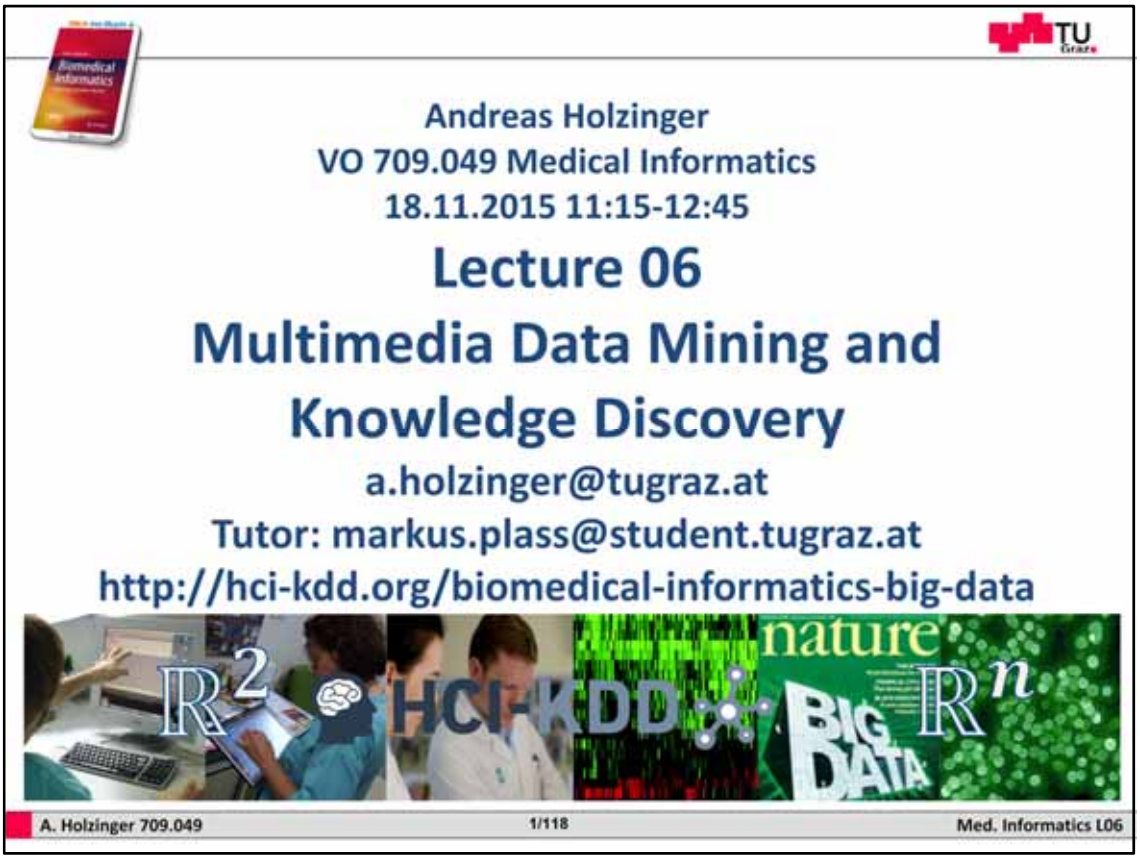


Status as of 16.11.2015 06:00

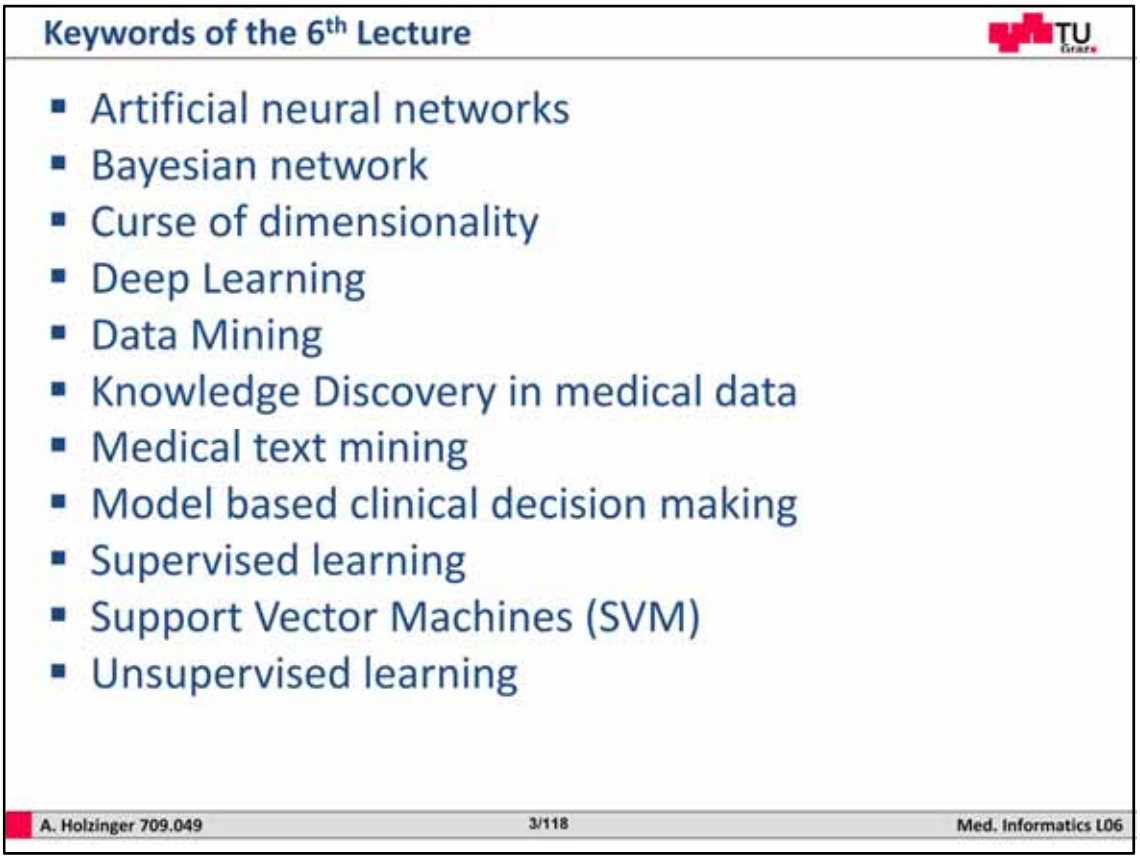
Dear Students, welcome to the 6th lecture of our course. Please remember from the last lecture the difference between well-structured, weakly-structured, standardized, non-standardized; XML as semi-structured format for the electronic patient record (epr); networks = graphs; important network metrics, small-world network; how to get point-cloud data sets from natural images; examples for a medical knowledge space, graph entropy measures.

Please always be aware of the definition of biomedical informatics (Medizinische Informatik):
Biomedical Informatics is the inter-disciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health (and well-being).

Schedule of the course

- 1. Intro: Computer Science meets Life Sciences, challenges, future directions
- 2. Back to the future: Fundamentals of Data, Information and Knowledge
- 3. Structured Data: Coding, Classification (ICD, SNOMED, MeSH, UMLS)
- 4. Biomedical Databases: Acquisition, Storage, Information Retrieval and Use
- 5. Semi structured and weakly structured data (structural homologies)
- **6. Multimedia Data Mining and Knowledge Discovery**
- 7. Knowledge and Decision: Cognitive Science & Human-Computer Interaction
- 8. Biomedical Decision Making: Reasoning and Decision Support
- 9. Intelligent Information Visualization and Visual Analytics
- 10. Biomedical Information Systems and Medical Knowledge Management
- 11. Biomedical Data: Privacy, Safety and Security
- 12. Methodology for Info Systems: System Design, Usability & Evaluation

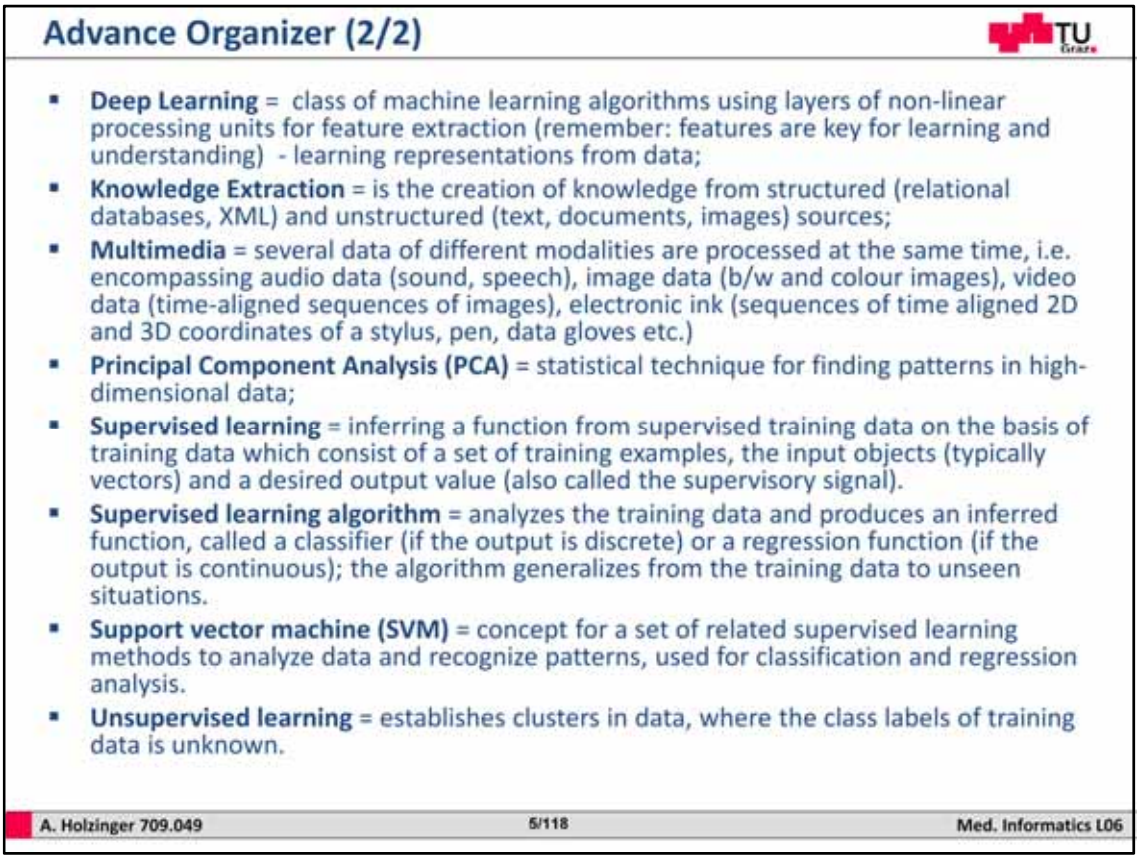
A. Holzinger 709.049 2/118 Med. Informatics L06


