

# Machine Learning

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)  
Institute for Business Economics and Information Systems  
& Institute for Computer Science  
University of Hildesheim  
<http://www.isml.uni-hildesheim.de>

---

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

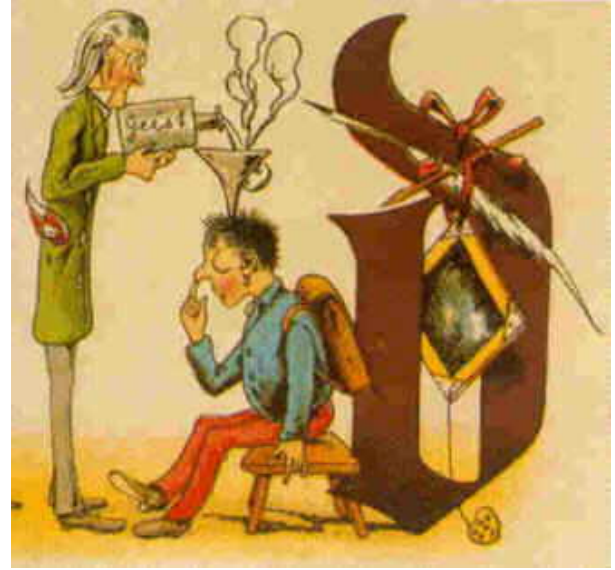
1/34

## 1. What is Machine Learning?

## 2. Overview

## 3. Organizational stuff



# What is Machine Learning?



# What is Machine Learning?

## 1. Information Systems: predict what customers will buy.

→ **Person**  
Für wen ist das Geschenk?

weiblich männlich









---

→ **Lebensphase?**  
In welcher Lebensphase befindet sich der/die Beschenkte?

Kind    jugender Erwachsener    reifer Erwachsener    Senior

---





→ **Interesse?**  
Welches Hauptinteresse hat die/der zu Beschenkende?

Business & Karriere    Design & Eleganz    Kultur & Kunst    Leben & Genießen    Natur & Entspannung    Körper & Schönheit    Sport & Fitness    Technik & Basteln

---

→ **Weitere Interessen?**  
Welche weiteren Interessen hat sie/er?

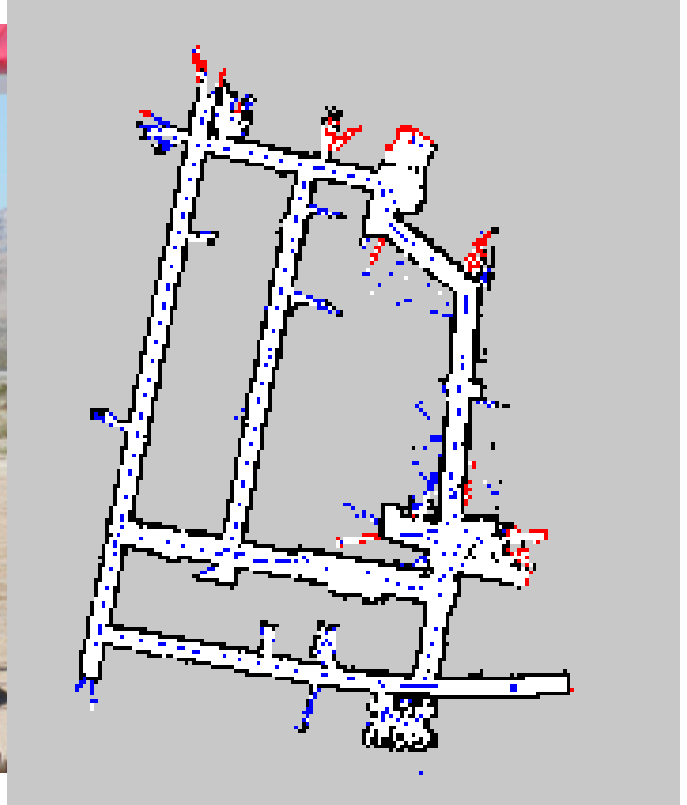
   

Business & Karriere    Design & Eleganz    Kultur & Kunst    Leben & Genießen

Done

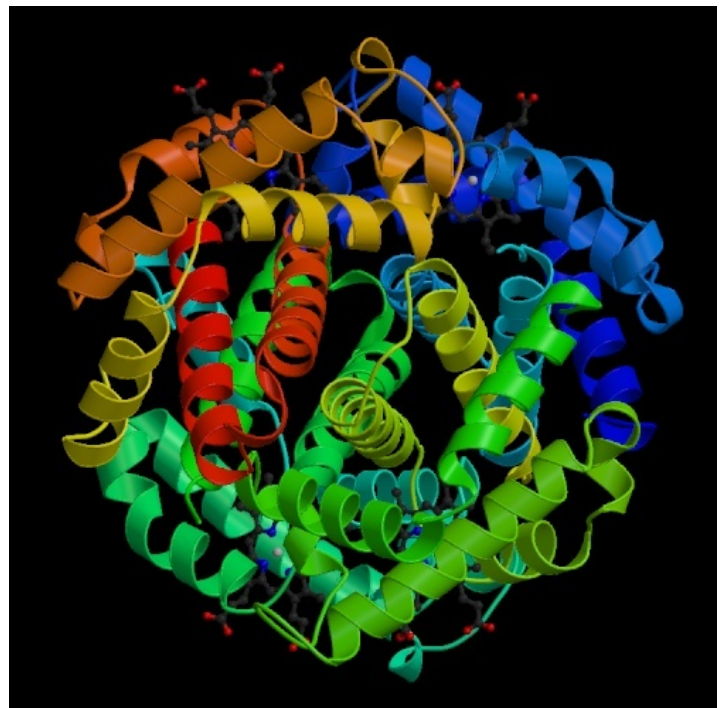
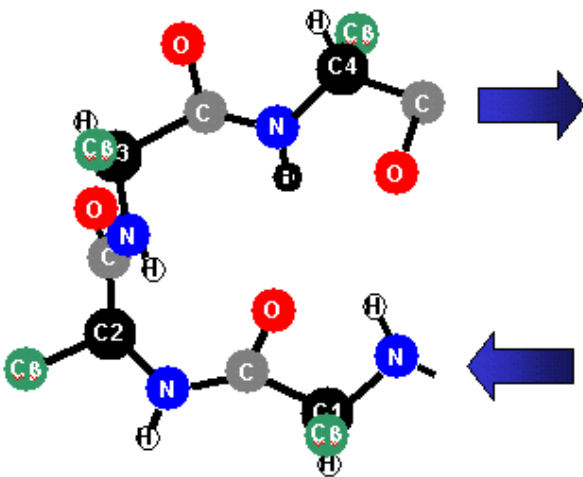
# What is Machine Learning?

