Proposal for Experimental Mathematics

(Tentative) Title of the project: Introduction to Statistical Learning.

Description: In the discipline of statistical sciences, the term *statistical learning* is used to refer to the vast set of methods that are available for analyzing data and making predictions. Probably the most simple and oldest of these methods is the simple linear regression, which is covered in any introductory statistics course. In this project our objective to understand various methods which might not have been covered in a introductory statistics course. For example, we can study multiple linear regression, which is a generalization of the simple linear regression; we can study classical non-linear models such as logistic or probit regression. We can study various techniques that are used in the sparse setting i.e., where we have large number of parameters but not all of them are expected to have significant effect. Among these methods the most popular are the LASSO (Least absolute absolute shrinkage and selection operator) and the ridge regression. We might also delve a little bit on the topic of neural networks.

Goal: Our goal will be to implement various methods, as describe above, on either simulated data sets or real data sets available freely. Our guide will be the two books listed below:

1. The elements of statistical learning: by Jerome Friedman, Trevor Hastie, and Robert Tibshirani.

2. An introduction to statistical learning: by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

Both are available freely online; prospective students are encouraged to glance through the books. The second book has a set of "labs" (computational projects) at the end of each chapter and we might follow these during the project.

Tools: We will use R or Python for doing the statistical analysis.

Links: The names of the books in the above list have links embedded, click on the names to access the books.