Weather forecasting with Markov chains

Math 42 Final Project
Department of Mathematics, UCLA

Junwon Choi
(UID: 605 317 213 / Email: cjunwon@g.ucla.edu)
Brian Chau
(UID: 109 548 540 / Email: bchau3648@g.ucla.edu)
Yifan Jiang
(UID: 706 037 292 / Email: jiangyifan0421@g.ucla.edu)
Mingyeong Kim
(UID: 905 974 280 / Email lunamkbruin@g.ucla.edu)
Nalin Chopra
(UID: 405 746 993 / Email: nalinchopra@g.ucla.edu)

December 7, 2022

Contents

Abstract

Weather forecasting typically relies on analyzing physical patterns and environmental conditions. This paper attempts to predict future weather conditions without relying on current physical/environmental conditions. This process was modeled purely based on the Markov chain and prediction results were coded and generated with Python. The Markov chain methodology was applied across a total of 20 years worth of training and testing weather data mainly in the campus of the University of California, Los Angeles. Predictions were tested across multiple subsections of our dataset. A simplified model categorized each day as 'Rain' and 'No Rain', and predictions using a simplified model produced an average accuracy of 77% across different datasets. A more complex model that categorized each day as one of six different weather conditions had predictions producing an average accuracy no greater than 35%.

All code and datasets used in this report can be found here: `https://github.com/cjunwon/Math_42_Final_Project`.

1 Problem description

Weather plays a crucial role in our lives, influencing clothing choices, travel plans, choice of transportation, etc. Naturally, weather forecasting is an essential tool that helps individuals make decisions on a daily basis. Hence most people are usually interested in short-term forecasts, which are predictions of the weather for up to a week in the future. [1] According to the American Geosciences Institute, "In earlier times, before the telegraph and the telephone were invented, weather observations from faraway places could not be collected in one place soon after they were made. In those times, the only way of predicting the weather was to use your local experience." [1] Analyzing weather patterns became feasible in the 1800's, with more advanced technology being introduced in the 1900's, allowing for more accurate weather forecasting methods.

Based on such information, we were interested in predicting the weather from the perspective of people before advanced technologies were developed.

This research paper attempts to forecast weather conditions purely through mathematical modeling that does not depend on the physical conditions of our current environment or any other tools developed for weather forecasting. The paper addresses the following questions:

- Which mathematical model is most appropriate for weather prediction?
- How can we fit the selected model to predict weather in a particular location?
- How accurate/significant is the model? How does it perform when fitted to different locations?

Thus we would need to develop and apply a mathematical model that only required past weather data to predict future weather conditions. The model could be tested by applying it to predict in the same location as the origin of the dataset and further validated by applying the same model to different locations that are distant from the initial dataset.

This method would simulate a similar thought process that humans have. For example, a Los Angeles native that is used to weather patterns of Southern California might lack necessary information and experience about weather patterns of New York, and thus their assumptions or predictions about New York's weather would be based on Southern California's patterns, not New York's.

2 Simplifications

2.1 Data collection and train/test Splitting

Weather data was collected from "https://www.visualcrossing.com" with location restricted to the University of California, Los Angeles campus (zip code - 90024) [2]. Weather data from the year 2000 and 2022 were downloaded.

We used two different time frame schemes to find the most efficient and accurate model of our data. The Markov chain discrete models were broken into two: one with only five years' worth of data and one with the entire data, that can assist in predicting if it is more accurate to predict using short term or long term datasets. We broke annual data into quarters (as defined by UCLA's quarter system), so we collected data for Winter, Spring, Summer, and Fall quarters, allowing us to cater more for each specific quarter. For example, winter would contain the months January, February, and March, which parallels UCLA's winter quarter. We can compare it with models built on seasonal data only vs annual data to see which is more accurate.

For a *long time frame* analysis, data collected daily from 2002 to 2017 were used as a training dataset to create the transition matrix that represents a Markov chain. Data collected from 2018 to 2021 were used as a testing dataset to fit the transition matrix obtained with the training data. Data from 2022 were omitted as it did not contain a full year's worth of data. Data was therefore split into 80% training data and 20% testing data.

For a *short time frame* analysis, data collected daily from 2017 to 2020 were used as a training dataset to create the transition matrix that represents a Markov chain. Data collected from 2021 was used as a testing dataset to fit the transition matrix obtained with the training data. Entries from 2022 were omitted as it did not contain a full year's worth of data. Data was therefore split into 80% training data and 20% testing data.

2.2 Analysis steps

The analysis of our problem was split into 6 main steps:

1. All seasons + Simplified (long time frame)

   This takes into account every day of each year, and categorizes each day as 'Rain' or 'No Rain'. A total of 20 years worth of training and testing data were used.

2. All seasons + Simplified (short time frame)

   This takes into account every day of each year, and categorizes each day as 'Rain' or 'No Rain'. A total of 5 years worth of training and testing data were used.

