Selected Topics of Pervasive Computing

Stephan Sigg

Georg-August-University Goettingen, Computer Networks

06.11.2013

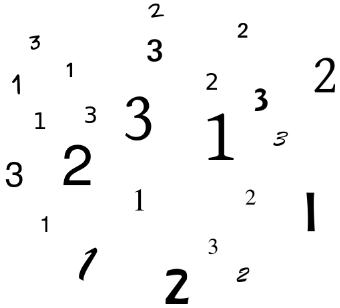
Overview and Structure

30.10.2013 Organisational
30.10.3013 Introduction
06.11.2013 Classification methods (Basic recognition, Bayesian, Non-parametric)
13.11.2013 Classification methods (Linear discriminant, Neural networks)
20.11.2013 –
27.11.2013 –
04.12.2013 –
11.12.2013 Classification methods (Sequential, Stochastic)
18.12.2013 Activity Recognition (Basics, Applications, Algorithms, Metrics)
08.01.2014 Security from noisy data (Basics, Entity, F. Commitment, F. Extractors)
15.01.2014 Security from noisy data (Error correcting codes, PUFs, Applications)
22.01.2014 Context prediction (Algorithms, Applications)
29.01.2014 Networked Objects (Sensors and sensor networks, body area networks)
05.02.2014 Internet of Things (Sensors and Technology, vision and risks)

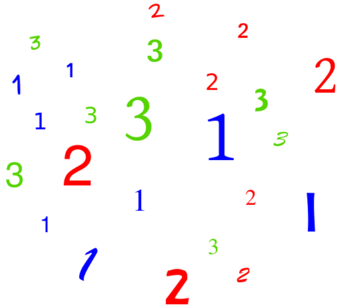
Outline

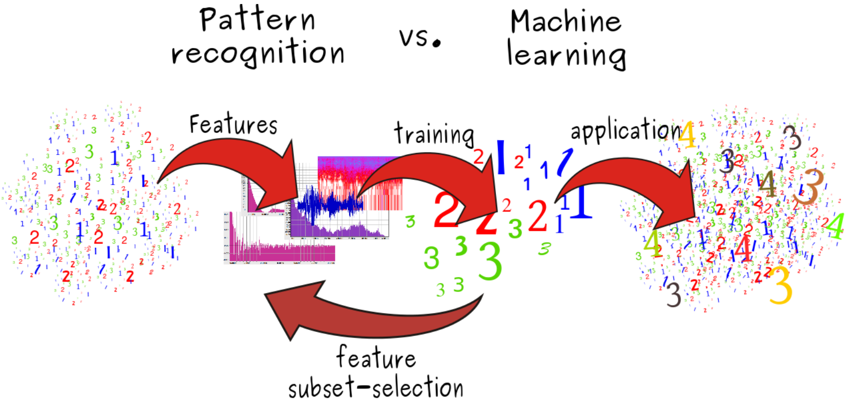
Pattern recognition vs. Machine learning

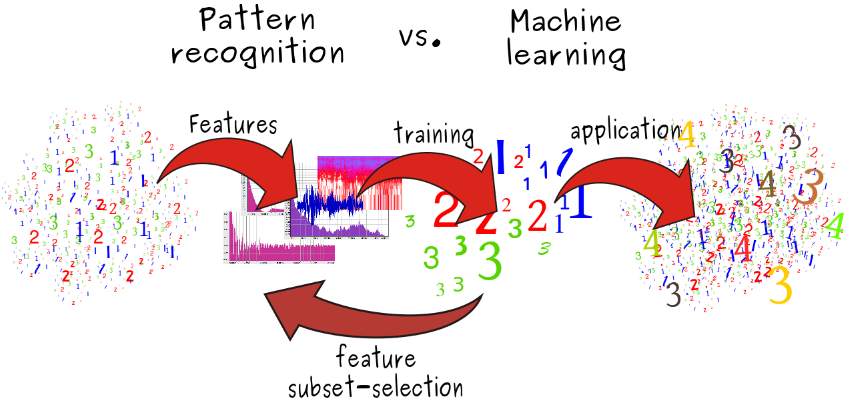
Pattern recognition vs. Machine learning



Pattern recognition vs. Machine learning



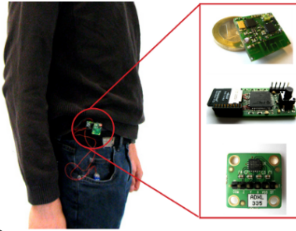

Pattern recognition
vs.
Machine learning
Features
training
application
feature subset-selection



- Mapping of features onto classes by using prior knowledge
- What are characteristic features?
- Which approaches are suitable to obtain these features?

