

Machine Learning

Exercise Sheet 9

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

January 7th, 2014
Submission until January 14th, 13.00 to wistuba@ismll.de

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.

Create a binary decision tree using the method introduced in the lecture („greedy strategy“).

Use the misclassification rate as the split quality criterion and report all possible splits during the process.

Graphical Procedures in R (4 Points)

Read Chapter 12 of „An Introduction to R“.

- Which data types accepts `plot` as parameters?
- Describe two command that can extend high-level graphics in R.
- Which output formats are supported by R?
- How can interactive graphics be created using R?