WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Machine Learning

- Basic concepts
- Tools

EFMD EQUIS ACCREDITED

???

and then!??

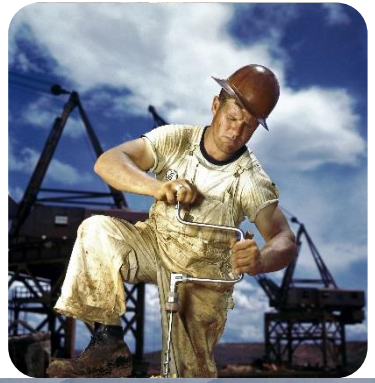
↓

prediction!

on Fridays:

Confidence 75%

probability

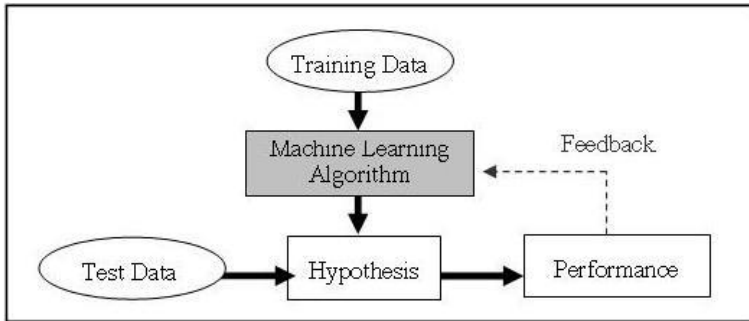


Walmart

But… why?

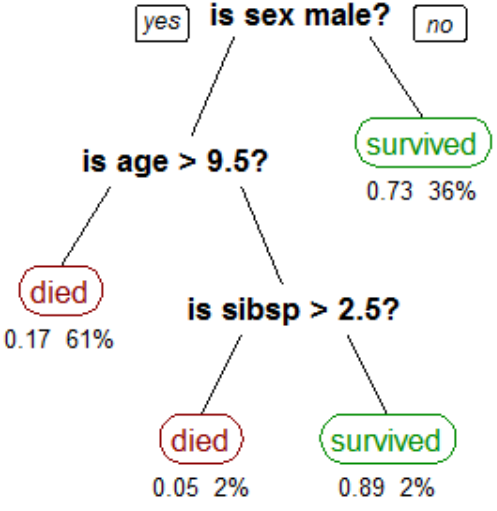
ML: Basic Concepts

- how to construct computer programs that automatically **improve** with experience
  - TASK: recognize handwritten words
  - Performance: % words correctly classified
  - Training data: a set of handwritten words, with given classification



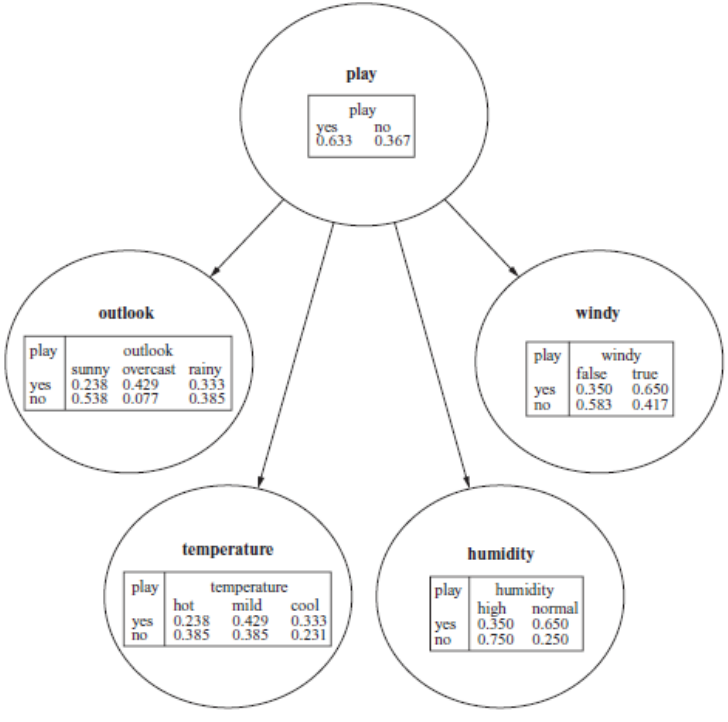
ML: Basic Concepts

- Algorithms
  - Supervised (labelled by a supervisor)
    - **Decision Trees, Decision Rules**
    - Bayesian Classifiers
      - to which of a set of categories a new observation belongs
  - Unsupervised (finding interesting groups into data)
    - Clustering
    - Association rules



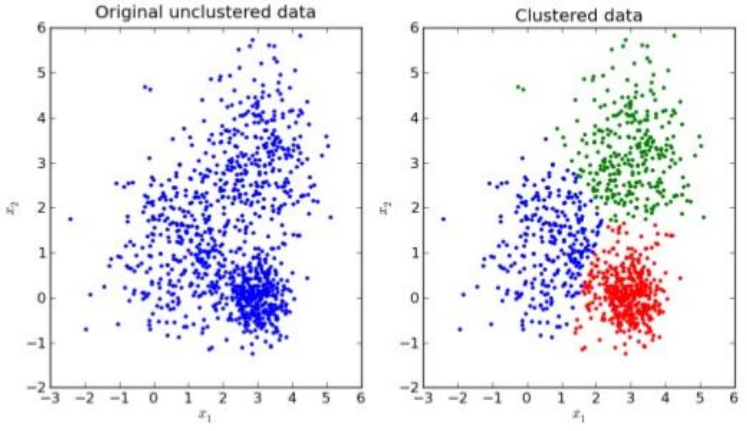
ML: Basic Concepts

- Algorithms
  - Supervised (labelled by a supervisor)
    - Decision Trees, Decision Rules
    - **Bayesian Classifiers**
      - to which of a set of categories a new observation belongs
  - Unsupervised (finding interesting groups into data)
    - Clustering
    - Association rules



