Lab 11

Exercise 1 - Principal Components Regression

For this lab we will use the `Auto.rda` dataset. In addition we should use `library(pls)` and thus we have to install the `pls` package.

a) Perform Principal Components Regression (PCR) using all variables except `name`. For this use `pcr.fit = pcr(mpg~.-name-origin+as.factor(origin),data=Auto)`. In addition, perform usual linear regression using the `lm` function. What do you observe in terms of the R^2 if we use all variables?

b) Investigate the principal components and see how much of the variance they explain. Use the function `princomp` to perform principal components analysis on the numeric matrix `X` and use the function `screeplot` to plot the variances for each of the PCA components.

c) As you can see from (b), the 1st principal component contains a huge portion of the total variation of the `X` variables and it is dominated by the variable `weight`. To eliminate this effect, standardize the variables on `X`. This simply means that you have to subtract each `X`-variable's mean and divide by the st.deviation.

d) Use K-fold cross-validation with $K = 10$ for `pcr`. How many components do we need to include in our regression?

Exercise 2 - Partial Least Squares

Similar to exercise 1, just replace `pcr` with `plsr` and see what is the number of components that provides the lowest prediction MSE.