2. Robotics: Build a map of the environment based on sensor signals.



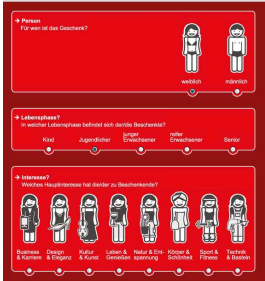
# What is Machine Learning?

3. Bioinformatics: predict the 3d structure of a molecule based on its sequence.



# What is Machine Learning?

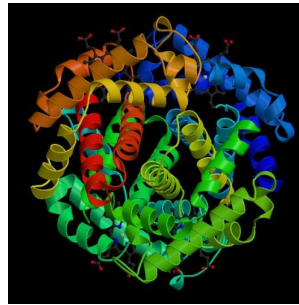
## Information Systems



## Robotics

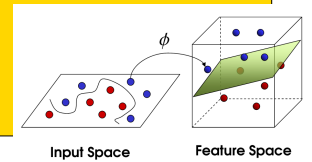


## Bioinformatics



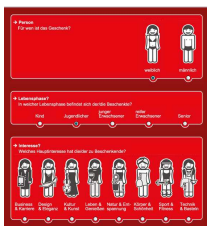
Many Further Applications!

**MACHINE LEARNING**



# What is Machine Learning?

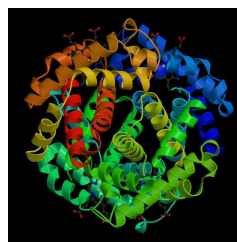
## Information Systems



## Robotics



## Bioinformatics



Many Further Applications!

**MACHINE LEARNING**

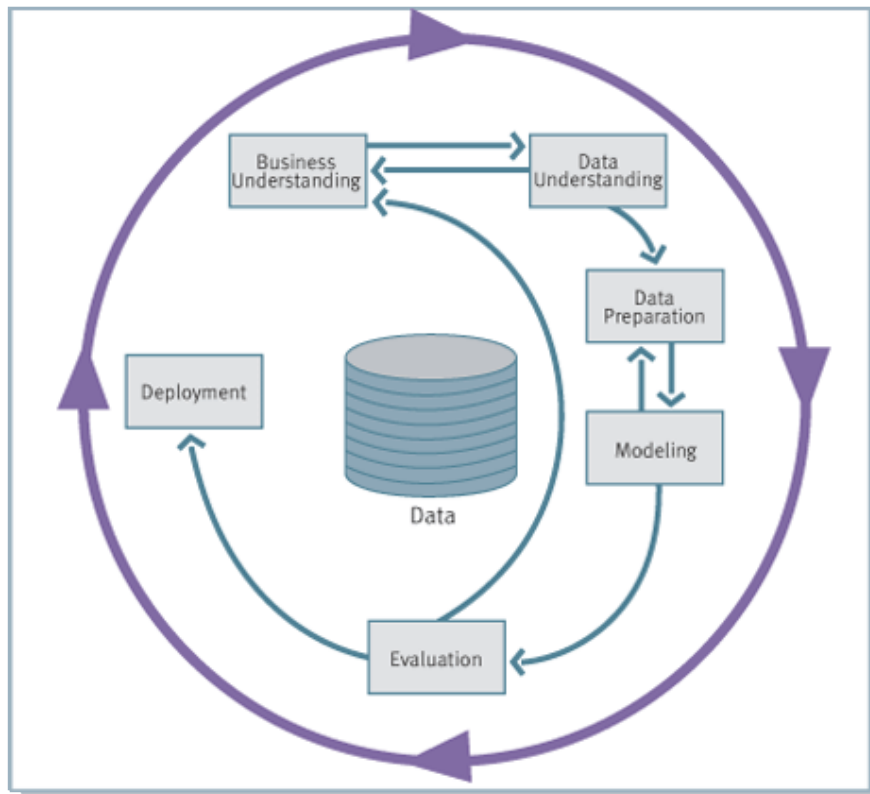
**OPTIMIZATION**

**NUMERICS**

**ALGORITHMICS**



## Process models



Cross Industry Standard Process for Data Mining (CRISP-DM)

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

7/34

## Machine Learning / 1. What is Machine Learning?

One area of research, many names (and aspects)

**machine learning**

historically, stresses learning logical or rule-based models  
(vs. probabilistic models).

**data mining**

stresses the aspect of large datasets and complicated tasks.

**knowledge discovery in databases (KDD)**

stresses the embedding of machine learning tasks in  
applications, i.e., preprocessing & deployment; data mining is  
considered the core process step.

**data analysis**

historically, stresses multivariate regression methods and  
many unsupervised tasks.

**pattern recognition**

name preferred by engineers, stresses cognitive applications  
such as image and speech analysis.

**applied statistics**

stresses underlying statistical models, testing and methodical  
rigor.

