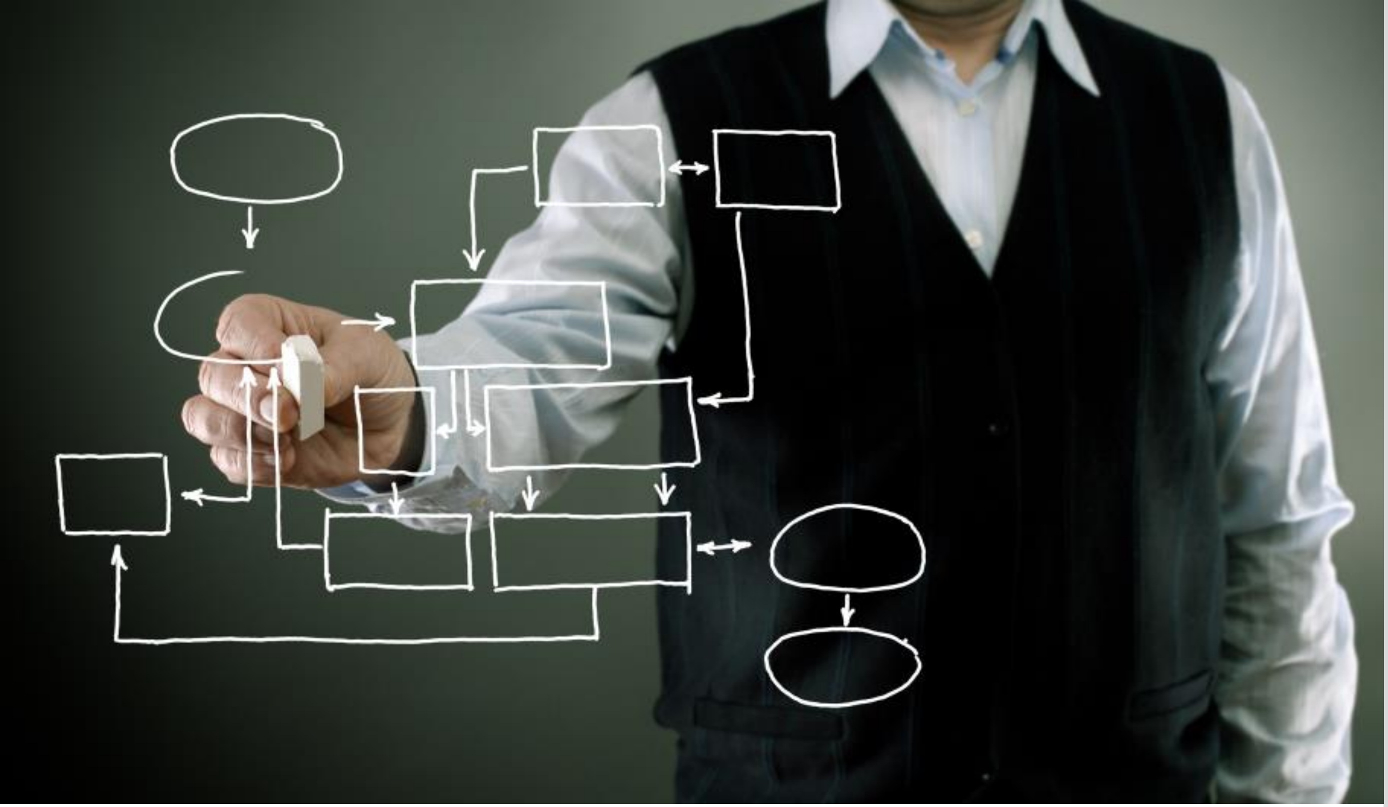


Basics of machine learning

Contact:
Dr. Michael Affenzeller
FH OOE - School of Informatics, Communications and Media
Heuristic and Evolutionary Algorithms Lab (HEAL)
Softwarepark 11, A-4232 Hagenberg

e-mail:
michael.affenzeller@fh-hagenberg.at
Web:
http://heal.heuristiclab.com
http://heureka.heuristiclab.com

HeuristicLab

OBERÖSTERREICH

What is machine learning/ data mining?

MACHINE LEARNING

“The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience“

(Tom Mitchell, 1997)

DATA MINING

“Data Mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data “

(Fayyad *et al*, 1996)

What is learning?

"Learning is making useful changes in our minds."
(Marvin Minsky, 1985)

"Learning denotes changes in a system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently next time."
(Herbert Simon, 1983)

"Learning means behaving better as a result of experience."
(Stuart Russell & Peter Norvig, 1995)

"Learning is constructing or modifying representations of what is being experienced."
(Ryszard Michalski, 1986)

Basics of machine learning

- **What is a system with the ability to learn?**

- **System (Program) with input and output *f*:**

$$x \longrightarrow \boxed{f} \longrightarrow y$$

- Search & Sort
- Shortest path
- Linear programming
- …

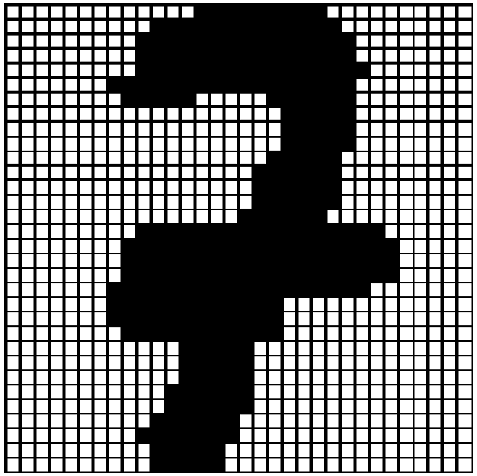
- Algorithmic solutions
- How to solve a problem?

- Pattern matching
- Robot playing soccer
- Diagnosis
- …

- Automatic adaption
- How to learn to solve a problem?

Basics of machine learning

Task: Recognize handwritten numbers



$\rightsquigarrow$ 7

Input: $X \in \{0, 1\}^{1024}$

Output: $Y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

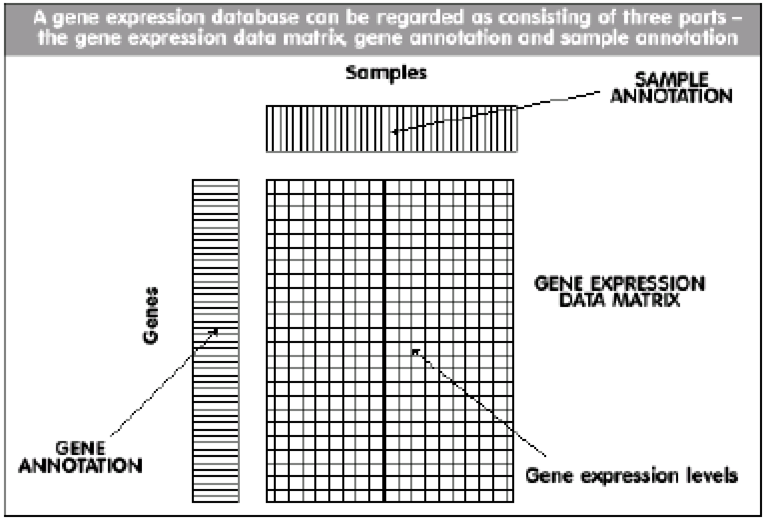
Task: Classification of man or woman?

Input: Picture of a person (as Bitmap)
Output: Sex of the person

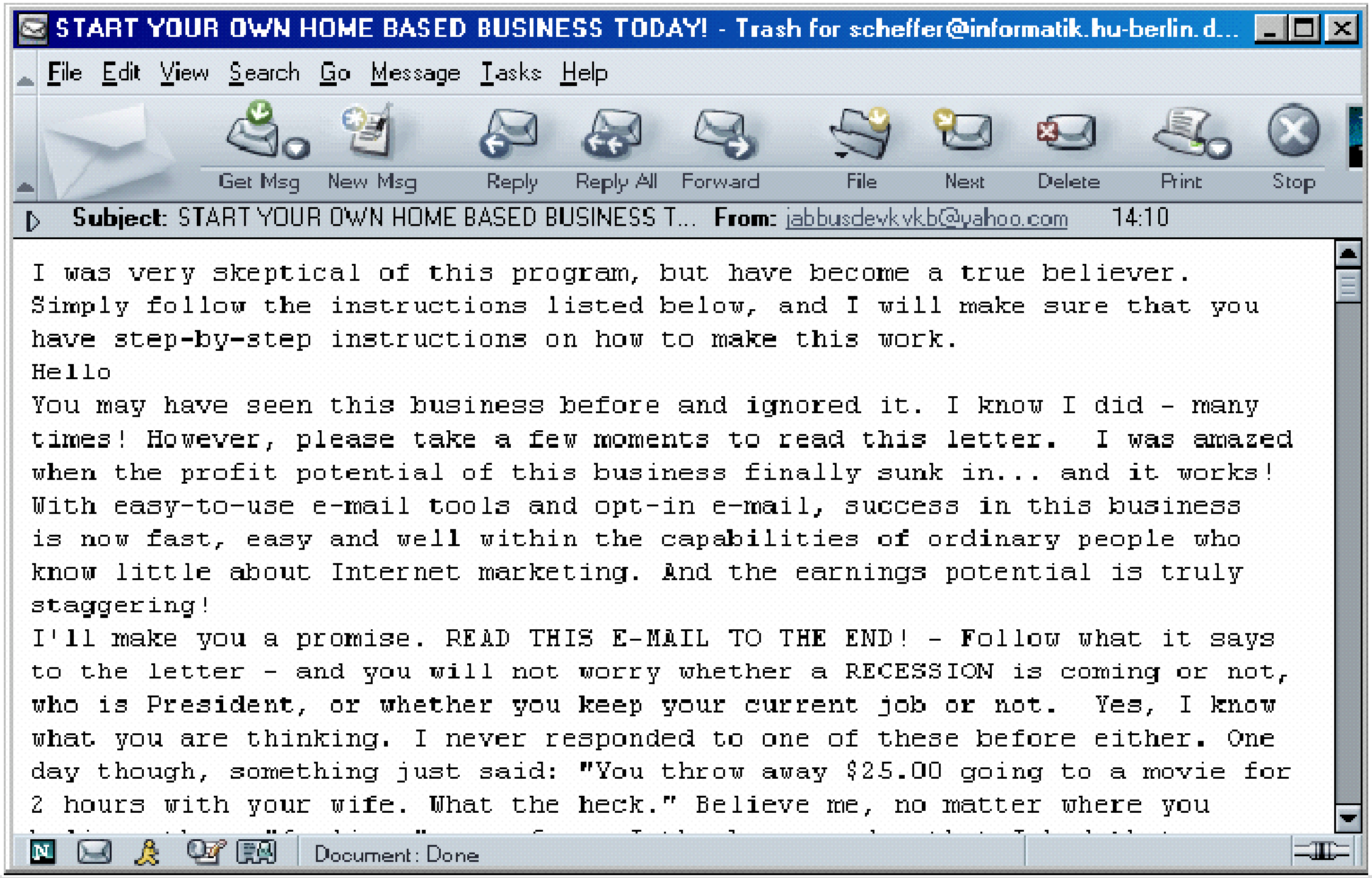
Task: Analyzing of credit risk (credit rating)

Task: Mining of gene expression data

- Analysis of gene databases
- Search for relations of the genes and the functions of the proteins



Task: Spam Filter



Training data: Input/Output

$X = ($0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0$)$

$Y = 7$

HeuristicLab

OBERÖSTERREICH

Training data: Input/Output

$$
\begin{array}{l}
X = (0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1 \\
\quad 1\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,1\,1\,1 \\
\quad 1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1 \\
\quad 1\,1\,1\,1\,1\,0\,0\,1\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1 \\
\quad 1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1 \\
\quad 1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1 \\
\quad 1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1 \\
\quad 1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\
\quad 0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0)
\end{array}
$$

$Y = 4$

HeuristicLab

OBERÖSTERREICH

Prognosis: Input given / Output wanted

$X = ($ 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0
0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1
1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 $)$

$Y = ???$

Feature Selection: Which features can be used to classify the data?

