Lab 3

Exercise 1 - Polynomial Regression. Predict US population

a) Load the data `USpop.csv` and plot `Population` as a function of `Year`.

b) Use a linear model to fit the data. Does a linear model provide a good fit?

c) Calculate R^2. What do you observe? Does the value of R^2 imply that a linear model is a good choice?

d) Using the linear model, predict the US population for the year 2030. Is this a good prediction?

e) Produce appropriate residual plots and decide whether or not an important predictor has been omitted. What do you observe?

f) Fit a quadratic model and plot the fitted curve. Is this a good fit?

g) Predict the population for the year 2030. Is this a resonable prediction?

Exercise 2 - Regression Diagnostics

a) Load the `Auto.rda` dataset. Predict `mpg` using the predictors `year`, `accelerator`, `horsepower`, `weight`. Generate different residual plots.

b) Which of these residuals can be considered as outliers? Compare with the Bonferroni-adjusted quantile from t-distribution.

c) Test Normality using Shapiro-Wilk normality test. Also look at the Normal Q-Q plot above. Shapiro-Wilk statistic W measures how close the graph is to a straight line.

d) Test HOMOSCEDASTICITY (constant variance) using the Breausch-Pagan test.

e) Check for INFLUENTIAL DATA.