Introduction to Bayesian Data Analysis

Mentor: Nicks Irons

Mentee: Xuweiyi Chen

My project for this quarter supervised by Nick Irons is Introduction to Bayesian Analysis. We went over many basic concepts about Bayesian Statistics and MCMC. In the first Bayesian Statistics part, we went over concepts like: Bayes' rule, the Beta Binomial Bayesian Model, balance and sequentially in Bayesian Analysis, Conjugate family and how to estimate posterior.

I wrote answers for Bayesian analysis and Metropolis – Hastings and Gibbs sampler for questions. At the second half of the quarter, we decided to investigate one specific model: Latent Dirichlet Allocation. LDA is a model for topic modeling in documents based on Bayesian Statistics. I coded the model in R and presented in word clouds. I really enjoyed the project because it is simple and elegant. I have also presented the project in front of the group. During the project, I spent about one week to figure out what is LDA. We also spent some time trying to come up a very simple way to make LDA intuitive. Eventually, we borrowed Rstan to get the best performance for LDA model.

My favorite takeaways in this project are the distinguishes between frequentist statistics and Bayesian statistics. My background is probability series, so I am very comfortable about how Bayesian Methods work because it is very intuitive. However, I always wonder questions like why statistics work. I discovered some interesting findings from my mentor and this project. I really do enjoy literatures from Ludwig Wittgenstein. Thus, I always find it hard for me to believe abstract ideas and mathematics could unfold the truth in the real world. However, Bayesian Statistics requires only finite amount of data, which really makes me enjoy this project.

LDA GitHub Repository: https://github.com/XuweiyiChen/LDA_twitter_sentiment