3. All seasons + 6 different weather conditions (long time frame)

   This takes into account every day of each year, and categorizes each day as 'Clear', 'Partially cloudy', 'Overcast', 'Rain', 'Rain, Partially cloudy', 'Rain, Overcast'. A total of 20 years worth of training and testing data were used.

4. All seasons + 6 different weather conditions (short time frame)

   This takes into account every day of each year, and categorizes each day as 'Clear', 'Partially cloudy', 'Overcast', 'Rain', 'Rain, Partially cloudy', 'Rain, Overcast'. A total of 5 years worth of training and testing data were used.

5. By season + 6 different weather conditions (long time frame)

   This separates each year's data into four different seasons, and categorizes each day as 'Clear', 'Partially cloudy', 'Overcast', 'Rain', 'Rain, Partially cloudy', 'Rain, Overcast'. A total of 20 years worth of training and testing data were used.

6. All seasons + Simplified applied to University of California, San Diego (UCSD)/New York University (NYU) (long time frame)

   This takes into account every day of each year, and categorizes each day as 'Rain' or 'No Rain'. A model created using the University of California, Los Angeles' (UCLA) weather data is applied to a nearby location (UCSD) and a more distant location (NYU) for validation purposes. A total of 20 years worth of training and testing data were used.

2.3 Assumptions and limitations

Our model will assume that the weather condition for a certain day depends on the weather conditions of the day before. The model is simplified by categorizing each day as one of the weather conditions mentioned above. In reality, weather is significantly more complex and has various factors affecting it, such as temperature, humidity, wind speed, etc. Therefore, solely using the probability of categorical conditions without taking into account other natural variables that affect weather conditions can be a limiting factor that hinders the accuracy of the model.

3 Mathematical Model

3.1 Explanation of Markov chains

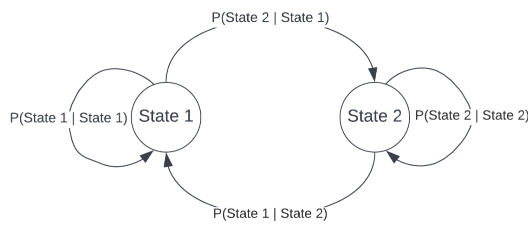

Figure 1: Visualization of General Markov Chain with Two States

A Markov chain is a mathematical system that changes from one state to another based on statistical criteria. The defining feature of a Markov chain is that the possible future states are fixed regardless of how the process got to its current state. In other words, the chance of transitioning to any particular condition is purely determined by the current state and the amount of time elapsed. As we can see from Figure 1, the probabilities from transitioning from one state to another are conditional probabilities that only depend on the information about the previous state. While Markov chains with any size of state space can be discussed, the initial theory and most implementations are focused on scenarios with a finite (or countably infinite) number of states [5].

There are three properties which identify a state model as being a Markov model. We considered these properties to build up our own Markov chain model and equations [3].

- The Markov assumption: the probability of one's moving from state i to state j is independent of what happened before moving to state j and of how one got to state i.
- Conservation: the sum of the probabilities out of a state must be one.
- The vector $X(t)$ is a probability distribution vector which describes the probability of the system's being in each of the states at time n.

One of the Markov chain properties that we focused on is that knowledge of the previous state is all that is necessary to determine the probability distribution of the current state [2]. We concentrate on finding the transition matrix of the Markov chain, whose entries are $P(s_i = X|s_{i-1} = Y)$, where s_i is the state of the chain at the current time step, s_{i-1} is the state at the previous time step, and X, Y are contained in the respective state space of the model. The transition matrix for a Markov chain is then a matrix of probabilities of moving from one state to another. Thus, if T is the transition matrix with n states,

$$T = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \dots & \dots & \ddots & \dots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

such that the rows would represent a state at time t and the columns would represent the state at time $t + 1$.

3.2 Application of Markov chains to weather forecasting

As we have different combinations of models, we have two different types of state spaces. We define the universal state space to be S = {Clear; Partially cloudy; Overcast; Rain; Rain, Partially cloudy; Rain, Overcast}, which is used to create the more complicated model. We also define the state space A = {Rain, No Rain}. This is the simplest version we use to predict the daily weather conditions. From the state space S, {Clear; Partially cloudy; Overcast} are conditions that fall under {No Rain} in state space A whereas the other three conditions are considered under {Rain} conditions. We utilize these state spaces to generate probabilities as well as the transition matrix and predict the sequence of states of future weather conditions.

