2. **(40 points)** Implement the Viterbi algorithm in a programming language of your choice (we recommend python). Please submit your code and the output files archived together with the latex report. The output files need to be in the same format as the given example output file.
The HMM has three states S_1, S_2, S_3, and four observations W, X, Y, Z. In the output files, the states are labeled only by the numbers 1, 2, 3. The state transition probability matrix $\mathbf{A}$ and the observation probability matrix $\mathbf{B}$ are given as follows:

$$\mathbf{A} = \begin{pmatrix} 0 & 0.1 & 0.9 \\ 0.6 & 0.1 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.6 & 0.1 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.2 & 0 & 0.5 & 0.3 \end{pmatrix}$$

(a) **(5 Points)** Draw the state graph of the Markov chain. Note: you can draw it by hand and include as graphics in your latex report.

(b) **(5 Points)** Determine the total number of possible state sequences of length $T = 5$ for this particular HMM.

(c) **(10 points)** For a given probability distribution over the states at time t

$$\mathbf{p}_t = \begin{pmatrix} p_t(S_1) \\ p_t(S_2) \\ p_t(S_3) \end{pmatrix}$$

the distribution at the next time step is obtained by

$$\mathbf{p}_{t+1} = \mathbf{A}^T \mathbf{p}_t \tag{1}$$

Calculate the stationary distribution $\mathbf{p}_s$ defined by the equation

$$\mathbf{p}_s = \mathbf{A}^T \mathbf{p}_s \tag{2}$$

Note: You can do an approximate solution by iterating equation (1) with arbitrary starting distribution until it converges, or you can do an exact solution by solving equation (2) with the additional constraint that the sum of the probabilities must be 1.
Hint: $p_s(S_1) = 0.32352941$

(d) **(20 points)** Use the calculated stationary distribution as the initial distribution π_0. Run your Viterbi implementation on the given 5 input files. Use the test input and output to check if your algorithm is correct and then submit your output files (each correctly predicted file is worth 4pts).

3. **(10 points)** Consider the Baum-Welch re-estimation procedure of learning an HMM to model some observation sequence. The number of states of the HMM can in theory be chosen arbitrarily. How do you expect the number of states to affect the performance of the learned HMM? Consider the performance on the training data as well as on unseen test data.
