Final Project

Fall 2016 (20% points)

In this project, you are given a sample with 6 variables in the file mydata.txt on the moodle. The 1st variable V1 is the response, and the other variables (V2-V6) are covariates. There are totally 200 observations in the sample. Use the statistical methods introduced in class to build a model for describing the relationship between the response and covariates. Possible techniques you need to use are cross-validation, hypothesis testing, nonparametric smoothing, regression diagnostics, regularization, and bootstrap, etc.

In addition to providing your arguments, you are asked to provide your R codes and answer the following questions.

1. Write the dataset to a file in excel format. (1pts)

2. Find sample correlation between the response and each of the covariates. (1pts)

3. Propose an initial model to fit the dataset and check its appropriateness using regression diagnostics. (3 pts)

4. Refine your model based on your discovery in step 3 and check if it is appropriate. (3 pts)

5. Can you improve the model you fit in step 4? Why? (3 pts)

6. Write down the final model which is best for fitting the dataset and provide 95% confidence intervals/bands to your estimators of the parameters/curves. (1 pts)