

Project Proposal

Team:

Sunny Lau, Jerry Gai

Project Theme:

Going further on time series and state-space models

Public Repo:

https://github.com/GerGerGai/Bayesian_TimeSeries_StateSpace_Modeling#

Datasets:

Ozone Level Detection: <https://archive.ics.uci.edu/dataset/172/ozone+level+detection>

The first dataset is about ozone level detection available from the UCI Machine Learning Repository, comprises two distinct subsets:

Eight-hour peak set: This subset focuses on eight-hour peak ozone levels.

One-hour peak set: This subset centers on one-hour peak ozone levels.

Both sets were collected between 1998 and 2004 in the Houston, Galveston, and Brazoria areas. They are characterized as multivariate, sequential, and time-series data, suitable for classification tasks. Each dataset contains 2,536 instances with 72 real-valued features focusing on conditions like humidity, wind speed and temperatures.

California Earthquake Data:

Direct access:

http://www.socr.ucla.edu/docs/resources/SOCR_Data/SOCR_Data_Earthquakes_Over3.html

Detailed Explanations:

https://wiki.socr.umich.edu/index.php/SOCR_Data_021708_Earthquakes

The SOCR Earthquake Dataset provides comprehensive information on earthquakes with magnitudes greater than 3.0, primarily focusing on events in California. The dataset spans from 1800 to 2008 and includes various parameters for each recorded earthquake including date and time, location, magnitude, seismic stations etc.

Potential Approaches:

We will begin by performing exploratory data analysis on the selected dataset and preparing it in a format suitable for analysis. Next, we will develop and select an appropriate time-series prediction model based on the dataset's characteristics. Depending on the dataset chosen, the model may be designed for classification (ozone level) or regression (earthquake data).

Our current plan includes implementing a Bayesian state-space model using Markov Chain Monte Carlo (MCMC), as covered in this course. Additionally, we will explore and develop a more advanced Bayesian state-space model beyond the scope of the class, which hopefully can get some advice from the teaching team on potentially good directions to go. To evaluate model performance, we will use a testing set for comparison and, where possible, provide explanations based on the mathematical foundations of both models.

Short plans to ensure equal contributions:

We have a public repo to track everyone's contributions.

Sunny and Jerry will hold brief meetings before each major phase of the project—exploratory data analysis, building simpler models, developing advanced models, and writing the report—to discuss ideas and allocate tasks in greater detail.

In general, we plan to collaborate on coding and implementation. For the report, responsibilities will be divided so that each person contributes to different sections of the main project steps, ensuring a balanced workload. Before the final submission, we will meet again to review, refine, and finalize the project together.