

HMM on Weather Prediction

Group Member:

Ziyuan Wang (wangzi@uw.edu)

Ye Jin (yjin2@uw.edu)

Program Description:

The main purpose for this program is to determine the weather (Rainy, Cloudy or Sunny day) by observing the previous days' temperature, humidity and Pressure.

Brief Description on our Technique:

Data Collection:

Collect weather data of Seattle from the website, <https://www.wunderground.com/>, including the date, temperature, humidity, pressure and weather.

Data Process:

Firstly, the transition probability matrix and emission probabilities are defined by our experience. The place we chose is Seattle, so the probability of rainy days should be much greater than the cloudy and sunny days.

Secondly, since our intuition was not that accurate, we decided to get the transition probability matrix and emission probabilities by process the data we have.

For the transition probability matrix, because we had the today's weather and next day's weather, the times of each transition states can be calculated and then divided the times of the condition state. In this way, we can get the Transition probability.

For example, firstly, the 'Rainy|Rainy', 'Cloudy|Rainy' and 'Sunny|Rainy' states appear 3, 4, 3 times each. 'x|y' means today's weather is 'y' and next day's weather is 'x'. The total times of Rainy days is $3 + 4 + 3 = 10$. Therefore, $P(Rainy|Rainy) = 3/10 = 0.3$. By this way, the transition probability matrix can be calculated.

For the emission probability, we firstly set a threshold to determine the three level of the temperature, humidity and pressure.

For temperature:

$$1 : [-\infty \quad 50.0] \quad 2 : [50.0 \quad 68.0] \quad 3 : [68.0 \quad +\infty]$$

For humidity:

$$1 : [-\infty \quad 39.0] \quad 2 : [39.0 \quad 66.0] \quad 3 : [66.0 \quad +\infty]$$

For pressure:

$$1 : [-\infty \quad 29.3] \quad 2 : [29.3 \quad 29.9] \quad 3 : [29.9 \quad +\infty]$$

By this way, the temperature, humidity and pressure are divided into three level separately. And then, the emission probabilities can be calculated by getting the times of the level 1, 2 or 3 under certain state and then dividing the times of that state.

For example, when the weather is rainy day, for the temperature, there are 4 times on level 1, 3 times on level 2, and 3 times on level 3. The total times of rainy days is 10. Therefore, $P(1|Rainy) = 4/10 = 0.4$. $P(1|Rainy)$ means when the weather is Rainy, the probability of level 1 temperature. Because we only considered one observed state, temperature, humidity or pressure, we did not define a character to represent them.

MMH Algorithm:

The Forward Algorithm and Viterbi Algorithm are all achieved in this project.

The reference we read provided the pseudo code for these two algorithm.

The Forward Algorithm:

function FORWARD(*observations* of len T , *state-graph* of len N) **returns** *forward-prob*

create a probability matrix *forward*[N, T]

for each state s **from** 1 **to** N **do** ; initialization step

$forward[s, 1] \leftarrow \pi_s * b_s(o_1)$

for each time step t **from** 2 **to** T **do** ; recursion step

for each state s **from** 1 **to** N **do**

$$forward[s, t] \leftarrow \sum_{s'=1}^N forward[s', t-1] * a_{s', s} * b_s(o_t)$$

$$forwardprob \leftarrow \sum_{s=1}^N forward[s, T] \quad ; \text{termination step}$$

return *forwardprob*

The Viterbi Algorithm:

function VITERBI(*observations* of len T , *state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix $viterbi[N, T]$

for each state s **from** 1 **to** N **do** ; initialization step

$viterbi[s, 1] \leftarrow \pi_s * b_s(o_1)$

$backpointer[s, 1] \leftarrow 0$

for each time step t **from** 2 **to** T **do** ; recursion step

for each state s **from** 1 **to** N **do**

$viterbi[s, t] \leftarrow \max_{s'=1}^N viterbi[s', t-1] * a_{s', s} * b_s(o_t)$

$backpointer[s, t] \leftarrow \operatorname{argmax}_{s'=1}^N viterbi[s', t-1] * a_{s', s} * b_s(o_t)$