# 1. What is Machine Learning?

## 2. Overview

### 3. Organizational stuff

## Machine Learning Problems

1. Density Estimation
  2. Regression
  3. Classification
  4. Optimal Control
  5. Clustering
  6. Dimensionality Reduction
  7. Association Analysis
- } Supervised Learning
- } Reinforcement Learning
- } Unsupervised Learning

## 1. Density Estimation

Example 1: duration and waiting times for eruptions of the “Old Faithful” geyser in Yellowstone National Park, Wyoming (Azzalini and Bowman 1990).

continuous measurement from August 1 to August 15, 1985:

- duration (in min.),
- waiting time (in min.)

duration:

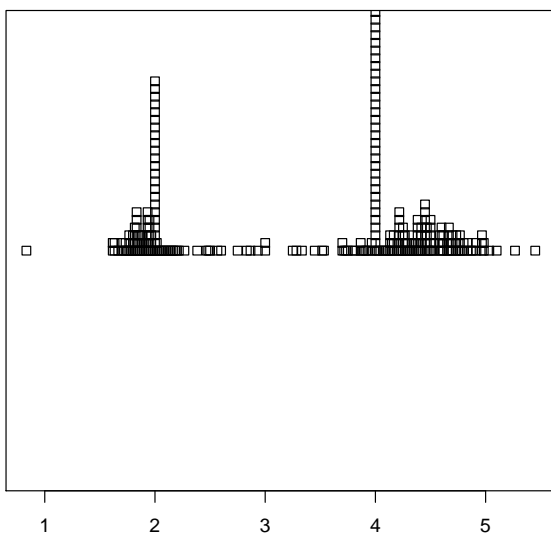
4.016667, 2.15, 4.0, 4.0, 4.0, 2.0,  
4.383333, 4.283333, 2.033333,  
4.833333, ...

What is a typical duration? waiting time?

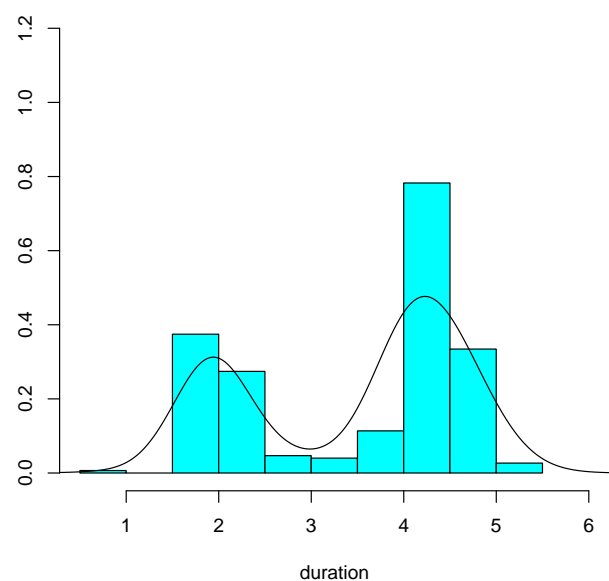


## 1. Density Estimation

durations: 4.016667, 2.15, 4.0, 4.0, 4.0, 2.0, 4.383333, 4.283333, ...