Focusing on the simplest model with the state space A = {Rain, No Rain}, the transition matrix T would be of the form

$$\mathrm{T}_{\text{simplified}} = \begin{array}{r} \\ \text{Rain (R)} \\ \text{No Rain (NR)} \end{array} \begin{array}{c} \begin{array}{cc} \text{Rain (R)} & \text{No Rain (NR)} \end{array} \\ \begin{pmatrix} P(R|R) & P(NR|R) \\ P(R|NR) & P(NR|NR) \end{pmatrix} \end{array}$$

where P(A_i, A_j) is the conditional probability of the current condition being A_i given that the previous day had a condition of A_j and $A_i, A_j \in A$. In the transition matrix, for the rows i and columns j, $A_{i,j}$ represents the probability of the weather condition on day $t+1$ to be the weather condition assigned to column j, where row i is the weather assigned on day t.

Similarly, for the model with six conditions, the transition matrix would be

$$\mathrm{T}_{\text{six cond.}} = \begin{array}{c|cccccc} & \text{C} & \text{PC} & \text{OV} & \text{R} & \text{RPC} & \text{ROV} \\ \hline \text{C} & P(C|C) & P(PC|C) & P(OV|C) & P(R|C) & P(RPC|C) & P(ROV|C) \\ \text{PC} & P(C|PC) & P(PC|PC) & P(OV|PC) & P(R|PC) & P(RPC|PC) & P(ROV|PC) \\ \text{OV} & P(C|OV) & P(PC|OV) & P(OV|OV) & P(R|OV) & P(RPC|OV) & P(ROV|OV) \\ \text{R} & P(C|R) & P(PC|R) & P(OV|R) & P(R|R) & P(RPC|R) & P(ROV|R) \\ \text{RPC} & P(C|RPC) & P(PC|RPC) & P(OV|RPC) & P(R|RPC) & P(RPC|RPC) & P(ROV|RPC) \\ \text{ROV} & P(C|ROV) & P(PC|ROV) & P(OV|ROV) & P(R|ROV) & P(RPC|ROV) & P(ROV|ROV) \end{array}$$

Where: Clear - (C), Partially Cloudy - (PC), Overcast - (OV), Rain - (R), Rain, Partially Cloudy - (RPC), Rain, Overcast - (ROV)

4 Solution of Mathematical Problem

To create our Markov chain model, the parameters inside the transition matrix were first calculated. The parameters inside the transition matrix need to be the probabilities of a particular weather condition shall succeed a particular weather condition the following day. For this process, we utilize the standard method of calculating conditional probabilities for possible events. We counted the number of times a state on one day would succeed a state on the next day, the states representing all the combinations of weather conditions. Then these counts were organized in a matrix. After normalizing the rows so that the sum of each row would be equal to one, we get our transition matrix.

Following this process, we get the following matrices for our models:

1. All seasons + Simplified (long time frame):

$$\mathrm{T} = \begin{array}{r} \\ \text{Rain (R)} \\ \text{No Rain (NR)} \end{array} \begin{array}{c} \begin{array}{cc} \text{Rain (R)} & \text{No Rain (NR)} \end{array} \\ \begin{pmatrix} 0.467 & 0.533 \\ 0.096 & 0.904 \end{pmatrix} \end{array}$$

2. All seasons + Simplified (short time frame):

$$\mathrm{T} = \begin{array}{r} \\ \text{Rain (R)} \\ \text{No Rain (NR)} \end{array} \begin{array}{c} \begin{array}{cc} \text{Rain (R)} & \text{No Rain (NR)} \end{array} \\ \begin{pmatrix} 0.495 & 0.505 \\ 0.085 & 0.915 \end{pmatrix} \end{array}$$

3. All seasons + 6 different weather conditions (long time frame):

$$T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array} \begin{array}{c} \begin{array}{cccccc} C & PC & OV & R & RPC & ROV \end{array} \\ \begin{pmatrix} 0.690 & 0.199 & 0.001 & 0.034 & 0.073 & 0.004 \\ 0.250 & 0.640 & 0.028 & 0.009 & 0.067 & 0.006 \\ 0.028 & 0.743 & 0.156 & 0.0 & 0.046 & 0.028 \\ 0.366 & 0.214 & 0.0 & 0.145 & 0.248 & 0.028 \\ 0.190 & 0.323 & 0.0242 & 0.029 & 0.335 & 0.098 \\ 0.055 & 0.339 & 0.0630 & 0.008 & 0.260 & 0.276 \end{pmatrix} \end{array}$$

4. All seasons + 6 different weather conditions (short time frame):

$$T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array} \begin{array}{c} \begin{array}{cccccc} C & PC & OV & R & RPC & ROV \end{array} \\ \begin{pmatrix} 0.734 & 0.173 & 0.0 & 0.013 & 0.079 & 0.0 \\ 0.249 & 0.636 & 0.037 & 0.004 & 0.064 & 0.011 \\ 0.0 & 0.765 & 0.206 & 0.0 & 0.029 & 0.0 \\ 0.5 & 0.056 & 0.0 & 0.222 & 0.222 & 0.0 \\ 0.183 & 0.254 & 0.0296 & 0.018 & 0.420 & 0.095 \\ 0.087 & 0.565 & 0.087 & 0. & 0.217 & 0.043 \end{pmatrix} \end{array}$$