- Amount of ones
- Length of the longest sequence of ones
- Increase of the density of ones
- ……….

Lernen ist jede Veränderung eines Systems, die es ihm erlaubt, eine Aufgabe bei der Wiederholung derselben Aufgabe oder einer Aufgabe derselben Art besser zu lösen [Simon, 1983].

Lernen ist das Konstruieren oder Verändern von Repräsentationen von Erfahrungen [Michalski, 1986].

Formalization of learning

- Supervised learning:

  The system gets questions (Inputs) with correct answers (Outputs) with the goal that after the training the system can give correct answers to new questions.
  (Distinction between reinforcement and correcting learning)

- Unsupervised learning:

  System has to identify structure of data itself.
  Example: Clustering

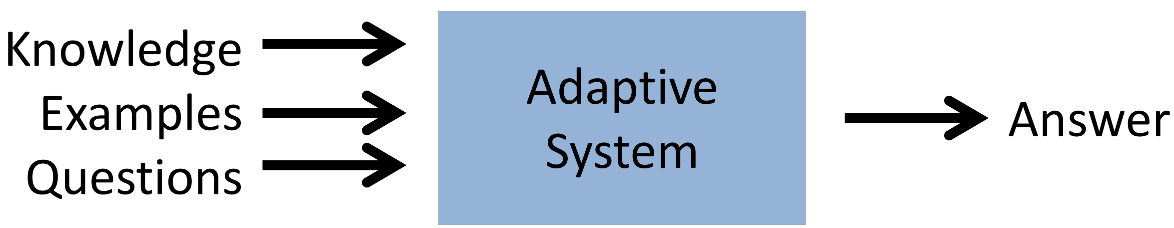
Basics of machine learning

Supervised Learning (Learn with Examples)

- Given: A set of examples (x_i, y_i)

$$S = \{(x_1, y_1), (x_2, y_2), \ldots , (xn, y_n)\} \subseteq x \times y$$

- Transduction: Output y for input x0 ??
- Induction: Complete functionally relation $f : X \rightarrow Y$??



Basics of machine learning

Unsupervised Learning (Example Clustering)

- Given: Set of data of patients
- Wanted: cluster patients with similar features (Symptoms)

Typical applications of SOMs (Self organizing maps)

Classification of variables/ attributes

- nominal/categorical: finite range, no order
  - Example: color ∈ {red, blue, green}
- ordinal: finite range, linear order
  - Example: hotel category ∈ {economy, mid scale, luxury}
- numerical: numerical (intercal or proportional) range
  - Example: temperature, size

Distinction of of task based on output

- Classification: Output is categorical
  - binary, multi-class, multi-label classification
- Ordinal classification: Output is ordinal
- Regression: Output is numerical

Examples:

- Two-class classification
- Multi-class classification
- Regression
- Time-series analysis

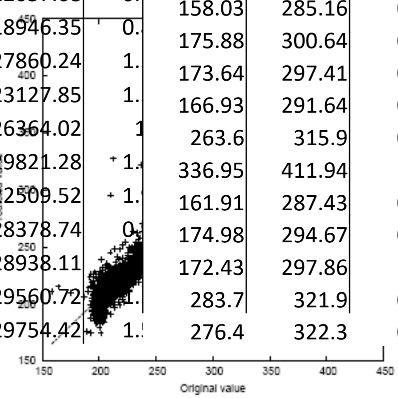
Data Based Modeling: Process

RP02						
199.6						
169.99	291.74	0.243	1.2125	75.75	13314.67	1
168.6	294.46	0.243	1.7025	70.75	13110.49	1.
171.8	296.7	0.243	1.5175	57	24066.61	1.
179.17	293.89	0.227	1.5225	58.5	24625.71	1
172.76	296.34	0.235	1.365	58.75	29176.93	1.
233.7	307.6	0.		54	29869.	
165.43	294.29	0.		54.25	24438.	
151.08	282.85	0.261	1.6475	59.5	3309.304	0.
158.8	287.48	0.251	1.5325	65	12057.68	0.
158.03	285.16	0.258	2.265	64	18946.35	0.
175.88	300.64	0.237	1.445	50.5	27860.24	1.
173.64	297.41	0.238	1.315	50.5	23127.85	1.
166.93	291.64	0.244	1.33	53.25	26364.02	1
263.6	315.9	0.206	1.37	49.25	29821.28	1.
336.95	411.94	0.15	1.885	52	22509.52	1.
161.91	287.43	0.251	1.4125	57.25	28378.74	0.
174.98	294.67	0.229	1.38	59	28938.11	1.
172.43	297.86	0.24	1.395	55.25	29560.72	1.
283.7	321.9	0.201	1.3325	57.75	29754.42	1.

LR, ANN, SVM, GP, ...

Modeling (Training)

Est



Application

Prognosis

2	ZUGFESTIGKEIT	VERFESTIGUNG	RAUHEITUS	SPITZENZAHLUS	MASSE	DRESSIERGRAD_MW	STRECKGRAD_MW	GESAMTVERFORMUNGSGRAD_MW	WALZKRAFT_MW
.64	359.12	0.25	1.3225	102.75	12035.59	0.6609	0	0.6612	?
169.99	291.74	0.243	1.2125	75.75	13314.67	1.041	0.0773	1.0407	?
168.6	294.46	0.24	1.7025	75	13110.49	1.2131	0.0232	1.2134	?
171.8	296.7	0.243	1.51	57	24066.61	1.0876	0.0712	1.0882	?
179.17	293.89	0.227		58.5	24625.71	1.311	0.0245	1.311	?
172.76	296.34	0.235	65	58.75	29176.93	1.2165	0	1.2168	?
233.7	307.6	0.214	1.325	54	29869.29	1.5809	0.0009	1.5789	?
165.43	294.29	0.251	1.325	54.25	24438.0		0.0383	1.0663	?
151.08	282.85	0.261	1.6475	59.5	3309.3		0	0.9504	?
158.8	287.48	0.251	1.5325	65	12057.		0.0409	0.8482	?
158.03	285.16	0.258	2.265	64	18946.		0.0518	0.8651	?
175.88	300.64	0.237		50.5	27860.		0.0166	1.302	?
173.64	297.41	0.238	1.	50.5	23127.		0.0143	1.3158	?
166.93	291.64	0.244	1.33	5	26364.		0.0002	1.1291	?
263.6	315.9	0.206	1.37		29821.		0.0231	1.563	?
336.95	411.94	0.15	1.885	52	22509.		0.0112	1.9382	?
161.91	287.43	0.251	1.4125	57.25	28378.		0.063	0.9263	?
174.98	294.67	0.229	1.38	59	28938.		0.018	1.2498	?
172.43	297.86	0.24	1.395	55.25	29560.		0.0258	1.2421	?
283.7	321.9	0.201	1.3325	57.75	29754.		0.1024	1.512	?
276.4	322.3	0.209	1.3525	54.75	17823.		0.1069	1.46	?

WALZKRAFT_MW
1498.92
1899.42
1704.12
2102.28
1955.39
2198.06
2003.71
1698.28
1697.23
1760.48
1848.21

Occam's Razor: **If a models M_1 explains a phenomena as good as another model M_2 and M_1 is simpler, then we choose M_1 (Preference of the simpler model)**

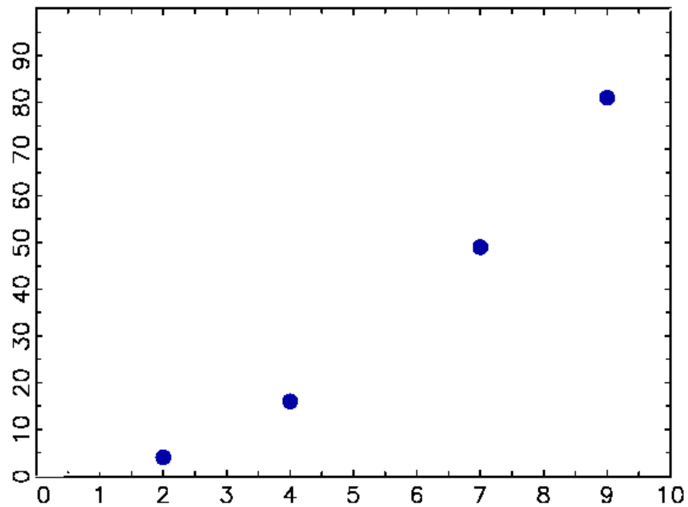
BSPL $f : \Re \rightarrow \Re$

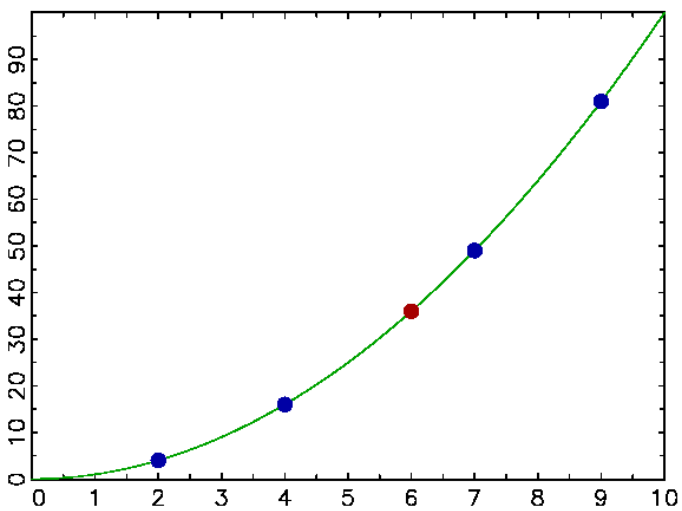
$$S = \{ (1,1), (3,5), (7,13), (10,19) \}$$

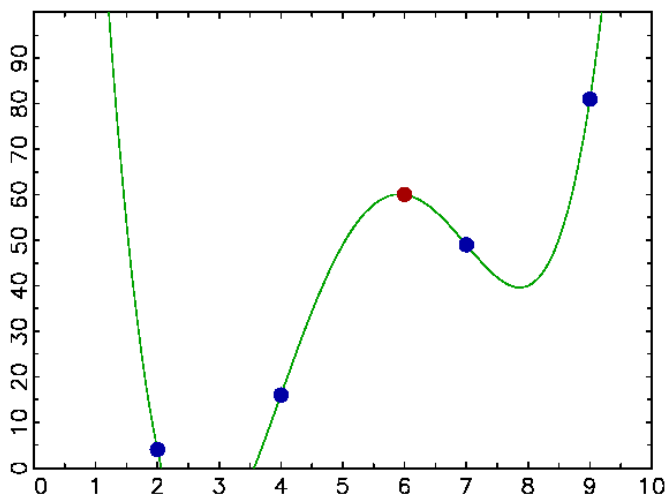
Hypothesis 1: $h : x \mapsto 2x - 1$

Hypothesis 2: $h : x \mapsto x^4 - 21x^3 + 141x^2 - 329x + 209$

Overfitting vs. Simplification (Concept)

