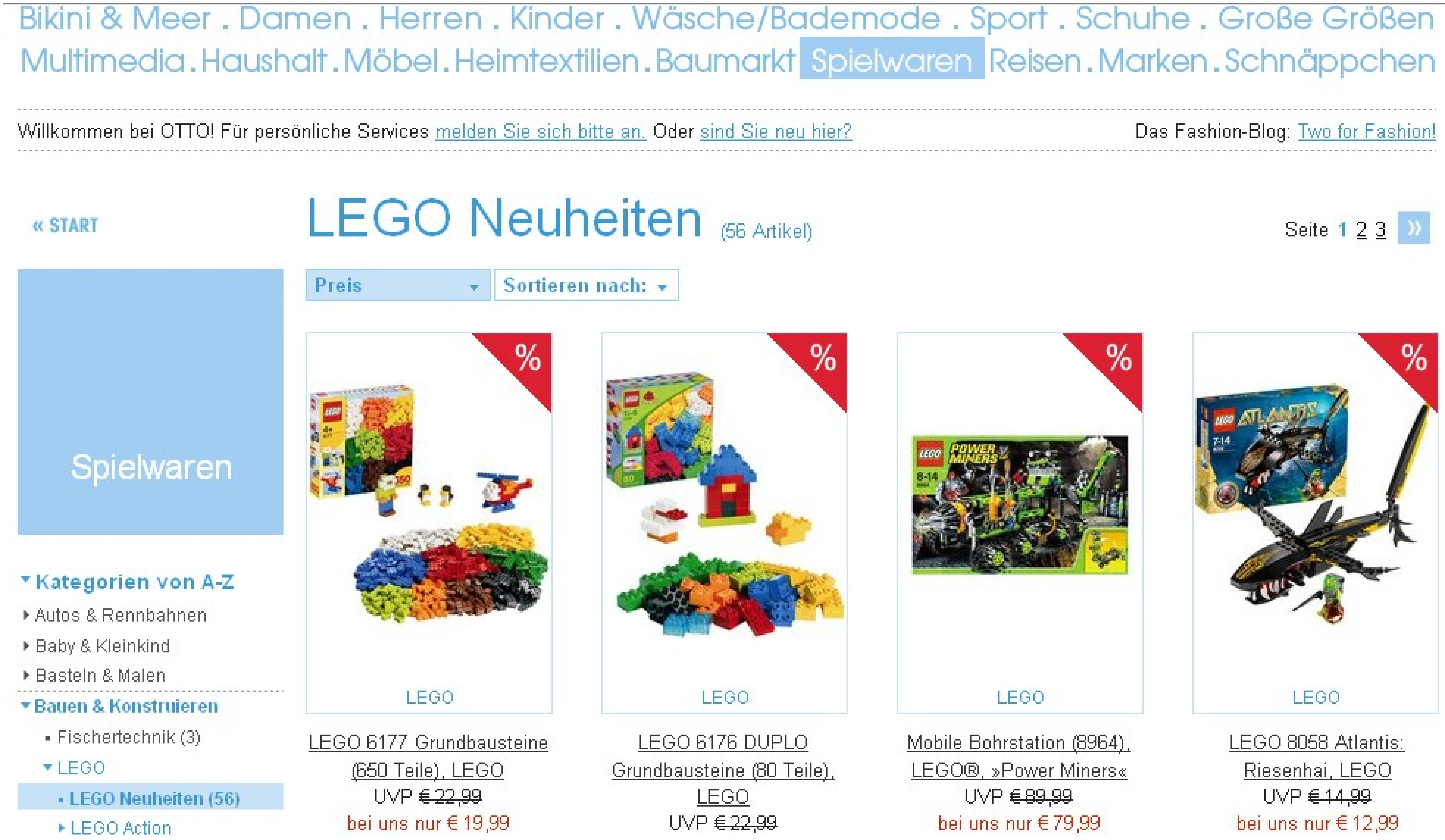
Tree-like structures

break down high complexity into smaller parts

Drill-Down Search Menu

- Structures complexity of product variants



Willkommen bei OTTO! Für persönliche Services melden Sie sich bitte an. Oder sind Sie neu hier?

Das Fashion-Blog: Two for Fashion!

LEGO Neuheiten (56 Artikel)

LEGO 6177 Grundbausteine (650 Teile), LEGO
UVP ~~€ 22,99~~
bei uns nur € 19,99

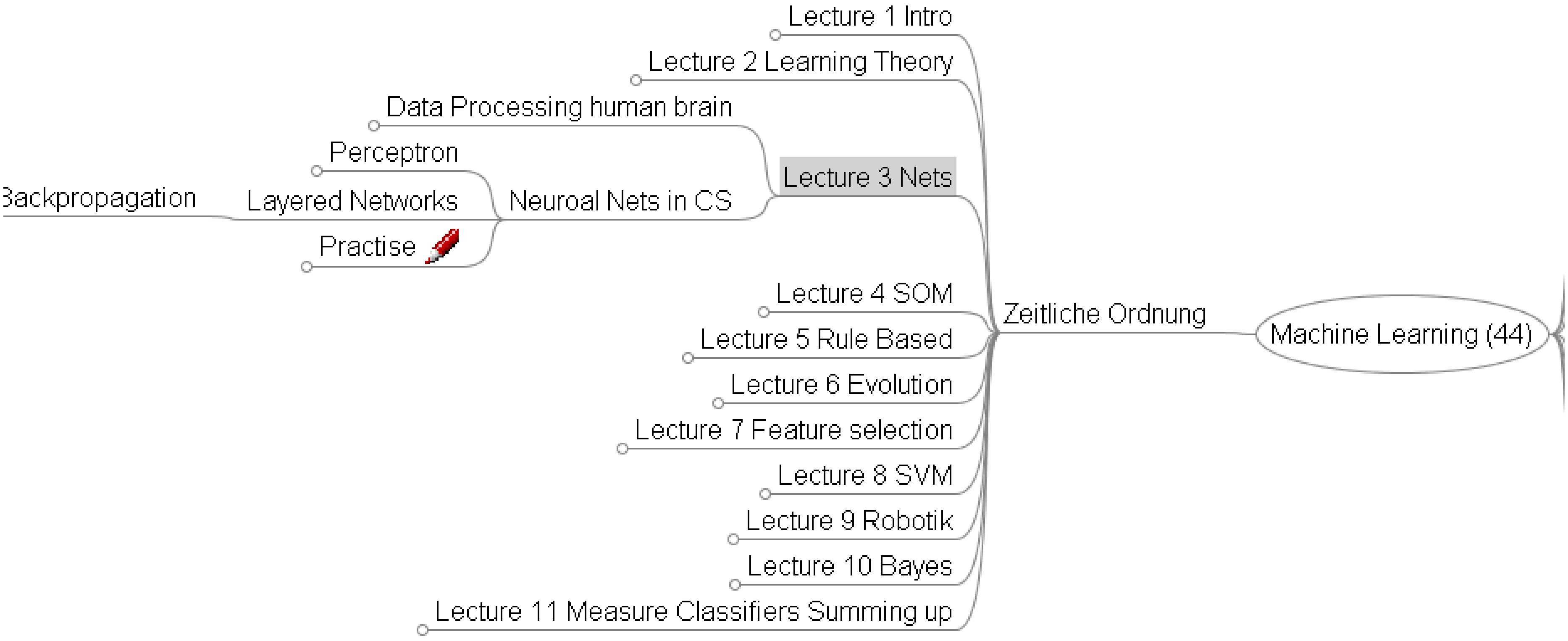
LEGO 6176 DUPLO Grundbausteine (80 Teile), LEGO
UVP ~~€ 22,99~~