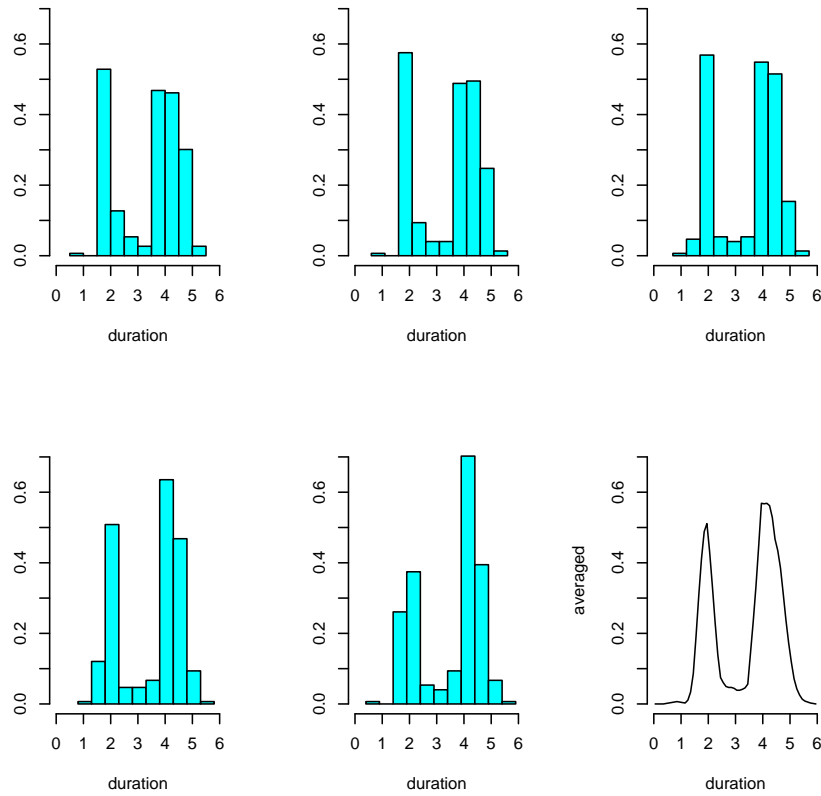


strip chart

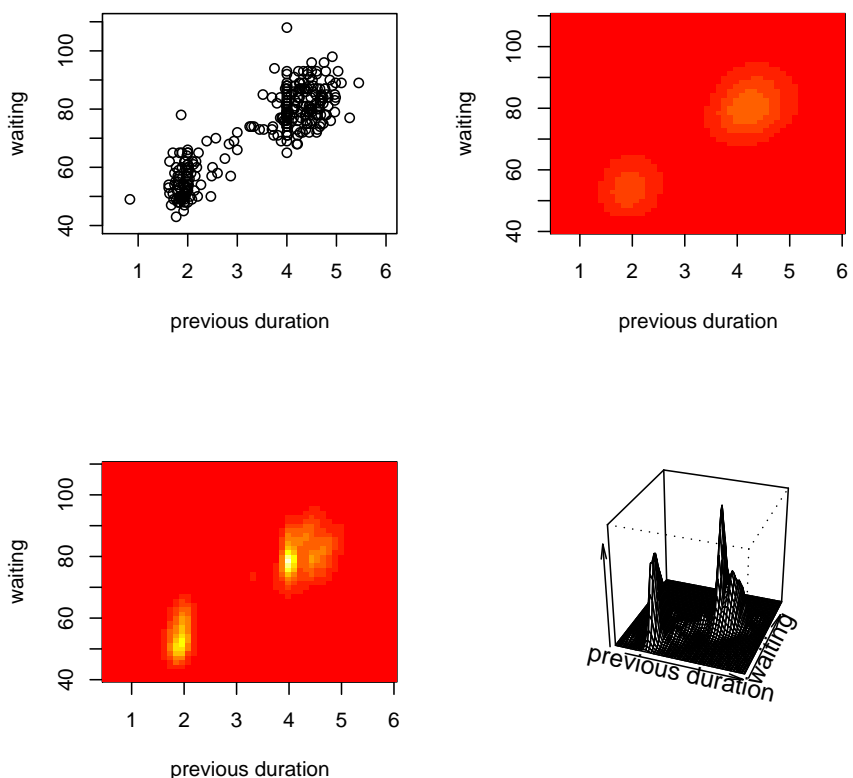


histogram

# 1. Density Estimation



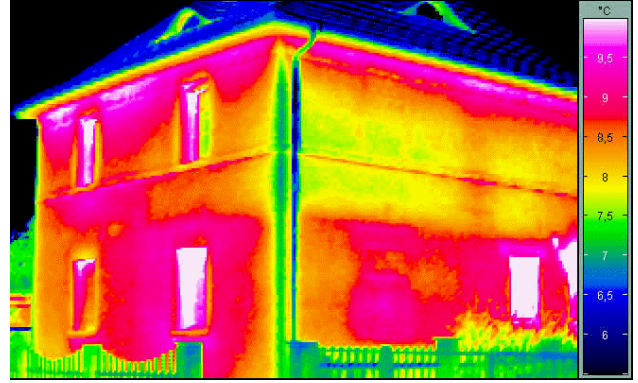
# 1. Density Estimation





## 2. Regression

Example 2: how does gas consumption depend on external temperature?  
(Whiteside, 1960s).



weekly measurements of

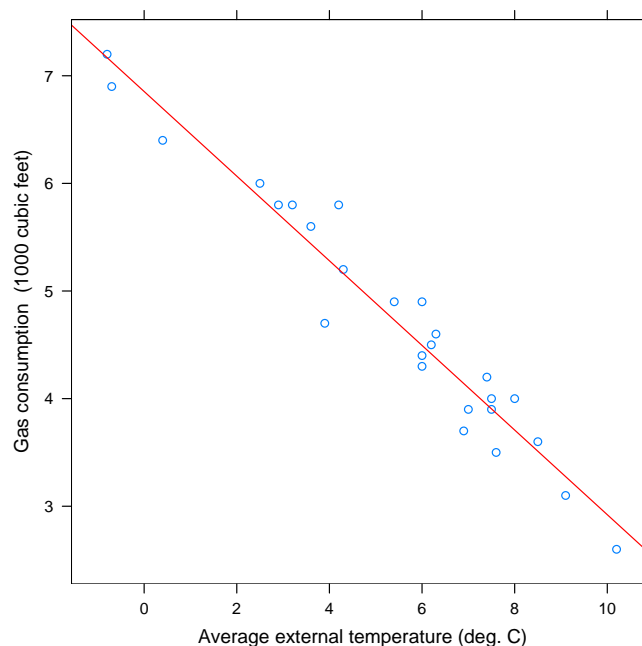
- average external temperature
- total gas consumption  
(in 1000 cubic feet)

A third variable encodes two heating seasons, before and after wall insulation.

How does gas consumption depend on external temperature?

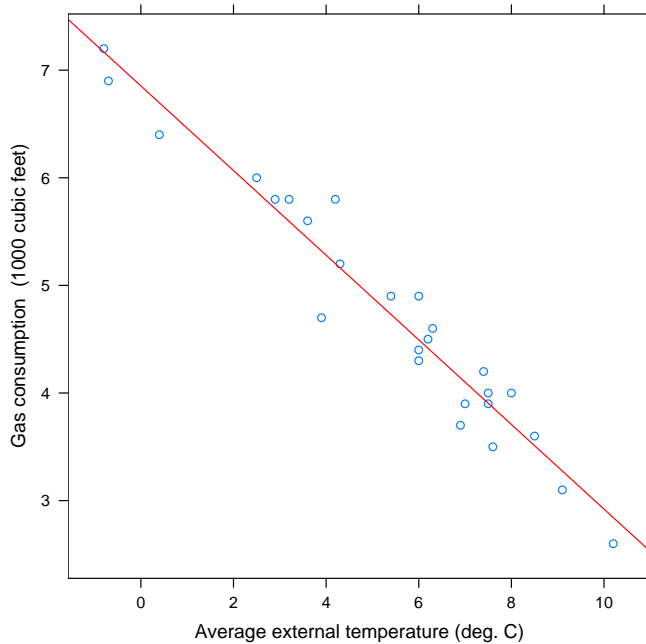
How much gas is needed for a given temperature ?

