Machine Learning
Exercise Sheet 7

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

December 5th, 2016
Submission until December 12th, 13.00 via learnweb!

Exercise 11: Decision Trees (12 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.

Create two binary decision trees using the method introduced in the lecture („greedy strategy"). You can stop after the first two splits.

Use the a) Information Gain and b) Gini Index as the split quality criterion, respectively.
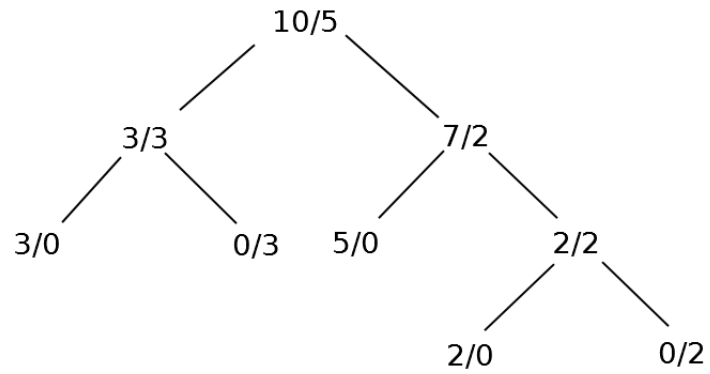
Abbildung 1: Decision tree for Exercise 12

Exercise 12: Decision Trees - Regularization (8 Points)

The decision tree in Figure 1 was learned without regularization. How would the tree look like if one of the following regularization methods was applied.

- Minimum number of points per cell is set to 4.

- Maximum number of cells is set to 2.

- Maximum depth is set to 3.

Draw all three resulting trees.