Mobile Bohrstation (8964), LEGO®, »Power Miners«
UVP ~~€ 89,99~~
bei uns nur € 79,99

LEGO 8058 Atlantis: Riesenhai, LEGO
UVP ~~€ 14,99~~
bei uns nur € 12,99

www.otto.de

Lecture planning



Machine Learning (44)

Zeitliche Ordnung

Lecture 1 Intro

Lecture 2 Learning Theory

Lecture 3 Nets

Data Processing human brain

Neuroal Nets in CS

Perceptron

Layered Networks

Backpropagation

Practise

Lecture 4 SOM

Lecture 5 Rule Based

Lecture 6 Evolution

Lecture 7 Feature selection

Lecture 8 SVM

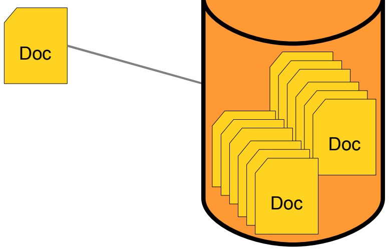
Lecture 9 Robotik

Lecture 10 Bayes

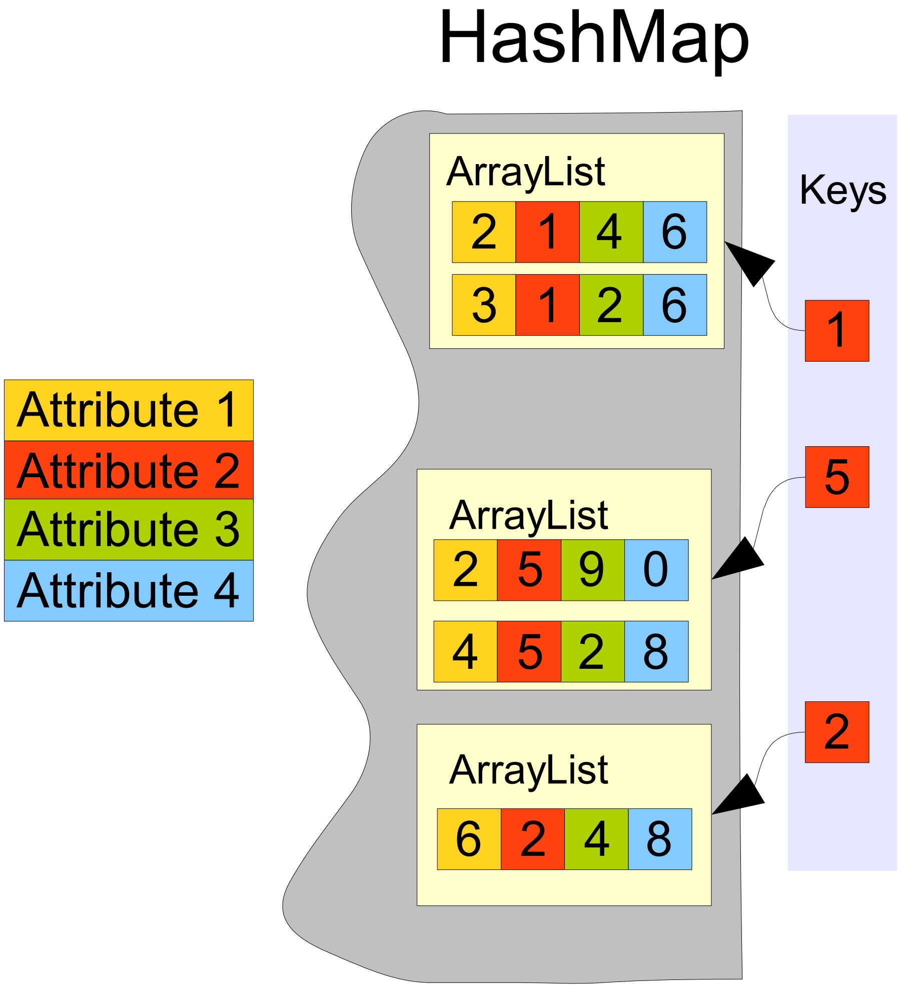
Lecture 11 Measure Classifiers Summing up

Index into Databases

- Maintain low access times even if the database is “loaded”



- Hash-Maps
  - One-step access
  - Is **not** applicable for comparable conditions (>,<)