5. By season + 6 different weather conditions (long time frame):

- Fall:

$$T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array} \begin{array}{c} \begin{array}{cccccc} C & PC & OV & R & RPC & ROV \end{array} \\ \begin{pmatrix} 0.713 & 0.139 & 0.0 & 0.048 & 0.090 & 0.010 \\ 0.346 & 0.488 & 0.016 & 0.026 & 0.111 & 0.013 \\ 0.0 & 0.588 & 0.176 & 0.0 & 0.176 & 0.059 \\ 0.288 & 0.205 & 0.0 & 0.233 & 0.247 & 0.027 \\ 0.213 & 0.269 & 0.028 & 0.046 & 0.347 & 0.097 \\ 0.107 & 0.268 & 0.036 & 0.018 & 0.214 & 0.357 \end{pmatrix} \end{array}$$

- Winter:

$$T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array} \begin{array}{c} \begin{array}{cccccc} C & PC & OV & R & RPC & ROV \end{array} \\ \begin{pmatrix} 0.677 & 0.158 & 0.0 & 0.043 & 0.116 & 0.002 \\ 0.343 & 0.491 & 0.019 & 0.009 & 0.123 & 0.012 \\ 0.076 & 0.846 & 0.076 & 0.0 & 0.0 & 0.0 \\ 0.375 & 0.15 & 0.0 & 0.05 & 0.375 & 0.05 \\ 0.210 & 0.261 & 0.012 & 0.019 & 0.377 & 0.121 \\ 0.038 & 0.396 & 0.019 & 0.0 & 0.302 & 0.245 \end{pmatrix} \end{array}$$

- Spring:

$$T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array} \begin{array}{c} \begin{array}{cccccc} C & PC & OV & R & RPC & ROV \end{array} \\ \begin{pmatrix} 0.649 & 0.272 & 0.004 & 0.016 & 0.057 & 0.0 \\ 0.196 & 0.693 & 0.044 & 0.002 & 0.056 & 0.006 \\ 0.017 & 0.736 & 0.192 & 0.0 & 0.035 & 0.017 \\ 0.454 & 0.363 & 0.0 & 0.0 & 0.181 & 0.0 \\ 0.115 & 0.519 & 0.038 & 0.009 & 0.240 & 0.076 \\ 0.0 & 0.375 & 0.312 & 0.0 & 0.187 & 0.125 \end{pmatrix} \end{array}$$

- Summer:

$$
T = \begin{array}{c} \\ C \\ PC \\ OV \\ R \\ RPC \\ ROV \end{array}
\begin{array}{c}
\begin{array}{cccccc} \text{C} & \text{PC} & \text{OV} & \text{R} & \text{RPC} & \text{ROV} \end{array} \\
\begin{pmatrix}
0.693 & 0.262 & 0.001 & 0.021 & 0.021 & 0.0 \\
0.211 & 0.734 & 0.021 & 0.006 & 0.025 & 0.0 \\
0.136 & 0.727 & 0.090 & 0.0 & 0.0 & 0.045 \\
0.619 & 0.285 & 0.0 & 0.047 & 0.047 & 0.0 \\
0.162 & 0.534 & 0.046 & 0.046 & 0.186 & 0.023 \\
0.0 & 0.5 & 0.0 & 0.0 & 0.5 & 0.0
\end{pmatrix}
\end{array}
$$

6. All seasons + Simplified applied to UCSD/NYU (long time frame):

- UCSD:

$$
T = \begin{array}{c} \\ \text{Rain (R)} \\ \text{No Rain (NR)} \end{array}
\begin{array}{c}
\begin{array}{cc} \text{Rain (R)} & \text{No Rain (NR)} \end{array} \\
\begin{pmatrix} 0.467 & 0.532 \\ 0.095 & 0.904 \end{pmatrix}
\end{array}
$$

- NYU:

$$
T = \begin{array}{c} \\ \text{Rain (R)} \\ \text{No Rain (NR)} \end{array}
\begin{array}{c}
\begin{array}{cc} \text{Rain (R)} & \text{No Rain (NR)} \end{array} \\
\begin{pmatrix} 0.467 & 0.532 \\ 0.095 & 0.904 \end{pmatrix}
\end{array}
$$

A visual representation of these transition matrices can be found in the appendix 8.

Alongside this matrix, an initial condition is needed. We take the first day of the testing data and make a column vector whose entries are all 0 except for the entry that corresponds to the weather outlook condition of the first day, which would be equal to 1. The reasoning behind this is that given the information about the first day (the initial condition), we know that the probability of it happening will be 100%. This will allow our model to be specific to the test data, as opposed to making a column vector of the proportion of days that are of each condition.

With the transition matrix and the initial condition, we are now able to make predictions. Since we know the condition of the first day, we take the row of the transition matrix corresponding to that condition. We then use the probabilities in that row as weights to randomly select the condition of the next day. This process can be repeated for any number of days.

