Saarland University

The Elements of Stastical Learning

Assignement 6

Due Date: 17.01.2018

Thibault Schowing *Mat. 2571837*
Sarah Mcleod *Mat. 2566398*
January 16, 2018

Problem 1

Question How do you have to change the procedure of generating a MARS model to make a decision tree?

Answer If you replace the piecwise linear basis functions used in MARS with step functions $I(x - t > 0)$ and $I(x - t < 0)$ and add a resriction that a node may not be split more than once, then you can make a descision tree.

Question Can you argue on the basis of the relationship between MARS and decision trees revealed in (a) what is an advantage of MARS over decision trees and what is an advantage of decision trees over MARS?

Answer For numeric data, MARS would tend to be better than Decision Trees because the reflected pairs (hinges) would adapt more accurately/simply to the underlying structure of the data (for instance local linearity) than the constant segmentation realised by decision trees. Think about a diagonal set of points that would have to be separated in many sub-trees, with MARS a single line and two knots can estimate it accurately.

The decision trees are more faster to create and, especially for categorical data, the fits of partitioning would be better. Even if MARS adapts to non linear and categorical data, depending on the exact shape of the data, a decision tree could separate more accurately the different regions.

Problem 2

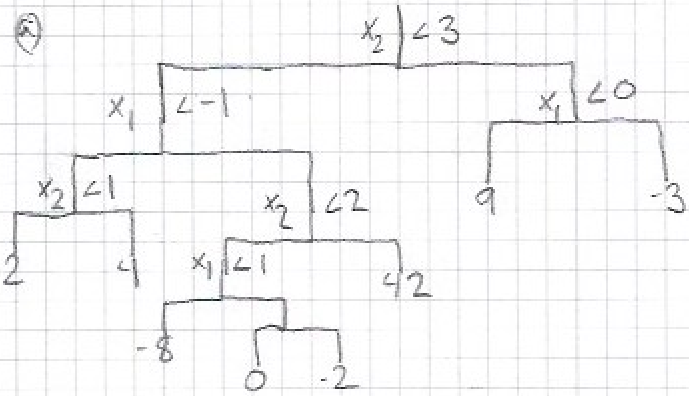
Our new loss function that is differentiable at zero and is not as susceptible to outliers is:

$$L(y - f(x)) = \begin{cases} (y - f(x))^2 & \text{if } |y - f(x)| < \delta \\ 2\delta|y - f(x)| & otherwise \end{cases}$$

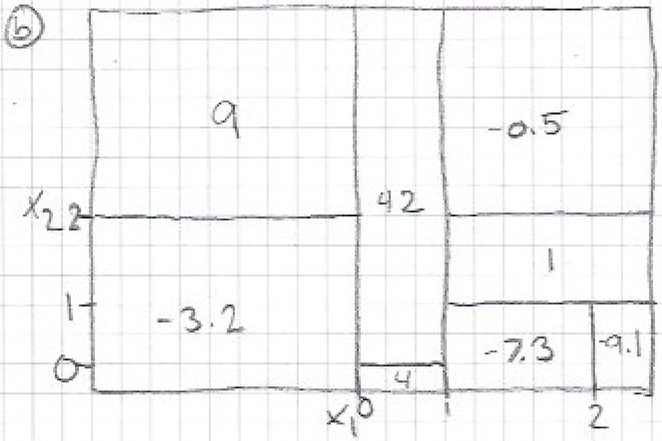
Problem 3

problem 3

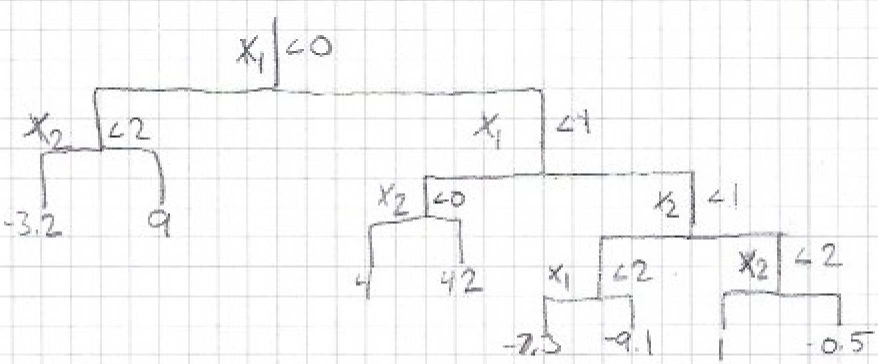
(a)



(b)