Keywords of the 6th Lecture
TU Graz
Artificial neural networks
Bayesian network
Curse of dimensionality
Deep Learning
Data Mining
Knowledge Discovery in medical data
Medical text mining
Model based clinical decision making
Supervised learning
Support Vector Machines (SVM)
Unsupervised learning
A. Holzinger 709.049
3/118
Med. Informatics L06

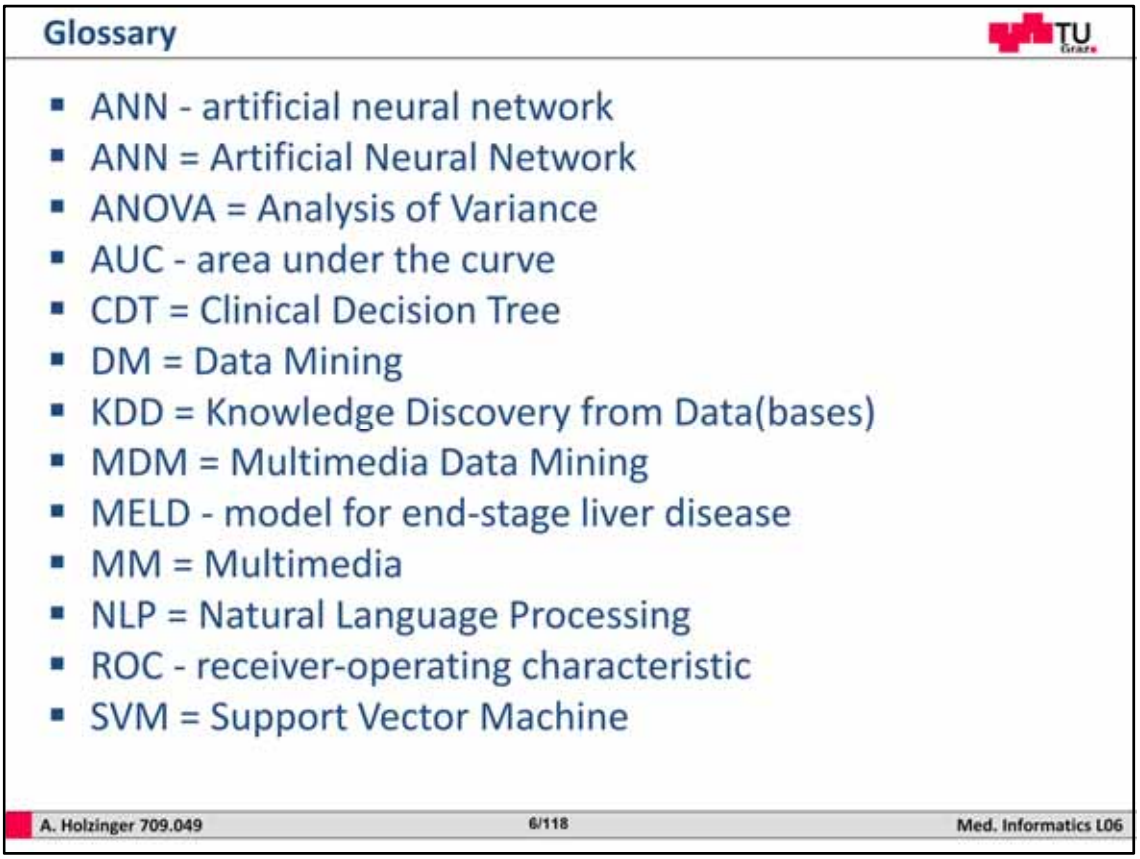
Advance Organizer (1/2)

- **Artificial neural network (ANN)** = a computational adaptive model (inspired by biological neural networks), consisting of interconnected groups of artificial neurons; processes information using a connectionist approach.
- **Association rule learning** = a set of techniques for discovering interesting relationships, i.e., "association rules," among variables in large databases used for data mining;
- **Classification** = a set of techniques to identify the categories in which new data points belong, based on a training set containing data points that have already been categorized; these techniques are often described as supervised learning because of the existence of a training set; they stand in contrast to cluster analysis, a type of unsupervised learning; used e.g. for data mining;
- **Cluster analysis** = statistical method for classifying objects that splits a diverse group into smaller groups of similar objects, whose characteristics of similarity are not known in advance; a type of unsupervised learning because training data are not used - in contrast to classification; used for data mining.
- **Data mining** = a set of techniques to extract patterns from large data by combining methods from statistics and machine learning with database management (e.g. association rule learning, cluster analysis, classification, regression, etc.);
- **Knowledge Discovery (KD)** = process of identifying valid, novel, useful and understandable patterns out of large volumes of data

A. Holzinger 709.049 4/118 Med. Informatics L06



Summarized, we have the following challenges:
1) Cross-disciplinary cooperation with domain experts
2) Data-driven challenges including a) Massive data sets; b) Heterogeneous Data; c) Streaming Data (e.g. from large sensor networks, Multimedia, etc.); d) Graph Data (e.g. Protein Network data, etc.); e) Data restrictions (accessibility, privacy, legal restrictions, etc.);
3) Context (Data Mining occurs always in a particular context)
4) Interpretability
5) Computational Resources
6) Benchmarking against Gold-Standards
8) Embedded data mining
Cross-disciplinary cooperation with domain experts
http://deeplearning.net/deep-learning-research-groups-and-labs/



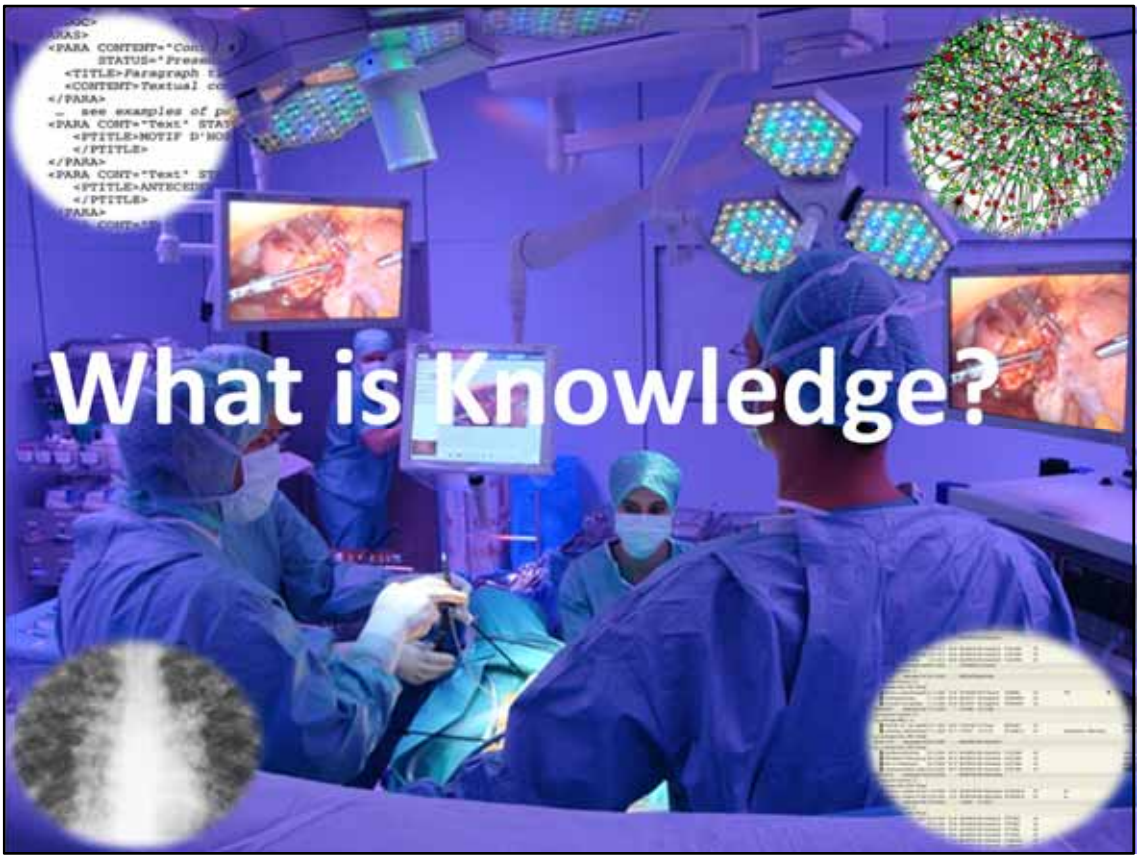
A big issue is the so called knowledge acquisition bottleneck. Many researchers from the data mining community expected that machine learning techniques would automate the knowledge acquisition process and thus would exclude experts form the process of building models – but the opposite is true, we need the knowledge of the domain expert, hence data mining approaches must put the human intelligence into the loop (Holzinger, 2013).

3) Data-driven challenges including
a) Massive data sets
b) Heterogeneous Data
c) Streaming Data (e.g. from large sensor networks)
d) Graph Data
Many data mining methods are designed for collections of objects well-represented in rigid tabular formats. However, we are increasingly finding ourselves confronting collections of interrelated objects whose natural representation is in graphs, in particular biomedical structures.
e) Data Restrictions
including data accessibility, temporal limits, legal restrictions, privacy, data provenance (such as in situations where copyright or patents may be relevant). We face a range of research challenges in developing data mining methods capable of honoring various restrictions that are not directly about the data but rather about the use of the data.