For example, with an all season simplified model (long time frame):
Let Day 1 of the model be a 'No Rain' day, which would give a vector $V = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We take transition matrix $T = \begin{pmatrix} 0.467 & 0.533 \\ 0.096 & 0.904 \end{pmatrix}$.

We multiply V and the transpose of T using a Python function. This outputs a vector with two elements containing probabilities of 'Rain' or 'No Rain' given the past day is 'No Rain'. Since the Markov chain model's key feature is randomness, we use a Python function to use the two probabilities of the transposed second row (which add to 1.0) as inputs to make a random choice. This choice will determine whether Day 2 will be 'Rain' or 'No Rain'. It is safe to assume that the random function has a higher probability of choosing 'No Rain' given the previous day was 'No Rain', since that has a probability of 90.4%. We repeat this step by a desired amount.

The prediction outputs would be appended to the testing dataset containing actual values to later calculate the accuracy of the predictions.

As an example, the first five entries of the 'All Seasons 6 Weather Conditions (long time frame)' and its accuracy at one iteration is shown below:

```
   index    datetime         condition predicted_condition
0   6575  2018-01-01             clear               clear
1   6576  2018-01-02             clear               clear
2   6577  2018-01-03             clear               clear
3   6578  2018-01-04  partially_cloudy    partially_cloudy
4   6579  2018-01-05  partially_cloudy                rain
0.3750855578370979
```

Figure 2: First five entries of a predictive iteration for 'All Seasons 6 Weather Conditions' and its accuracy

This is different from the algorithm used in class, which repeatedly multiplied the transition matrix by the column vector to get a distribution of the population. Because we are trying to find an exact solution for each day, and not a distribution or probability vector, we decided on the method detailed above. In addition, due to the long testing period, the algorithm used in class would tend towards a stationary distribution. Our method ensures that this does not happen and allows our model to be more dynamic, which would fit the problem better since weather is constantly changing.

5 Results

In order to compare the models against each other, we needed to quantify the accuracy of our model. We considered statistical tests, such as the chi-squared test, as well as other options, like calculating the distribution of days with different weather outlooks predicted by the model and seeing how much the distribution differs from reality. However, this only took into account the proportion of the types of conditions and ignored their order. We believe that the sequential arrangement of days is an important aspect of the Markov chain. As a result, we decided to directly calculate the proportion of days where the results from our model matched the test data, although the other approaches could possibly have lead to higher accuracy. It is important to note that because our process involved some randomness, we generated 100 predictions and averaged their accuracy.

We list out the accuracy for all 10 of our models

1. All seasons + Simplified (long time frame): 77%
2. All seasons + Simplified (short time frame): 79%
3. All seasons + 6 different weather conditions (long time frame): 37%
4. All seasons + 6 different weather conditions (short time frame): 39%
5. By season + 6 different weather conditions (long time frame)
   - Fall: 37%
   - Winter: 35%
   - Spring: 40%
   - Summer: 45%
6. All seasons + Simplified applied to UCSD/NYU (long time frame)
   - UCSD: 73%
   - NYU: 43%

Concentrating on models numbered one to four, it is clear that the simplified model performed much better than the model with six conditions and would be more useful for applications to real world scenarios. The low accuracy of the more complicated model is expected, considering the simplicity of the Markov chain. But, it is unexpected that the models trained on a shorter time frame had about a 2% higher accuracy than the ones trained on a longer period. This could be due to the fact that the models trained on a shorter time frame only considered years that were closer to the testing data. On the other hand, the models trained with a longer time frame included data that was irrelevant to the testing data. As a result, the shorter time frame could feasibly be more accurate. This highlights one of our model's advantages – our model does not need an extensive amount of data to be accurate, at least for the simplified one.

Now we focus on model number five, where we make a distinction between seasons. From the results, we can see that most of the uncertainty from the model lies in the fall and winter seasons. This is reasonable because according to the New York Times, California "typically gets 75 percent of its annual rainfall between November and March" [4]. As a result, it is to be expected that fall and winter would have more of a mix between rainy and clear days while spring and summer would have more consistent weather.

For model number six, we apply our simplified model that spans all seasons (model number 1) to UCSD and NYU. Although we get slightly worse performance at UCSD, there is a drastic decrease in accuracy with NYU. This is expected because California has a relatively temperate climate and the weather would not vary widely when comparing UCLA's weather to that of UCSD. However, NYU is on the other side of the coast and also includes snowy days. Consequently, we get significantly worse accuracy, which explains our results.

6 Improvement

6.1 Implementation of time-inhomogeneous Markov chains

The transition matrix of our model stays constant throughout the whole process (i.e. at each time step, the same initial transition matrix is applied to the current step). A possible improvement is to use time-inhomogeneous Markov Chains. Its definition is broader, as it allows for non-stationary transition probabilities, that is, as time goes on, the probability of moving from one state to another may change. With time-inhomogeneous Markov chains, the transition matrix is generated again at each time step. Since this method adjusts the probabilities at each timestep, we assume that it provides a more accurate result.

