Homework # 8: Variable Selection and Shrinkage (Chap. 6)
Due April 4 by noon. No quiz on April 4.

1. (**Real data analysis - Chap. 6, # 9, p.263**)

   Predict the number of applications received based on the other variables in the **College** data set. This data set is from our textbook. To access it, you can type `library(ISLR);` `attach(College)`. Fit

   (a) Least squares regression, selecting the best model;
   (b) Ridge regression, with λ chosen by cross-validation;
   (c) Lasso, with λ chosen by cross-validation;
   (d) PCR model, with M, the number of principal components, chosen by cross-validation;
   (e) PLS model, with M chosen by cross-validation.

   Evaluate performance of each method in terms of prediction accuracy. Report prediction mean squared errors obtained by cross-validation.

   Comment on the results obtained. How accurately can we predict the number of college applications? Is there much difference among the test errors resulting from these five approaches? Which method appears most accurate?

A few announcements:

- April 30 – Graduate student presentations, in DMTI-111, approximately 2-4 pm, lunch included.

- June 10 – One-Day Course "Big data analytics: dealing with structured, semi-structured, and unstructured data", by Dr. Choudur Lakshminarayan, Teradata Corp.

  COURSE DESCRIPTION. The course will cover theory and methods dealing with structured, semi-structured, and unstructured data based on real-world scenarios. Examples will include application of mathematical statistics and machine learning to numeric, clickstream, and text data from the real world. The range of algorithms will cover supervised and unsupervised learning spanning projection methods such as Discriminant Analysis, Principal Component Analysis, Independent Component Analysis and Regression Analysis (exponential families), Deep Neural Networks, and Statistical Clustering. We will use the R/SAS programming languages for analyzing the data and demos.

  I should have funds to cover your registration, lunch included.

  See https://www.american.edu/cas/qprc/One-Day-Course.cfm

- June 11-13 – The 36th Annual Quality and Productivity Research Conference. I should have funds to cover your registration, especially if you present a poster. I can also use your help at the conference, for extra pay. Lunches included (and more).

  See https://www.american.edu/cas/qprc/