Learning Goals: At the end of this 6th lecture you …

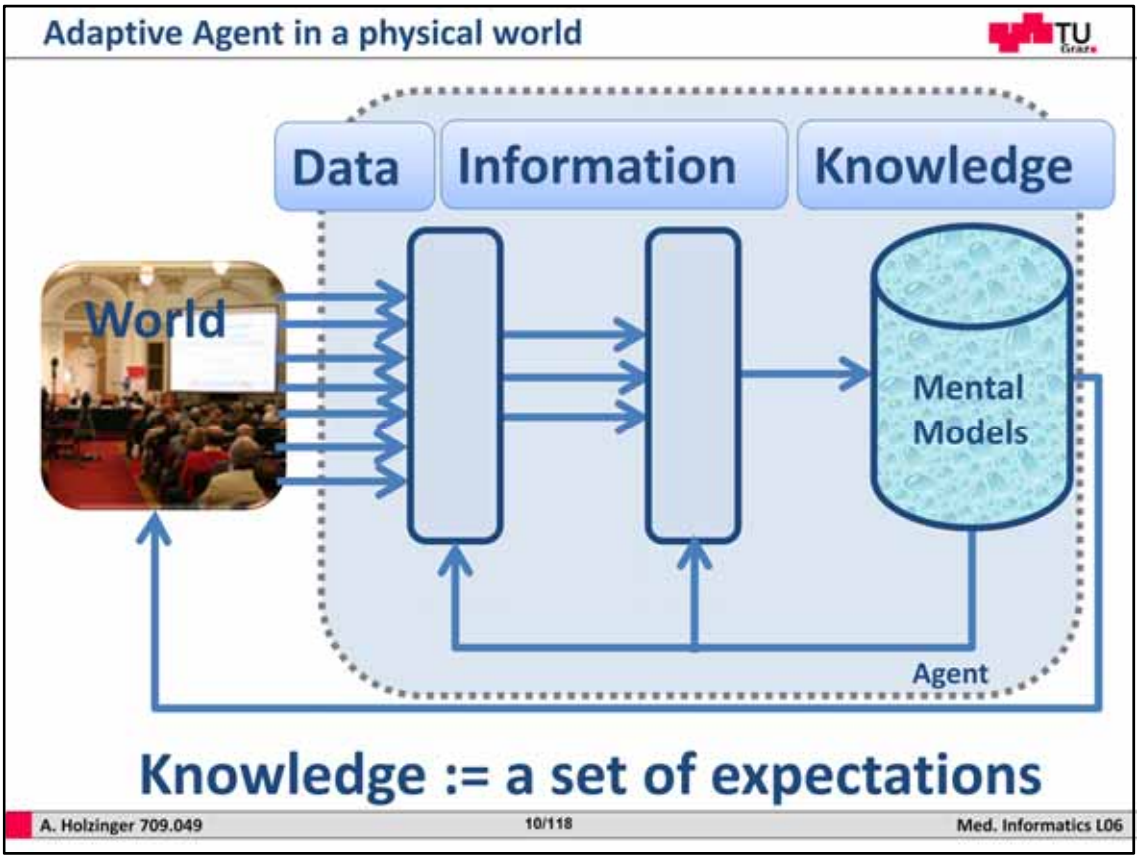
- … are aware of the importance of gaining **knowledge from (big) data**;
- … know the differences between Data Mining and Knowledge Discovery;
- … understand the basic process of knowledge discovery from data(bases) (KDD-chain);
- … have an overview on some data mining algorithms used in biomedical informatics;
- … have seen some examples of data mining applied in the biomedical domain;

3) Context
In biomedicine we need experts who understand the domain, the problem, and the data sets, hence the context (Berka, Rauch & Tomecková, 2007).
4) Interpretability
Much data mining research focuses on well-defined metrics such as classifier accuracy, which does not always match the goals of particular data mining tasks. One broad example of this is a case where there is a need for interpretability in the results of data mining. In biomedicine, customer relationship management, or domains with appropriate regulatory characteristics (where you may need to explain the results of a decision), data mining is often irrelevant if it does not produce results that can be explained to others. 5) Computing Resources
We are beginning to see the development of largescale data-intensive computing clusters by major Internet companies, and such architectures can change the nature of how data mining will be performed.
6) Benchmarking against Gold-Standards
Data Some areas of data mining rely heavily on benchmark data sets (see e.g., (Kreuzthaler et al., 2010), (Kreuzthaler et al., 2011)), which allow us to compare results across competing methods. However, benchmarks are a means to an end, not the end in itself. These benchmark problems are intended to be representative of the sorts of problems our algorithms will see in practice, but what we will see in practice will change over time. We thus make sure that our data sets stay timely as technological and scientific advances allow our ambitions to grow. For example, we do not have widely available benchmark 'massive' data sets. There is value in enabling widespread access to benchmark massive data sets, such as snapshots of the World Wide Web, social networks, etc. Even so, we must remain vigilant, realizing that even these benchmarks will outlive their value as research tools as our capabilities and ambitions grow.
7) Reproducibility, both to verify the results of others and to build off of them. Unfortunately, we often do not see results documented with the rigor found in other noncomputing experimental sciences. How can we replicate results generated on proprietary or otherwise restricted data? How can results be replicated if they require the use of specialized equipment, such as massive data clusters found in only a handful of commercial enterprises? How do we make sure academic research addresses problems that are important in practice when most academic institutions cannot match the resources available in the most heavily data-driven enterprises? More fundamentally, even when a researcher has access to identical resources, including both data as well as computing and software platforms, it is often difficult to replicate results identically in the absence of details about data cleaning, algorithm parameter selections, and the like. As a scholarly community, we need to be able to document experimental results in sufficient detail to allow reproducibility.
8) Embedded data mining
A future goal would be to have KDD tools not separately but integrated in the workbenches, e.g. integrated into the clinical workplace in a hospital. Advanced biomedical technologies have made systems biology a promising area in which data mining faces new challenges and has an increasing importance (Hirsh, 2008).



Here a typical scenario in the hospital with a heterogeneity of different data.

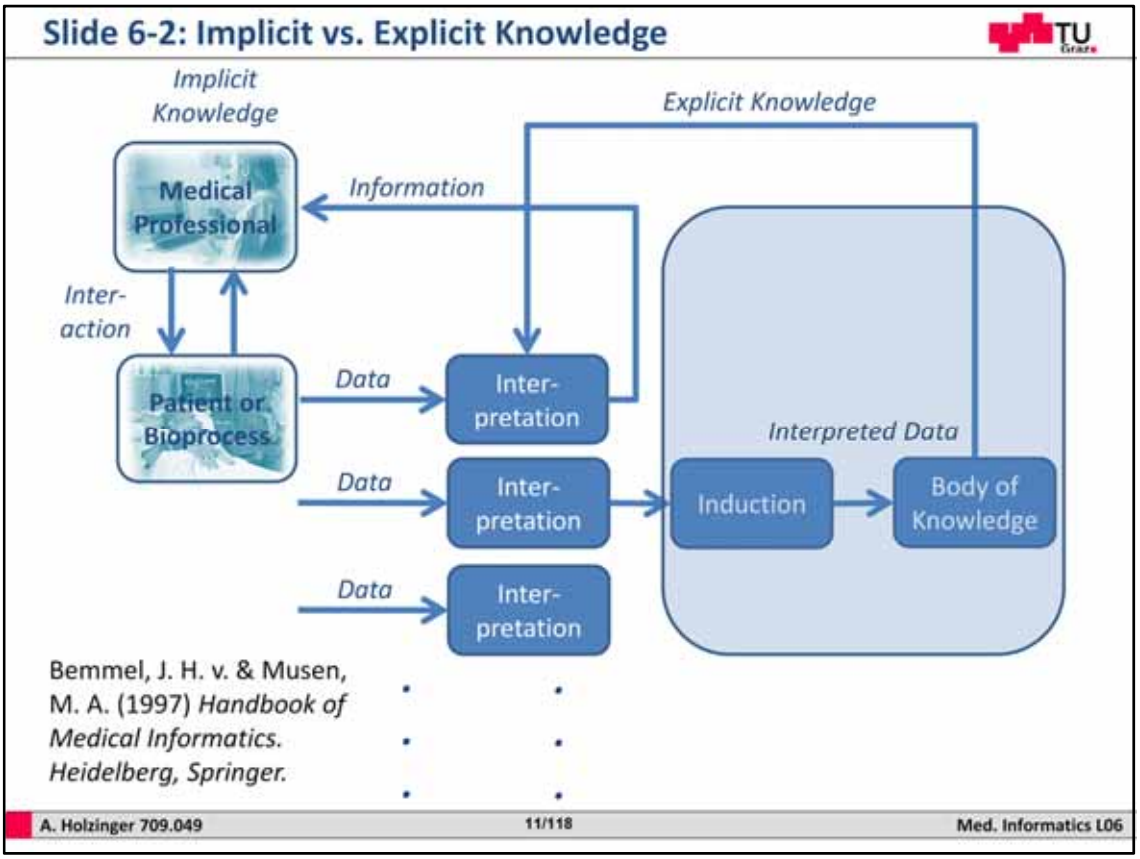
You should always remember the differences between data, information and knowledge; and in particular the difference between structured and unstructured data, the latter mixed up with unmodelled data – called unstructured information !



Agent (lat. für “Vermittler”, “Handelnder”); Software-Agent ist autonomem Verhalten fähig ist. Das bedeutet, dass abhängig von verschiedenen Zuständen (Status) ein bestimmter Verarbeitungsvorgang abläuft, ohne dass von außen ein weiteres Startsignal gegeben wird oder während des Vorgangs ein äußerer Steuerungseingriff erfolgt;

In English do not mix up the word „agent“ with biological agent or pharmacological agent,
we mean an intelligent agent (IA) which is defined as an autonomous entity which observes through sensors and acts upon an environment using actuators
Intelligent agents may also learn or use knowledge to achieve their goals.

One of the fundamental goals of artificial intelligence is to understand and develop intelligent agents that simulate
human-like intelligence.
Li, N., Matsuda, N., Cohen, W. W. & Koedinger, K. R. 2015. Integrating representation learning and skill learning in a human-like intelligent agent. Artificial Intelligence, 219, 67-91.

