

Exercise 5

We will in this exercise consider the use of GAM (generalized additive models). We will consider a dataset giving measurements of daily ozone concentration on 111 days from May to September 1973 in New York. Also three covariates are given. The variables are

ozone daily ozone concentration (ppb)

radiation solar radiation (langleys)

temperature daily maximum temperature (degrees F)

wind wind speed (mph)

The dataset is available from the course home page.

In R, commands for performing GAM is made available by the command

```
library(mgcv)
```

You will in this exercise need to use the **help** command extensively in order to understand the commands used. For instance,

```
help(gam)
```

provides a lot of information regarding the **gam** command.

- (a) Fit an additive model to the cube root of ozone concentration as a function of temperature, wind speed and radiation with the command

```
oz.gam <- gam(ozone^(1/3)~s(radiation)+s(temperature)+s(wind),data=ozone)
```

Try to understand what this command does.

Perform the commands

```
summary(oz.gam)
plot(oz.gam)
gam.check(oz.gam)
vis.gam(oz.gam,view=c("radiation","wind"))
```

and try to understand what these commands do.

- (b) GAM include linear models as a special case:

```
oz.lm <- gam(ozone^(1/3)~radiation+temperature+wind,data=ozone)
```

Compare with the results obtained from the gam model.

- (c) Try out the command

```
anova(oz.lm, oz.gam, test="F")
```

and try to understand the output (note that `anova` is a generic function and that the actual function that is used is `anova.gam`). Is it worthwhile to use additive models in this case?

- (d) By use of the `anova` command, could the non-linear structure be removed on any of the covariates?
- (e) Try also out prediction using the `predict` command. In particular, find out how to obtain standard errors of the predictions.

Hint: Note that also the `predict` command is generic, and that the actual function used is `predict.gam`.