IoT Analytics- Project 5: HMMs

Wen-Han Hu(whu24)

Task 1: Data Set Generation

In the first task of project, we need to generate the transition matrix P and event matrix B by drawing random numbers with personal student id as seed. My student id is 200263453 so the seed for these random number is 200263453.

Since we have 4 states and 3 objects, we need to generate a 4x4 random matrix as our transition matrix P and 4x3 random matrix as our event matrix B. After that, we need to normalize each row to make sure the sum of each row is equal to 1. The following graph are the matrices after generation.

Transition Matrix P:

```
Gernerating p matrix...............
Checking each row's sum of prob. is equal to 1.
Sum of row[0] is 1.0
Sum of row[1] is 1.0
Sum of row[2] is 1.0
Sum of row[3] is 1.0
[[0.25691376 0.17040213 0.46721004 0.10547407]
 [0.10696801 0.16903803 0.48485879 0.23913518]
 [0.17139729 0.37094994 0.33560144 0.12205134]
 [0.08770787 0.29766955 0.01358607 0.60103652]]
```

Event Matrix B:

```
Gernerating b matrix...............
Checking each row's sum of prob. is equal to 1.
Sum of row[0] is 1.0
Sum of row[1] is 1.0
Sum of row[2] is 1.0
Sum of row[3] is 1.0
[[0.30500857 0.11587838 0.57911305]
 [0.30627907 0.59700434 0.09671658]
 [0.51807384 0.12990698 0.35201919]
 [0.3496476  0.35823218 0.29212022]]
```

We can observe that the most probable next state when state = 1 is state = 3 (p = 0.4672), the most probable next state when the most probable next state when state = 2 is also state = 3 (p = 0.4848), the most probable next state when state is 3 is state = 2 or 3 (p = 0.3709 and p = 0.3356), and lastly the most probable next state when state is 4 is state = 4 (p = 0.6010). On the other hand, the most probable observation when state = 1 is 3 (p = 0.5791), the most probable observation when state = 2 is 2 (p = 0.5970), the most probable observation when state = 3 is 1 (p = 0.5181), and lastly the observation is quite equally distributed when state = 4. (p = 0.3496, p = 0.3583, p = 0.2921).

We generate 1000 observations O and sequence of states Q using the method stated on the project description. The partial result will show on the following tasks but the primary purpose for this 1000 observations is to help us train the HMM model to estimate parameters on Task 4. Please refer to the source code for the detailed implementation.

Task 2: Estimate $p(O|\lambda)$

This task actually asks us to solve the Problem 1 of HMM model which is to calculate the probability that this sequence came from the given HMM. The given observation sequence O = 123312331233. In this task, we implement the **forward algorithm** to compute the probability of the observation sequence O. The forward and backward algorithm both are better performance than brute-force solution and same time complexity. The result as following.

```
The Orginal Observation Sequence O: [3, 3, 2, 3, 1, 2, 1, 2, 1, 1, 3, 3]
The probability p(O|λ) is 7.759208566964474e-07 with O: [1, 2, 3, 3, 1, 2, 3, 3, 1, 2, 3, 3]
```

The result of $p(O|\lambda) = 7.76 \times 10^{-7}$ is nearly 0 which means it is almost impossible to have this given observation sequence with our parameters.

The first line of result shows the first 12 observation we generate from the Task 1. The generated observations O = [3, 3, 2, 3, 1, 2, 1, 2, 1, 1, 3, 3] comparing to given observation O = [1, 2, 3, 3, 1, 2, 3, 3, 1, 2, 3, 3] have much difference. This helps to explain why we have such low probability result. Still, the primary fact is due to the probability we generated in Task 1.

Task 3: Estimate the Most Probable Sequence Q

This task asks us to solve the Problem 2 of HMM which is to find the most probable sequence of states that gives rise to a sequence of observations O = [1, 2, 3, 3, 1, 2, 3, 3, 1, 2, 3, 3]. In this task, we use the Viterbi algorithm because it is an optimization procedure comparing to the solution 1 which mentioned in the textbook section 12.5. The result as following.

```
The Orginal Sequence Q: [1, 3, 2, 3, 2, 2, 3, 2, 3, 2, 3, 3]
The Most Probable Sequence Q: [1, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4] with O: [1, 2, 3, 3, 1, 2, 3, 3, 1, 2, 3, 3]
```

The sequence of states Q we generated in Task 1 is [1, 3, 2, 3, 2, 2, 3, 2, 3, 2, 3, 3]. Someone may think the result is wrong because we should have 4 states [1, 2, 3, 4] as part of our sequence Q. However, it does make sense when we tract back to the transition matrix P in the Task 1. We can that both state 1 and 2 have high possibility to shift to state 3 and state 3 is possible to stay in state 3 or back to state 2. Hence, it is probable the sequence of states Q will back and forth between state 2 and 3. That actually matches our result of Q.

The result of most probable Q is [1, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]. Considered that the probabilities of different objects are very close in state = 4 and the given observation is relatively even comparing to our generated observation. Also, once the state shift to 4, it has high probability the state would stay at state = 4 (p = 0.6). Thus, we can also infer that is the most probable Q by the matrices we have. This result also helps us to explain why we have such low probability in Task 2. First, the choice to shift to 4 is relatively low. Second, the state = 4 would stay the same and it is not able to shift to the state with higher probability in desired observation of even matrix B.

Task 4: Train the HMM

This task asks us to solve the Problem 3 of HMM. The estimated parameters along with their corresponding *p*-values as following.

```
The estimated transition matrix P:
 [[0.24089929 0.2433524  0.25590455 0.25984376]
 [0.2338624  0.24164307 0.25549448 0.26900004]
 [0.23552676 0.24905907 0.2443094  0.27110476]
 [0.22807157 0.24670467 0.24436792 0.28085584]]

The estimated event matrix B:
 [[0.29599125 0.25341334 0.45059541]
 [0.28618943 0.36068252 0.35312805]
 [0.46970822 0.29850055 0.23179123]
 [0.41090241 0.48293571 0.10616188]]

The estimated start probability pi:
 [9.03856202e-01 9.45757724e-02 1.56734096e-03 6.84299128e-07]
```

```
p-value of transition matrix P: 0.9999964230241212
p-value of event matrix B: 0.9999536646387183
p-value of start probability pi: 0.9910621095427448
```

The estimated transition matrix P has more equally distributed state probability comparing to generated transition matrix. We can also observe that our event matrix B is very similar to generated event matrix B. For example, the state 2 is very similar to the state 4 in generated event matrix B and other three states have one probability of object is higher than other objects. For instances, state 1 observation object 3 is higher, state 3 observation object 1 is higher, state 4 observation. With the estimated results above, we can suggest that the generated observation O is relative equal amount on each object. Because the estimated matrix P and B tell us, the next state is not important, and each observation object has quite fair opportunity to be observed. In order to confirm that, we print out the number of observations on each object.

```
Number of observation object 1 is 368
Number of observation object 2 is 353
Number of observation object 3 is 279
```

The number of observations on each object quite much our prediction. The number of each object is within 250 to 350, that is why we draw the result from estimation and also the p-values on transition matrix P and event matrix B and start probability pi indicate that the actual parameters and estimated parameter are not significant difference. Thus, our estimation is quite good fit for our generated 1000 observations.