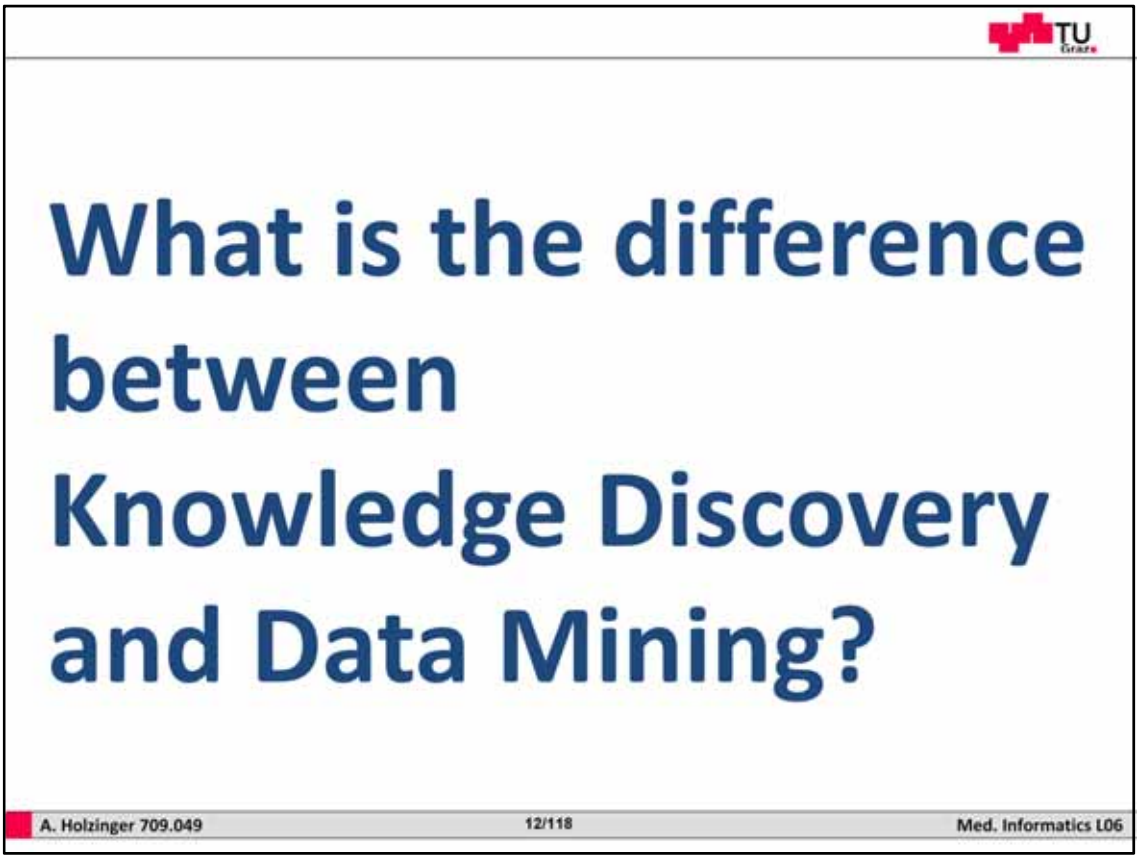


Data are the physical entities at the lowest abstraction level which are, e.g. generated by a patient (patient data) or a biological process (e.g. Omics data). According to (Bemmel & Musen, 1997) data contain no meaning.

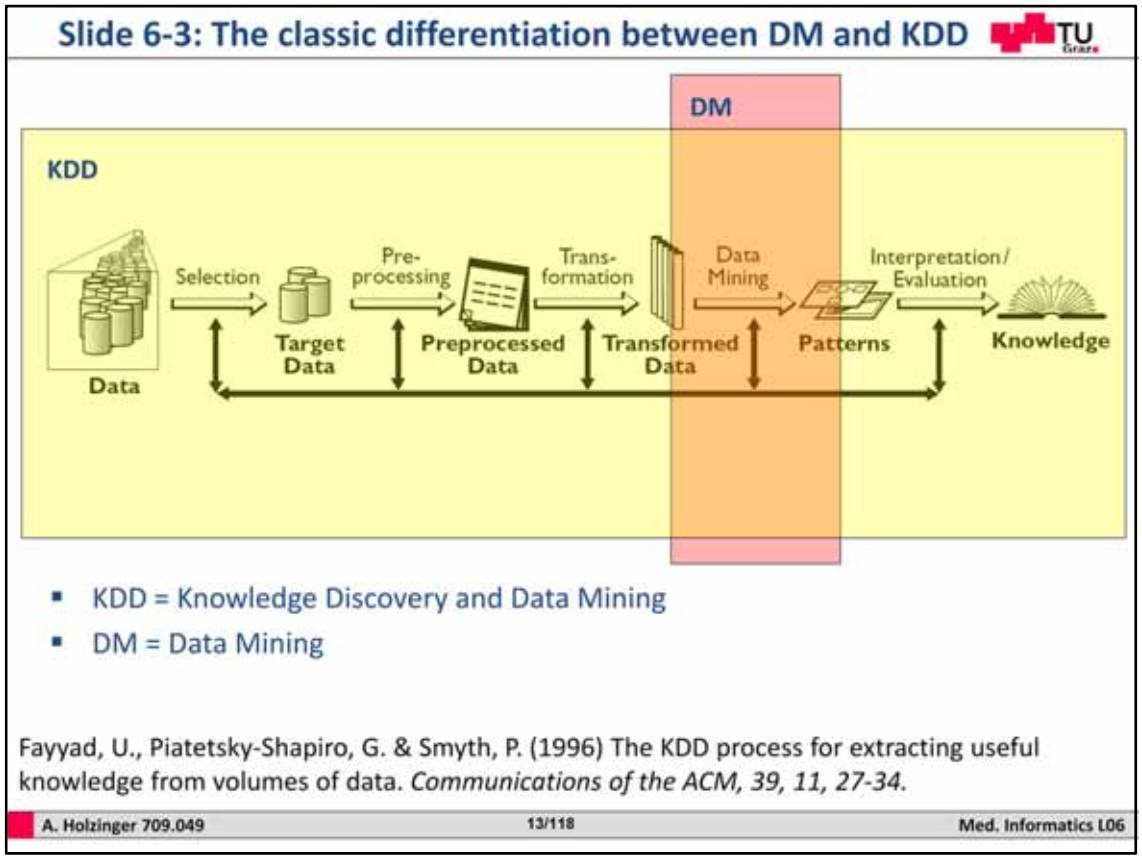
Information is derived by interpretation of the data by a clinician (human intelligence).

Knowledge is obtained by inductive reasoning with previously interpreted data, collected from many similar patients or processes, which is added to the so called body of knowledge in medicine, the **explicit knowledge.** This knowledge is used for the interpretation of other data and to gain **implicit knowledge** which guides the clinician in taking further action.

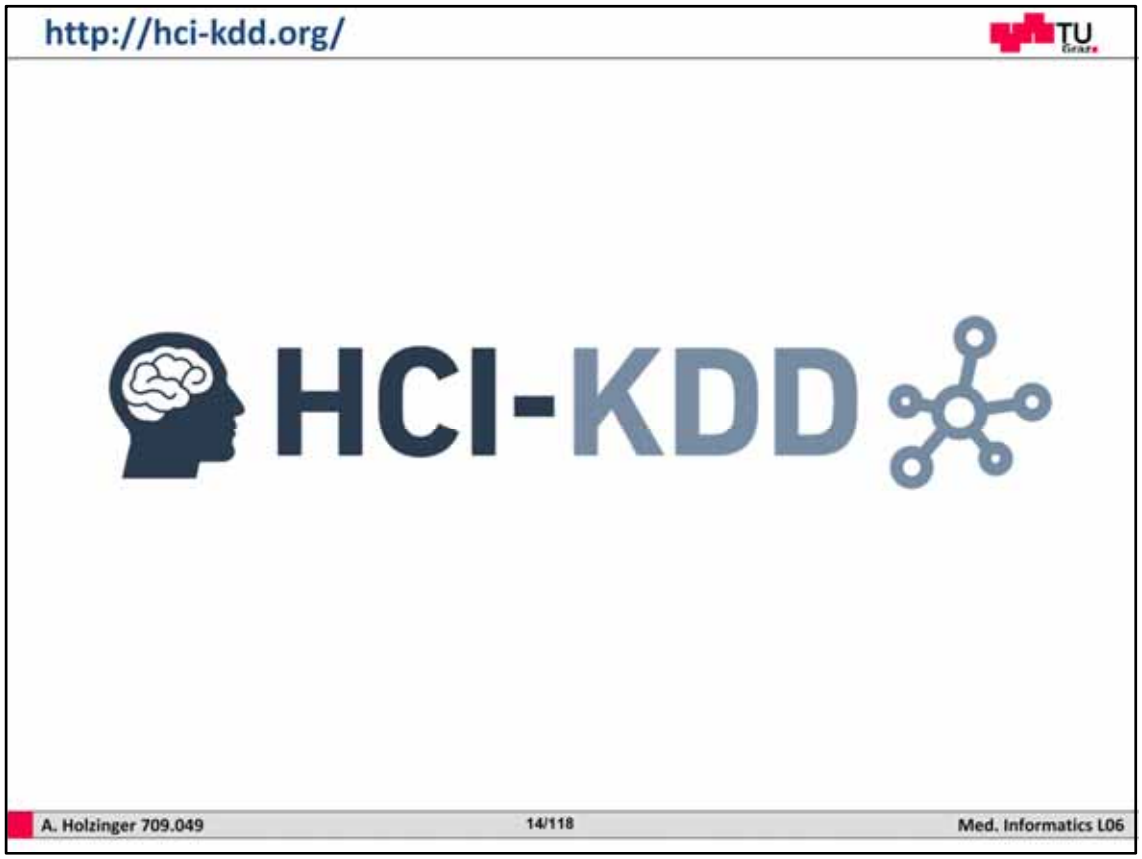
Explicit Knowledge is obtained by inductive reasoning with previously interpreted data, collected from many similar patients or processes, which is added to the so-called "body of knowledge".
Implicit Knowledge is gained by the interpretation of other data on the basis of explicit knowledge and guides the clinician in making decisions and taking further action.