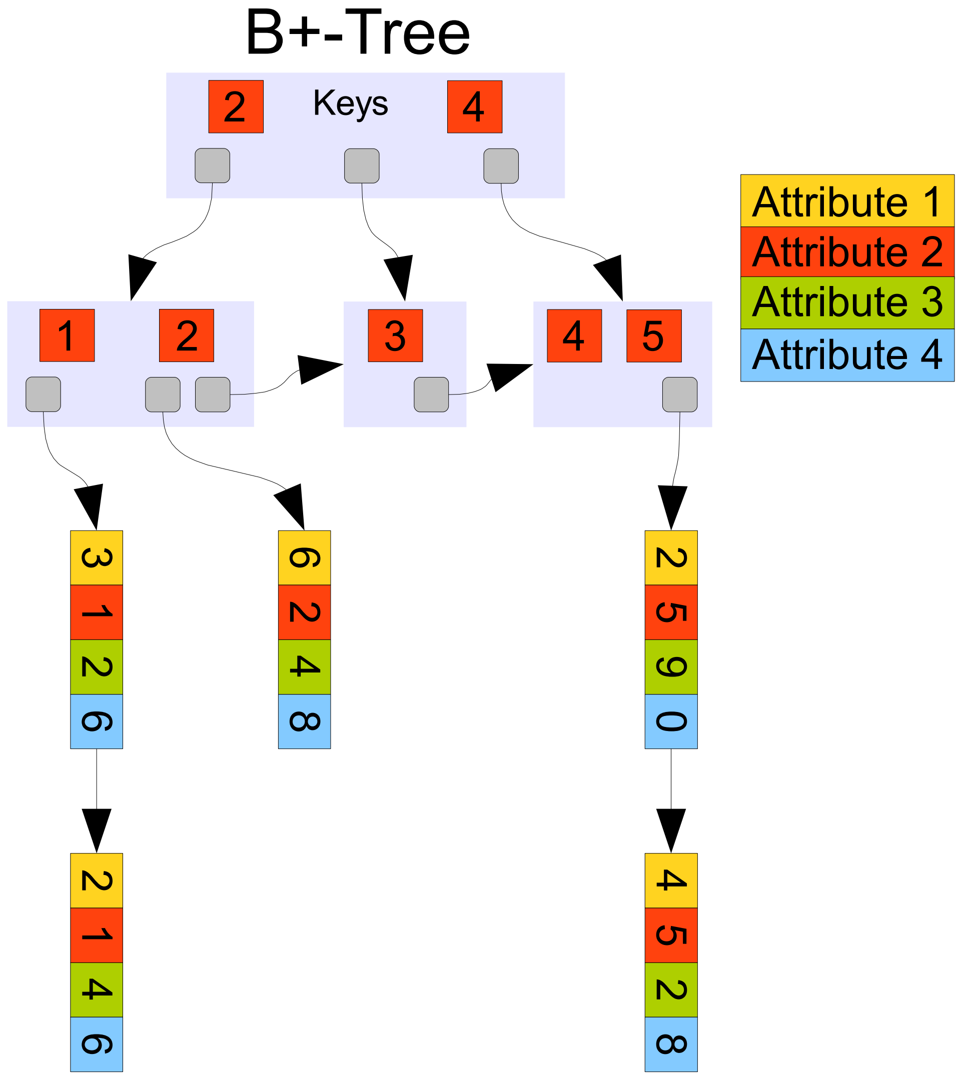
2 1 4 6
3 1 2 6

2 5 9 0
4 5 2 8

6 2 4 8

Support for comparison

- B+tree
    - Keys are sorted
- Samples are sorted at time of insertion
- For char/text attributes use a index length



B+Tree

Bucket

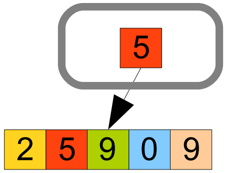
- Contains buckets
  - The tree consists of
    - Inner nodes
      - Contain only keys and pointers-to-buckets
    - End nodes (the leaves)
      - Contain keys, pointers-to-buckets and pointers-to-data
  - Each buckets contains
    - between n/2 ≤ m ≤ n keys
    - m+1 pointers-to-buckets or m pointers-to-data (only leaves)
    - The keys in the bucket are ordered

Creating a B+Tree

- Create a B+Tree for the following data using attribute 2 (red) and n=3
  - The insertion order is from top to bottom

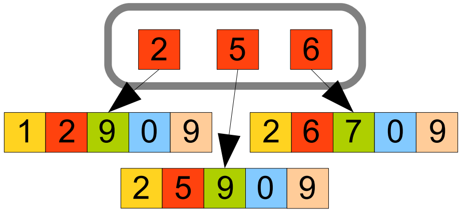
Samples					Keys
2	5	9	0	9	5
2	6	7	0	9	6
1	2	9	0	9	2
1	4	5	0	9	4
4	8	2	0	1	8
4	9	9	0	1	9

1. Insert “5”



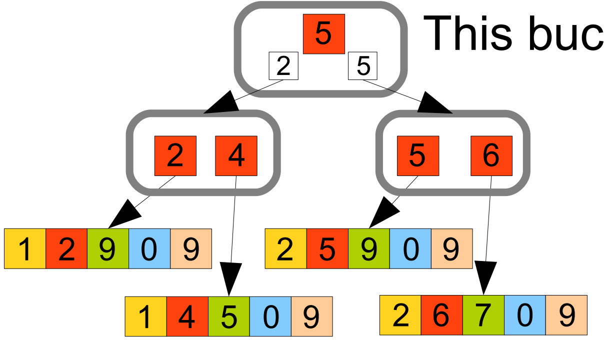
This bucket contains a key and a pointer to a data sample

2. Insert “6”,”2”



Maximal size for the bucket is reached!

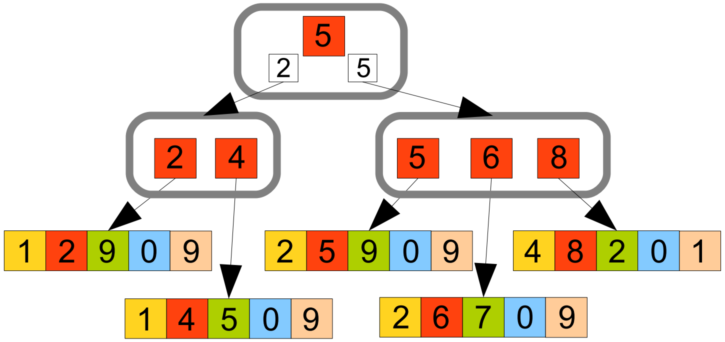
3. Insert “4”



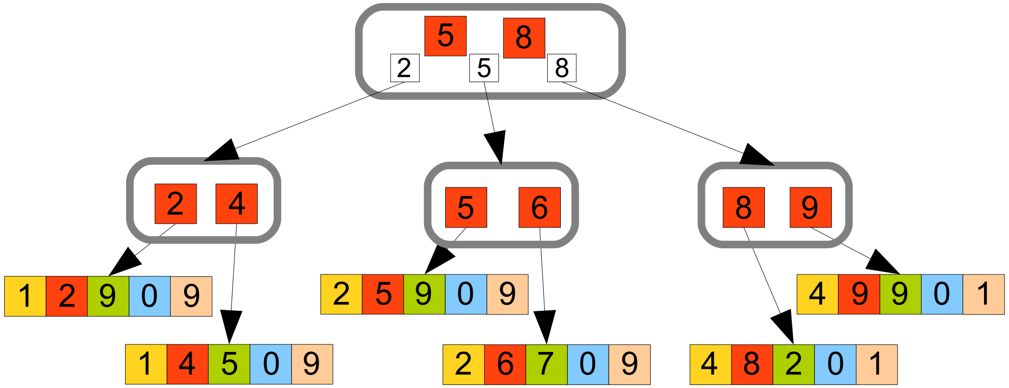
This bucket contains a pointers to other buckets