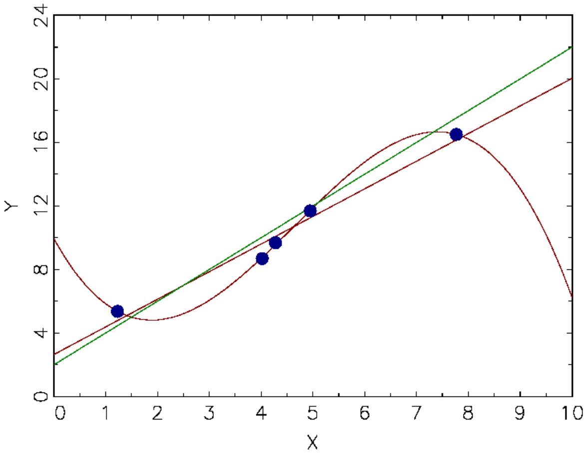






Problem: Overfitting vs. Underfitting

- Overfitting: Approximation of “noise“ in training partition leads to bad results in estimation partition

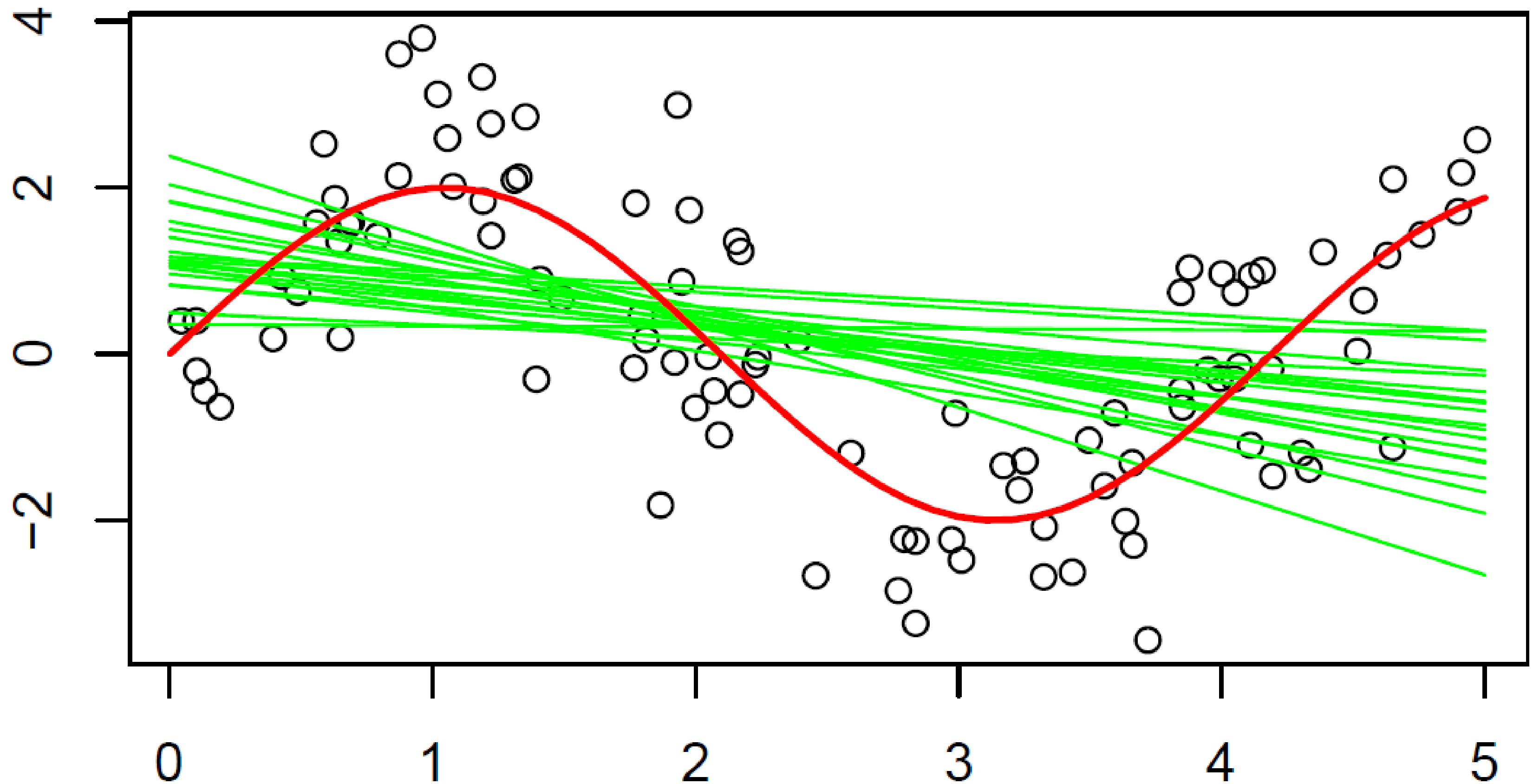


- Underfitting:
  If the complexity of the training model is to low to fully understand the problem, e.g. trying to solve nonlinear problems (a priori not known) with linear model

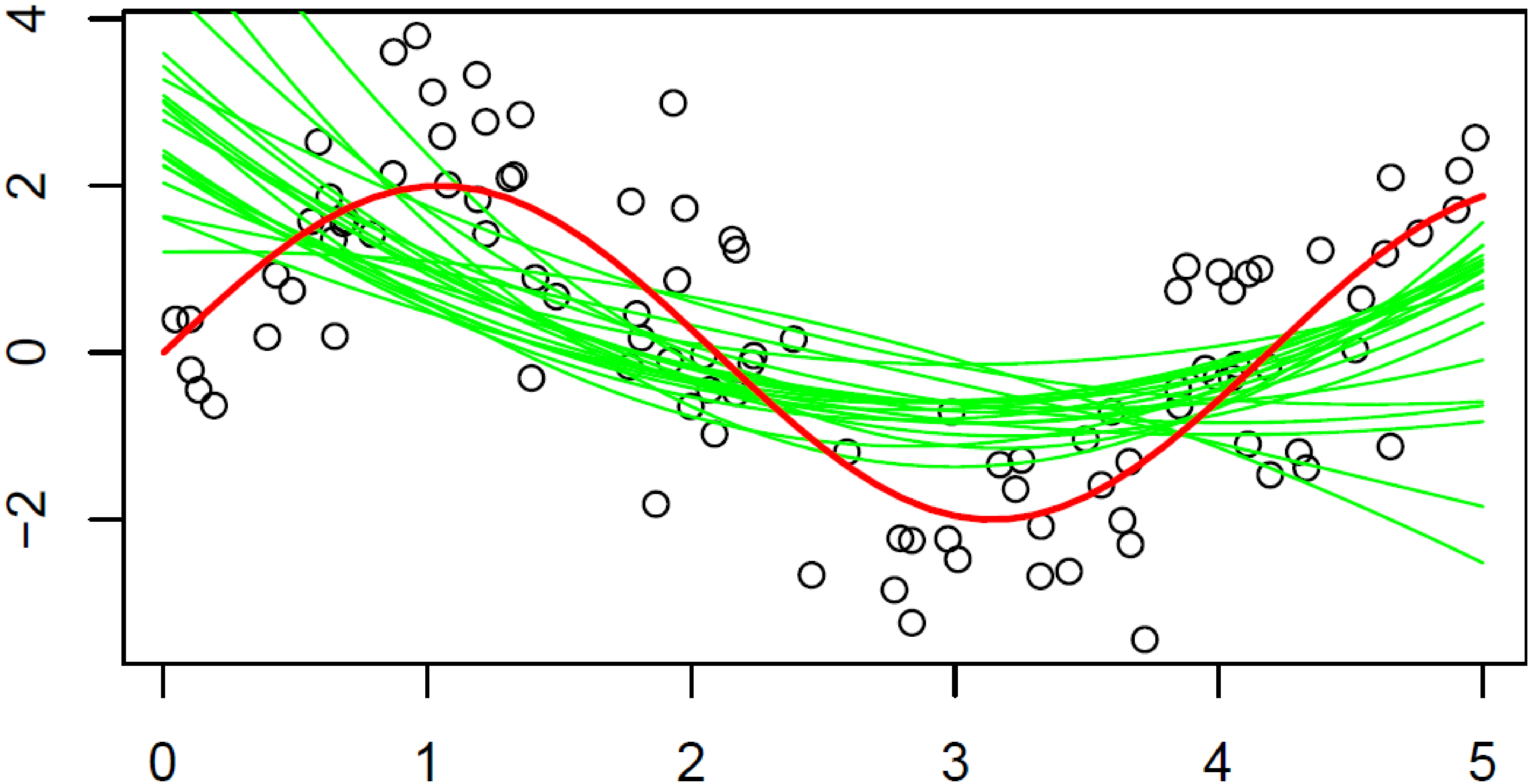
Basics of machine learning

Overfitting vs. Simplification (Concept)

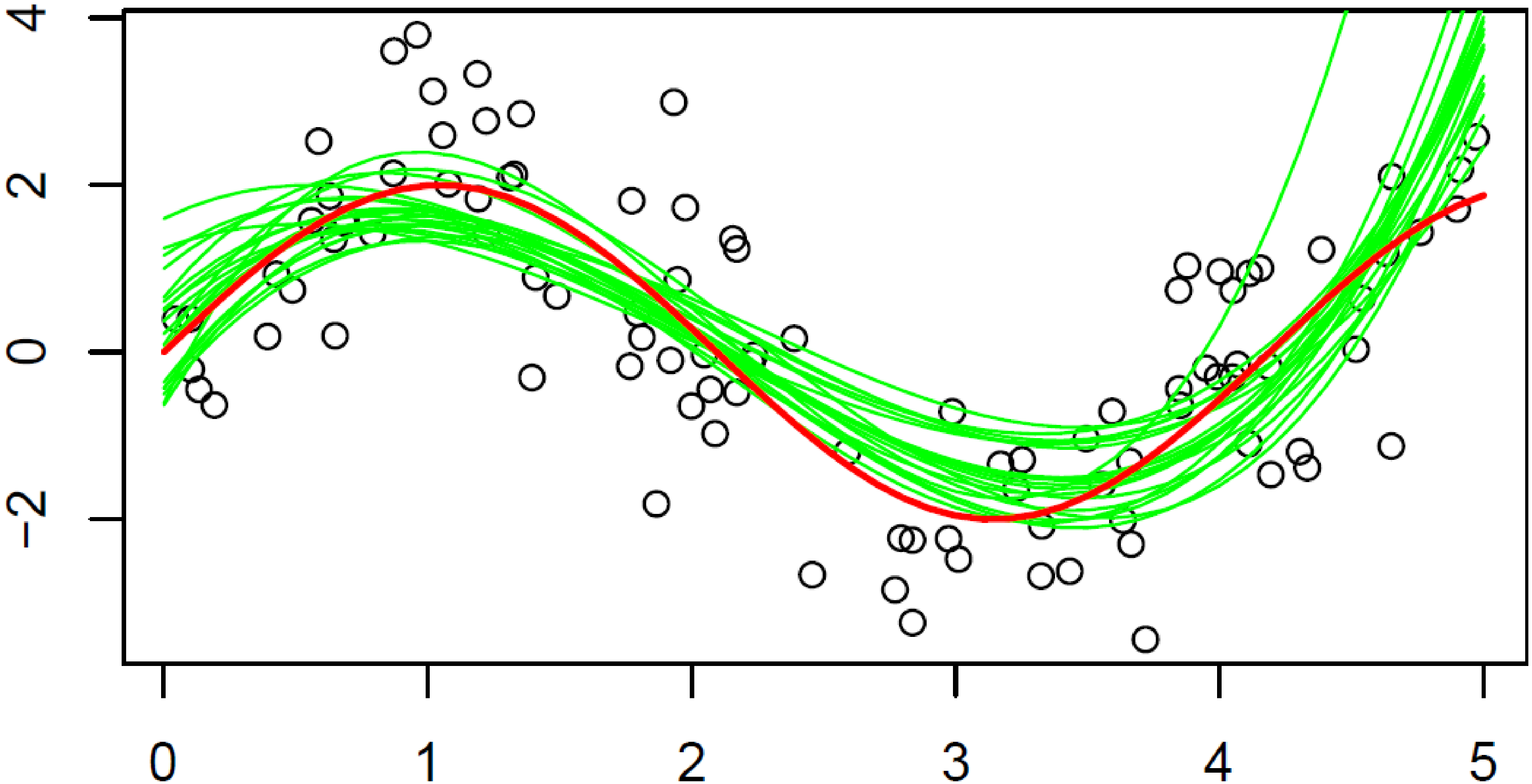
- Bias Variance compromise: Polynom (Degree 1)



