Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Thomas Seidl
Julian Busch, Evgeniy Faerman,
Florian Richter, Klaus Schmid

Knowledge Discovery in Databases
SS 2016

Exercise 12: Numerical Prediction

We are looking for Tutors for „Einführung in die Programmierung" for WS2016/17. If you are interested please write a short email to Florian Richter (richter@dbs.ifi.lmu.de).

Exercise 12-1   Regression Trees

Consider the following data samples of the form (x, y), where the input value is $x \in R$ and the output value is $y \in R$:

$$P_1 = (-3, -1), P_2 = (-2, 0), P_3 = (-1, 1), P_4 = (1, 1), P_5 = (2, 0), P_6 = (3, -1)$$

Search for the first best split. If the decision is obvious, you don't have to compute all possible splits. Then decide whether the split is significant or not by using the impurity ratio with $\tau_0 = 0.5$.

(a) Use as impurity measure the variance of the output.

(b) Use as impurity measure the variance of the residuals.

Exercise 12-2   QA-Session

The last half of the last exercise slot will be dedicated to a QA-session, which is intended to give you an opportunity to ask questions about the lecture and exercise contents and also benefit from the discussions of other students' questions. Note that we cannot answer any specific questions regarding exam contents.

During the session, we will answer and discuss the questions, which you send to me via e-mail in advance (busch@dbs.ifi.lmu.de). Please hand in your questions before next week, so I have some time to prepare them.

All submitted questions will be treated in all of the exercise sessions, so you don't have to show up to all sessions in order not to miss anything.