Creating a B+Tree

4. Insert “8”

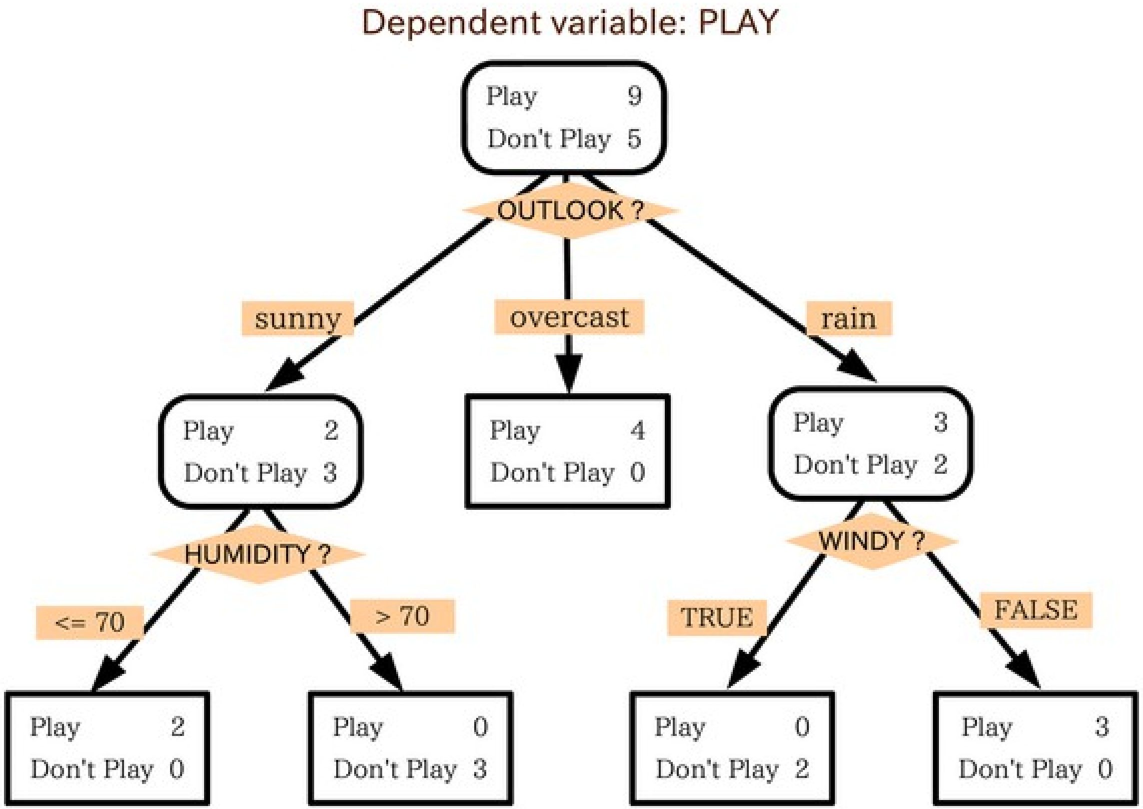


5. Insert “9”



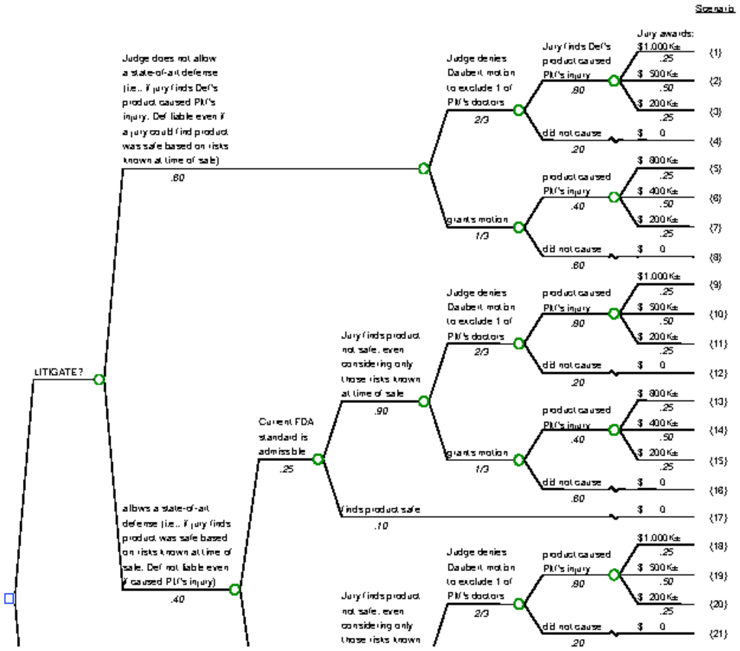
- If a bucket splits
  - the middle key travels “up” to the parent
  - The pointers are named according to the smallest key in the bucket to which the pointer points