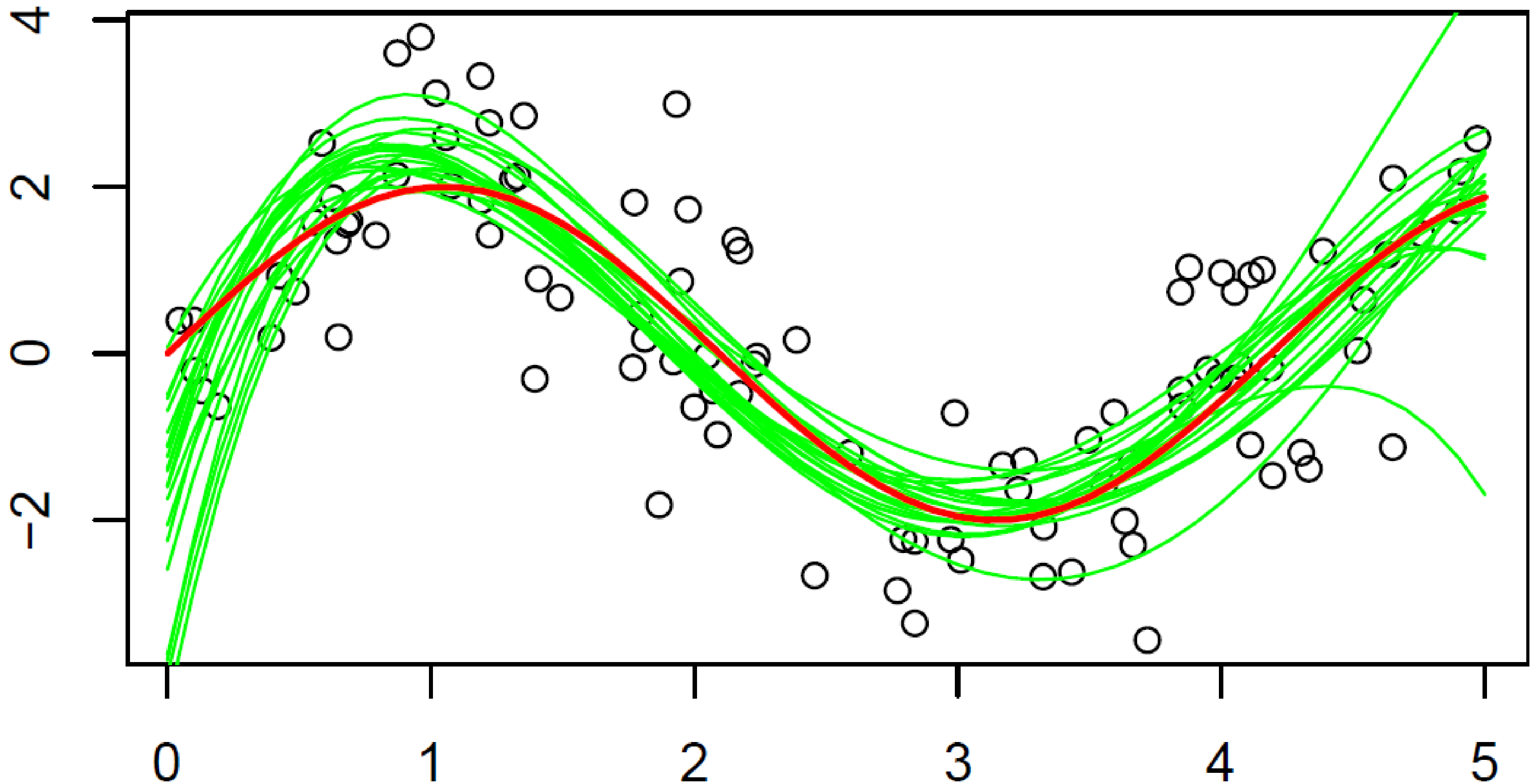
- Bias Variance compromise: Polynom (Degree 2)



- Bias Variance compromise: Polynom (Degree 3)

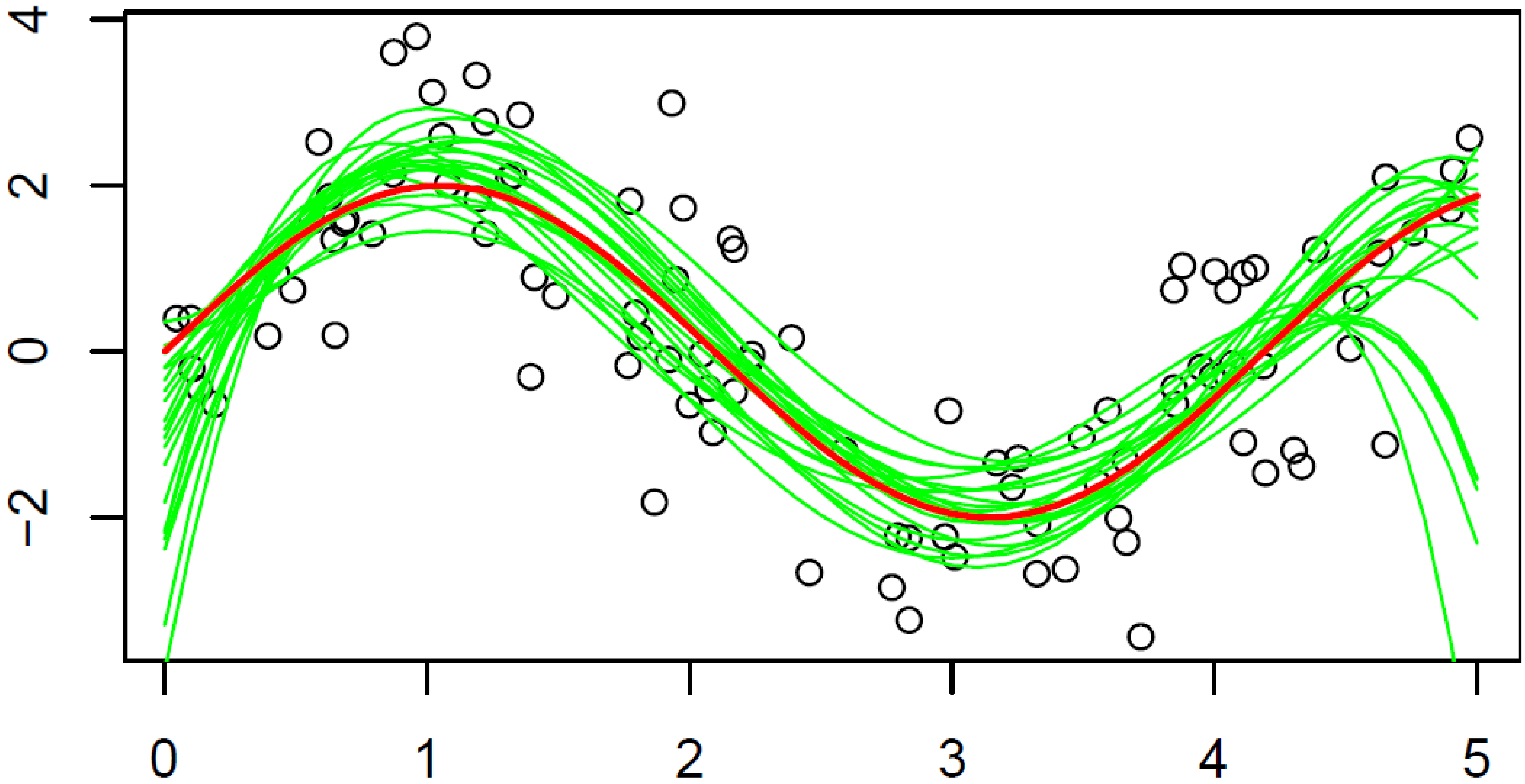


- Bias Variance compromise: Polynom (Degree 4)



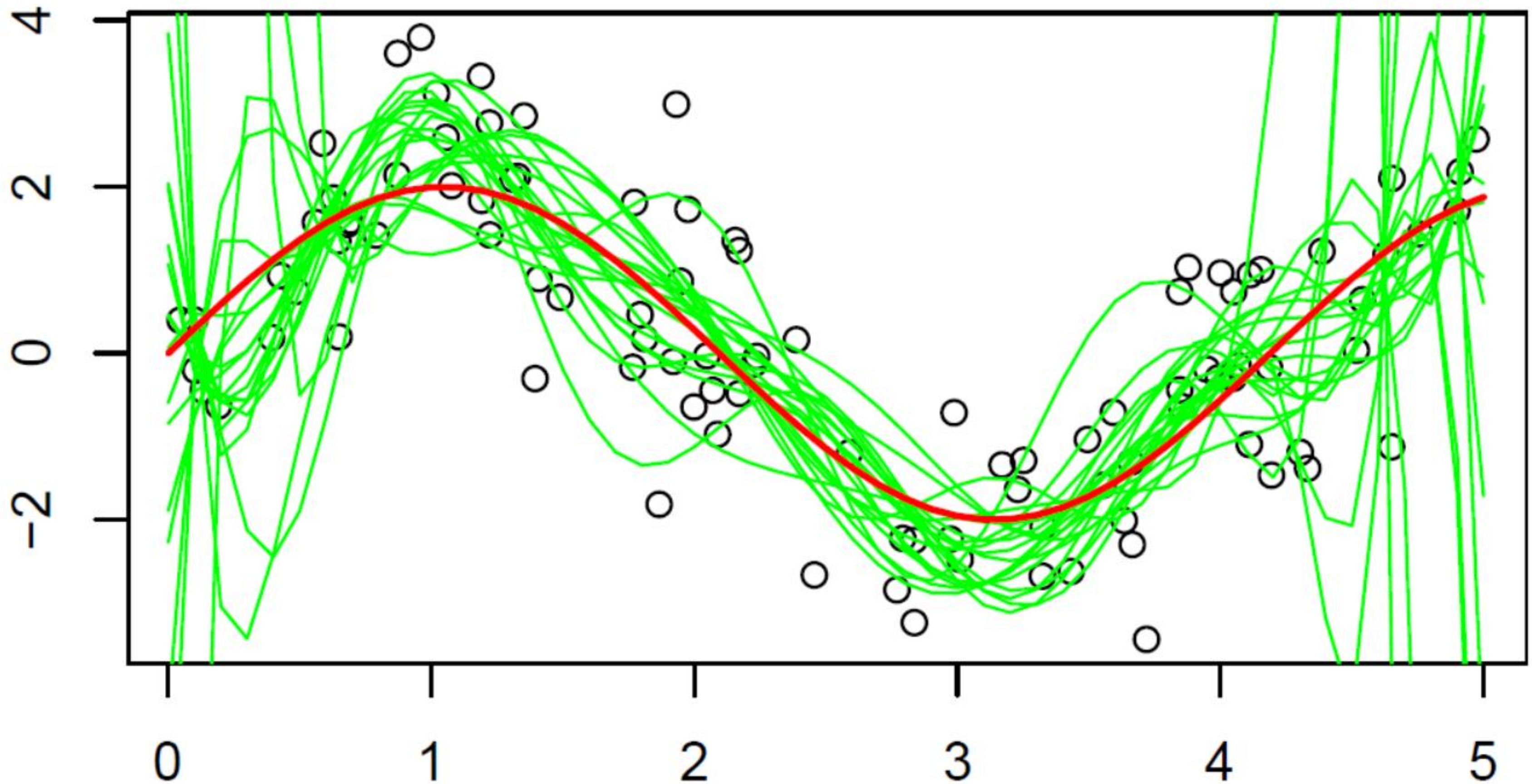
Basics of machine learning
Overfitting vs. Simplification (Concept)

