In-class Assignment # 4

Due: Thursday, March 22, 2018, 11:59 p.m.
Total Points: 50

This assignment is designed to help you better understand concepts that were presented during class. You must complete this assignment on your own, but feel free to ask questions to your classmates and the instructor. **Each student is responsible for submitting their own solutions**. Be sure to include your name in your work. Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Your assignment must be submitted as a single pdf document on Canvas. For each question, your submission should include text for the question, your MATLAB results/answers, and your MATLAB code. The questions and your answers must be typed; for example, in Latex or Microsoft Word. Images may be scanned and inserted into the document if it is too complicated to draw them properly. Be sure to show all work. **All assignments must be submitted on time to receive credit.** No late work will be accepted, unless you have a prior arrangement with the instructor.

Question 1. [50 POINTS]

Two data sets are posted on Canvas. The first has an $n \times d = 1000 \times 50$ data matrix (X) "pred1.dat" with a 1000×1 response vector (y) in "resp1.dat." The second has a 1000×500 data matrix "pred2.dat" with a response vector in "resp2.dat." These data sets were generated according to the standard linear regression model.

(a) For each data set, use the first half of the data (observations $i = 1, \ldots, n/2$, all d predictors) to estimate values for $\boldsymbol{W}$, (e.g. $\hat{\boldsymbol{W}}$). Plot the estimated values for $\hat{\boldsymbol{W}}$, one from the first data set and one from the other.

(b) For each data set, use your estimate of $\boldsymbol{W}$ on the 2nd half of the data set $(n/2 + 1, ..., n)$, to get your estimated response variables, $\hat{\boldsymbol{Y}}$. Compute and report your total squared error: $SSE = \sum_{i=n/2+1}^{n} (\hat{y}_i - y_i)^2$

Be sure to include the results and your code in the solutions report.