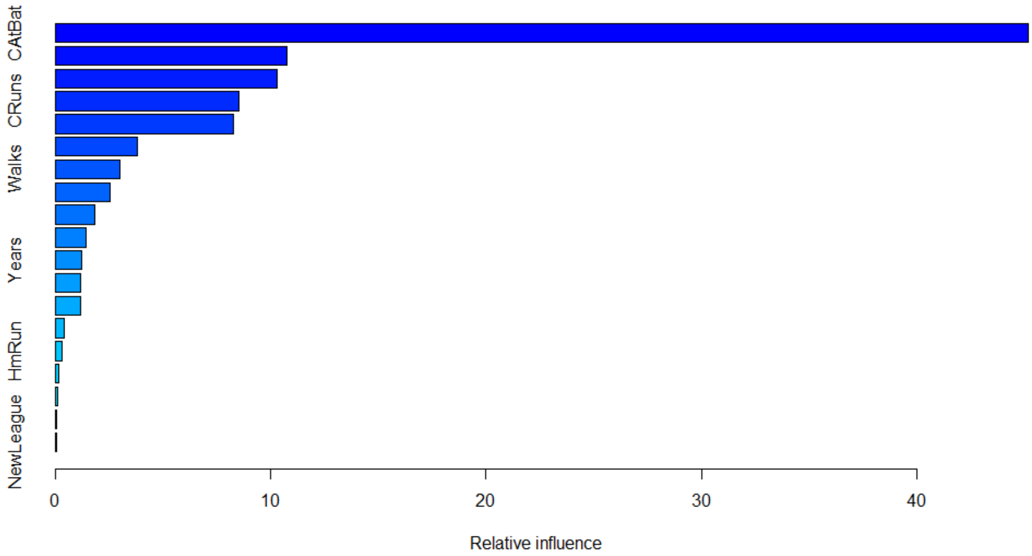


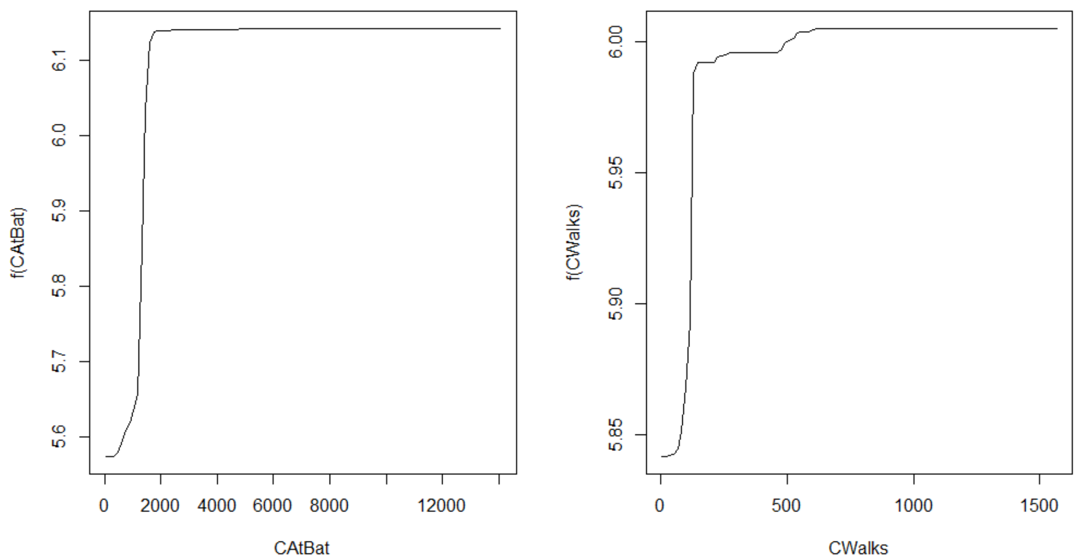
(c)



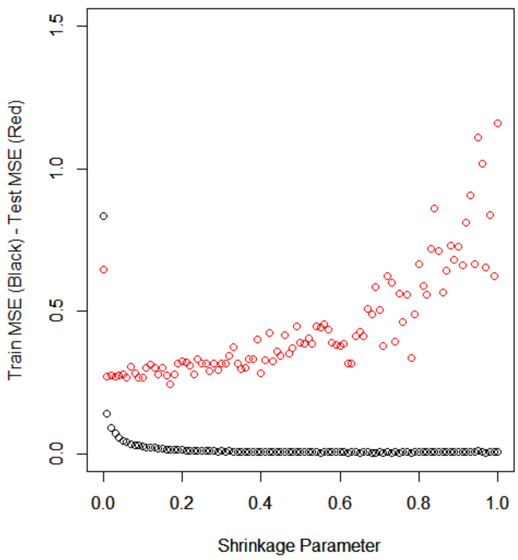
Problem 4

b With the default shrinkage parameter (0.001) we obtained the following Relative Influence Plot and Partial Dependence Plot:





We then execute for a δ from 0.01 to 1 and plot the train MSE and test MSE:



We choose $\delta = 0.2$

c We have the following values:

Algorithme	MSE
Boosting	**0.24**
Least Squares	0.49
Ridge	0.56

d Variable importance:

For boosting the graphing on point b show the actual following values:

For Ridge:

For linear regression, the two highest coefficient are AtBat and Hits

Table 1: Variable importance for Boosting

var	rel.inf
CAtBat	45.18862170
CWalks	10.76140033
CHits	10.27594487
CRuns	8.49553266
CRBI	8.27788387
AtBat	3.77926831
Walks	2.97904456
CHmRun	2.50711182
Hits	1.81700963
RBI	1.39305145
Years	1.18646093
PutOuts	1.18241928
Runs	1.15779063
Assists	0.38518856
HmRun	0.29914573
Errors	0.15584311
Division	0.07288507
League	0.04818061
New League	0.03721686