Machine Learning
Exercise Sheet 4

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

November 11th, 2014
Submission until November 18th, 13.00 to wistuba@ismll.de

Exercise 13: Decision Trees (5 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 levels (root plus children).
Use the a) Information Gain and b) Gini Index as the split quality criterion, respectively.

Exercise 14: Decision Trees - Regularization (5 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.

Exercise 15: Perceptron (5 Points)

a)

x_1	x_2	x_3	class
4	3	6	negative
2	-2	3	positive
1	0	-3	positive
4	2	3	negative

Apply the perceptron learning algorithm until convergence on the given data. Use a step length $\alpha = 1$ and start with $\beta = 0$, $\beta_0 = 1$. Use the algorithm with a small difference: choose the training instances sequentially instead randomly (line 6).

b)

x_1	x_2	class
1	1	positive
1	0	negative
0	0	positive
0	1	negative

Show that the problem given in the table above cannot be solved with a single perceptron.

Exercise 16: SVM (5 Points)

D	a	b	c	d	e	f	g	h	i
x	-3	-2	-1	-0.5	0	0.5	1	2	3
Klasse	-1	-1	+1	+1	+1	+1	+1	-1	-1

a) Plot the data D. Is it possible to seperate

1. Which shape does a hyperplane have in the 1-dimensional space? Which in the 2-dimensional, which in the 3-dimensional space?
2. Plot the data D.
3. Is the data D linear separable? If yes, sketch the maximum margin hyperplane, the margin planes and the support vectors. If not explain briefly.

b) Given is the mapping function $h : \mathbb{R} \to \mathbb{R}^2$:

$$h(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$$

1. Apply h to the data D.
2. Plot the transformed data.
3. Is the data linear separable in the transformed space? If yes, sketch the maximum margin hyperplane, the margin planes and the support vectors. If not explain briefly.

c) Explain in your own words how a SVM is optimized by the submanifold minimization.

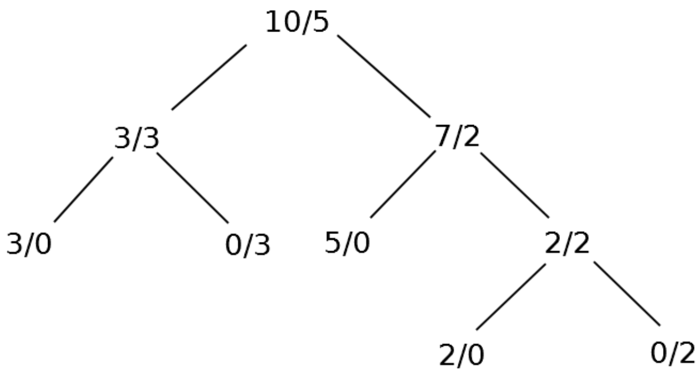

Abbildung 1: Decision tree for Exercise 14