## 2. Regression

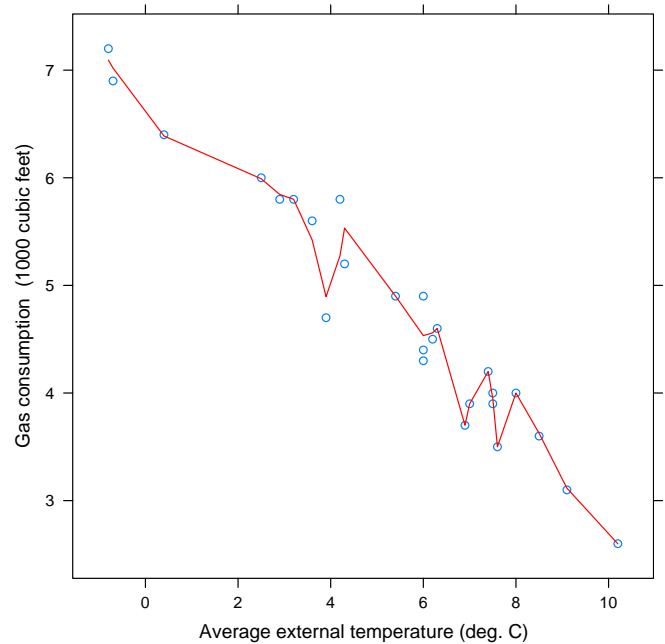


linear model

## 2. Regression



linear model



more flexible model

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

16/34

## 3. Classification / Supervised Learning

Example 3: classifying iris plants  
(Anderson 1935).

150 iris plants (50 of each species):

- species: setosa, versicolor, virginica
- length and width of sepals (in cm)
- length and width of petals (in cm)



iris setosa



iris versicolor



iris virginica

Given the lengths and widths of sepals and petals of an instance, which iris species does it belong to?

See iris species database  
(<http://www.badbear.com/signa>).

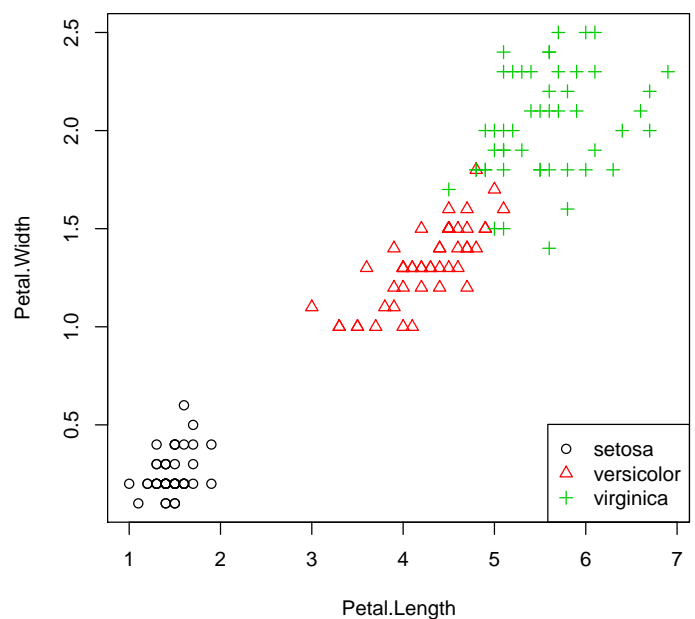
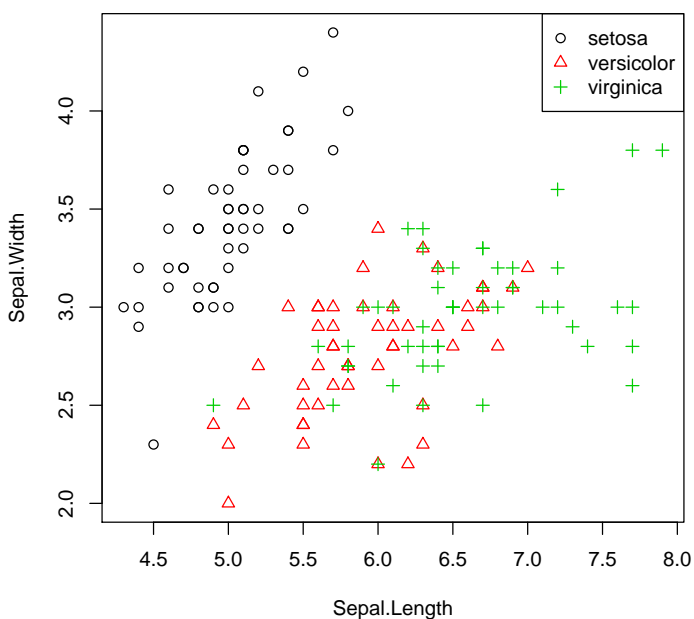
## 3. Classification / Supervised Learning

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.10	3.50	1.40	0.20	setosa
2	4.90	3.00	1.40	0.20	setosa
3	4.70	3.20	1.30	0.20	setosa
4	4.60	3.10	1.50	0.20	setosa
5	5.00	3.60	1.40	0.20	setosa
⋮	⋮	⋮	⋮	⋮	⋮
51	7.00	3.20	4.70	1.40	versicolor
52	6.40	3.20	4.50	1.50	versicolor
53	6.90	3.10	4.90	1.50	versicolor
54	5.50	2.30	4.00	1.30	versicolor
⋮	⋮	⋮	⋮	⋮	⋮
101	6.30	3.30	6.00	2.50	virginica
102	5.80	2.70	5.10	1.90	virginica
103	7.10	3.00	5.90	2.10	virginica
104	6.30	2.90	5.60	1.80	virginica
105	6.50	3.00	5.80	2.20	virginica
⋮	⋮	⋮	⋮	⋮	⋮
150	5.90	3.00	5.10	1.80	virginica

