Introduction to Machine Learning in R  
(NPFL054)

Easy HW – ROC curve and Cross-validation  
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.

Questions
1. Split the data d into a training set train and test set test 80:20. Develop a Logistic regression model and a Decision tree model on train to predict mpg01 and test the models on test. Plot their ROC curves and compare their AUCs.

2. Address the task of mpg01 prediction using SVM with Radial basis kernel. Use the data set d as a development data to run 8-fold cross validation. Use the function svm with kernel="radial". Report cross-validation error rates for various values of gamma and cost.

3. Address the task of mpg01 prediction using SVM with polynomial kernel. Use the data set d as a development data to run 8-fold cross validation. Use the function svm with kernel="polynomial" and gamma=1. Report cross-validation error rates for various values of cost and degree.

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.