

## **Bayesian Linear Regression and Applications**

*Instructor: Anupreet Porwal Student: Yuchen Sun*

In this project, I started from simple linear regressions and ended up finding a predictive posterior for the bayesian linear regression. Overall, I really enjoyed this project. But at times it was challenging, especially when we moved from univariate to multivariate distributions. So, in the end, I not only learned Bayesian Linear Regression, but also got introduced to multivariate Gaussian, vector calculus, and many matrix algebra techniques.

My mentor began this project by briefly explaining to me what bayes' theorem is (i.e. posterior, prior, likelihood). Week one to three, we discussed in length on what makes finding the posterior difficult. We looked at using conjugate priors to simplify calculations. Then, we spent four weeks on multivariate Gaussian distribution. We spent most of the time deriving conditional, joint and marginal density functions, which became very useful later on. At this point, I was introduced to the covariance matrix and I was using many matrix algebra techniques such as completing the squares. After those four weeks, I learned maximum likelihood estimation and maximum a posteriori estimation. We looked at log likelihoods and why finding the maximum of one is equivalent to the other. At this point, we were ready to tie everything together with linear regression. We looked at using maximum likelihood to estimate the coefficients of a linear regression. We compared this estimation to a full posterior distribution of the coefficients, which we found by using the multivariate conditional, joint, and marginal density functions I've derived earlier. The posterior distribution together with the likelihood function forms the predictive posterior, which is used to predict the dependent variable if given the independent variables. Unlike the simple linear regression's estimate, which is assumed to have a constant error variance everywhere on the regression line, the predictive posterior will provide a varying error variance that changes depending on the position on the regression line.

After this project, I have a better understanding of what bayesian linear regression is. But this project is just a brief introduction, which makes me look forward to what I will learn in higher level statistics classes. I think this project is a great opportunity to expose undergraduate students to higher level statistical concepts, and I would imagine that I will have an easier time, after this, when I take a course on bayesian linear regression.