6.2 Extending time frame of analysis with second-order Markov chains

To improve the accuracy of our model, we attempted to implement a second-order Markov chain for both the simplified and six conditions models, which would use the past two states to predict the next state instead of one past state. However, during this implementation, we realized that there are still limitations that prevent us from doing so. For example, as we were trying to generate the transition matrix, we would get a row of all zeros because there aren't any observations where the two previous states were 'clear' $\rightarrow$ 'overcast'. Therefore, extending the time frame of our analysis when attempting to use the second-order Markov chains may resolve the problem that we faced, as it will have more instances that have all pairs of weather conditions.

7 Conclusions

7.1 Project achievements

Through this project, we were able to make a prediction model for weather that is independent of climate measurements and only relies on the condition of the previous day. By utilizing Markov chains, we were able to predict years worth of weather conditions with relatively high accuracy for a simplified weather model.
Additionally, we successfully expanded our model's scope by applying the transition matrices derived from UCLA-based data to predicting weather at UCSD (about 120 miles away from Los Angeles) with a relatively close accuracy to UCLA.

7.2 Project takeaways

This project highlighted a real world application of the Markov chain model. We learned that Markov chains are able to reflect our common thinking process in a more accurate and systematic manner.

References

[1] How has weather forecasting changed over the past two hundred years? *The American Geosciences Institute*, 2022.

[2] Visual Crossing Corporation, 2022.

[3] Randall Swift Douglas D. Mooney. *A Course in Mathematical Modeling*, page 122. The Mathematical Association of America, 1999.

[4] Soumya Karlamangla. California is expected to enter a fourth straight year of drought. *The New York Times*, October.

[5] Henry Maltby, et al. Markov chains. *Brilliant.org*, December 2022.

8 Appendix

8.1 Visual Representation of Transition Matrices

1. All seasons + Simplified (long time frame):

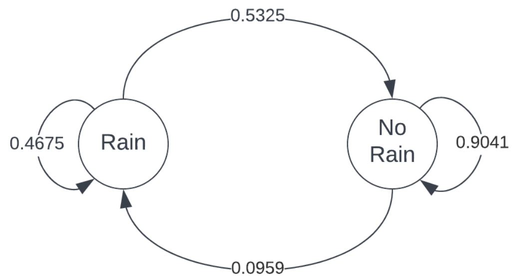

Figure 3: Visualization of Simplified Markov Model Trained Over Long Time Frame with All Seasons

2. All seasons + Simplified (short time frame):

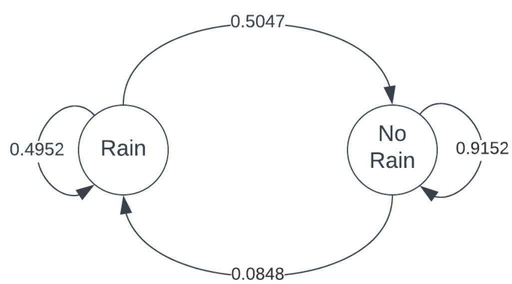

Figure 4: Visualization of Simplified Markov Model Trained Over Short Time Frame with All Seasons

3. All seasons + 6 different weather conditions (long time frame):

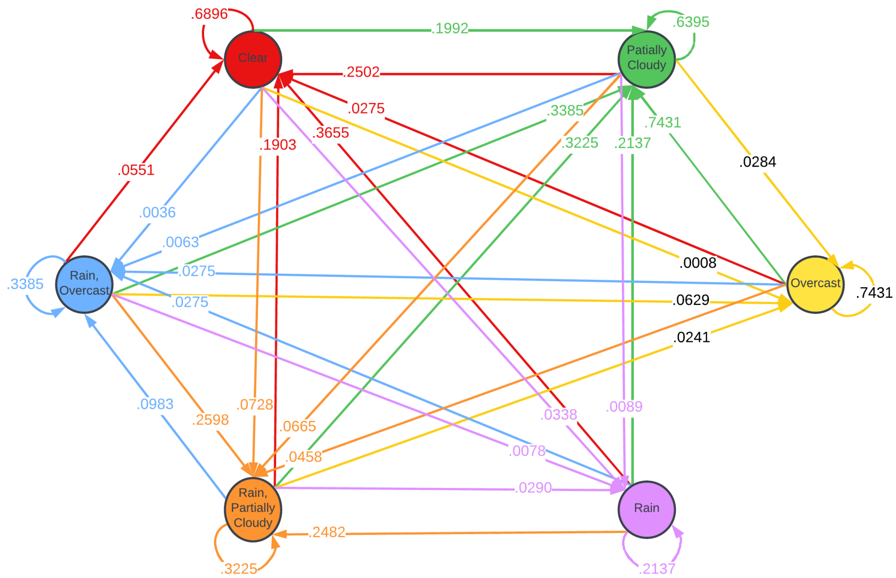

Figure 5: Visualization of Six States Markov Model Trained Over Long Time Frame with All Seasons