Decision making

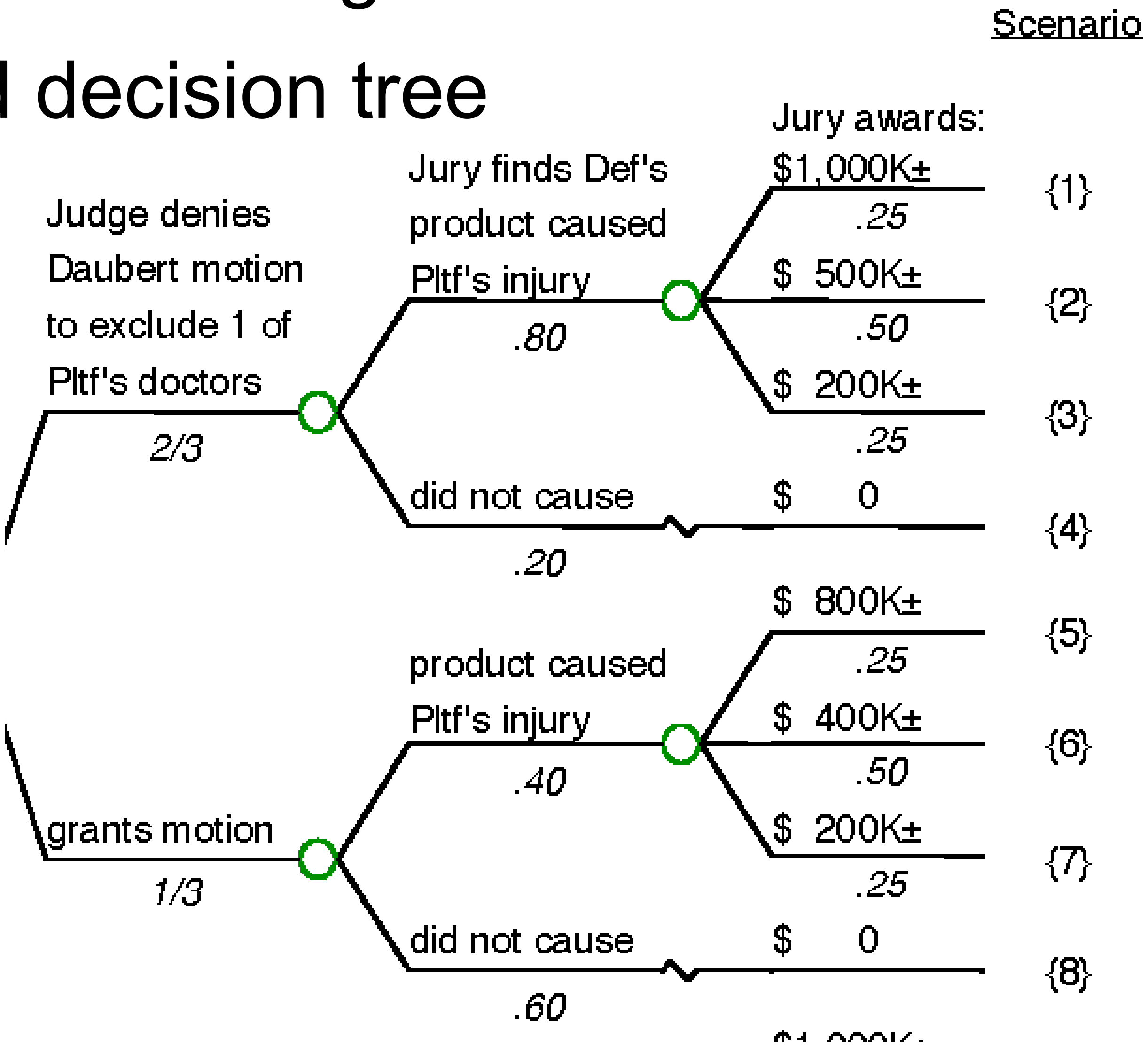


Litigation

- Strategical decision making
- A fully developed decision tree



Scenario
LITIGATE?
Judge does not allow a state-of-art defense (i.e., if jury finds Def's product caused Plf's injury, Def liable even if a jury could find product was safe based on risks known at time of sale)
.60
allows a state-of-art defense (i.e., if jury finds product was safe based on risks known at time of sale, Def not liable even if caused Plf's injury)
.40
Current FDA standard is admissible
.25
Jury finds product not safe, even considering only those risks known at time of sale
.90
finds product safe
.10
Jury finds product not safe, even considering only those risks known
Judge denies Daubert motion to exclude 1 of Plf's doctors
2/3
grants motion
1/3
Jury finds Def's product caused Plf's injury
product caused Plf's injury
.80
.40
did not cause
.20
.60
Jury awards:
$1,000K±
$ 500K±
$ 200K±
$ 800K±
$ 400K±
$ 0
.25
.50
{1}
{2}
{3}
{4}
{5}
{6}
{7}
{8}
{9}
{10}
{11}
{12}
{13}
{14}
{15}
{16}
{17}
{18}
{19}
{20}
{21}

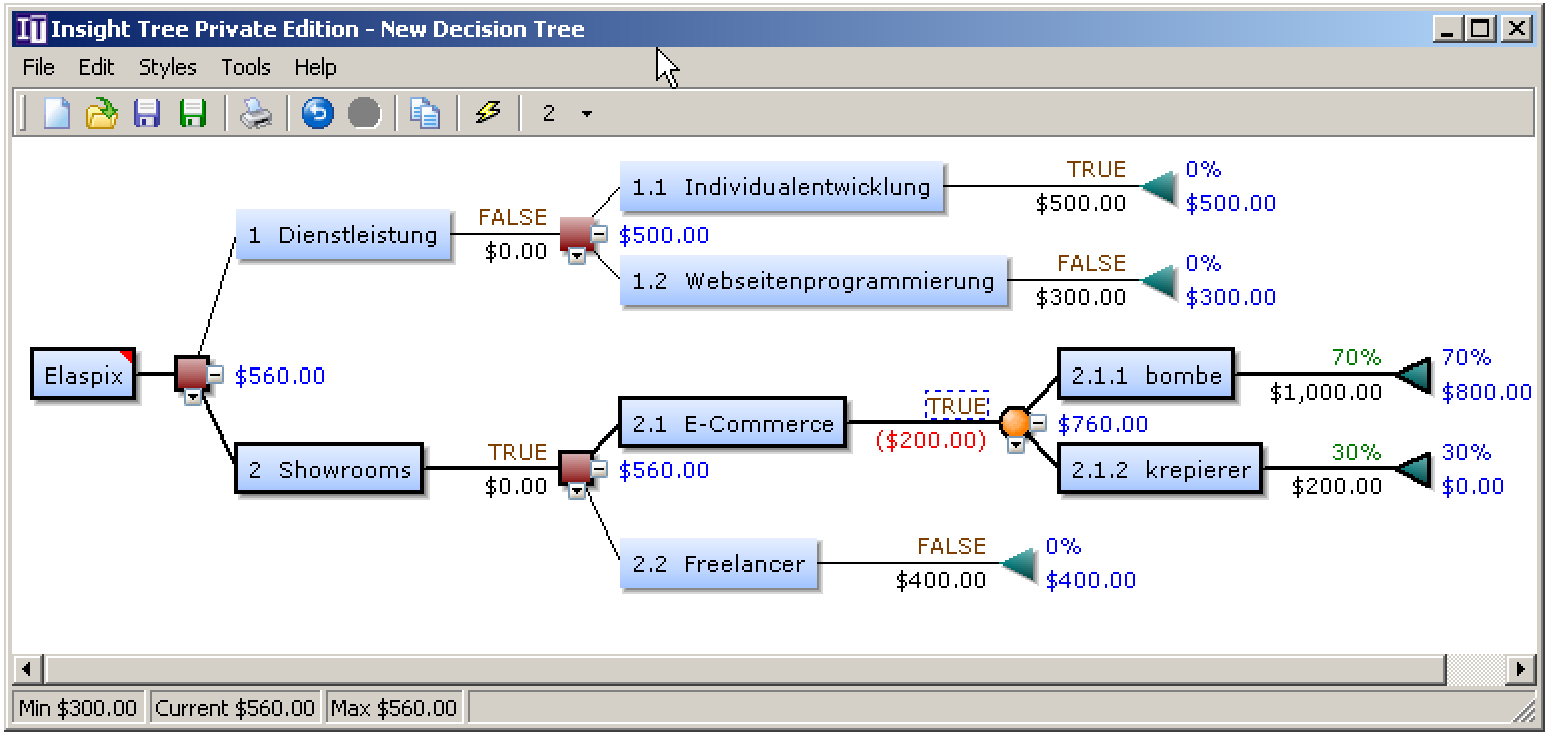


Decision Tree Analysis

- Find answers to a multiple-choice question with several option for each question
- Incrementally develop the tree
  - From Top to Bottom
  - Think local is much easier than thinking global
- Easy to use and understand
- Deal with uncertainty