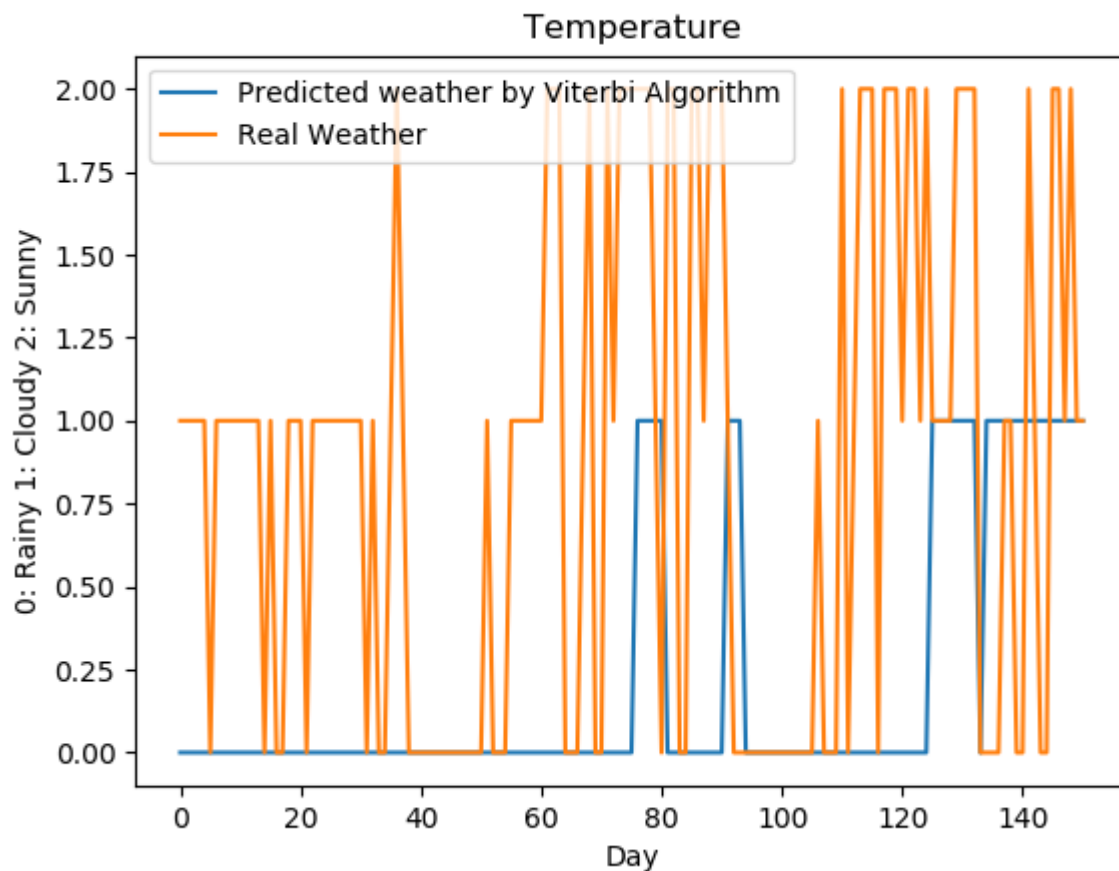
$bestpathprob \leftarrow \max_{s=1}^N viterbi[s, T]$; termination step

$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^N viterbi[s, T]$; termination step

$bestpath \leftarrow$ the path starting at state $bestpathpointer$, that follows $backpointer[]$ to states back in time