4. All seasons + 6 different weather conditions (short time frame):

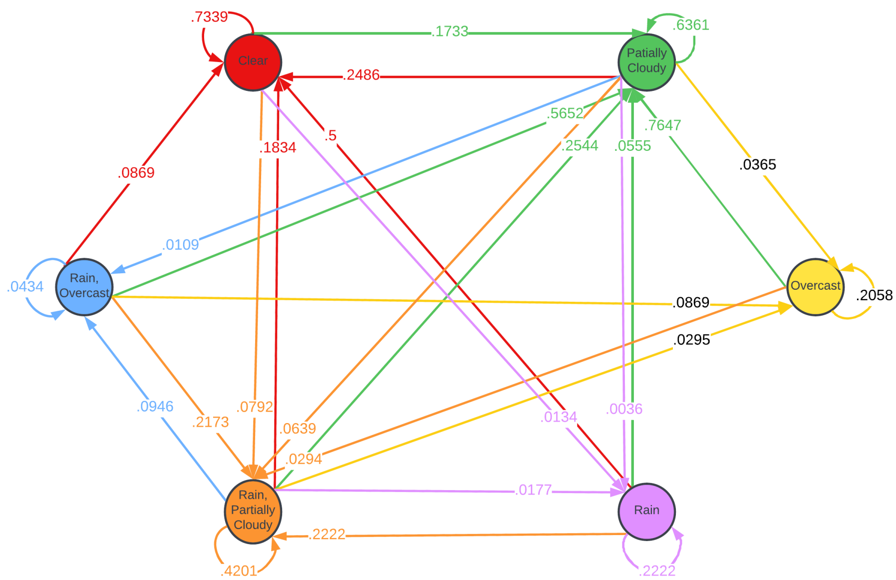

Figure 6: Visualization of Six States Markov Model Trained Over Short Time Frame with All Seasons

5. By season + 6 different weather conditions (long time frame)

- Fall:

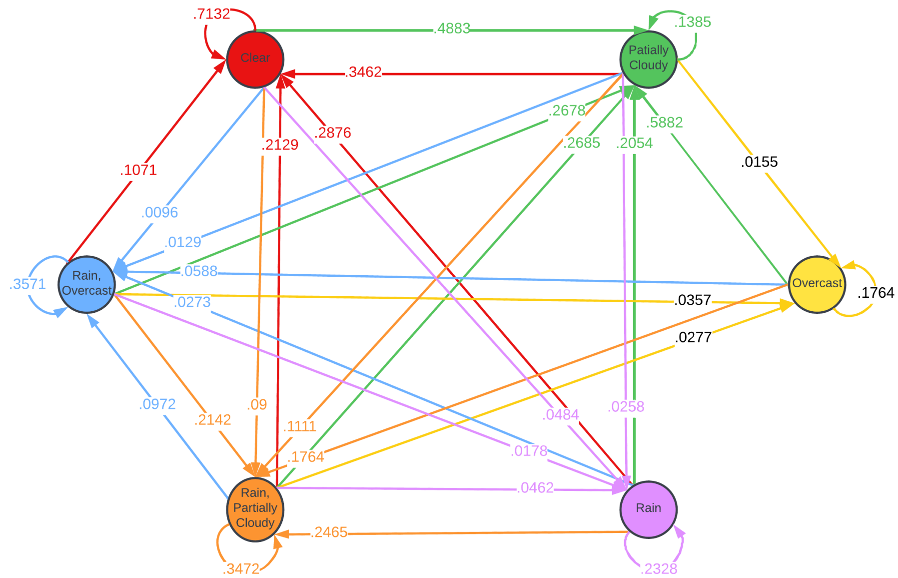

Figure 7: Visualization of Six States Markov Model Trained Over Long Time Frame for Fall

- Winter:

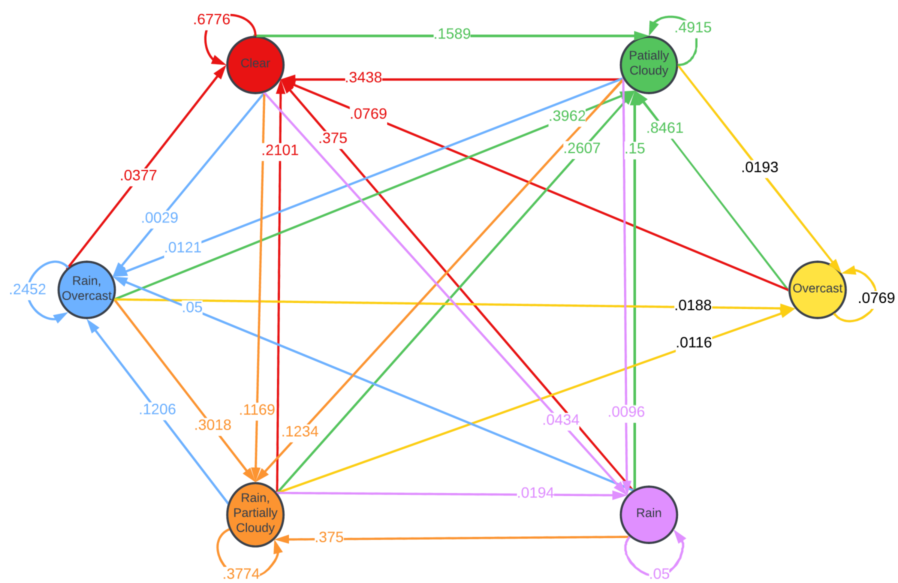

Figure 8: Visualization of Six States Markov Model Trained Over Long Time Frame for Winter

- Spring:

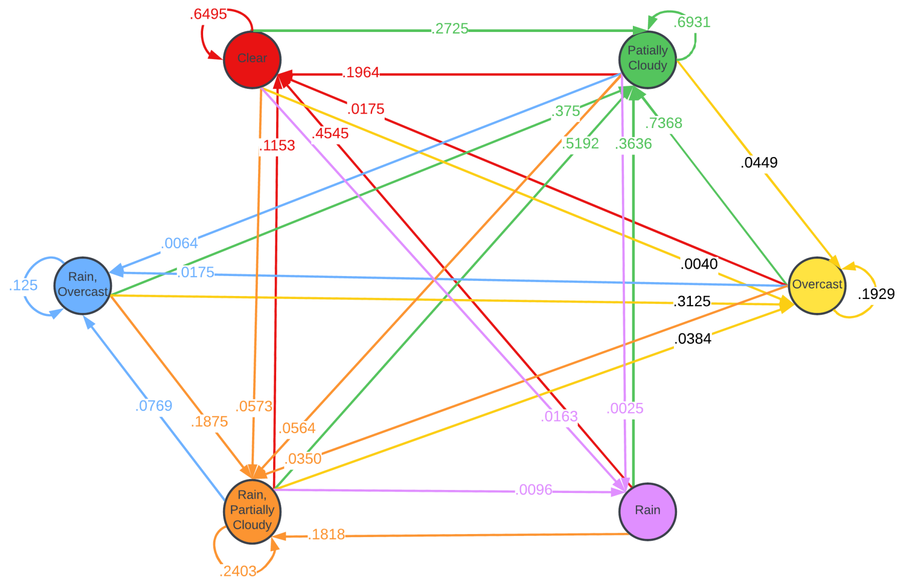

Figure 9: Visualization of Six States Markov Model Trained Over Long Time Frame for Spring

- Summer:

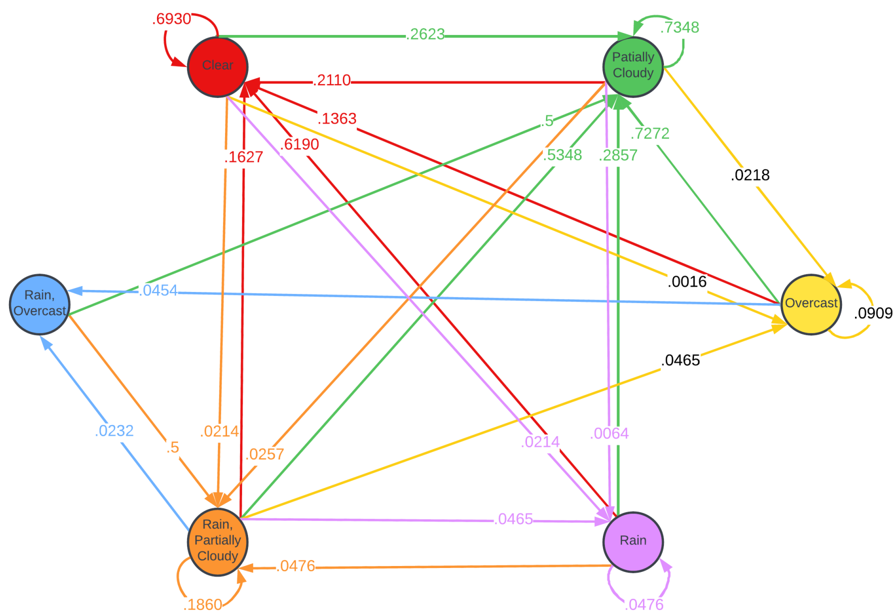

Figure 10: Visualization of Six States Markov Model Trained Over Long Time Frame for Summer

6. All seasons + Simplified applied to UCSD/NYU: Note that the same transition matrix as Figure 3 was applied to UCSD and NYU