ML: Basic Concepts

- Algorithms
  - Supervised (labelled by a supervisor)
    - Decision Trees, Decision Rules
    - Bayesian Classifiers
      - to which of a set of categories a new observation belongs
  - Unsupervised (finding interesting groups into data)
    - **Clustering**
    - Association rules



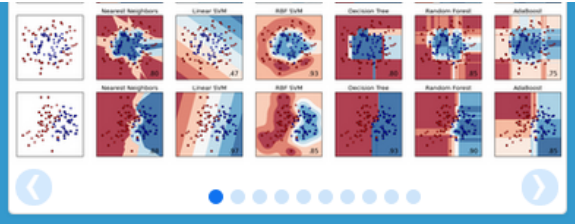
ML: Basic Concepts

- Algorithms
  - Supervised (labelled by a supervisor)
    - Decision Trees, Decision Rules
    - Bayesian Classifiers
      - to which of a set of categories a new observation belongs
  - Unsupervised (finding interesting groups into data)
    - Clustering
    - **Association rules**
      - interesting relations between variables

$$
\begin{aligned}
&\{I1, I2\} \Rightarrow I5, && confidence = 2/4 = 50\% \\
&\{I1, I5\} \Rightarrow I2, && confidence = 2/2 = 100\% \\
&\{I2, I5\} \Rightarrow I1, && confidence = 2/2 = 100\% \\
&I1 \Rightarrow \{I2, I5\}, && confidence = 2/6 = 33\% \\
&I2 \Rightarrow \{I1, I5\}, && confidence = 2/7 = 29\% \\
&I5 \Rightarrow \{I1, I2\}, && confidence = 2/2 = 100\%
\end{aligned}
$$

Machine learning in python

- http://scikit-learn.org/stable/index.html



- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.
Algorithms: *SVM*, *nearest neighbors*, *random forest*, ... — *Examples*

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: *SVR*, *ridge regression*, *Lasso*, ... — *Examples*

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: *k-Means*, *spectral clustering*, *mean-shift*, ... — *Examples*

Dimensionality reduction

Reducing the number of random variables to

Model selection

Comparing, validating and choosing

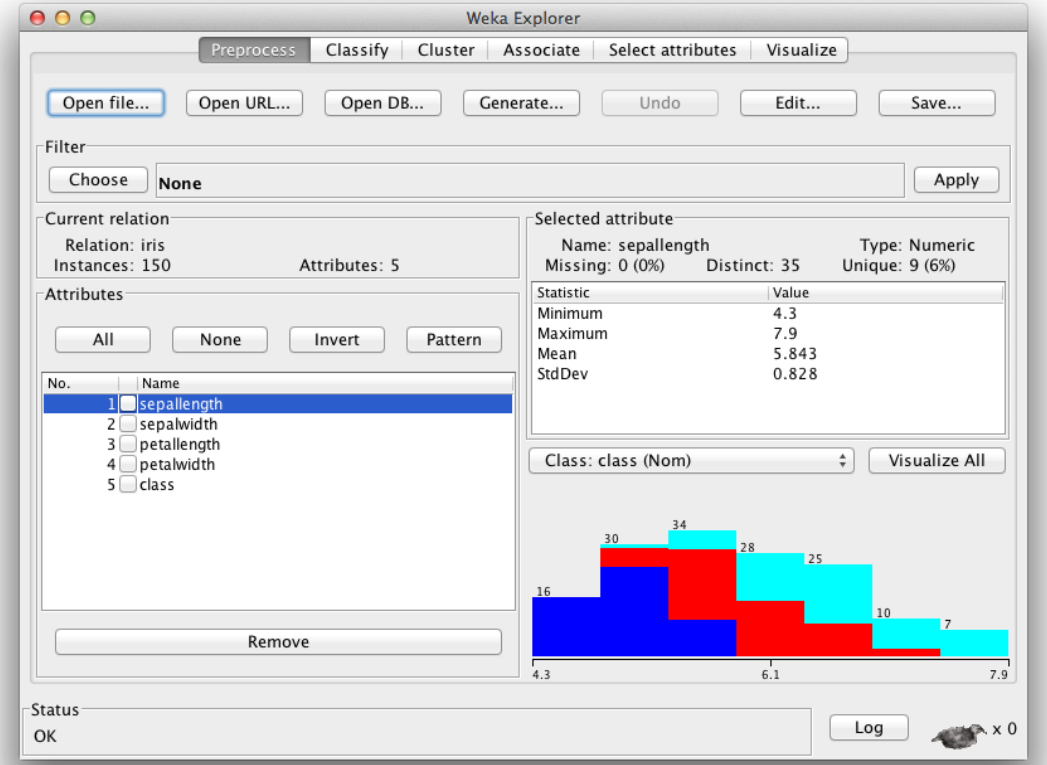
Preprocessing

Feature extraction and normalization.

xmllint

Machine learning in Weka

- http://www.cs.waikato.ac.nz/ml/weka/



xmllint

Other tools

- Matlab
- R
- **Orange**
- RapidMiner

- Quick tutorial: http://de.slideshare.net/liorrokach/introduction-to-machine-learning-13809045