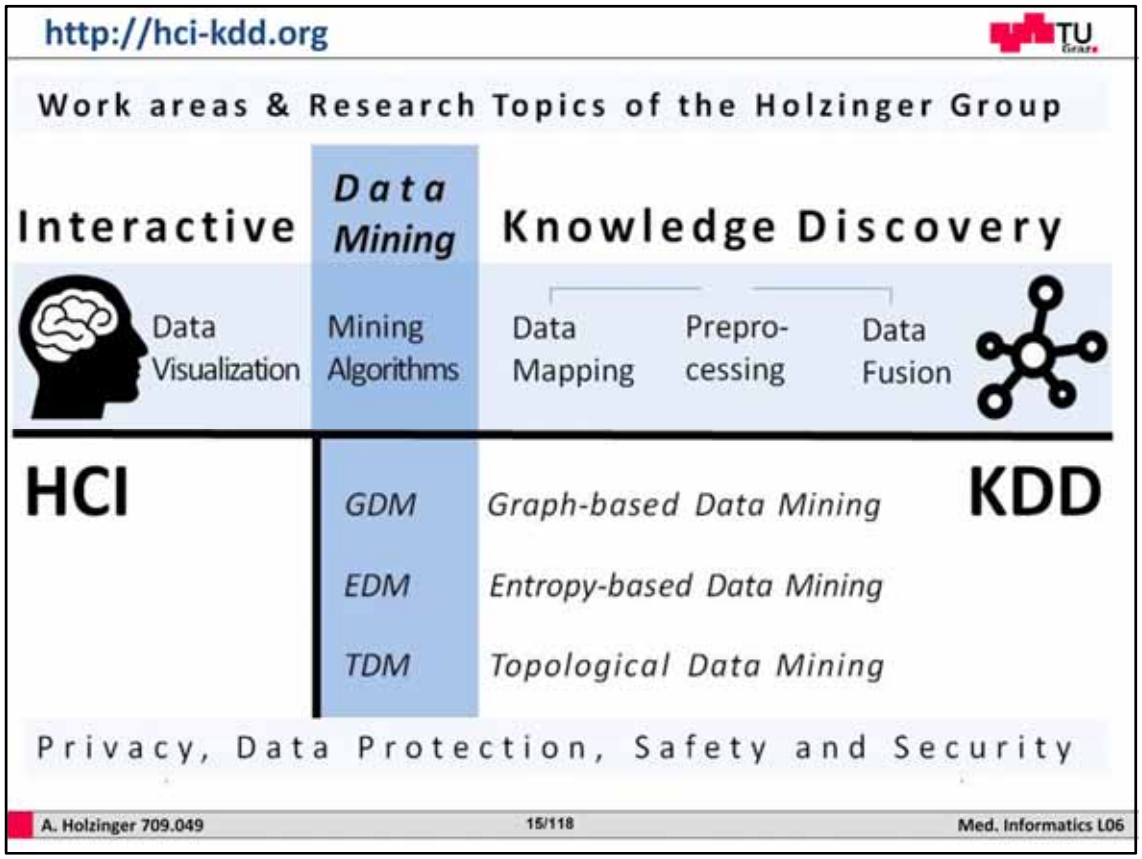


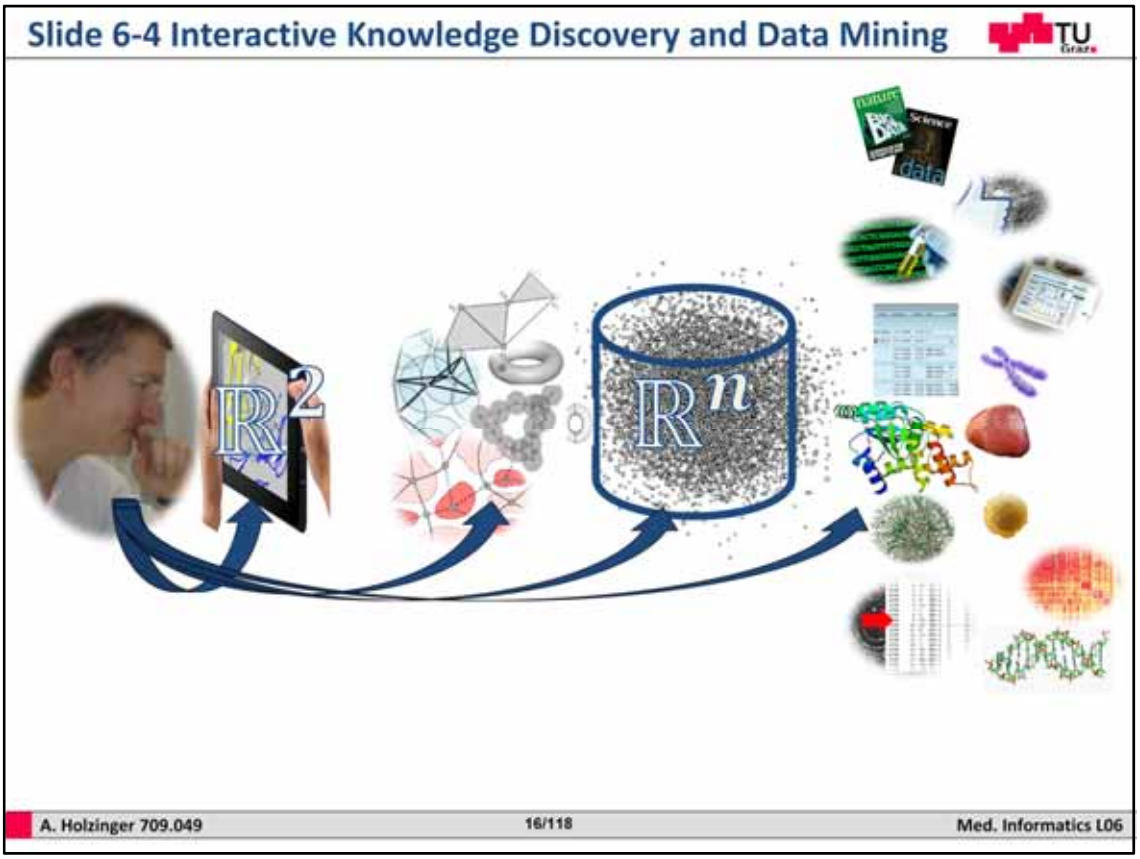
We can understand the difference between DM and KDD best, when looking at the classic KDD process chain by (Fayyad, Piatetsky-Shapiro & Smyth, 1996): DM is a subset of KDD. (see next slide)



1. Learning from the application domain: includes understanding relevant previous knowledge, the goals of the application and a certain amount of domain expertise;
2. Creating a target dataset: includes selecting a dataset or focusing on a subset of variables or data samples on which discovery shall be performed;
3. Data cleansing (this is not a typo, it is called cleansing, not cleaning ;-) and preprocessing: includes removing noise or outliers, strategies for handling missing data etc.;
4. Data reduction and projection: includes finding useful features to represent the data, dimensionality reduction etc.; 5. Choosing the function of data mining: includes deciding the purpose and principle of the model for mining algorithms (e.g., summarization, classification, regression, and clustering);
6. Choosing the data mining algorithm: includes selecting method(s) to be used for searching for patterns in the data, such as deciding which models and parameters may be appropriate (e.g., models for categorical data are different from models on vectors over reals) and matching a particular data mining method with the criteria of the KDD process;
7. Data mining: searching for patterns of interest in a representational form or a set of such representations, including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis;
8. Interpretation: includes interpreting the discovered patterns and possibly returning to any of the previous steps, as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns, and translating the useful ones into terms understandable by users;
9. Using discovered knowledge: includes incorporating this knowledge into the performance of the system, taking actions based on the knowledge, or documenting it and reporting it to interested parties, as well as checking for and resolving potential conflicts with previously believed knowledge.


http://hci-kdd.org/
HCI-KDD
A. Holzinger 709.049
14/118
Med. Informatics L06

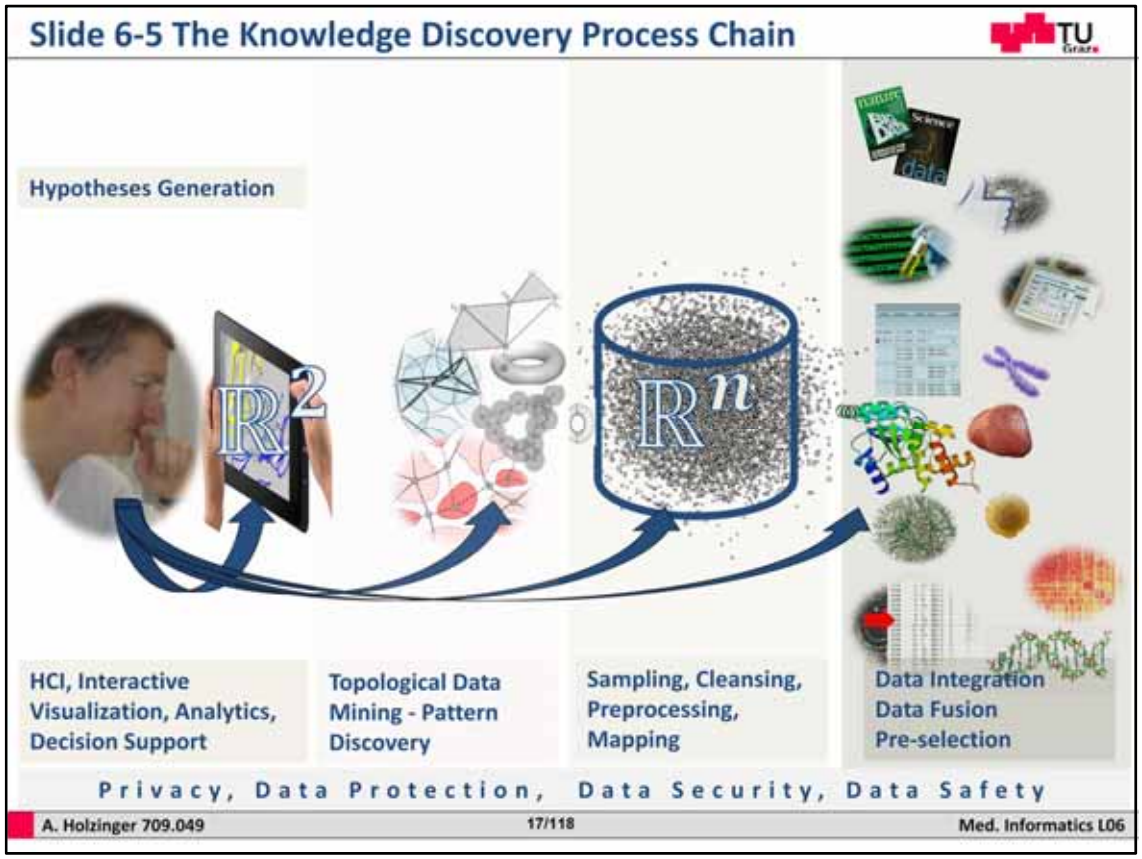

http://hci-kdd.org
TU Graz
Work areas & Research Topics of the Holzinger Group
Interactive
Data Mining
Knowledge Discovery
Data Visualization
Mining Algorithms
Data Mapping
Prepro-cessing
Data Fusion
HCI
KDD
GDM
Graph-based Data Mining
EDM
Entropy-based Data Mining
TDM
Topological Data Mining
Privacy, Data Protection, Safety and Security
A. Holzinger 709.049
15/118
Med. Informatics L06