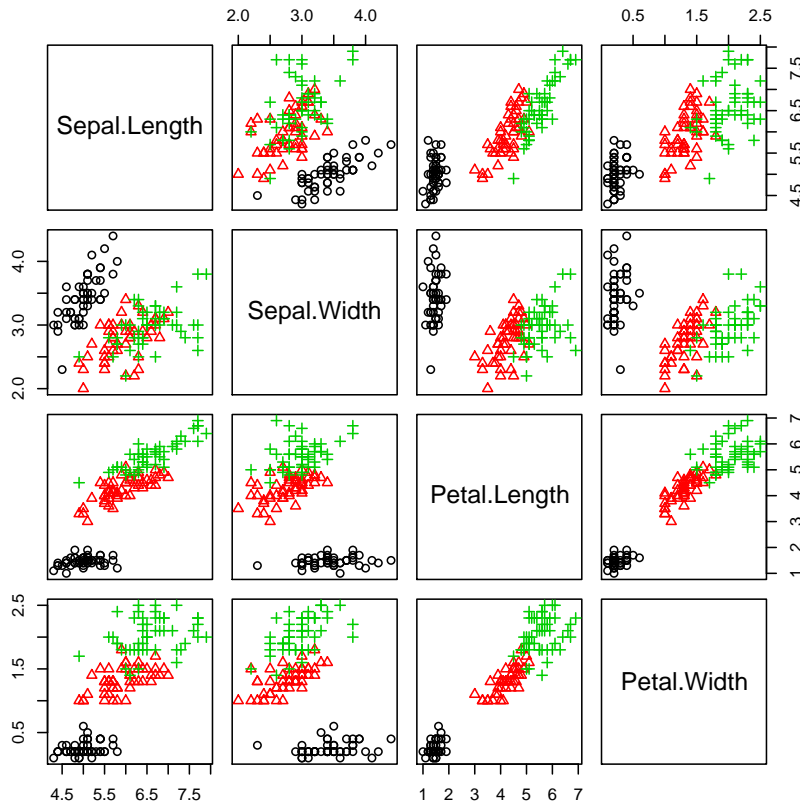
Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
 Course on Machine Learning, winter term 2012/13

18/34

## 3. Classification / Supervised Learning



### 3. Classification / Supervised Learning



Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

20/34

### 3. Classification / Supervised Learning

#### Example 4: classifying email (lingspam corpus)

Subject: query: melcuk (melchuk)

does anybody know a working email  
(or other) address for igor melcuk  
(melchuk) ?

legitimate email ("ham")

Subject: '

hello ! come see our naughty little  
city made especially for adults  
[http://208.26.207.98/freeweek/  
enter.html](http://208.26.207.98/freeweek/enter.html) once you get here, you  
won't want to leave !

spam

How to classify email messages as spam or ham?

## 3. Classification / Supervised Learning

Subject: query: melcuk (melchuk)

does anybody know a working email  
(or other) address for igor melcuk  
(melchuk) ?

⇒

a	1
address	1
anybody	1
does	1
email	1
for	1
igor	1
know	1
melcuk	2
melchuk	2
or	1
other	1
query	1
working	1

## 3. Classification / Supervised Learning

lingspam corpus:

- email messages from a linguistics mailing list.
- 2414 ham messages.
- 481 spam messages.
- 54742 different words.
- an example for an early, but very small spam corpus.

### 3. Classification / Supervised Learning

All words that occur at least in each second spam or ham message on average (counting multiplicities):

	!	your	will	we	all	mail	from	do	our	email
spam	14.18	7.45	4.36	3.42	2.88	2.77	2.69	2.66	2.46	2.24
ham	0.38	0.46	1.93	0.94	0.83	0.79	1.60	0.57	0.30	0.39

	out	report	order	as	free	language	university
spam	2.19	2.14	2.09	2.07	2.04	0.04	0.05
ham	0.34	0.05	0.27	2.38	0.97	2.67	2.61

example rule:

if  $\text{freq}("!") \geq 7$  and  $\text{freq}(\text{"language"}) = 0$  and  $\text{freq}(\text{"university"}) = 0$  then spam,  
else ham

Should we better normalize for message length?

### 4. Reinforcement Learning

A class of learning problems where the correct / optimal action never is shown, but only positive or negative feedback for an action actually taken is given.

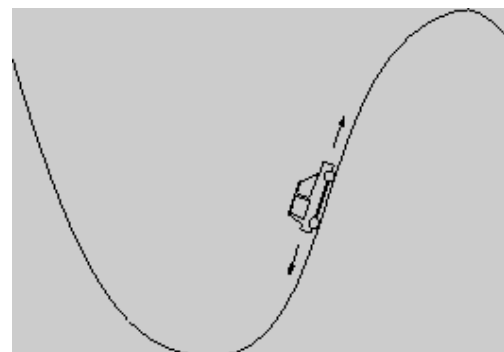
Example 5: steering the mountain car.

Observed are

- x-position of the car,
- velocity of the car

Possible actions are

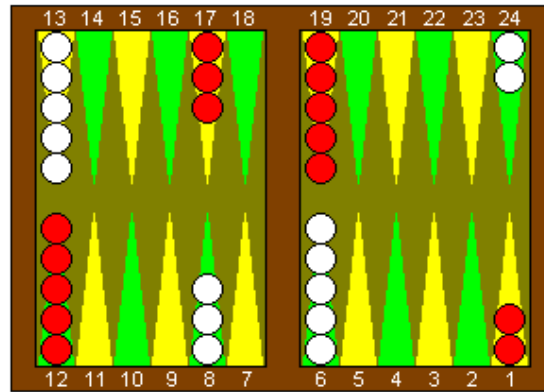
- accelerate left,
- accelerate right,
- do nothing



The goal is to steer the car on top of the right hill.



## 4. Reinforcement Learning / TD-Gammon



**Figure 2.** An illustration of the normal opening position in backgammon. TD-Gammon has sparked a near-universal conversion in the way experts play certain opening rolls. For example, with an opening roll of 4-1, most players have now switched from the traditional move of 13-9, 6-5, to TD-Gammon's preference, 13-9, 24-23. TD-Gammon's analysis is given in Table 2.

