MVE055/MSG810, Matematisk statistik och diskret matematik, 2016/17
Assignment 3
Due October 17, 2016

Analysis of Wine Quality Data

Data description:

We consider a set of observations on a number of red and white wine varieties involving their chemical properties and ranking by tasters. The two datasets are related to the red and white variants of the Portuguese "Vinho Verde" wine. The information about the wines are available as a Excel file, where the first sheet contains the white wine and have 4898 different varieties and the other contains the red wine and have 1599 varieties. For more information, read [Cortez et al., 2009].
There are 12 different properties (attribute) of the wines as follows:

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar

5. chlorides
6. free sulfur dioxide
7. total sulfur dioxide
8. density
9. pH
10. sulphates
11. alcohol
12. quality (has score between 0 and 10).

The response "Quality" is assumed to be a continuous variable and is predicted by the independent predictors, all of which are continuous.

Analysis:

Linear regression should be fitted to the data (**You should consider both data sets**). Choose the correct type of regression analysis. You need to specify the model. Model specification consists of determining which predictor variables to include in the model and whether you need to model curvature and interactions between predictor variables. Are all predictors significant? Compare the fitted models for both data sets? What is your conclusion?

Written rapports

Present your reports in a nice readable form in English. It is not enough just to give answers and references to the formulas you use. Your report should contain:

- Description of the problem.
- Statistical model.
- Your statistical analysis.
- Conclusions.
- As an appendix give the program code you used.
- Email your report to **rozam@chalmers.se**.

Reference

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.