CHE384 Data to Decisions
Chris Mack, University of Texas at Austin

Homework #8 – Multicollinearity and PCA

Turn in your solution with the answers to the questions below. Also, email to me the supporting R script that you used to perform the analysis. (Please name the file using this format: HW8_yourname.R).

1. Using the alcohol data set found in DataSet4.xlsx, you wish to create a model to predict logSolubility using one or more of the calculated molecular properties of various alcohols.
    a) Search through all first-order linear models (no interactions) and find the model with the smallest BIC. Report that model.
    b) Run a correlation matrix on all of the variables. What can you conclude? Can you think of two variables that should definitely be excluded from the model? Which ones are best to eliminate?

2. Using the same data set from problem 1, perform a principle component analysis on the regressor variables. Describe what you find, and what your next actions should be.