Course 27411 Exercises 8/4 2013

Classification Exercise A (to be presented by Team 7 Monday 15/4)

Part 1: (Definitely)
Use the wine data previously used. It is uploaded under file sharing with a couple of additional variables included: y and y2 coding for the Barolo and Grigno wines respectively.

A: For this part use the LDA feature of Unscrambler: Predict the wine-class using, LDA and/or QDA and/or Mahalanobis (the three options). Try with/without the PCA-LDA option. What are the results?

B For this part use the PLS-feature of Unscrambler:
Try to set up a regression model that predicts either of the two binary response variables – start with the Barolo-indicator:

1. Set up the model – chose the number of components.
2. Look at the predictions (by Cross validation) and construct the classification table: (true states vs. predicted states). Use a threshold of 0.5 for the predictions.
3. What are the error rates?
4. Try to select two good variables and do the analysis for these two: (not of major importance which you chose, though, for the point of the exercise)
    a. Look at the sample plots: does the linear model used seem to be OK? (Are the two within group covariances approximately equal?)
    b. Try to define and include the squared variables.
    c. Try also to include the product of the two variables.
    (The approach of b) and c) resembles the Quadratic Discriminant Analysis (QDA) – although NOT exactly equivalent)
    d. Compare with a QDA analysis using the LDA feature of Unscrambler.
5. (IF time) Do the same for the other two wines AND/OR try a PLS2 to do a 3-class PLS discrimination.

Part 2: (potentially)
Work with the hand ("manual computer") calculation exercises uploaded (including solutions) at the web site. (and not necessarily as detailed and thorough as suggested in the exercise writing)