Program	Hidden Units	Training Games	Opponents	Results
TD-Gam 0.0	40	300,000	Other Programs	Tied for Best
TD-Gam 1.0	80	300,000	Robertie, Magriel, ...	-13 pts / 51 games
TD-Gam 2.0	40	800,000	Var. Grandmasters	-7 pts / 38 games
TD-Gam 2.1	80	1,500,000	Robertie	-1 pts / 40 games
TD-Gam 3.0	80	1,500,000	Kazaros	+6 pts / 20 games

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

26/34

## 5. Cluster Analysis

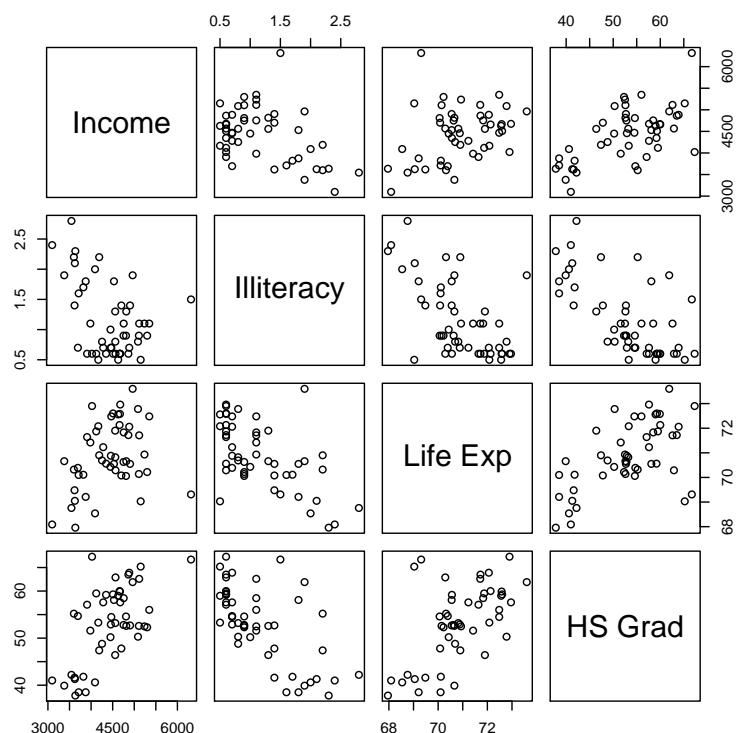
Finding groups of similar objects.

Example 6: sociographic data of the 50 US states in 1977.

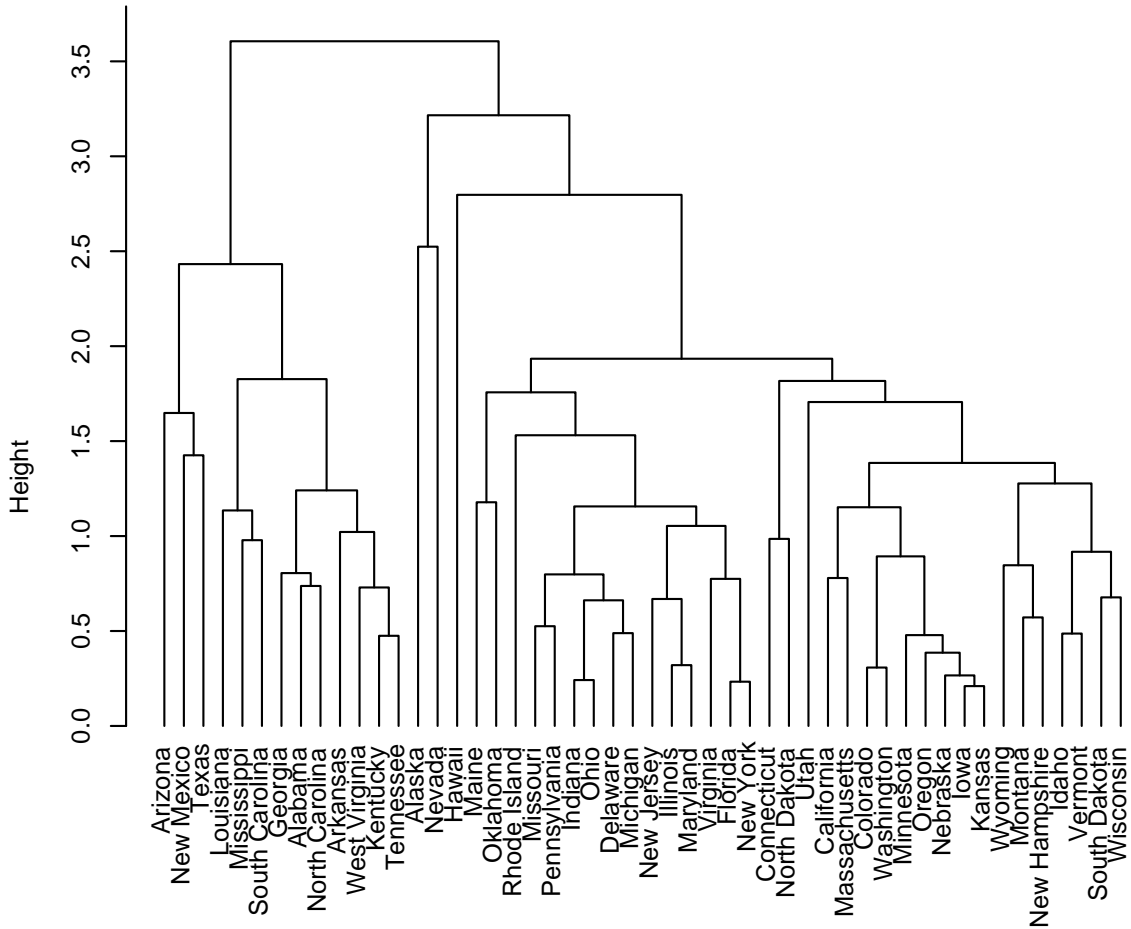
state dataset:

- income (per capita, 1974),
- illiteracy (percent of population, 1970),
- life expectancy (in years, 1969–71),
- percent high-school graduates (1970).

and some others not used here.