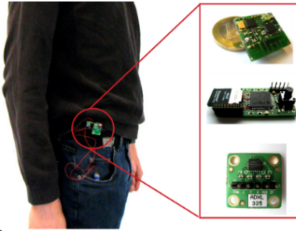
Data sampling

- Record <u>sufficient</u> training data
    - Annotated! (Ground-truth)
    - Multiple subjects
    - Various environmental conditions (time of day, weather, ...)



Data sampling

- Record <u>sufficient</u> training data
  - Annotated! (Ground-truth)
  - Multiple subjects
  - Various environmental conditions (time of day, weather, ...)

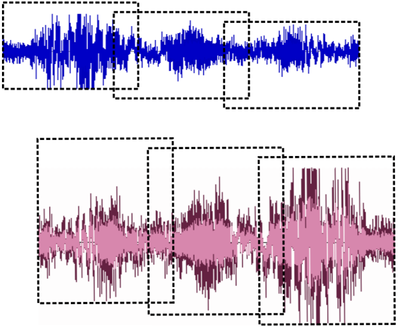


Example

- Electric supply data over 15 years covers 5000 days but only 15 christmas days
- Especially critical events like accidents (e.g. plane, car, earthquake) are scarce

Feature subset-selection

- Pre-process data
  - Framing
  - Normalisation



Feature subset-selection

- Pre-process data
  - Framing
  - Normalisation

Domain knowledge?
-> better set of ad-hoc features

Features commensurate?
-> normalise

Pruning of input required?
-> if no, create disjunctive features or weithted sums of features

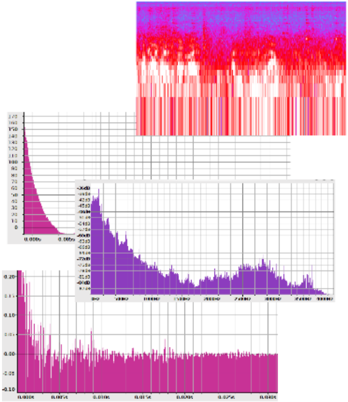
Independent features?
-> construct conjunctive features or products of features

Is the data noisy?
-> detect outlier examples

Do you know what to do first?
-> If not, use a linear predictor

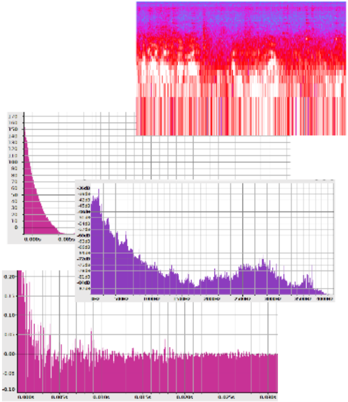
Feature extraction

- Identify meaningful features
  - remove irrelevant/redundant features



Feature extraction

- Identify meaningful features
  - remove irrelevant/redundant features
- Features can be contradictory!



Feature subset-selection

Simple ranking of features with correlation coefficients

Example: Pearson Correlation Coefficient

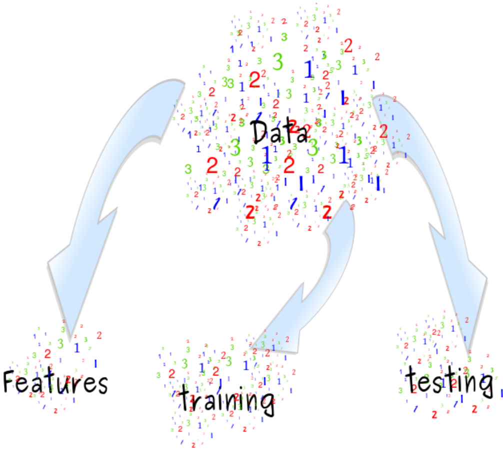
$$\varrho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{1}$$