- Bias Variance compromise: Polynom (Degree 5)



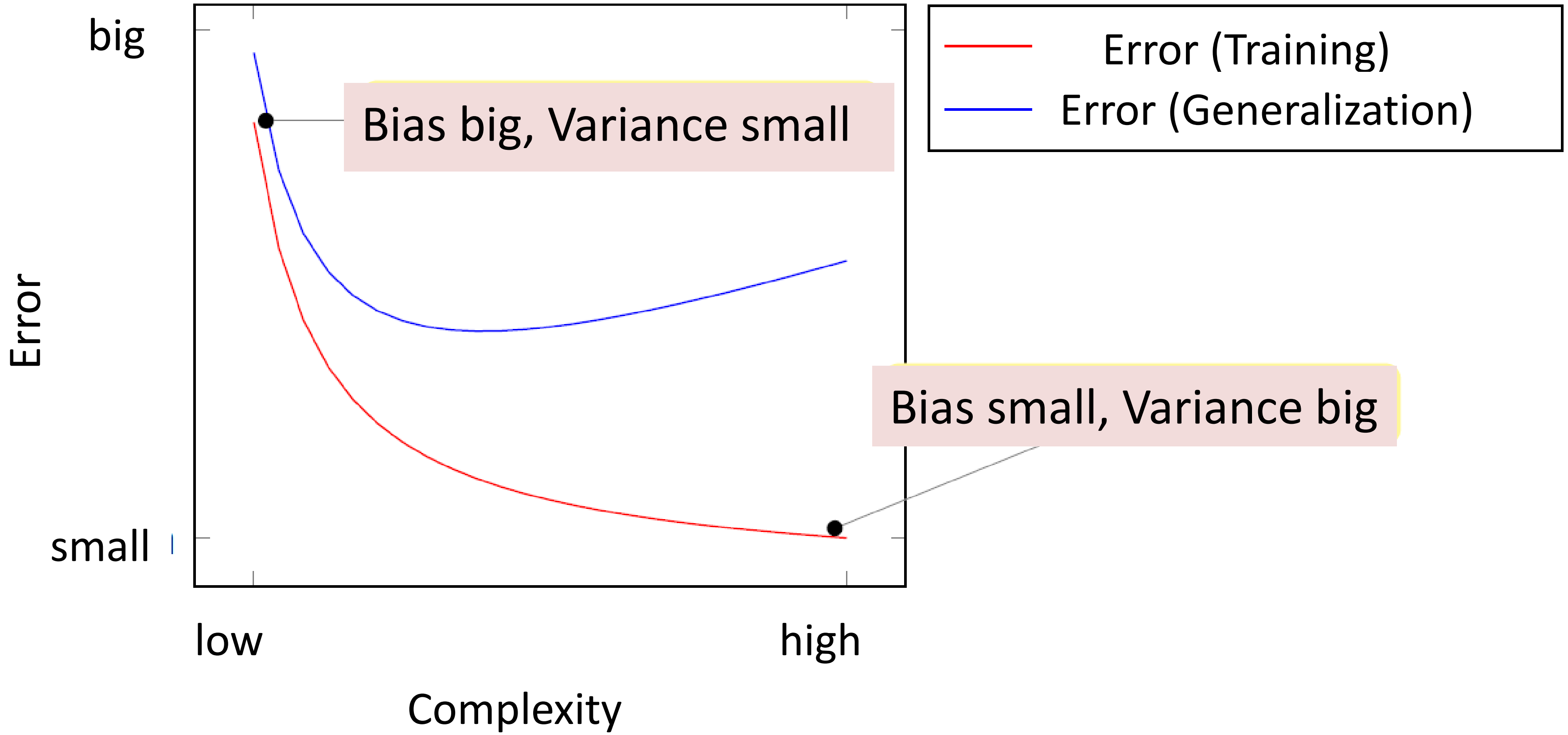
Basics of machine learning
Overfitting vs. Simplification (Concept)

- Bias Variance compromise: Polynom (Degree 10)

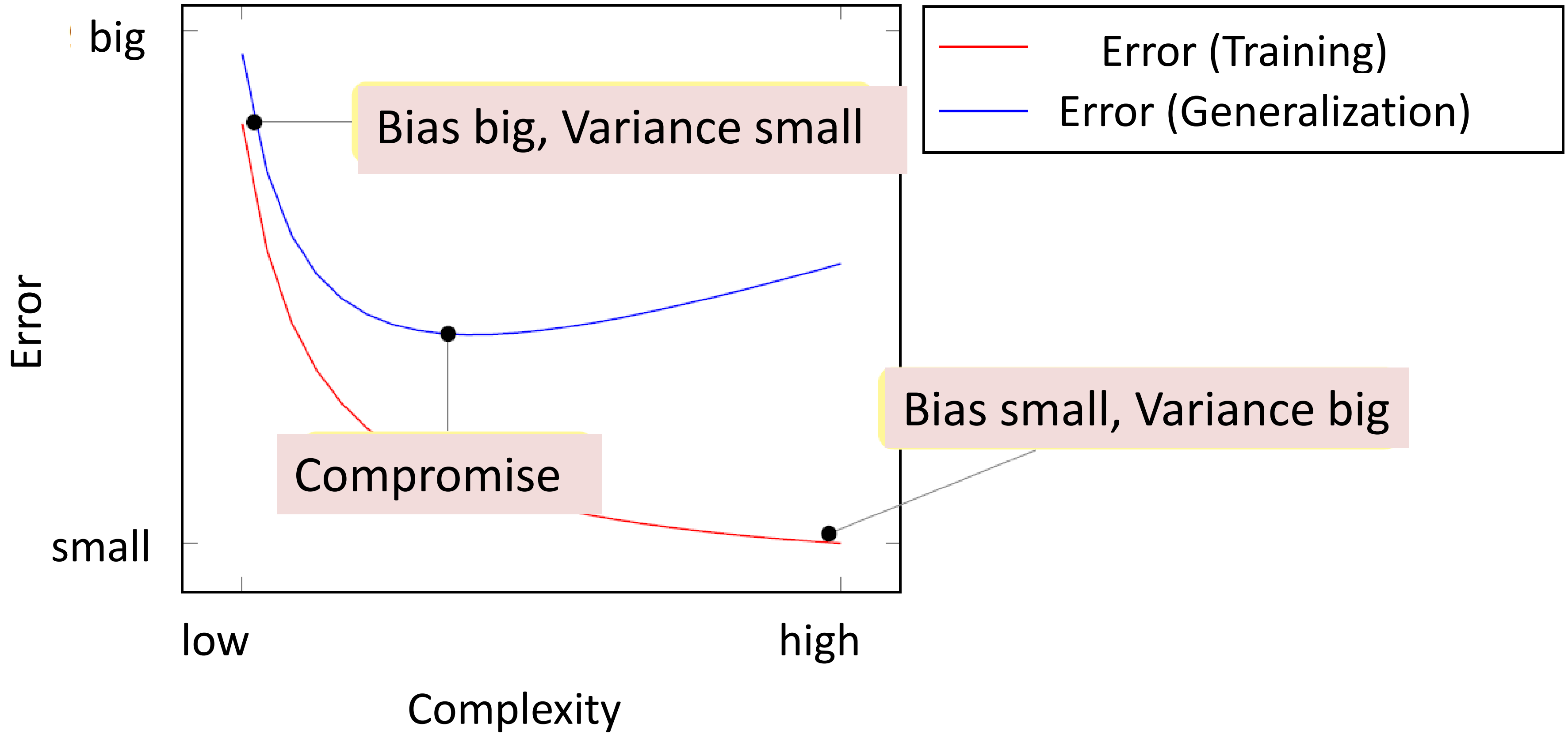


Basics of machine learning

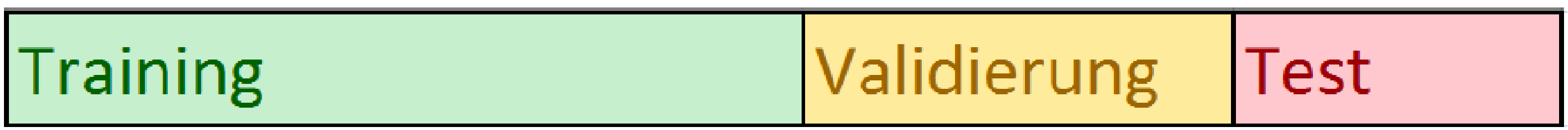
Overfitting vs. Simplification (Concept)



Overfitting vs. Simplification (Concept)



- **Error of generalization not known**
- **Estimation of error**
  - Reserved data partition for validation / test
  - Training on restricted partition



Problems: non-deterministic functions

Day	Month	Coin	Result
Monday	March	1 Euro	Tail
Wednesday	November	2 Euro	Tail
Saturday	November	2 Euro	Head
Sunday	December	1 Euro	Tail
Sunday	August	1 Euro	Tail
Monday	July	1 Euro	Head
Monday	June	2 Euro	Tail
Friday	May	2 Euro	Head

Seemingly explanation of examples with non-relevant features

- A 1 euro coin tossed on a Sunday results in tail.

Error probability

- Assumption: Instances becomes generated with a probability distribution p(x)
- Error probability: Probability of an instance, which has been misclassified by the hypothesis
- Can’t be measured, because p(x) and the result are unknown (otherwise we wouldn’t have to learn)
- Estimation of empirical error rate with data

Training error rate

- Number of misclassified training examples / number of all trainings examples
- Problem: Hypothesis is adjusted to the trainings data
- We would like to know how good the hyopthesis is for unknown instances
- Idea: Partition of trainings examples in tranings partition and test partition

Error estimation

- Algorithm (training and test)
  - 80% of examples: trainings partition
  - 20% of examples: test partition
  - h1 = learn algorithm trainings partition)
  - Determine E with test partition
  - h = learn algorithm (all examples)
  - Get hypothesis h with error estimator E

- Error estimation
- **Additionally a validation set can be used:**
  - Partition of data set in 3 groups (training, validation, test)
  - Search of hypothesis with trainings partition (as usual)
  - Different kind of usage of validation set:
    - Selection of best model eveluated with validation set (robustify)
    - Detection of overfitting
    - Regulation of hypothesis complexity

- Error estimation

- Better than Training/Test: cross-validation
- n-fold crossvalidation:
  - Partition data set in n groups
  - Use (n-1) of these groups to train
  - Use remaining one to test
  - Repeat procedure n-times, until every partition was used once for testing
  - Return value is the arithmetic mean of results in the test partition

Hypothesis space H:

- Influences the training success
- Allows contribution of knowledge
- The more complex, the greater the risk of overfitting
- Perhabs it doesn‘t contain the wanted function

