Linear Regression

STA 325: Lab 5, Fall 2017

Today's agenda: Linear Regression

Programming partner's: You should have a programming partner for each lab, and you should switch off who is programming, and use each other for help. We will spend about 30–50 minutes per week on lab exercises and you will be expected to bring you laptops to class to work on these exercises in class. Myself and the TA's will be in class to help you.

Lab Tasks

1. Your goal is to build a regression model for Gross National Product (GNP) based on two input variables: number of people employed and the total population using the longley data in R. Start by performing graphical exploratory data analysis: create univariate density estimates and scatterplots to understand the bivariate features of the data. Do you see anything interesting?
2. Build the regression model, and provide the coefficients of the model and details on their significance using xtable. Interpret your coefficients. Is the intercept meaningful? What can you do to make the intercept more meaningful?
3. Perform regression diagnostics via graphical methods: Assess normality of your residuals, constant variance, independence as well any potentially influential points using Cook's Distance with a threshold value of $\frac{4}{n}$. Be sure to detail what you see. Do you need to transform your data? If you were to transform your data, how would it impact the interpretation of your model?
4. Create a plot of Population against GNP that shows the fitted regression line holding Employment at its mean value. Add prediction and confidence intervals to your plot based on the same assumption in different colors. Where are the intervals narrowest? What do you expect will happened to the intervals as $n \to \infty$.
5. Please complete the following exercises in ISL: 2 and 5.