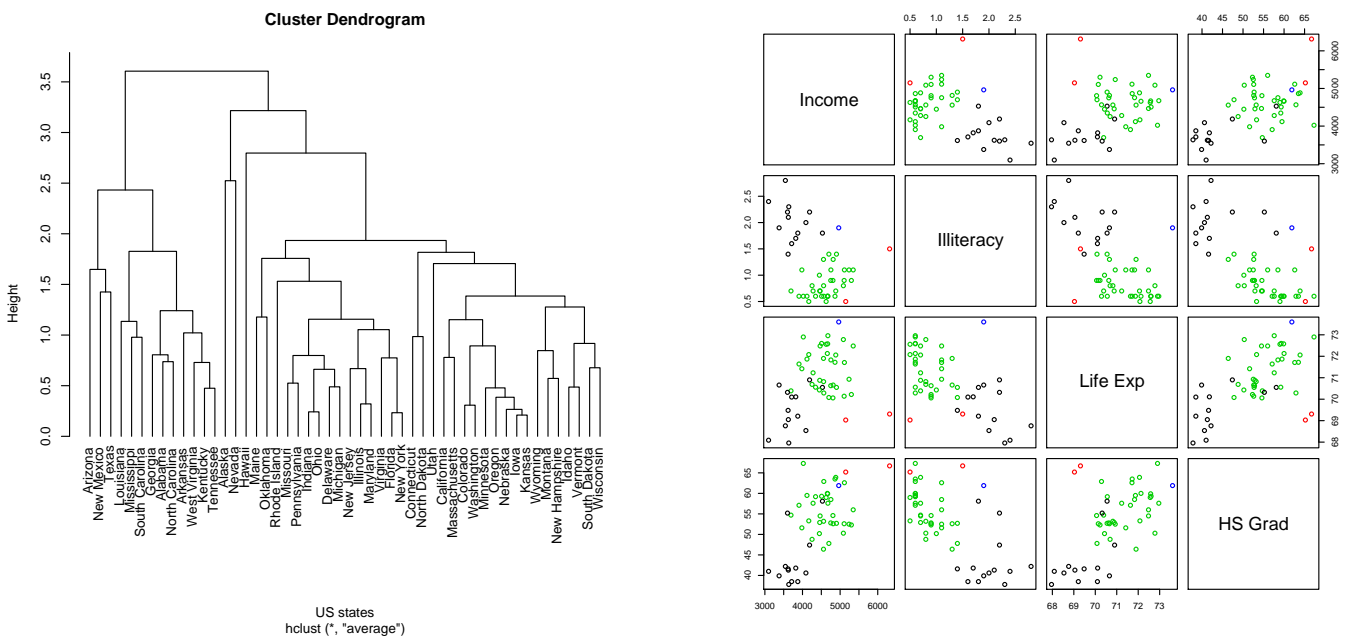


### 5. Cluster Analysis



Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISML), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
 Course on Machine Learning, winter term 2012/13 28/34

### 5. Cluster Analysis



black: Arizona et al., red: Alaska & Nevada, green: California et al., blue: Hawaii.

## 7. Association Analysis

Association rules in large transaction datasets:

- look for products frequently bought together (**frequent itemsets**).
- look for rules in buying behavior (**association rules**)

Examples:

- {beer, pampers, pizza} (support=0.5)  
  {bread, milk} (support=0.5)
- If beer and pampers, then pizza (confidence= 0.75)  
  If bread, then milk (confidence=0.75)

cid	beer	bread	icecream	milk	pampers	pizza
1	+	-	-	+	+	+
2	+	+	-	-	+	+
3	+	-	+	-	+	+
4	-	+	-	+	-	+
5	-	+	+	+	-	-
6	+	+	-	+	+	-

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

30/34

Machine Learning

### 1. What is Machine Learning?

### 2. Overview

### 3. Organizational stuff

## Exercises and tutorials

- There will be a weekly sheet with two exercises handed out **each Tuesday** in the lecture.  
1st sheet will be handed out next Tue. 30.10.
- Solutions to the exercises can be submitted until **next Tuesday before the lecture**  
1st sheet is due Wed. 6.11.
- Exercises will be corrected.
- Tutorials **each Thursday 16–18**,  
1st tutorial at Mon. 1.11.
- Successful participation in the tutorial gives up to 10% bonus points for the exam.

## Exam and credit points

- There will be a written exam at end of term  
(2h, 4 problems).
- The course gives 8 ECTS (3+2 SWS).
- The course can be used in
  - Wirtschaftsinformatik MSc / Informatik / Gebiet KI & ML
  - IMIT MSc. (neu) / Informatik / Gebiet KI & ML
  - IMIT MSc. (alt) / IT Machine Learning,
  - IMIT MSc. (alt) / BW Business Intelligence,
  - as well as in any BSc program.

### Some books

- Richard O. Duda, Peter E. Hart, David G. Stork (<sup>2</sup>2001):  
*Pattern Classification*, Springer.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (<sup>2</sup>2009):  
*The Elements of Statistical Learning*, Springer.  
Also available online as PDF at <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- Christopher M. Bishop (2007):  
*Pattern Recognition and Machine Learning*, Springer.
- W. N. Venables, B. D. Ripley (2002):  
*Modern Applied Statistics with S*, Springer.
- Tom Mitchell (1997):  
*Machine Learning*, McGraw-Hill.
- Christopher M. Bishop (1996):  
*Neural Networks for Pattern Recognition*, Oxford University Press.

---

Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute BW/WI & Institute for Computer Science, University of Hildesheim  
Course on Machine Learning, winter term 2012/13

33/34

### Some First Machine Learning / Data Mining Software

- R (v2.15.1, 22.6.2012; <http://www.r-project.org>).
- Weka (v3.6.8, 15.8.2012; <http://www.cs.waikato.ac.nz/~ml/>).
- SAS Enterprise Miner (commercially).

#### Public data sets:

- UCI Machine Learning Repository  
(<http://www.ics.uci.edu/~mlearn/>)
- UCI Knowledge Discovery in Databases Archive  
(<http://kdd.ics.uci.edu/>)