Learning = Search for “good“ Hypotheses in H

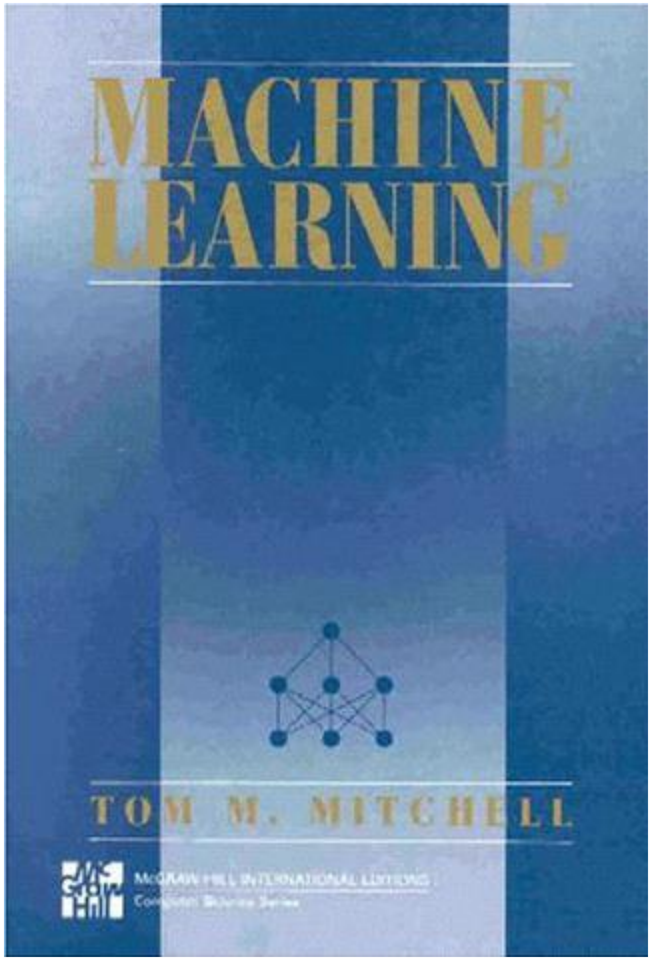
Example:

Class of degree of polynomial m:

$$\mathcal{H} = \left\{ x \mapsto \sum_{i=0}^{m} \alpha_i \cdot x^i \mid \alpha_0, \alpha_1, \ldots, \alpha_m \in \Re \right\}$$

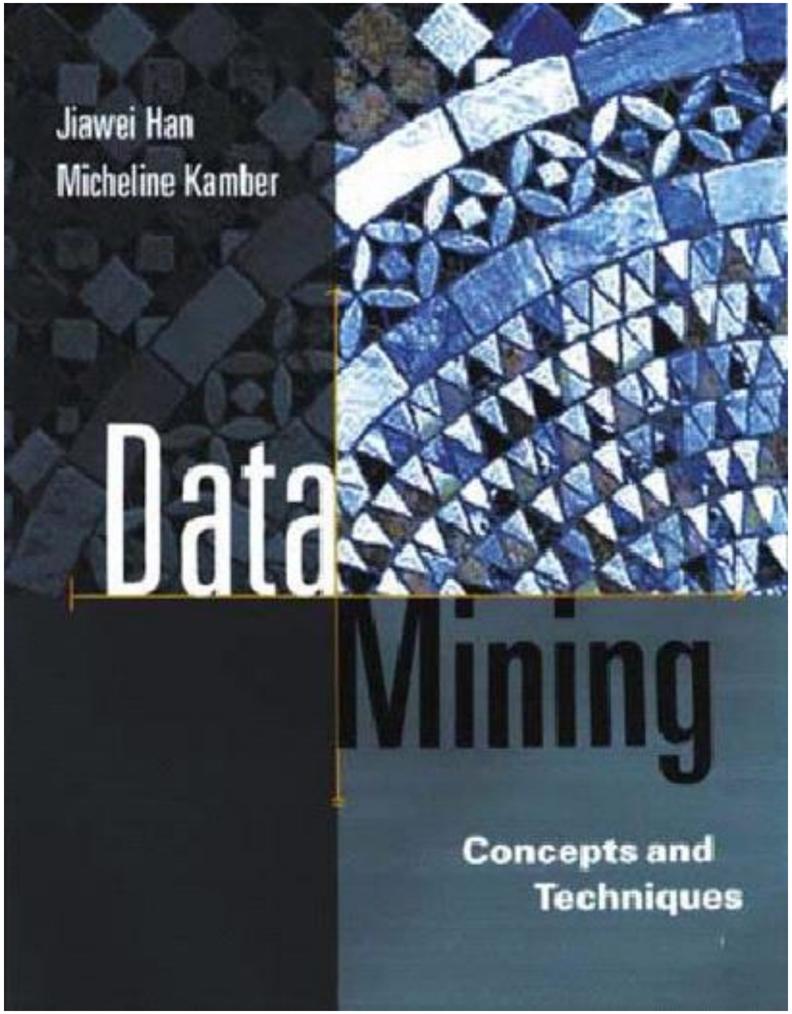
Literature:

- The classic
- Not up to date



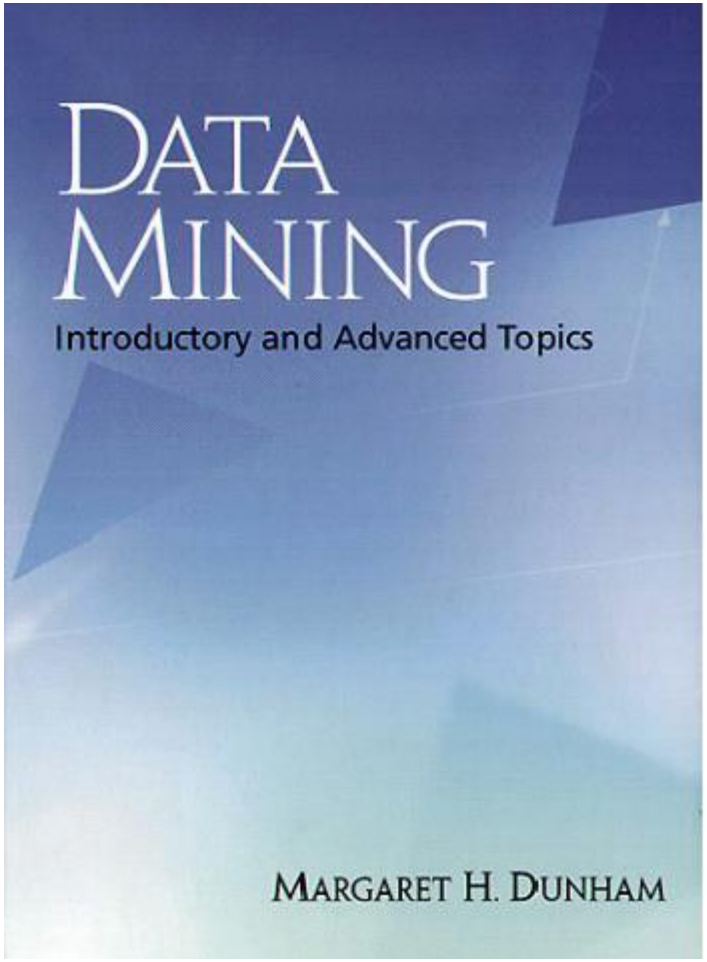
Literature:

- Comprehensive
- 60 € at Amazon.de



Literature:

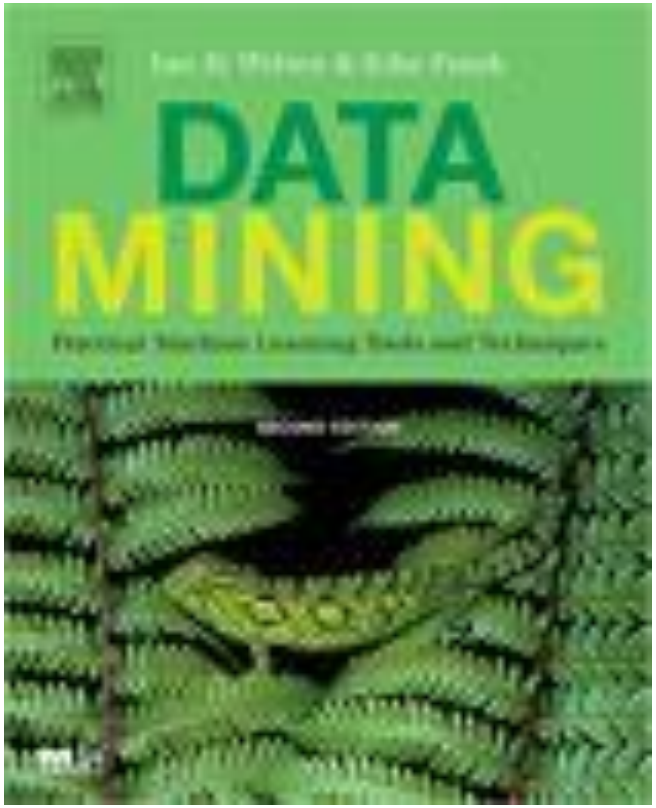
- Similar to Han, Kamber
- Not that detailed
- 58€ at Amazon.de



Basics of machine learning

Literature:

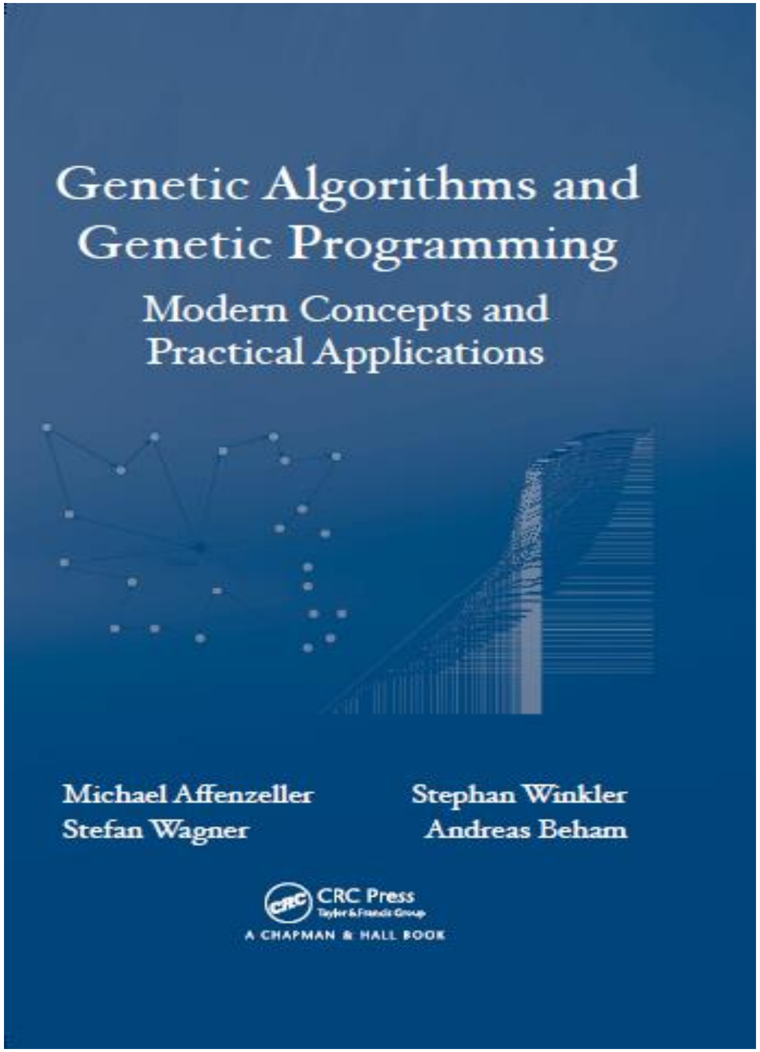
- Practice-oriented theory
- A lot of Tips and Hints about WEKA
- 48€ at Amazon.de



Basics of machine learning

Literature:

- Application of genetic programming for data analysis
- Advanced methods for non-linear system identification
- 64€ at Amazon.de