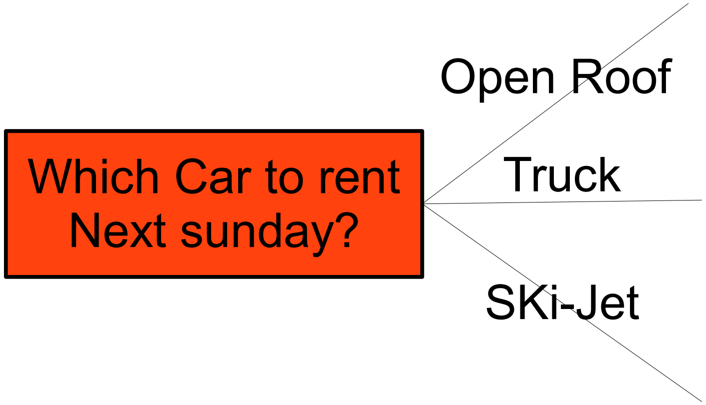
Example Software

- http://www.visionarytools.com/



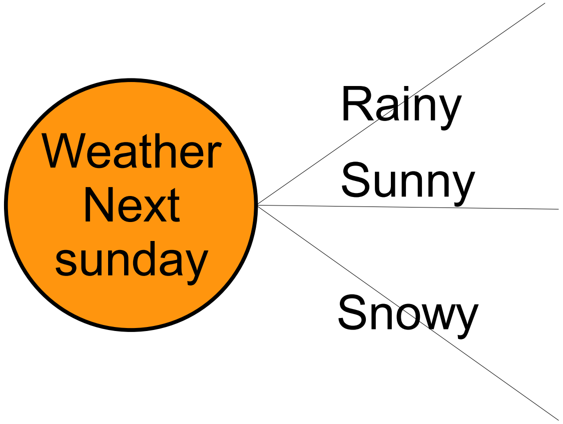
Your own Decisions

- A decision about a topic is drawn as square
- For each alternative create new sub-Trees



Uncertainties beyond your control

- Other decisions or Uncertainties are drawn as circle
  - How many outcomes may be possible?
    - Each form a new sub-tree



Make sub Trees until you are done

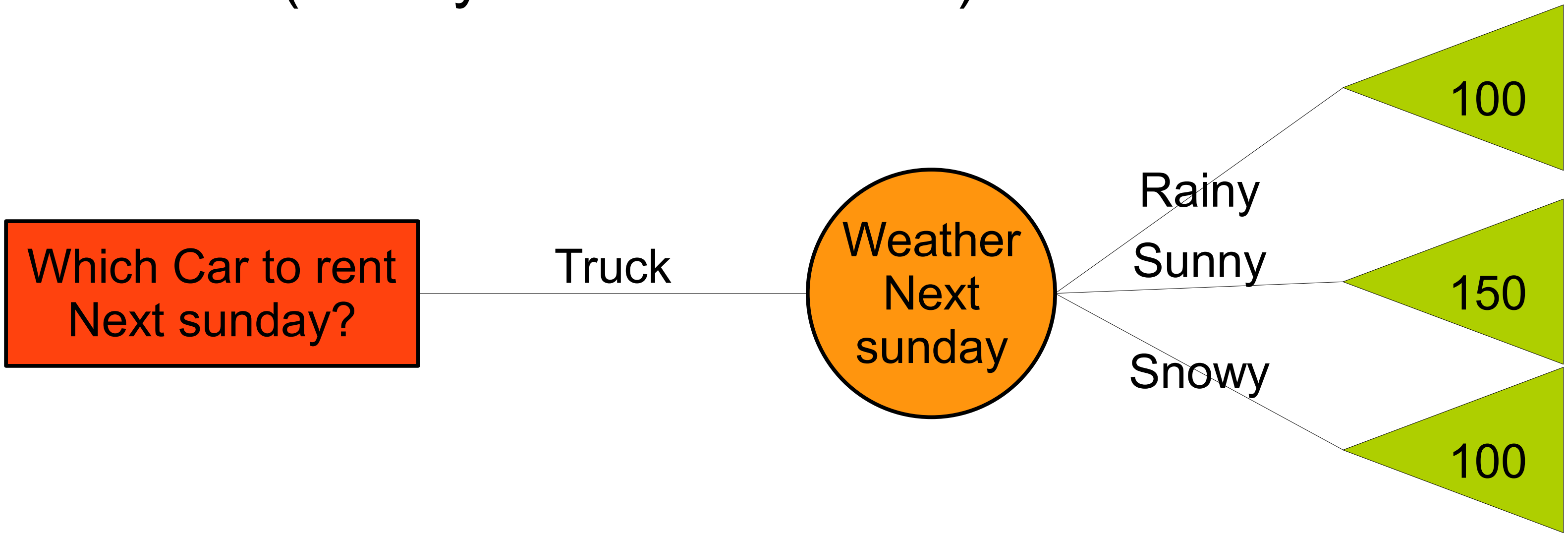
- At each line you have 3 options
  - Decision

Which Car to rent
Next sunday?

  - Uncertainty

Weather
Next
sunday

  - Outcome
    - may be postive or negative expressed in numerical values (mostly Euros or Dollars)



Which Car to rent
Next sunday?