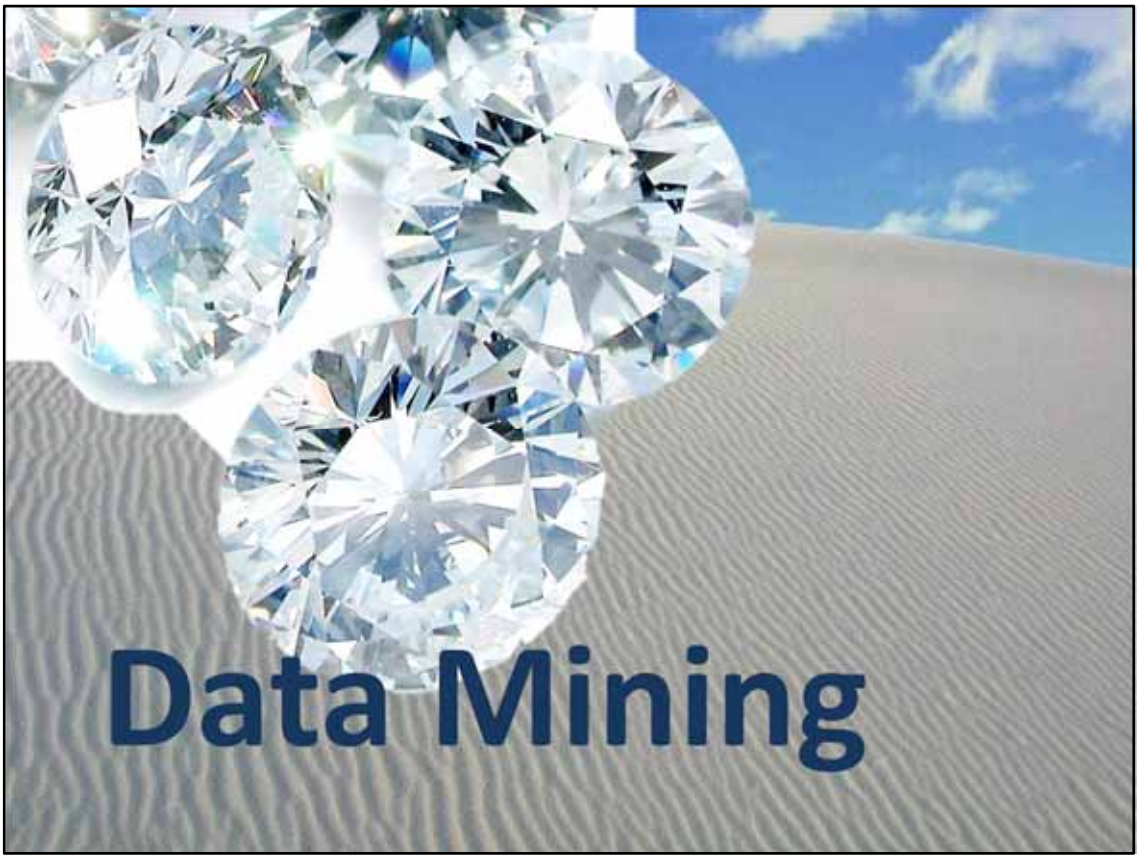
Interaction, communication and sensemaking are still missing, which are core topics in HCI (Blandford & Attfield, 2010). Consequently, a novel approach is to combine HCI & KDD (Holzinger, 2012) in order to enhance human intelligence by computational intelligence. The main contribution of HCI-KDD is to enable end users to find and recognize previously unknown and potentially useful and usable information. It may be defined as the process of identifying novel valid, and potentially useful data patterns, with the goal to understand these data patterns (Funk & Xiong, 2006). The domain expert in Slide 6-4 possesses explicit domain knowledge and by enabling him to interactively look at data sets, he may be able to identify, extract and understand useful information, to gain new, previously unknown knowledge (Holzinger et al., 2012).

Knowledge Discovery from Data: By getting insight into the data; the gained information can be used to build up knowledge. The grand challenge is to map higher dimensional data into lower dimensions, hence make it interactively accessible to the end-user (Holzinger, 2012), (Holzinger, 2013).
This mapping from $\mathbb{R}^n \rightarrow \mathbb{R}^2$ is the core task of visualization and a major component for knowledge discovery: Enabling effective interactive human control over powerful machine algorithms to support human sensemaking (Holzinger, 2012), (Holzinger, 2013).

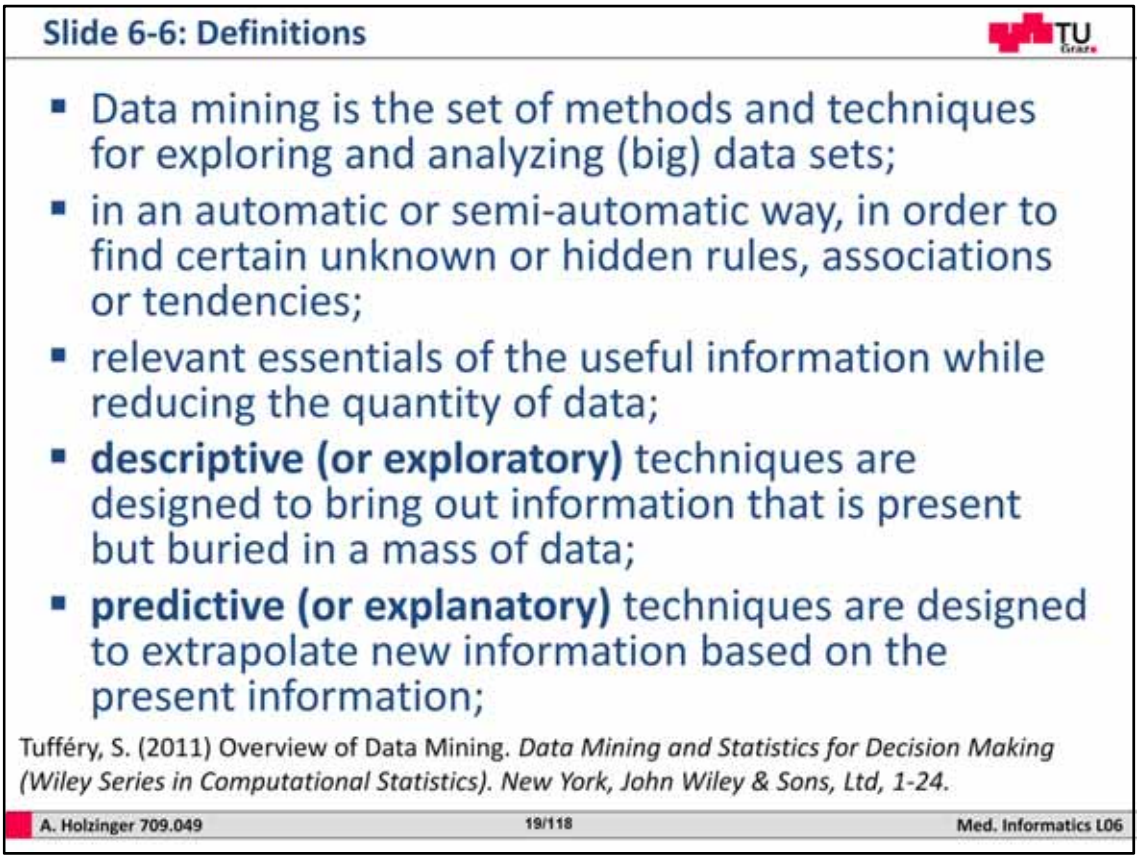
Holzinger, A. 2013. Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? *In: Alfredo Cuzzocrea, C. K., Dimitris E. Simos, Edgar Weippl, Lida Xu (ed.) Multidisciplinary Research and Practice for Information Systems, Springer Lecture Notes in Computer Science LNCS 8127. Heidelberg, Berlin, New York: Springer, pp. 319-328.*



KDD historically builds on 3 fields: computational statistics and machine learning; databases and artificial intelligence to construct tools that let the end users gain insight into the nature of massive data sets. Pazzani (2000) emphasizes that it is necessary to take human cognitive processes into account in order to increase the usefulness of KDD systems. End user's perceptions of novelty, utility, and understandability ultimately determine the acceptance of data mining. Figure 6-4 illustrates how these four fields in combination form the field of KDD. People have been learning representations of the environment for a long time and have been using these learned models to guide decision-making. Psychological investigation has revealed factors that simplify the learning, understanding, and communication of category information.



Knowledge Discovery and Data Mining is looking for diamonds in the mass of data



Originally, data mining was applied to laboratory research, clinical trials, actuarial studies and risk analysis. It is now applied from the infinitely small microscopic area (omics) to the infinitely large macroscopic area (astrophysics), from the most general (customer relationship management, CRM) to the most specialized (assistance to surgeries in operations), from the most open (e-commerce) to the most secret (fraud detection in mobile telephony and bank card applications), from the most practical (quality control, production management) to the most theoretical (human sciences, biology, medicine and pharmacology), and from the most basic (life science) to the most entertaining (audience prediction for television).

The most relevant fields are those where large volumes of data ("big data") have to be analysed, mostly with the aim of (rapid) decision making, which is in biomedicine still the key topic (Tufféry, 2011).

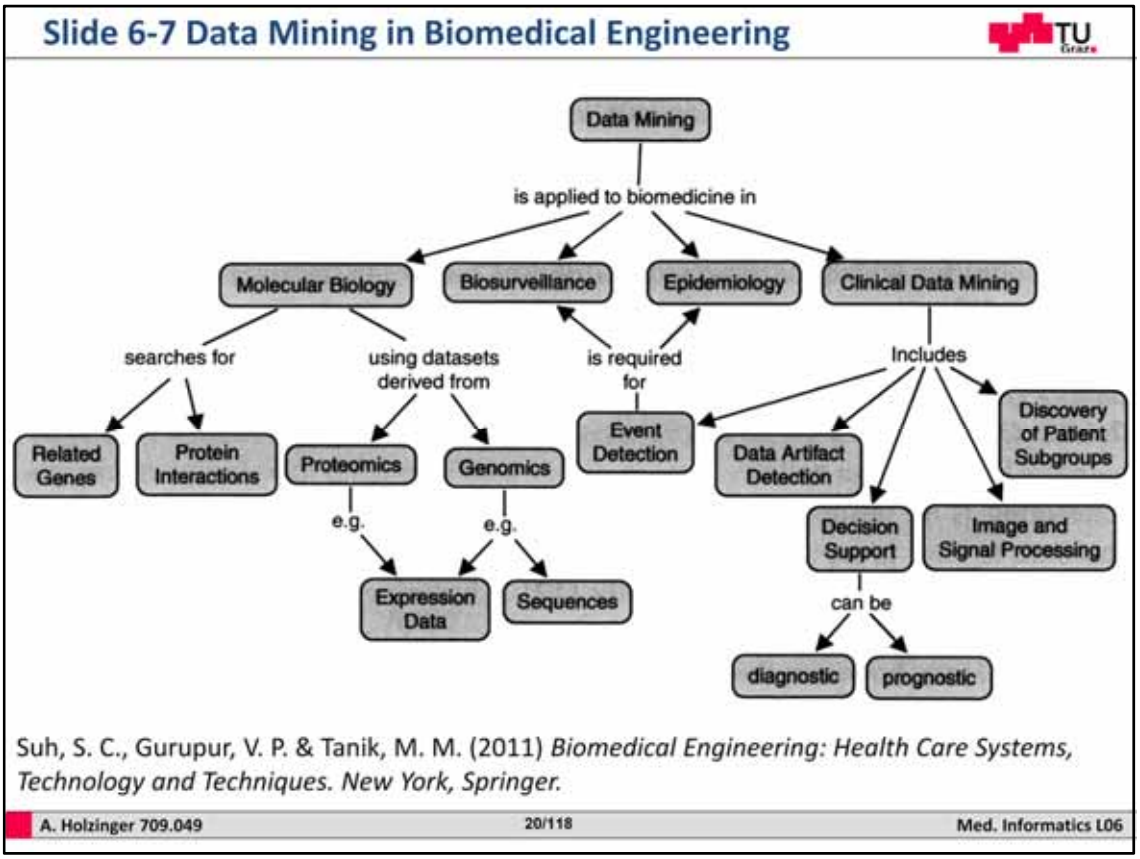
Data mining is defined as the set of methods and techniques for exploring and analysing (big) data sets, in an automatic or semi-automatic way, in order to find among these data certain unknown or hidden rules, associations or tendencies.

