Math 42
Final Project Proposal

Brian Chau
Yifan Jiang
Junwon Choi
Mingyeong Kim

Project Title: Predicting the weather with Markov chains

Project Category: Markov Chain

What is the problem that you will be investigating? Why is it interesting?

We will predict the weather in Los Angeles using the Markov chain model.

The weather significantly affects how we plan each day, which is why we check the weather forecast daily. Weather forecasting is conventionally done by analyzing the current physical environmental conditions - it will be interesting to forecast weather by just strictly using past weather data in a specific region.

What are the challenges of this project?

Deciding on the duration of our weather data may be a challenge. If the data collection process is not difficult, we will attempt to cover more years, since it will add accuracy (more training data) to our model.

Another challenge would be obtaining the values for the Markov matrix. We would have to sort through the raw data, choose the data that is relevant, and then determine our values for the matrix from there.

What dataset are you using? How do you plan to collect it?

We plan on using data from https://www.weather.gov/wrh/Climate?wfo=lox and https://www.wunderground.com/calendar/us/ca/los-angeles/KLAX. We will have access to historical temperature, precipitation, and general conditions in Los Angeles. This will provide flexibility in choosing which parameters to include in our model to predict the weather.

What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations?

We are proposing the Markov chain model to predict the expectation of weather. A Markov model is a stochastic model used to model randomly changing systems. It is assumed that future states depend only on the current state, not on the events that occurred before it. If we find some existing implementations, then we could add more parameters.

One way we could improve upon the Markov chain model would be to use an extension of it called the hidden Markov model. It would allow us to base our predictions based off of measurable values, such as temperature, wind speed, etc, and potentially give us a more realistic model.

What reading will you examine to provide context and background? If relevant, what papers do you refer to?

For context on Markov chains and their assumptions, we will refer to the following resource: TaylorB_module_f.pdf (ulisboa.pt).

In the case of the hidden Markov model, one research paper we will examine would be (https://ieeexplore.ieee.org/abstract/document/8284480). We will adapt their model to our specific region and make changes as needed.

We will also refer to our class textbook, “A Course in Mathematical Modeling” by Mooney and Swift - specifically chapter 3.8, which covers Markov Chains.

How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)

We will compare our model prediction with the range of actual data sets from our original data source. This would mean that we split our data into training and testing sets.

We will need to implement a method of statistical significance testing to evaluate our results.

We will use a python script to generate transition matrices and create a visual map of the Markov chain that is modeled.