Truck

Weather
Next
sunday

Rainy

Sunny

Snowy

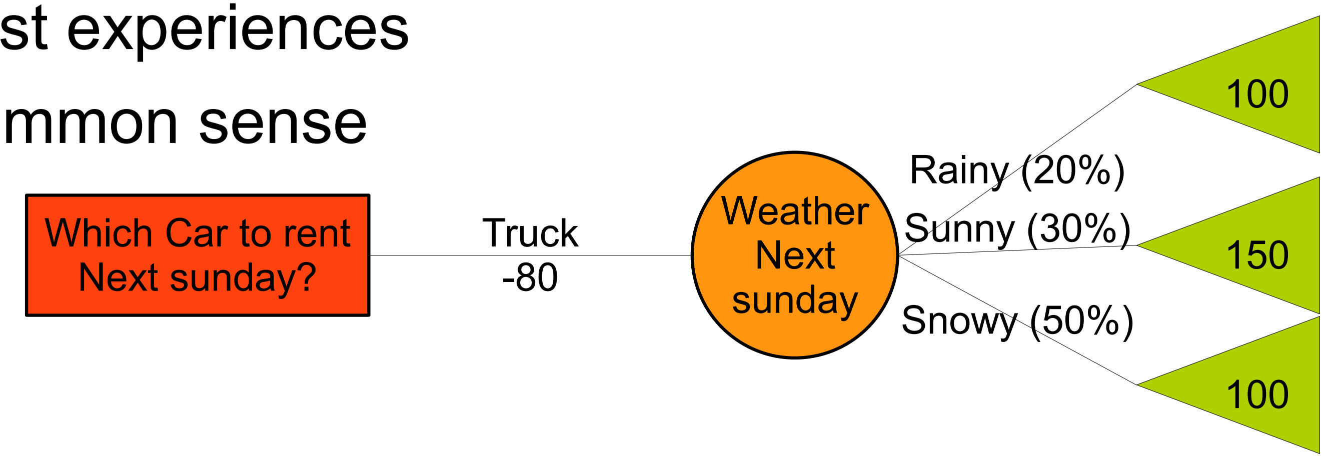
100

150

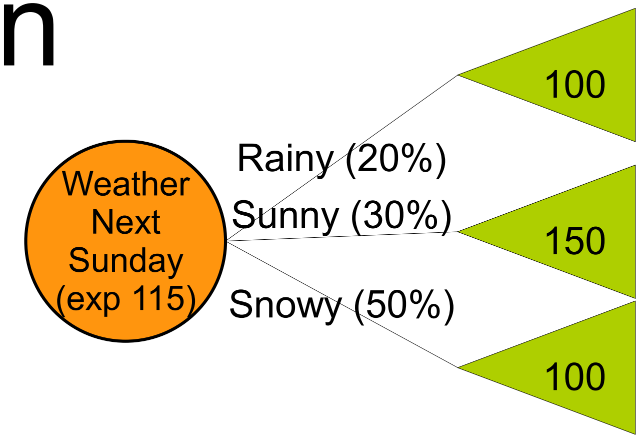
100

Make it more plastic using numbers

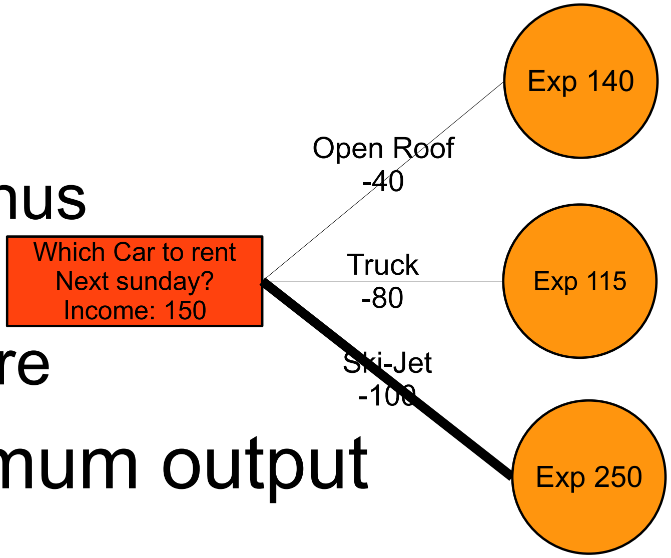
- To the branches of decisions add
  - Guaranteed Costs
- To the branches of uncertainties add
  - Likelihoods (sum up to 100%)
  - Those likelihoods are coming from
    - Market research
    - Past experiences
    - Common sense



Computation



- From right to left
  - For uncertainties
    - Multiply the outcome of each leaf with its probability and sum them up
    - Write the sum into the circle
  - For decisions
    - Compute: multiplication result minus the costs
    - Write the maximum into the square
  - Choose the path with the maximum output

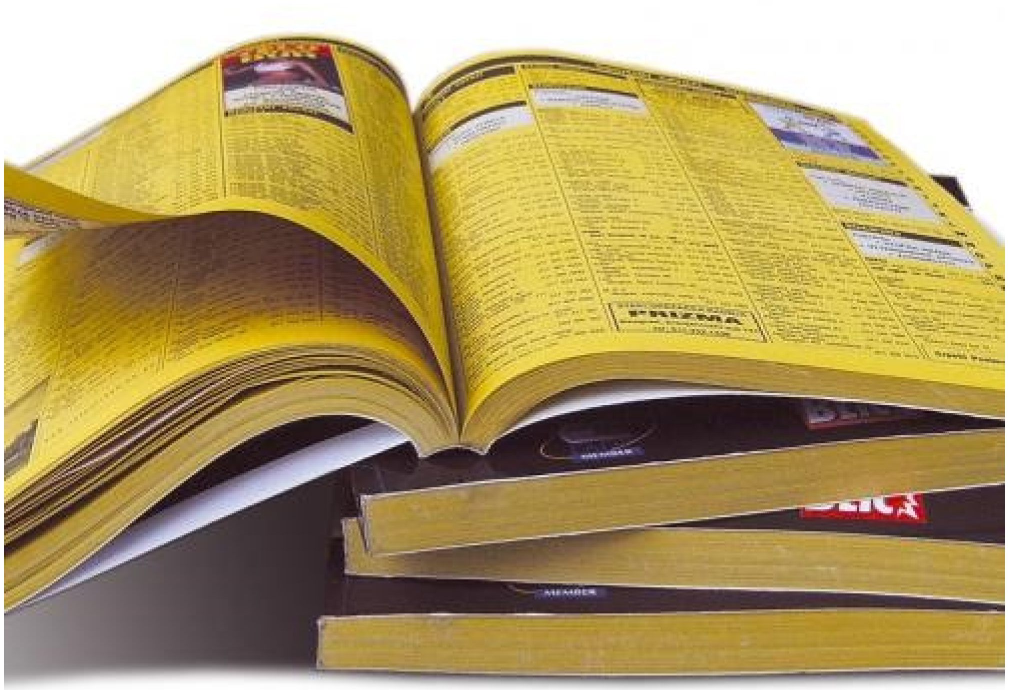


Induction of decision trees

- To understand large amounts of nominal data
  - Build an tree automatically and extract the paths as rules
  - Interprete the rules to understand the data
- To create prediction models from training data
  - Given: supervised samples with class and nominal attributes
  - Task: create a decision tree such that the class of new unseen samples can be predicted from their known attributes

Telephone book

- Stores “observations” in a alpha-numerical order
- What is the worst case number of queries if the telephone book is not ordered?
- What is the worst case number of queries when ordered?



http://www.calgary-city-maps.com/images/Calgary_phone_book.jpg

Concept of Entropy

- A measure of how much confusion (deutsch Unordnung) a system has
  - Entropy of zero means: no confusion
- A measure of how much uncertainty an event has
  - Entropy of zero has no uncertainty

http://www.unserekirche.de/data/images/520/245047771580.jpg

Information Entropy

- Derived from telecommunication engineering
  - How many bits are required to encode a message
  - e.g. The symbols transmitted over a network are only (a_1,a_2,a_3,a_4), (could mean: a_1=start,a_2=stop, a_3=reset, a_4=quit)
  - How many bits are required to encode the following messages? (the transmission time doesn't matter, only the number of different symbols are important, you have an extra signal for free that indicates that a new message is incoming)
    - Message 1: $\{a_1,a_1,a_1,a_1,a_1,a_1,a_1\}$
    - Message 2: $\{a_1,a_1,a_1,a_2,a_2,a_2\}$
    - Message 3: $\{a_1,a_2,a_3,a_4\}$

