

# Machine Learning

## Exercise Sheet 10

Prof. Dr. Dr. Lars Schmidt-Thieme, Martin Wistuba  
Information Systems and Machine Learning Lab  
University of Hildesheim

January 14th, 2014  
Submission until January 21th, 13.00 to wistuba@ismll.de

### Exercise 24: Decision Trees (6 Points)

Given is the following training data:

Day	Outlook	Temp.	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

The target variable *PlayTennis* with possible values *yes* and *no* needs to be predicted for different Saturdays depending on the attributes of the respective mornings.

Show the first two splits for creating a binary decision tree using the method introduced in the lecture („greedy strategy“). Hence, your result has only one root node and the left child of it as inner nodes.

Use the a) Information Gain and b) Gini Index as the split quality criterion.

### Exercise 24: Decision Trees - Pruning (4 Points)

Apply pruning on the decision tree in Figure 1 using the Cost Complexity Criterion

$$CCC(\hat{y}) = error(\hat{y}) + \lambda |leaves(\hat{y})|$$

with  $\lambda = 2$ . Show intermediate steps and calculations.

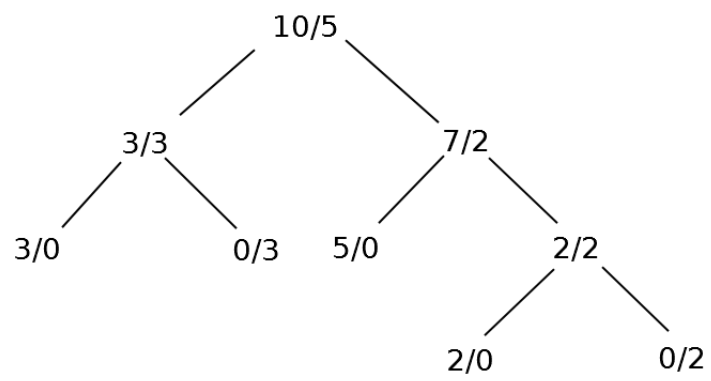


Abbildung 1: Decision tree for Exercise 24