- Identifies linear relation between input variables x_i and an output y

Feature subset-selection

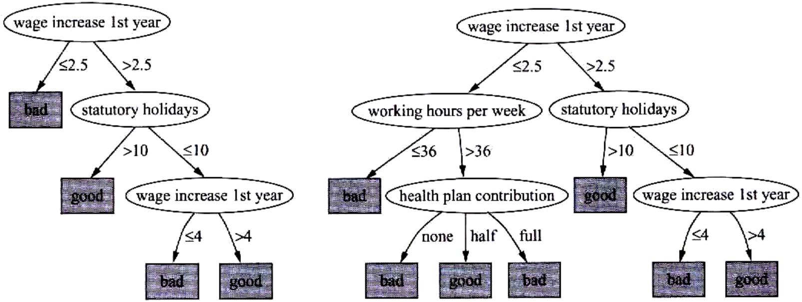
How to do reasonable feature selection

- Utilise dedicated test- and training- data-sets
- Pay attention that a single raw-data sample could not impact features in both these sets
- Don't train the features on the training- or test- data-set



Training of the classifier

A decision tree classifier



(a) (b)

Training of the classifier

Evaluation of classification performance

k-fold cross-validation

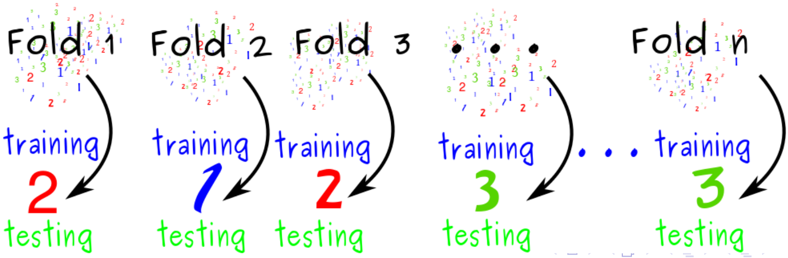
- Standard: k=10

Training of the classifier

Evaluation of classification performance

Leave-one-out cross-validation

- n-fold cross validation where n is the number of instances in the data-set
- Each instance is left out once and the algorithm is trained on the remaining instances
- Performance of left-out instance (success/failure)

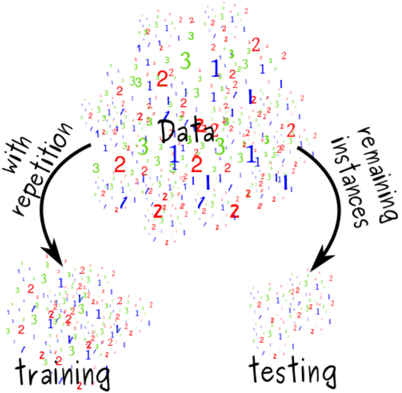


Training of the classifier

Evaluation of classification performance

0.632 Bootstrap

- Form training set by choosing n instances from the data-set with replacement
- All not picked instances are used for testing
- Probability to pick a specific instance:
  $1 - \left(1 - \frac{1}{n}\right)^n \approx 1 - e^{-1} \approx 0.632$



Training of the classifier

Evaluation of classification performance

Classification accuracy

- Confusion matrices
- Precision
- Recall

	Classification							
	Aw	No	To	Sb	Sl	Sr	St	$\sum$
Aw	**52**		3	6	0	17	22	100
No		**436**	25	7	6	17	9	500
To		40	**59**				1	100
Sb	15	22		**32**	4	22	5	100
Sl	12	11	1	6	**48**	8	14	100
Sr	4	15		6	1	**67**	7	100
St	3	18	1	1	24	10	**43**	100
$\sum$	92	551	86	65	94	129	83	

	Classification							
	Aw	No	To	Sb	Sl	Sr	St	recall
Aw	**.58**	.09		.13	.11	.05	.04	.58
No		**.872**	.05	.014	.012	.034	.018	.872
To		.4	**.59**				.01	.59
Sb	.15	.22		**.32**	.04	.22	.05	.32
Sl	.12	.11	.01	.06	**.48**	.08	.14	.48
Sr	.04	.15		.06	.01	**.67**	.07	.67
St	.03	.18	.01	.01	.24	.1	**.43**	.43
prec	.630	.791	.686	.492	.511	.519	.518	

