M358K - worksheet2 (multiple regression)

October 26th, 2017

Questions

Your goal is to predict the age of abalones based on various measurements.

The data file, abalone.csv, is on Canvas. For variable descriptions, see abalone-descrip.txt

Part 1: regression

1. Do a pairwise plot of all variables. For each plot, briefly describe what you see.
2. Run a linear regression model to predict age. Clearly show the R command that you use, and include the R's model summary.
3. Write down the equation that R gives you. Interpret all the coefficients and the p-values associated with the coefficients.
4. Report the R^2 and adjusted R^2 of your model. What are the meaning of these values?
5. Do a diagnostic plot for your model. Say which, if any, of the (a) independence (no mean trend) (b) normal distribution and (c) constant variance assumptions are violated.

Part 2: variable selection

1. Run variable selection using the command `stepAIC`. Write down the equation of the new model that R gives you, and compare that to the equation of the previous model.

2. Is the model found by stepAIC a good fit?

3. If your model is not a good fit, try some transformations in the X variables. Which transformations could be helpful?

4. Report the findings from your best model.