Data mining is the art of extracting information from data to gain insight into the data (sensemaking, knowledge, wisdom) – as a core element of the whole knowledge discovery process.

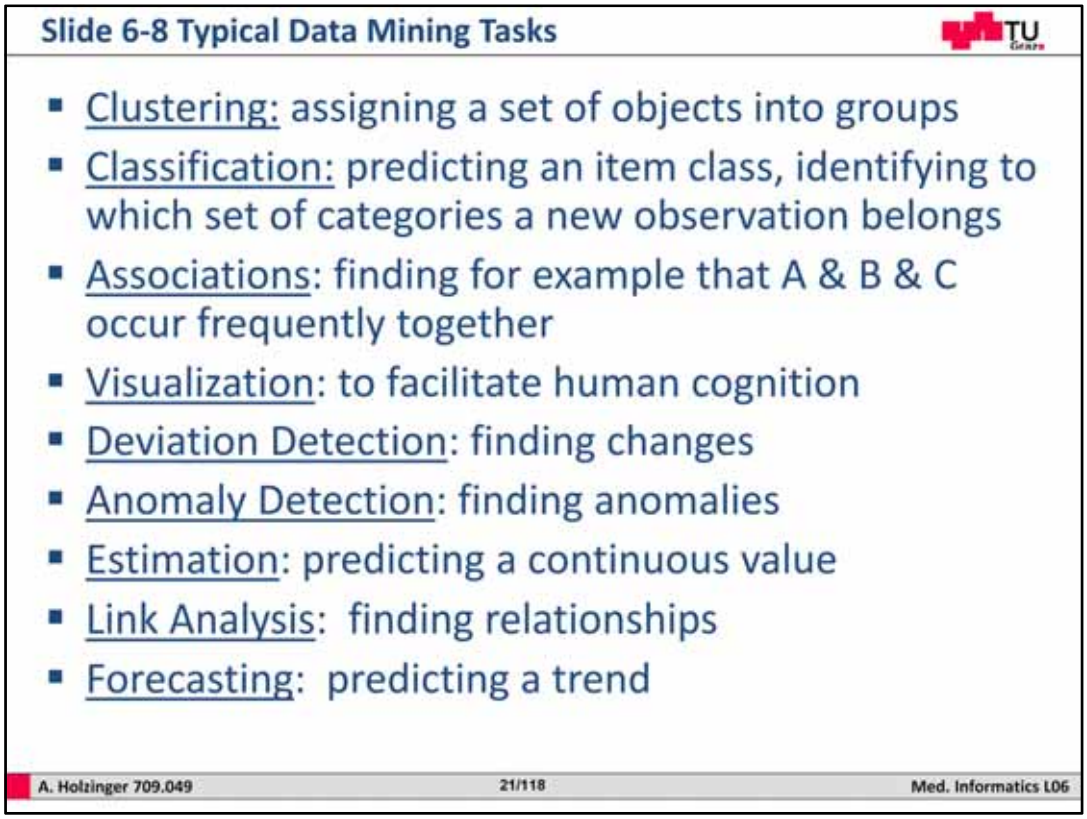
Data mining is therefore both descriptive and predictive:

a) Descriptive (exploratory) techniques discover information that is present but hidden in the mass of data (as in the case of automatic clustering of individuals and searches for associations between products or medicines), whilst

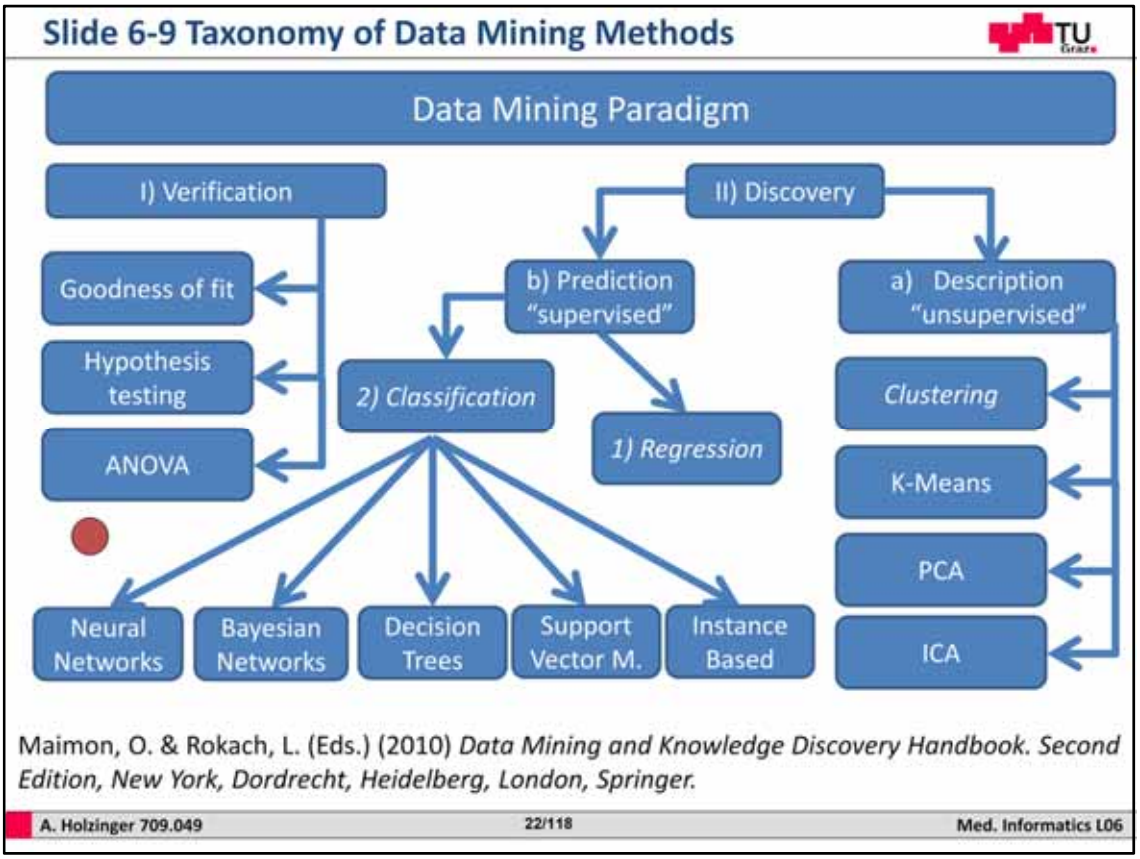
b) Predictive (or explanatory) techniques are designed to extrapolate new information based on present information. In the concept map in the next Slide we can see the main applications in the field of biomedical informatics and engineering.



In this slide we see a concept map of typical applications of data mining in biomedicine (Suh, Gurupur & Tanik, 2011): In Biomedicine data mining is applied to four main fields: molecular biology, biosurveillance, epidemiology and in the clinical field. In molecular biology researchers search for related genes and protein interactions, and they are using data sets derived from proteomics and genomics, e.g. from gene expression data or sequencing (e.g. High throughput sequencing). On the clinical side clinical data mining aims for discovery of patient subgroups, image and signal processing, data artifact detection and event detection (for biosurveillance and epidemiology) and decision support which can be diagnostic or prognostic.



Data mining tasks include:
Clustering (assigning a set of objects into groups); classification (predicting an item class, i.e. identifying which set of categories a new observation belongs);
Associations (finding that A&B&C occur frequently together);
Visualization (to facilitate human cognition);
Deviation detection (e.g. finding changes out of a given limit);
Anomaly detection (to find anomalies, irregularities or inconsistencies);
Estimation (e.g. predicting a continuous value);
Link analysis (finding relationships);
Forecasting (e.g. predicting a trend)



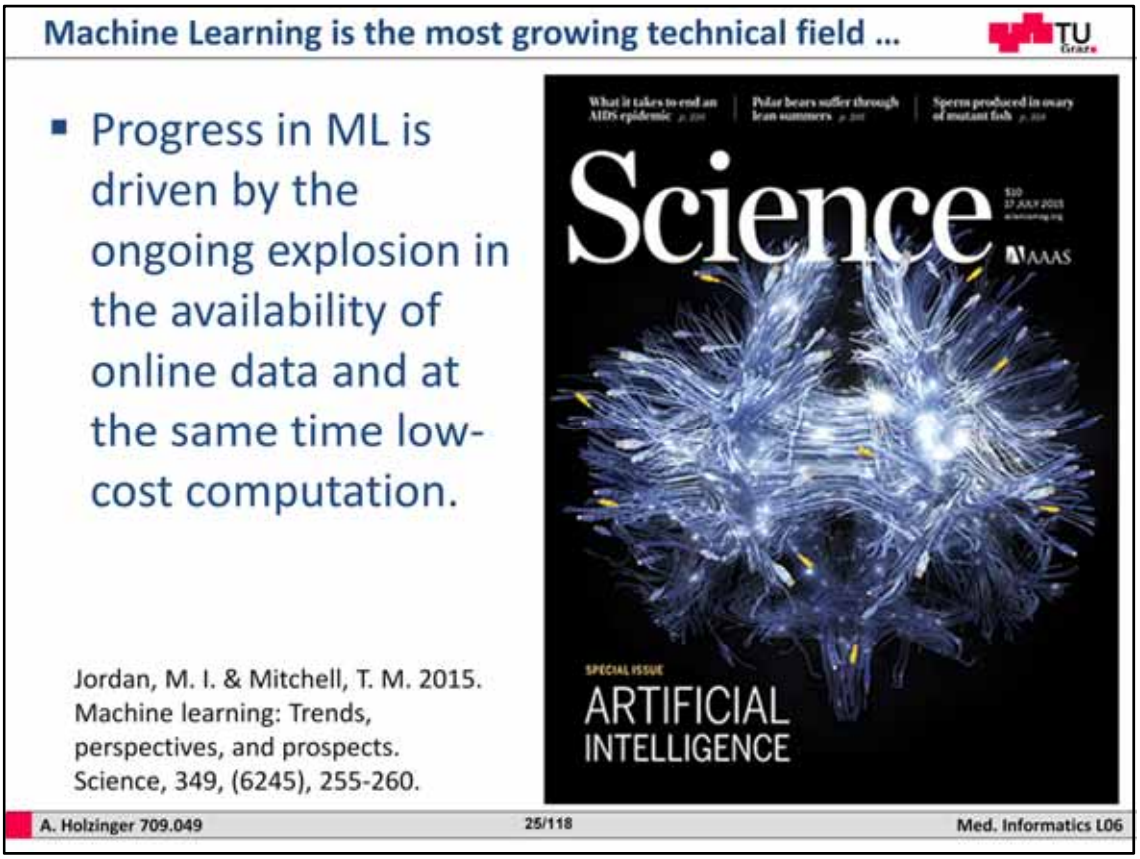
There is a large variety of methods used for very different purposes (Maimon & Rokach, 2010). This slide provides a rough overview on a possible taxonomy:
I) Verification Methods deal with the evaluation of a hypothesis proposed by an external source, e.g. a human expert. Methods include traditional statistics (e.g. Goodness of fit test), classic hypothesis testing (e.g. t-test) and most of all the Analysis of Variance (ANOVA).
II) Discovery Methods are for automatic identification of patterns in the data, mostly based on inductive learning, where a model is constructed (explicitly or implicitly), by generalization from a number of training samples. The assumption is to use the trained model and apply it for future (unforeseen) examples. We have to distinguish between two different types of such discovery methods:
a) Description methods aim to understand the process on how the underlying data are related to their parts. Typical methods include: Clustering (aka unsupervised learning), Summarization, Linguistic Summary, and Visualization. The comparison in Slide →6-10 and an example from gene expression in →Slide 6-11 shall make the differences of unsupervised vs. supervised learning clear.
b) Prediction methods (aka supervised learning) aim for automatic building of a behavioral model, which captures new and previously unseen samples and is able to predict values of one or more variables related to this sample. Prediction methods can be further categorized into:
1) Classification e.g. by application of Neural Networks, Bayesian Networks, Decision Trees, Support Vector Machines, and Instance Based methods and
2) Regression, e.g. in the simplest form (linear regression, or multiple regression);
The application of supervised learning to a real-world medical problem is described in Slide →6-12.