How many bits are needed

- How many bits are required to encode the following messages? (the transmission time doesn't matter)
  - Message 1: $\{a_1,a_1,a_1,a_1,a_1,a_1,a_1\}$ Bits=0
  - Message 2: $\{a_1,a_1,a_1,a_2,a_2,a_2\}$ Bits=1
  - Message 3: $\{a_1,a_2,a_3,a_4\}$ Bits=2
- Alternative question
  - How many Yes/No Questions are required to encode the message
  - How strong it is to guess / predict the message
    - The more bits the more difficult is the quest

The magic function

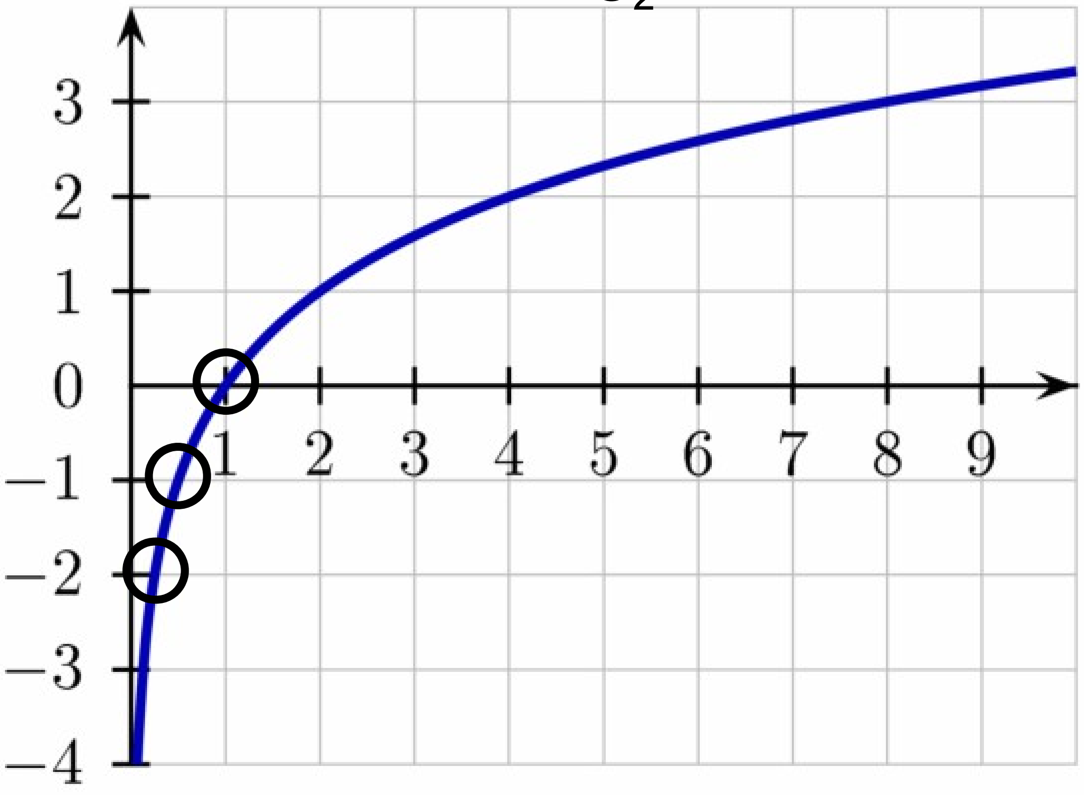
- Which function is best appropriate to compute the number of bits dependent on frequency
- Based on frequency freq(a_i) of message values
- Need a function f with properties
    - f(1)=0 and f(0.5)=1 and f(0.25)=2

Message	freq (a_1)	freq (a_2)	freq (a_3)	freq (a_4)	Entropy
$\{a_1,a_1,a_1,a_1,a_1,a_1,a_1\}$	1	0	0	0	0
$\{a_1,a_1,a_1,a_2,a_2,a_2\}$	0.5	0.5	0	0	1
$\{a_1,a_2,a_3,a_4\}$	0.25	0.25	0.25	0.25	2

... using the binary logarithm

- $f=-\log_2(freq)$

log$_2$



http://www.biocrawler.com/w/images/5/56/Binary_logarithm_plot.png

Computing the entropy

- E is the entropy (how many bits are needed)
- M is the message containing the symbols
- n is the number of defined symbols
- p_{ai} is the frequency of the a_i-th symbol

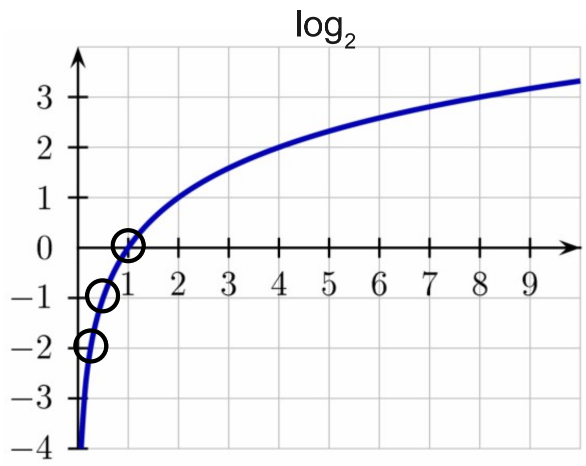
$$E(M) = -\sum_{i=1}^{n} p_{a_i} \cdot \log_2(p_{a_i})$$

e.g. E({a_1,a_1,a_1,a_2,a_2,a_2})=1 Bit

$$E(\{a_{1,} a_{1,} a_{1,} a_{2,} a_{2,} a_2\}) = -p_{a_1} \cdot \log_2(p_{a_1}) - p_{a_2} \cdot \log_2(p_{a_2})$$

Example

$$E(M) = -\sum_{i=1}^{n} p_{a_i} \cdot \log_2(p_{a_i})$$



http://www.biocrawler.com/w/images/5/56/Binary_logarithm_plot.png

Symbol	Absolute Häufigkeit	p(Symbol)
a_1	3 von 6	0.5
a_2	3 von 6	0.5

e.g. E({a_1,a_1,a_1,a_2,a_2,a_2})=1 Bit

$$E(\{a_1, a_1, a_1, a_2, a_2, a_2\}) = -p_{a_1} \cdot \log_2(p_{a_1}) - p_{a_2} \cdot \log_2(p_{a_2})$$

$$= -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5)$$

$$= 1\ Bit$$

Practise calculate number of bits for a set of messages

- Use the Matlab-method getEntropyForVector(aVector) contained in EntropyID3.zip
- Calculate the number of required bits for the vectors
  - [1;1;1;1;1;1;1;1;1]
  - [1;1;1;1]
  - [1;1;1;2;2;2]
  - [1;2;3;4;5;6]