Introduction to Machine Learning in R  
(NPFL054)

Easy HW – k-NN algorithm
Contact: Barbora Hladká (hladka@ufal.mff.cuni.cz)

Data

- Auto data set (ISLR package)

Get the data for the experiments

- Create a binary target attribute, mpg01, that contains a 1 if mpg contains a value above its median, and a 0 if mpg contains a value below its median. Create a single data set d containing both mpg01 and the other Auto attributes except mpg.

Question

1. Randomly split the data d into eight folds to perform 8-fold cross validation. Perform k-NN with several values of k in order to predict mpg01 using all the features except name. Plot 8-fold cross-validation error rate for different values of k. Use the knn() function, which is part of the class library. Compare the plots for unscaled and scaled data. Use the scale() function to scale the data.

Presentation

- Create a 5 min presentation.
- Present your answers. If you want to highlight something in your R code, please do it.
- Explain your answers clearly so that your audience understands your method well.