HONR 39900 – Homework 7

Justin A. Gould
gould29@purdue.edu

May 18, 2021

DUE DATE: 2021/11/01 23:59 EDT

Homework Instructions
To receive credit for the assignment, do the following:

1. Create an `.ipynb` file, and name it: `purduealias_honr39900_homework_number.ipynb` (e.g., `gould29_honr39900_homework_1.ipynb`)

2. Show all your work and follow the instructions below very carefully.

3. Please create your own Markov Chain Python utility from scratch. Use of any pre-built Markov Chain model (e.g., `sklearn`) will receive a 0 for both questions 1 and 2!

4. Submit a printout (e.g., as PDF) of your `.ipynb` file–and the file itself–to Brightspace by the due date.

5. You must show **all** your work and provide comments in your code explaining what you are doing.

6. Submit a PDF of your text answers to questions $3 - 8$.

For grading this assignment, I will not leverage unit tests. I will look at the printout of your `.ipynb` file. When in doubt, please show and comment all your work.

For questions 1 and 2, please create your own Markov Chain Python utility from scratch. Use of any pre-built Markov Chain model (e.g., sklearn) will receive a 0 for both questions 1 and 2!

Problem 1

Markov Chain #1: Weather – 16 points
The weather is West Lafayette varies between 3 possible states: sunny, rainy, or snowy. Use the following information to predict what the weather will be in 5 days if today is rainy:

- If the current day is sunny, there is a probability of 0.75 that the next day will be sunny, a probability of 0.15 that the next day will be rainy, and a probability of 0.1 that the next day will be snowy.
- If the current day is rainy, there is a probability of 0.24 that the next day will be sunny, a probability of 0.60 that the next day will be rainy, and a probability of 0.16 that the next day will be snowy.
- If the current day is snowy, there is a probability of 0.04 that the next day will be sunny, a probability of 0.15 that the next day will be rainy, and a probability of 0.84 that the next day will be snowy.

Input: The information stated above to create a transition matrix.

Desired Output: A string (`sunny`, `rainy`, or `snowy`) of the weather you are predicting.

Problem 2

Markov Chain #2: Coin Flip – 16 points
Suppose that $coin_1$ has a probability 0.7 of coming up heads if the previous flip is heads, and a probability of 0.4 of tails if the previous flip is tails. If $coin_1$ is flipped once per day, every day, and the first flip is heads, in what position will the coin be in 30 days?

Input: The information stated above to create a transition matrix.

Desired Output: A string (`head` or `tails`) of the coin side you are predicting.

For questions 3-8, determine if it is appropriate to use HMM? Provide a one-sentence explanation to your answer.

Problem 3

Stock market price data – 2 points
For example, using stock market price data to predict future prices.

Problem 4

Collaborative filtering on a database of movie reviews – 2 points
For example, Netflix challenge: predict about how much someone is going to enjoy a movie based on their and other users' movie preferences.

Problem 5

Daily precipitation data in Pittsburgh – 2 points
For example, using precipitation data to predict future rainfall.

Problem 6

Optical character recognition – 2 points
For example, multi-character recognition to determine if a word is `Tom` vs. `Tim`.

Problem 7

Situational #1 – 4 points
True or False: (if true, give a 1-sentence justification; if false, give a counterexample.)
When learning an HMM for a fixed set of observations, assume we do not know the true number of hidden states (which is often the case), we can always increase the training data likelihood by permitting more hidden states.

Problem 8

Situational #2 – 6 points
Show, in $1-3$ sentences, that if any elements of the parameters π (start probability) or A (transition probability) for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.