Training of the classifier

Evaluation of classification performance

Information score

Let C be the correct class of an instance and $\mathcal{P}(C)$, $\mathcal{P}'(C)$ be the prior and posterior probability of a classifier

We define:[1]

$$I_i = \begin{cases} -\log(\mathcal{P}(C)) + \log(\mathcal{P}'(C)) & \text{if } \mathcal{P}'(C) \geq \mathcal{P}(C) \\ \log(1-\mathcal{P}(C)) + \log(1-\mathcal{P}'(C)) & \text{else} \end{cases} \tag{2}$$

The information score is then

$$\mathrm{IS} = \frac{1}{n}\sum_{i=1}^{n} I_i \tag{3}$$

[1] I. Kononenko and I. Bratko: Information-Based Evaluation Criterion for Classifier's Performance, Machine Learning, 6, 67-80, 1991.

Training of the classifier

Evaluation of classification performance

Brier score

The Brier score is defined as

$$\text{Brier} = \sum_{i=1}^{n} (t(x_i) - p(x_i))^2 \tag{4}$$

where

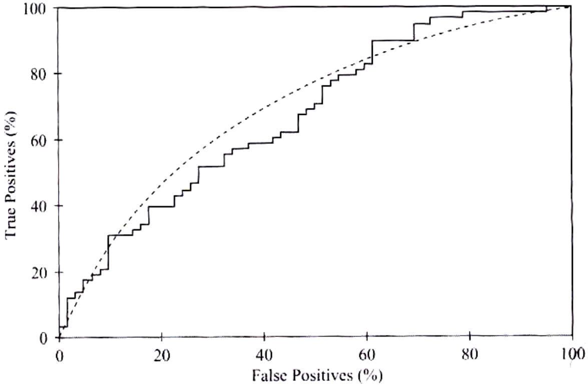
$$t(x_i) = \begin{cases} 1 & \text{if } x_i \text{ is the correct class} \\ 0 & \text{else} \end{cases} \tag{5}$$

and $p(x_i)$ is the probability the classifier assigned to the class x_i.

Training of the classifier

Evaluation of classification performance

Area under the receiver operated characteristic (ROC) curve (AUC)



Rank	Predicted	Actual Class
1	0.95	*yes*
2	0.93	*yes*
3	0.93	*no*
4	0.88	*yes*
5	0.86	*yes*
6	0.85	*yes*
7	0.82	*yes*
8	0.80	*yes*
9	0.80	*no*
10	0.79	*yes*
11	0.77	*no*
12	0.76	*yes*
13	0.73	*yes*
14	0.65	*no*
15	0.63	*yes*
16	0.58	*no*
17	0.56	*yes*
18	0.49	*no*
19	0.48	*yes*
...	...	...

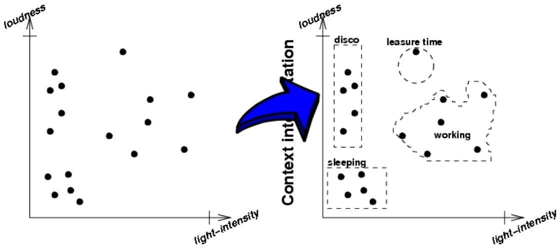
Pattern recognition and classification

Data mining frameworks

- Orange Data Mining (http://orange.biolab.si/)
- Weka Data Mining (http://www.cs.waikato.ac.nz/ml/weka/)

Pattern recognition and classification

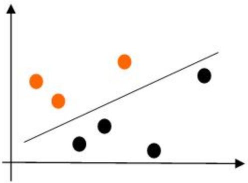
- From features to context
  - Measure available data on features
  - Context reasoning by appropriate method
    - Syntactical (rule based – e.g. RuleML)
    - Bayesian classifier
    - Non-parametric
    - Linear discriminant
    - Neural networks
    - Sequential
    - Stochastic



