Generalised linear and additive models

Extra exercise 9.1
NO_2 data from a road in Oslo

Copy the data set NO.dat to your computer. This data set consists of 500 hourly observations (from the years 2001-2002-2003) of NO_2 concentration at a road in Oslo with corresponding measurements of the number of cars and meteorological variables.

Read the data set into R by the `read.table` function,
`NO2dat<-read.table("NO.dat")`.
You may need to extend the file name with the correct directory.

Some information on the data:

- 1 response variable
  - logNO2: the (natural) logarithm of the NO_2 concentration
- 7 predictors
  - logCars: the (natural) logarithm of the number of cars
  - temp: temperature 2 m above ground (deg C)
  - tempDiff: temperature difference between 25 m and 2 m above ground (deg C)
  - windSpeed: wind speed (m/s)
  - windDir: wind direction (degrees between 0 and 360)
  - hour: time of day (hour)
  - dayNo: day number (counted from Oct. 1, 2001 - e.g., Oct.1 2001 = 1, Oct. 2 2001 = 2)

You can use the function `gam` from the R library `mgcv` both to fit linear, generalised linear and generalised additive models, so you can use this function in all tasks below, if you want to.

For each task below; print the summary of the estimation and plot the non-linear functions, if any.

a) Estimate a linear model with logNO2 as response. Assume that logNO2 is Gaussian.

b) Estimate an additive model with NO2 as response, where all the 7 predictors are included non-linearly. Which are the most important predictor variables? Do you think the estimated s-functions look reasonable?

c) The wind direction (in degrees) is a so called cyclic variable, which means that the lowest possible value of 0 is identical to he highest possible value of 360.

The hour variable (time of day) is also cyclic, and the value of 0 means the same as 24.

Take this into account such that s(wind direction=0) = s(wind direction=360) and s(hour=0) = s(hour=24)

This can be done in the `gam` function by writing `+s(windDir,bs="cp")+s(hour,bs="cp")` in the formula and include an extra argument `knots=list(windDir=c(0,360),hour=c(0,24))`.

d) Refit the model with more smoothing. In the `gam` function, this can be done by including an extra argument `gamma=c`, where the value of c controls the extra smoothing, such that the effective number of parameters is multiplied by c in the generalised cross validation criterion. If c=0.5*log(number of observations), the smoothing parameters are chosen so they roughly optimise BIC.