Homework to Week 3

Statistics: Principle, Methods and R (II)

Gao Fengnan

Week 4, 20 March 2017

The homework is due on Monday, 27 March 2017. Please hand in the solutions to the teaching assistant He Siyuan at the beginning of the lecture.

1. Get the passenger car mileage data from `http://lib.stat.cmu.edu/DASL/Datafiles/carmpgdat.html`. Please attach your R code.
   (a) Fit a multiple linear regression model to predict MPG (miles per gallon) from other variables. Summarize your analysis.
   (b) Use Mallows' C_p to select a best sub-model. To search through the models, try (i) forward stepwise, (ii) backward stepwise. Summarize your findings.
   (c) Perform all possible regressions. Compare $\overline{R}_p$, C_p, AIC. Compare the results.
2. Prove the Kullback-Leibler discrepancy is always non-negative. (Hint: Write the discrepancy as the expectation of something and apply Jensen's inequality.)