Friday, March 15, 2024

# STAT 447C Project Proposal

Team: Ayush Bharat

GitHub repository:

Theme: Comparing frequentist and bayesian methods for binary classification problems

---

Purpose:

One key component of bayesian data analysis is using data to modify existing beliefs about the state of the world. Frequentist methods assume nothing about the state of the world and use data to create a model. Deciding which approach to use can be challenging, given that both methods make very different assumptions. Frequentist methods have been the standard in statistics for a long time, although bayesian methods seem to be gaining more popularity. In this project, I will first use both methods on a dataset, then I will compare the results of the two methods and provide an analysis comparing the two methods.

Potential approaches:

The first problem would be picking the model to represent each method. Picking one model like logistic regression(frequentist) and bayesian logistic regression(bayesian; priors to be specified) to represent each group seems be an obvious way to go about comparing the two. Another way would be to specify different models for both groups and use the average predicted probability for classification.

In order to compare the two groups, I intend on splitting the data set into a training set and a testing set. The training set would be used to specify the model and the test set would be used to validate performance. This would provide a metric for comparing both groups.

Potential datasets:

https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction
- predicting customer satisfaction based on delay, flight duration

https://www.kaggle.com/datasets/gabrielsantello/cars-purchase-decision-dataset
- predicting purchasing decisions based on age, income