return $bestpath$, $bestpathprob$

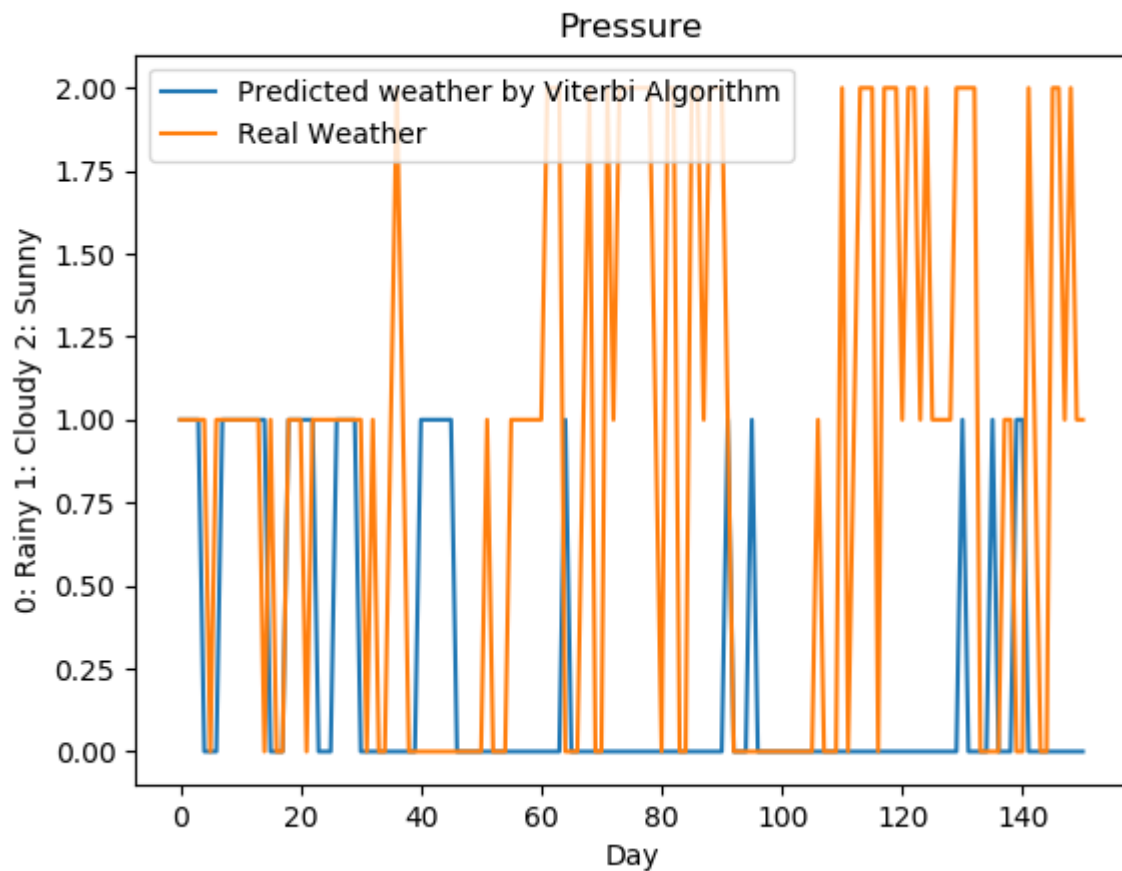
We programed these complex function and equation by matrix, so the code would be simple and be easy to understand.




```

-----Pressure-----
--Real Weather--
['Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Rainy',
', 'Cloudy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloud
y', 'Cloudy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Cloudy', 'Sunny', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Ra
iny', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy',
', 'Sunny', 'Sunny', 'Sunny', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Sunny', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Sunny', 'Cloudy', 'Sunny', 'Sunny', 'Sunny', 'Sunny', 'Sunny',
', 'Sunny', 'Sunny', 'Cloudy', 'Rainy', 'Sunny', 'Sunny', 'Rainy', 'Rainy', 'Sunny', 'Sunny', 'Cloudy', 'Sunny', 'Sunny', 'Sunny', 'Cloudy', 'Rainy
', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Sunny', 'Rainy', 'Cloudy', 'Sunny', 'Sunny', 'Sunny', 'Rainy', 'Sunny', 'Sunny', 'Sunny', 'Cloudy', 'Sunny', 'Sunny', 'Cloudy', 'Sunny
', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Sunny', 'Sunny', 'Sunny', 'Sunny', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Rainy', 'R
ainy', 'Sunny', 'Cloudy', 'Rainy', 'Rainy', 'Sunny', 'Sunny', 'Cloudy', 'Sunny', 'Cloudy', 'Cloudy']
--HMM Forward--
[2.23235832e-49 2.15621036e-49 2.10557611e-49]
--HMM Viterbi--
[2.23235832e-49 2.15621036e-49 2.10557611e-49]
--Predicted weather--
['Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy',
', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloudy', 'Cloud
y', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy',
', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Rainy', 'Rainy', 'Rainy', 'Cloudy', 'Cloudy', 'Rainy', 'Rainy', 'R
ainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy', 'Rainy']

```



For customized data,

Python main.py

Example Code

```
parser = argparse.ArgumentParser(description='Customized or Statistical Data')
parser.add_argument('--Customized', action='store_true')
args = parser.parse_args()
```

Using the parameters on the command to switch the mode, Customized or Statistical Data.

```
fig = plt.figure()
plt.title('Temperature')
plt.plot(timeDomain, Predict_State_Code, label='Predicted weather by Viterbi
Algorithm')
plt.plot(timeDomain, DATA.Code_realweather, label='Real weather')
plt.xlabel('Day')
plt.ylabel('0: Rainy 1: Cloudy 2: Sunny')
plt.legend()
```

Using the plot to show the result and compare with the real weather.

Learning in Project

Ziyuan Wang:

1. Learn how to program as a team.
2. Better understand the HMM.
3. Learn how to design a project and achieve the goal.

Ye Jin:

1. Learn the application of HMM.
2. Learn how to program as a team.

Future Design

1. Collect more accurate data.
2. Find a way to combine the three observation together to predict the weather.
3. For now, we just showed the result by plot (0: Rainy, 1: Cloudy, 2: Sunny). It is not that straightforward. Therefore, we are wondering if there is a better way to show it.
4. Design a GUI for this program.

Reference

Weather data: <https://www.wunderground.com/>

Algorithm reference: <https://web.stanford.edu/~jurafsky/slp3/A.pdf>

Partners' Reflections

Ziyuan Wang:

In this team, I was responsible for most of the coding work including the Forward algorithm, Viterbi Algorithm and Data Processing. During this experience, I found coding together is not simple, because it is hard to read and understand other's code. What we did was expressing our thought while coding, so it is easier to understand what the code means and follow the progress of developing the program. In this project, I am better understand how to do the project as a team and how to design a complete project.

Ye Jin:

In this team, I was responsible for part of coding work. Including the initial possibility setting and frame of the program, and collecting data from the internet. During this process, I found that, although large amount of data was on the Internet, it still needs to set a standard to collect the useful data. Also, I got more experience in teamwork of coding.