Example Assignment for Literate Statistical Analysis

Purpose

We show how to use Sweave and BibTeX for literate statistical analysis with an example assignment.

Literate statistical analysis with Sweave uses R to do the statistical computations including Figures and Tables. After you have perfected the R scripts these are embed in an Rnw file that is processed using R to produce a LaTeX file. Normally you use markups in the Rnw file which hide the R code in the final pdf output file.

Normally you should submit both the pdf file and the Rnw file as well as any other external files such as data that you have read in.

Running Sweave

After starting up R, click on **`File`** and then on **`Change dir...`**. . You should change directory to where your Sweave file, usually with extension **`FRnw`** is -- in this case the input file is **`Literate.Rnw`**. There are two functions available that are illustrated below. The function **`Sweave()`** is used to run R on the sweave file and produce the LaTeX output file, in this case it is **`Literate.Tex`**. The other function, **`Stangle()`**, provides an easy way to extract all the R code used in your sweave file into a separate file, in this case **`Literate.R`**.

See below for an example. Notice that it is very useful to give names to each code segment, such as **`LoadLibs`**, **`FigA`**, etc. In this way if something crashes on you, you can easier locate where the problem is.

```
R> Sweave("Literate.Rnw")
Writing to file Literate.tex
Processing code chunks ...
 1 : term verbatim (label=LoadLibs)
 2 : term verbatim (label=FigA)
 3 : term hide (label=preliminaries)
 4 : term verbatim (label=Regression)
 5 : term verbatim (label=LogisticRegression)
 6 : term tex (label=TREG)
 7 : term tex (label=TGLM)
 8 : term tex (label=ETA)
 9 : term verbatim pdf (label=FigA-repeat)
10 : term verbatim (label=LoadLibs)

You can now run LaTeX on 'Literate.tex'
R> Stangle("Literate.Rnw")
Writing to file Literate.R
```

Literate Statistical Analysis Assignment

HTF Mixture Data

The data from the textbook is included in the R library **`ElemStatLearn`** which may be downloaded from CRAN. The mixture data is available in the R list **`mixture.example`**. For more details see the documentation provided in the package and at the book website:

http://www-stat.stanford.edu/~tibs/ElemStatLearn/index.html

Construct a plot comparing the decision boundary for linear regression and logistic regression. Comment.

Compare the confusion tables and the misclassification rates on the training data for both methods.