6

Cross-Validation and Generalized additive models

You will in this exercise consider the South African Heart Disease data used in the book. Download the data from the book home page. Note that in the info for the data the R-code for reading the data correct is given. For the first 3 questions you should use the `gam` library.

Question 6.1. *Estimate a generalized additive logistic regression model using smoothing splines to these data with 4 degrees of freedom for each of the smoothers. Plot the effects. Do they look linear? Compare with Figure 5.4 in the book. Note that* `famhist` *is categorial and should thus not be smoothed.* `chd` *is the response.*

Question 6.2. *Partition the data randomly into 7 groups of equal size. Carry out 7-fold cross-validation where you estimate the GAM model 7 times each time excluding one of the groups and then predict the* `chd` *value on the last group. Use this to estimate the generalization error for the 0-1 loss function and for different choices of a single, common degrees of freedom for the smoothers.*

Question 6.3. *Do cross-validation as above to estimate the generalization error but use the likelihood loss instead.*

Question 6.4. *Compare your results with what you get by using* `gam` *in the* `mgcv` *library, that does automatic selection of tuning parameters.*

Note that the `mgcv` and `gam` libraries do not like each other particularly well. I have experienced several problems when trying to run one things from one of the libraries when they are both loaded. So the general advice is: Don't load both of them at the same time.