Pattern recognition and classification

- Allocation of sensor value by defined function
  - Correlation of various data sources
  - Several methods possible – simple approaches
  - Template matching
  - Minimum distance methods
  - 'Integrated' feature extraction
    - Nearest Neighbour
    - Neural Networks
- Problem
  - Measured raw data might not allow to derive all features required
  - Therefore often combination of sensors

Not enough features



Sufficient feature count

Pattern recognition and classification

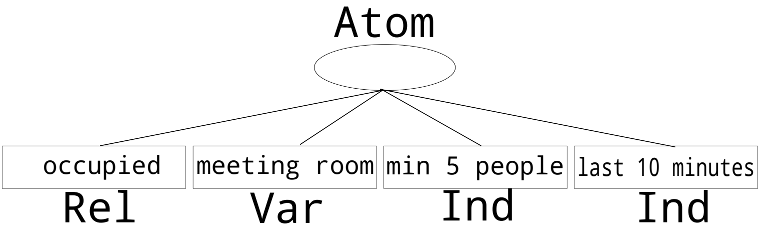
- Methods – Syntactical (Rule based)
  - Idea: Description of Situation by formal Symbols and Rules
  - Description of a (agreed on?) world view
  - Example: RuleML
- Comment
  - Pro:
    - Combination of rules and identification of loops and impossible conditions feasible

    Contra:
    - Very complex with more elaborate situations
    - Extension or merge of rule sets typically not possible without contradictions

Pattern recognition and classification

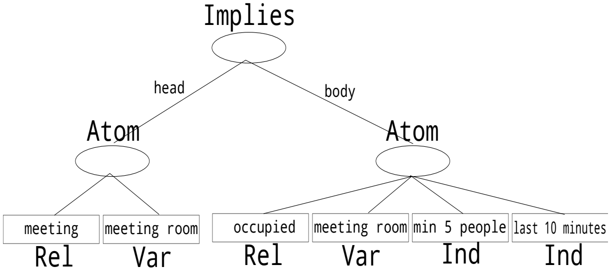
- Rule Markup Language: Language for publishing and sharing rules
- Hierarchy of rule-sub-languages (XML, RDF, XSLT, OWL)
- Example:
  - A meeting room was occupied by min 5 people for the last 10 minutes.

```
<Atom>
     <Rel> occupied </Rel>
     <Var> meeting room </Var>
     <Ind> min 5 persons </Ind>
     <Ind> last 10 minutes </Ind>
</Atom>
```



Pattern recognition and classification

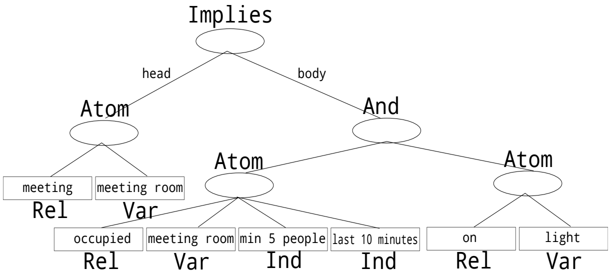
- Also conditions can be modelled
  - A Meeting is taking place in a meeting room when it was occupied by min 5 people for the last 10 minutes.




```
<Implies>
  <head>
    <Atom>
        <Rel> meeting </Rel>
        <Var> meeting room </Var>
    </Atom>
  </head>
  <body>
    <Atom>
        <Rel> occupied </Rel>
        <Var> meeting room </Var>
        <Ind> min 5 persons </Ind>
        <Ind> last 10 minutes </Ind>
    </Atom>
  </body>
</Implies>
```

Pattern recognition and classification

- Logical combination of conditions
  - A Meeting is taking place in a meeting room when it was occupied by min 5 people for the last 10 minutes and the light is on.




```
<Implies>
   <head>
      <Atom>
         <Rel> meeting </Rel>
         <Var> meeting room </Var>
      </Atom>
   </head>
   <body>
      <And>
         <Atom>
            <Rel> on </Rel>
            <Var> light </Var>
         </Atom>
         <Atom>
            <Rel> occupied </Rel>
            <Var> meeting room </Var>
            <Ind> min 5 persons </Ind>
            <Ind> last 10 minutes </Ind>
         </Atom>
      </And>
   </body>
</Implies>
```

Outline