Intelligent Data Analysis

Tutorial 6
Linear Models for Regression

Paul Prasse
Nuno Marques
Dr. Niels Landwehr
Prof. Tobias Scheffer

Date: 29-05-2015

Goals
In this tutorial we will implement an algorithm to learn a linear ridge regression model. As
the implementation of this model is very straightforward (one line of code), we will most
of the time in the tutorial coding a train-test splitting function. This function will be used
in every single machine learning project you work on.

Problem setting
We’ll use the Boston Housing Dataset. This dataset contains information collected by the
U.S Census Service concerning housing in the city of Boston in the state of Massachusetts
in 1978. Our goal is to predict the median value of the houses in a particular town in the
city of Boston given its attributes. Check the file ’housing.names’ for more information
on the attributes.

Task 1
Analyse the following MATLAB-function

function tutorial6

Our goal is to make this function work, by coding the functions it uses. By the end of
this class it will print the mean of the absolute differences between our predictions and
the real housing values.

Task 2
Write the following MATLAB-function

function [train,y_train,test,y_test] = split_train_test(data,seed,split,class_col)

Split the dataset in (split×100)% of the instances for training our model and the rest
for testing. Make sure you use the given random number generator seed so we all get the
same results. Don’t forget to remove the column that we are trying to predict (specified
in class_col) from both train and test.

You may find the following functions useful:
randperm, setdiff
Task 3
Write the following MATLAB-function

function [theta] = train_linear_reg((X,y_train,lambda)

Implement the ridge regression model (slide 24). The function should output the learned weight vector $\theta$.

You may find the following functions useful:
\texttt{pinv}, \texttt{eye}

Task 4
Write the following MATLAB-function

function pred = predict(theta,X)

which predicts housing values vector \texttt{pred} for a dataset X and a previously trained parameter vector $\theta$.

Task 4
Write the following MATLAB-function

function res = mean_abs_loss(pred,y)

which computes the mean of the absolute differences between our prediction vector \texttt{pred} and the real housing values y.