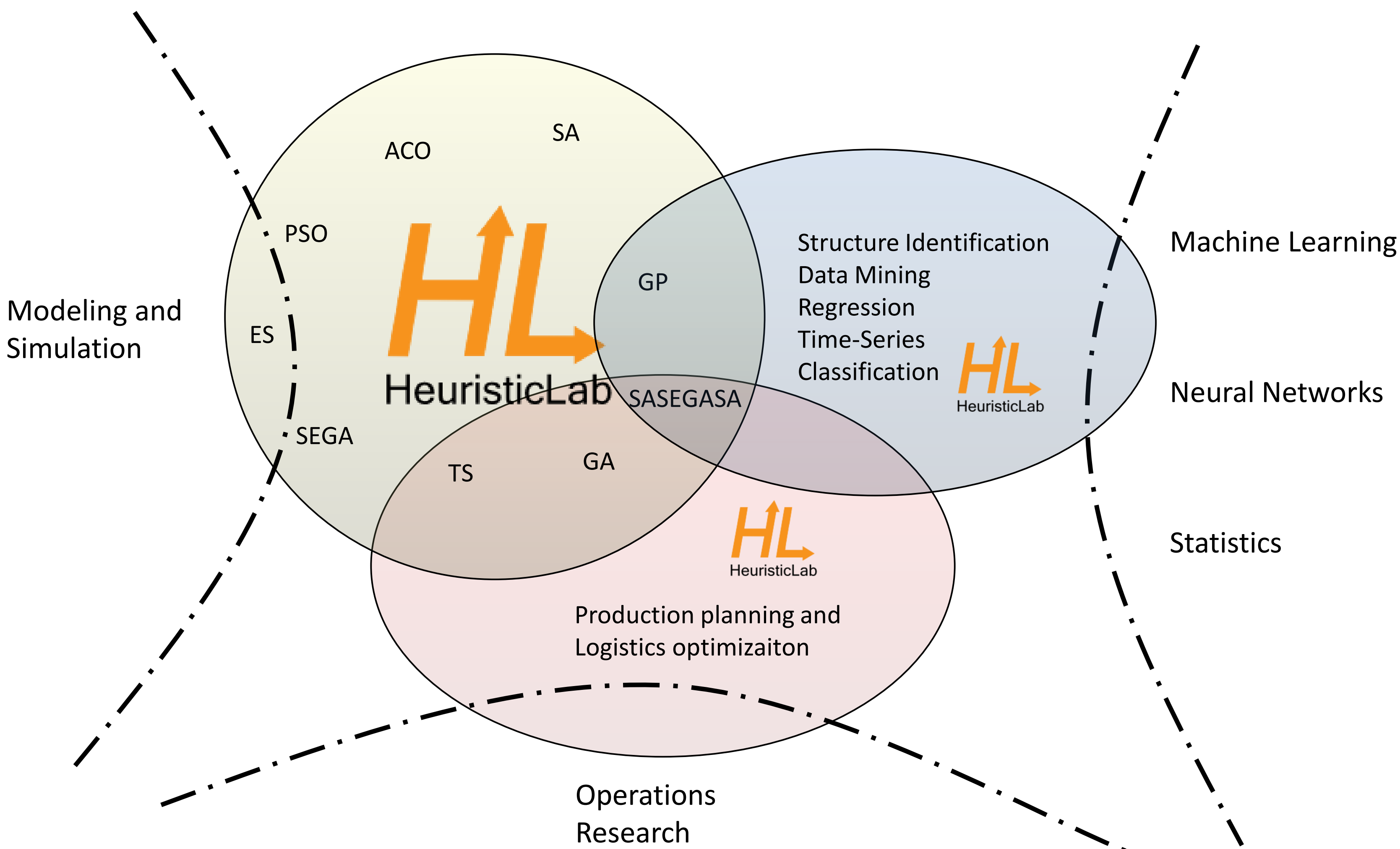


HeuristicLab
OBERÖSTERREICH

References:

- M. Affenzeller: Neuronale Netze (SS 04, JKU Linz)
- Eyke Hüllermeier: Methoden der Bioinformatik – Maschinelles Lernen (Uni Marburg)
- Rainer Malaka: Mustererkennung und Maschinelles Lernen (WS 04/05, Uni Karlsruhe)
- T. Scheffer/S. Bickel: Maschinelles Lernen und Data Mining (Humboldt Universität Berlin)
- Gerhard Widmer: Maschinelles Lernen (JKU Linz)
- GECCO 2005 – 2018, GPTP 2015 – 2018

Research Focus

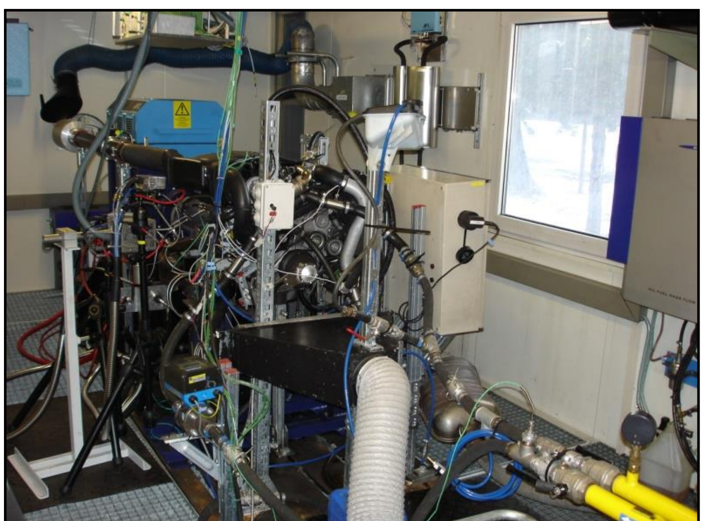


Example:
Virtual Sensors for Modeling Exhaust Gases

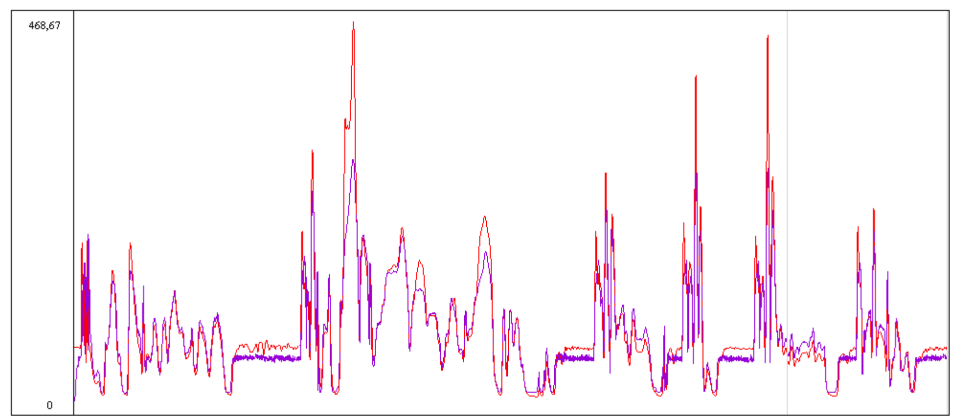
HeuristicLab
FH OBERÖSTERREICH

Motivation:

- High quality modeling of emissions (NO_x and soot) of a diesel engine
- **Virtual sensors:** (Mathematical) models that mimic the behavior of physical sensors
- Advantages: low cost and non-intrusive
- Identify variable impacts:
  - Injected fuel, engine frequency, manifold air pressure, concentration of O_2 in exhaustion etc.

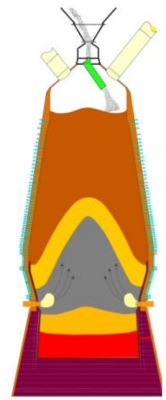


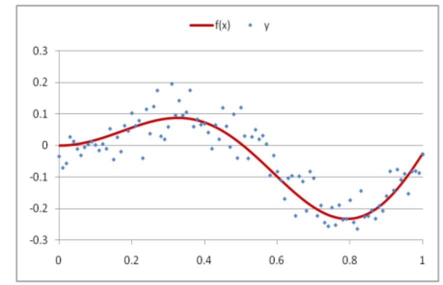
$NO_x(t) = f(x1_{(t-7)}, x2_{(t-2)}, \ldots)$



Example: Blast furnace modeling

x1	x2	x3	x4	x5	y
28.07845	13.93902	87.63394	20.07777	63.00267	250.4028
27.95657	12.75236	87.05083	19.95878	63.00894	440.0825
25.43135	23.03532	88.32881	21.98374	74.99575	292.6644
28.5034	36.71041	87.59461	20.55528	75.01106	100.8683
23.03413	46.5804	79.38985	18.67402	80.31421	435.7738
20.97957	41.52231	73.32074	21.49193	79.98517	288.5032
28.07431	28.49076	106.4166	27.38095	79.97826	?
28.00494	36.33813	104.7173	27.99428	75.00266	?
28.0274	31.84306	102.277	28.81878	78.1752	?
26.503	27.67078	93.81539	21.29002	62.99904	?
23.869	27.25298	93.67531	24.54099	80.00291	?

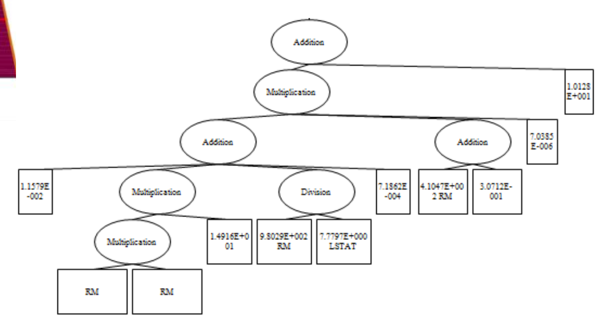




f(x)

Model

Prognosis



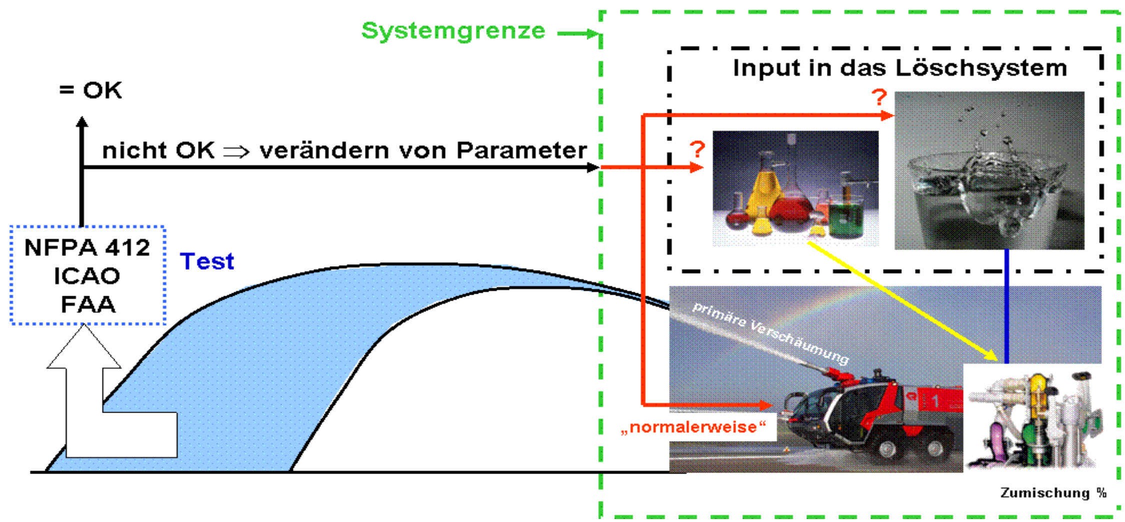
Innovations:

- Results as formulas → Domain experts can analyze, simplify and refine the models
- Integration of prior physical knowledge into modeling process
- Special interest of domain experts in model simplification and variable impact analysis

Example: Extinguishing systems

Goals

- Detect **relevant impact factors** and potential relationships between foam parameters with respect to throw range and foam quality
- **Model throw range and foam quality**
- Configure extinguishing systems for optimal throw range and foam quality



Example: Plasma Nitriding Modeling

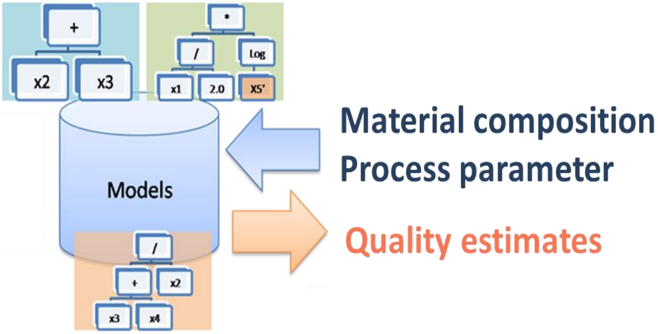
Motivation

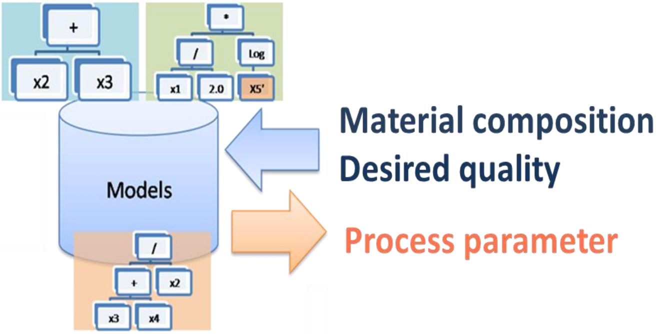
- Hardening of materials (e.g., transmission parts)
- Process parameter settings based on expert knowledge

Modeling Scenarios

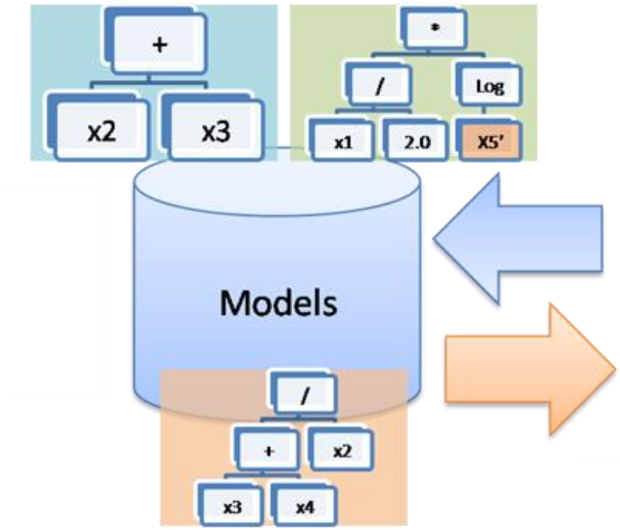
a) Prediction of quality values based on process parameters and material composition

b) Propose process parameter settings to reach the desired material characteristics

a) 

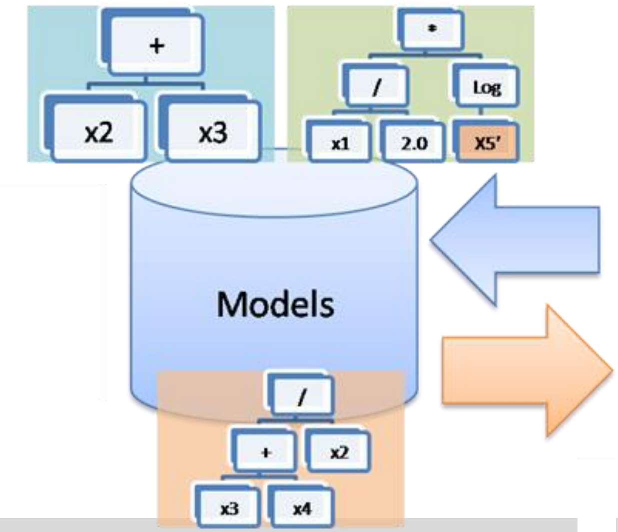
b) 

Example: Process Parameter Optimization



Material composition
Process parameter

Quality estimates



Material composition
Desired quality

Process parameter

Example: Medical Diagnosis

Motivation:

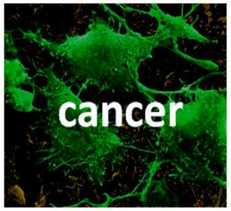
- **Research goal:** Identification of mathematical models for cancer diagnosis
- **Tumor markers:** substances found in humans (especially blood and / or body tissues) that can be used as indicators for certain types of cancer.

Data

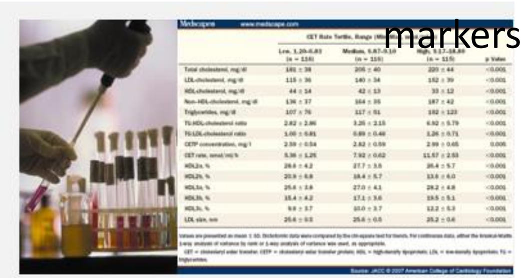
- Medical database compiled at the central laboratory of the General Hospital Linz, Austria, in the years 2005 – 2008
- Total: Blood values and cancer diagnoses for 20,819 patients

Modeling Scenarios

- Model virtual tumor markers using normal blood data
- Develop cancer diagnosis models using normal blood data
- Develop cancer diagnosis models using normal blood data and (virtual) tumor markers



Effects seen in data (blood examinations, tumor markers)



Cancer Diagnosis Estimation

Example: Medical Diagnosis: Data Preprocessing for Cancer Prediction

Patient	Date	[...Measured values...]	Diagnosis
1001	01.01.2004	[...]	-
1001	01.01.2005	[...]	-
1001	01.01.2006	[...]	-
1001	29.05.2007	[...]	-
1001	01.06.2007	[...]	-
1001	02.06.2007	[...]	-
1001	15.06.2007	[...]	**C61**
1001	02.07.2007	[...]	**C61**
1001	15.07.2007	[...]	**C61**
1001	02.09.2007	[...]	**C61**
1001	01.01.2008	[...]	**C61**
1001	05.02.2008	[...]	-
1001	01.03.2008	[...]	-
1001	17.03.2008	[...]	-
1001	01.01.2009	[...]	-

Healthy

Relevant measurements

Falsified (treatments, …)

Healthy (?)

Example: Medical Diagnosis: Symbolic Classification Ensembles

HeuristicLab

OBERÖSTERREICH

- Generate numerous forecasting models
- Reduce error caused by varianve
- Robustification of modeling results
- May be combined with various modeling techniques (e.g. random forests)

- May be combined with regression as well as with classification modeling
  - Averaging or prediction values for regression
  - Majority voting for classification

- Basis for enhanced confidence interpretation

Example: Medical Diagnosis: Symbolic Classification Ensembles

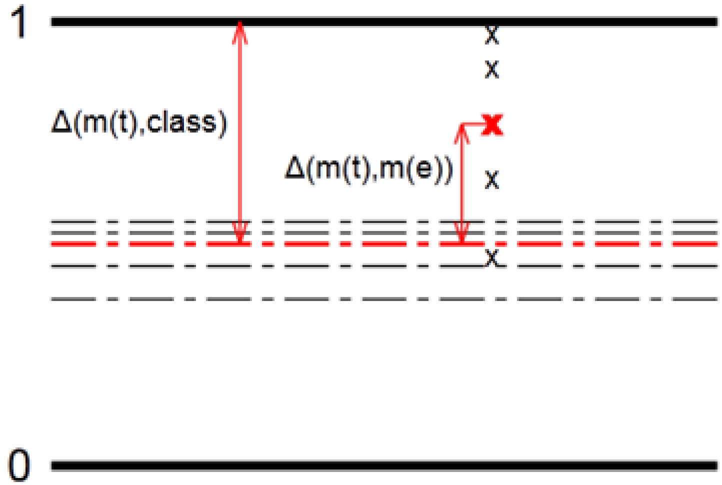
- **Based on clearness of majority voting**

$$cm_1 := 2\left(\frac{|votes(winning\ class)|}{|votes|} - 0{,}5\right) \in [0, 1]$$

- **Assuming for example 300 models for a 2-class classification problem:**
  - Vote of 300:0 results in $cm_1 = 1.0$
  - Vote of 150:150 results in $cm_1 = 0.0$
  - Vote of 200:100 results in $cm_1 = 1/3$
  - Vote of 250:50 results in $cm_1 = 2/3$

Example: Medical Diagnosis: Symbolic Classification Ensembles

- **Based on clearness of voting and on closeness of predictions**



$$cm_2 = \min\left(\frac{\Delta(m(t),m(e))}{\Delta(m(t),class)}, 1\right) \in [0, 1]$$

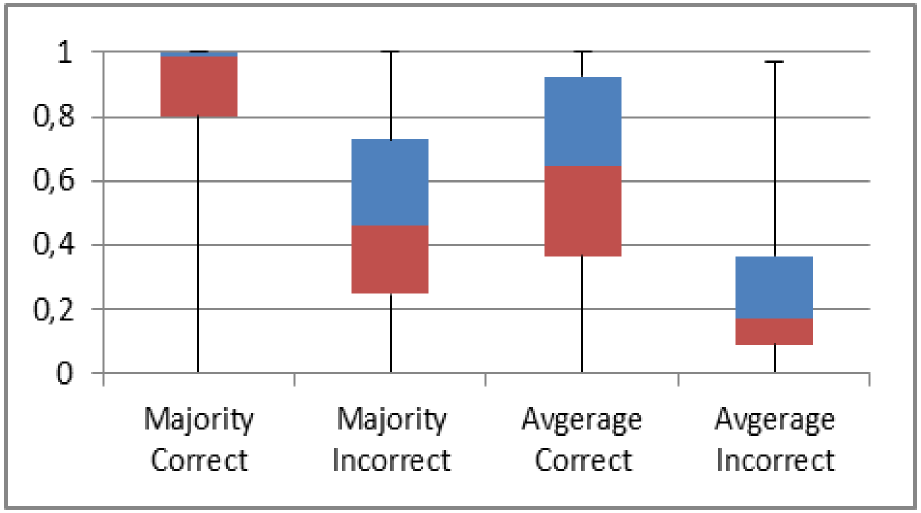
Example: Medical Diagnosis: Symbolic Classification Ensembles

Standard offspring selection GP modeling results

	Avg. training accuracy of 100 best models	Avg. test accuracy of 100 best models	Training accuracy of best model	Test accuracy of best training model
Breast with TM	83.17%	77.89%	84.74%	79.33%

Ensemble modeling results

	Majority Vote	Average Threshold
Accuracy training	84.99%	84.99%
Accuracy test	81.44%	81.44%
Average Confidence Correct Classified	$cm_1 = 0.8500$	$cm_2 = 0.6182$
Average Confidence Incorrect Classified	$cm_1 = 0.4806$	$cm_2 = 0.2449$
Confidence Delta	0.3694	0.3733



Example: Medical Diagnosis: Symbolic Classification Ensembles

- Test accuracy of best training model for breast cancer: 79.33%
- Test accuracy of ensemble model for breast cancer: 81.44%

Majority Vote

Test accuracy	Covered samples	Confidence
81.72%	95.33%	0.1
83.47%	88.24%	0.3
86.57%	78.05%	0.5
89.88%	69.97%	0.7
92.93%	56.09%	0.9
95.45%	46.74%	0.95
98.00%	35.41%	1