

Problem 1 (Barley data) Here we see how to fit additive models for the `barley` dataset of the R package `SMPracticals`. First install the package `mgcv`, which provides the `gam` function.

- (a) Fit the model $y \sim \text{Variety} - 1 + \text{Location}$, where `Location` is a smooth term, different for each block. You can use the following command line:

```
fit.gcv <- gam(y~Variety-1+s(Location,by=Block),data=barley),
```

the operator `s` specifies which terms are to be treated as basis functions and thus smoothed. The optional parameter `by` fixes blocks in which the curves are fitted.

- (b) By default the number of knots is computed automatically. Check the help to see how to specify it, and try varying it. Comment.
- (c) What is the smoothing basis used by default? How can you change it?
- (d) What method was used to fit the model?
- (e) Find the estimated degrees of freedom and the AIC of the fitted model.
- (f) Plot the estimated curve and the residuals. You will need to add a shift the curves, for instance by adding 4 to one and adding 8 to the other, so that you can see all of them on the same plot.

Hint: use `par(mfrow=c(1,3))` to display three plots in the same window and also look at the elements of `fit.plot.gcv <- plot(fit.gcv,seWithMean=TRUE)` for easier plotting of the fit.

- (g) To fit random effects for `Variety`, use

```
fit.gcv <- gam(y~s(Variety,bs="re")+s(Location,by=Block),data=barley)
```

- (h) Reproduce the previous analyses using the REML method for fitting.
- (i) Compare the two models. Which one seems preferable? Comment.