We can understand the difference between DM and KDD best, when looking at the classic KDD process chain by (Fayyad, Piatetsky-Shapiro & Smyth, 1996): DM is a subset of KDD. (see next slide)

Machine learning

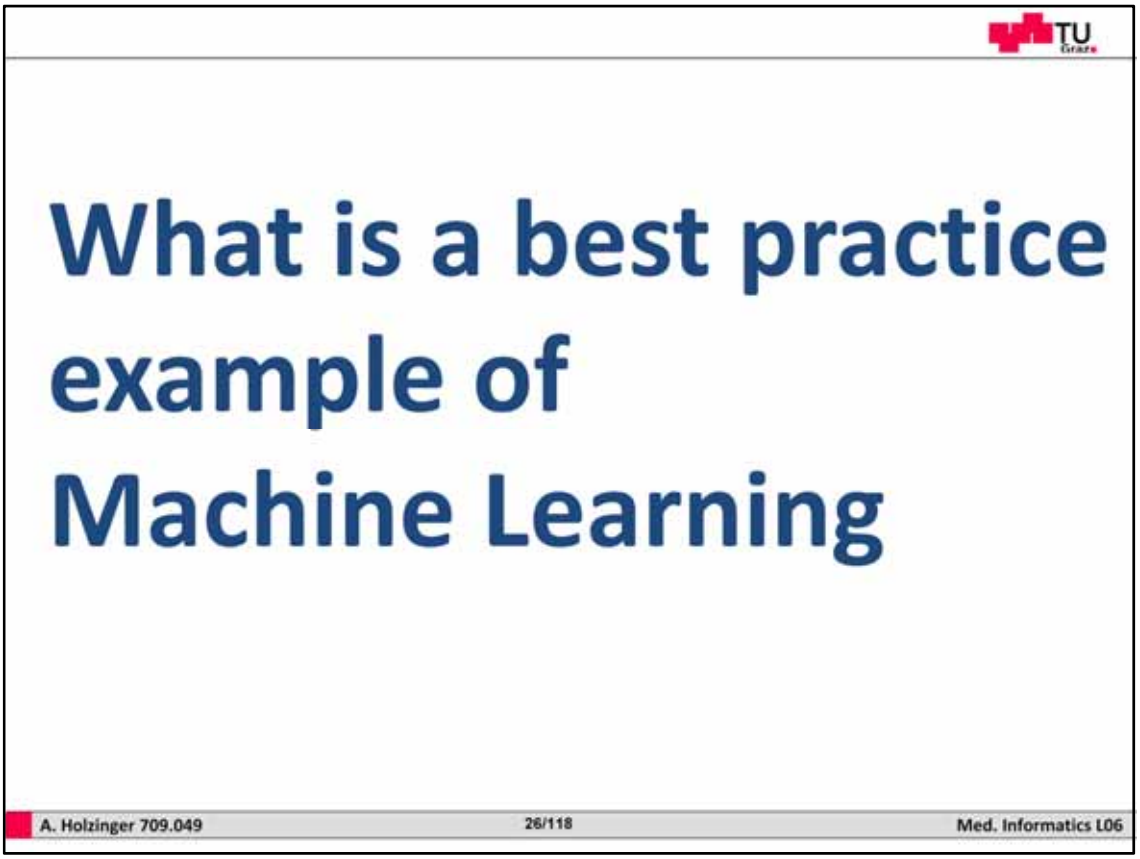
- Machine learning is NOT a well defined field: it refers to a broad range of various **algorithms** within a feature space; hence:
- **Features** are key to machine learning and knowledge discovery!
- Tom Mitchell: A scientific field is best defined by the **central questions it studies.**
- ML seeks to answer the question *"How can we build computer systems that **automatically** improve with experience, and what are the fundamental laws that govern all learning processes?"*

www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf



Machine learning addresses the question of how to build computers that improve automatically through experience. It is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science. Recent progress in machine learning has been driven both by the development of new learning algorithms and theory and by the ongoing explosion in the availability of online data and low-cost computation. The adoption of data-intensive machine-learning methods can be found throughout science, technology and commerce, leading to more evidence-based decision-making across many walks of life, including health care, manufacturing, education, financial modeling, policing, and marketing.
Science 17 July 2015:
Vol. 349 no. 6245 p. 333
DOI: 10.1126/science.349.6245.333-b

Science Podcast: 17 July 2015 Show



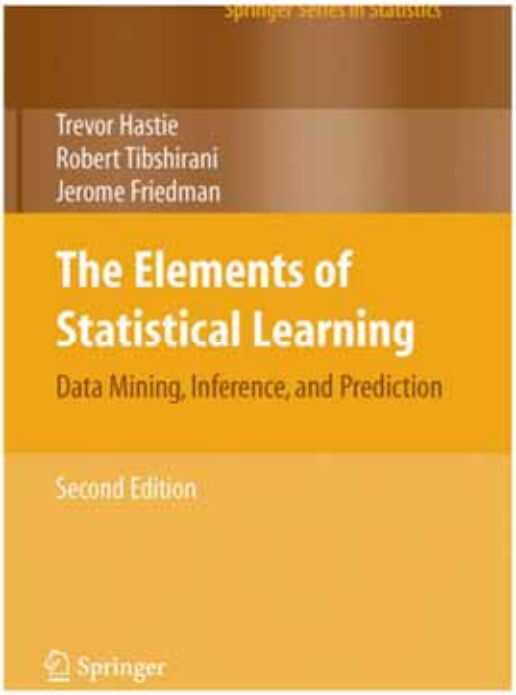
We can understand the difference between DM and KDD best, when looking at the classic KDD process chain by (Fayyad, Piatetsky-Shapiro & Smyth, 1996): DM is a subset of KDD. (see next slide)

http://dl.acm.org/citation.cfm?doid=2830674.2770869
http://dx.doi.org/10.1145/2770869

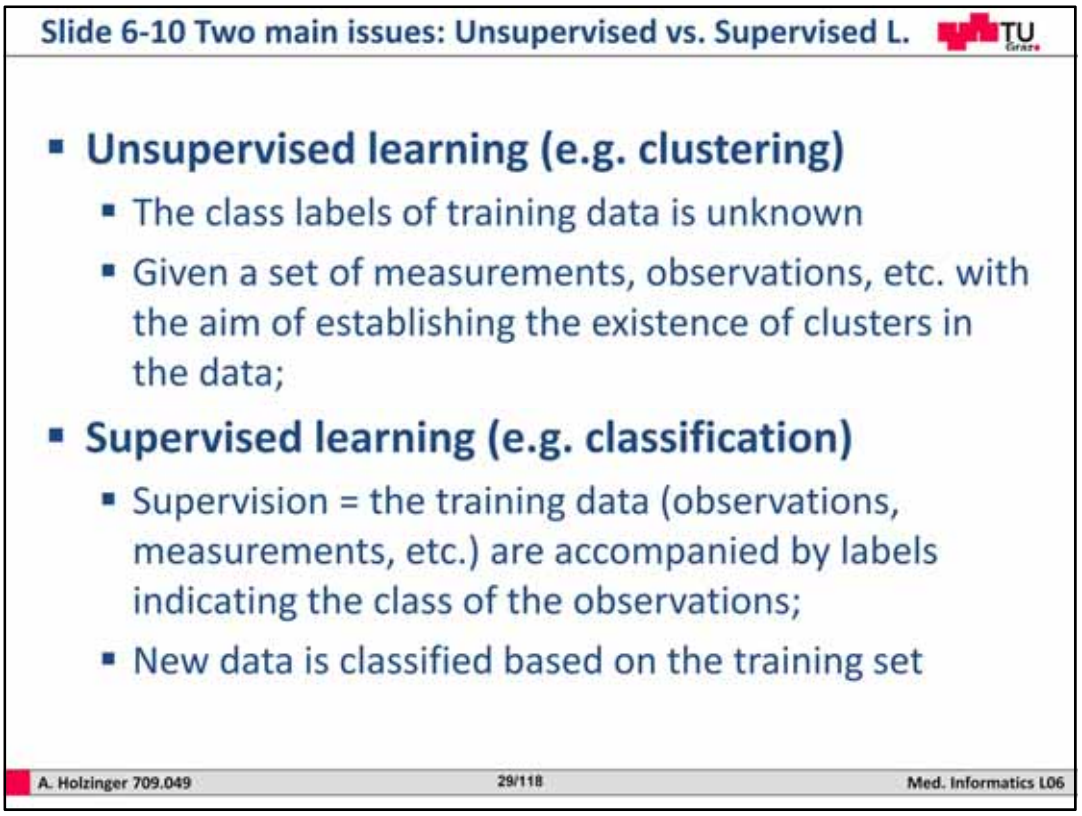
The unknown unknowns are the big challgenges

Machine Learning and Statistics are closely related

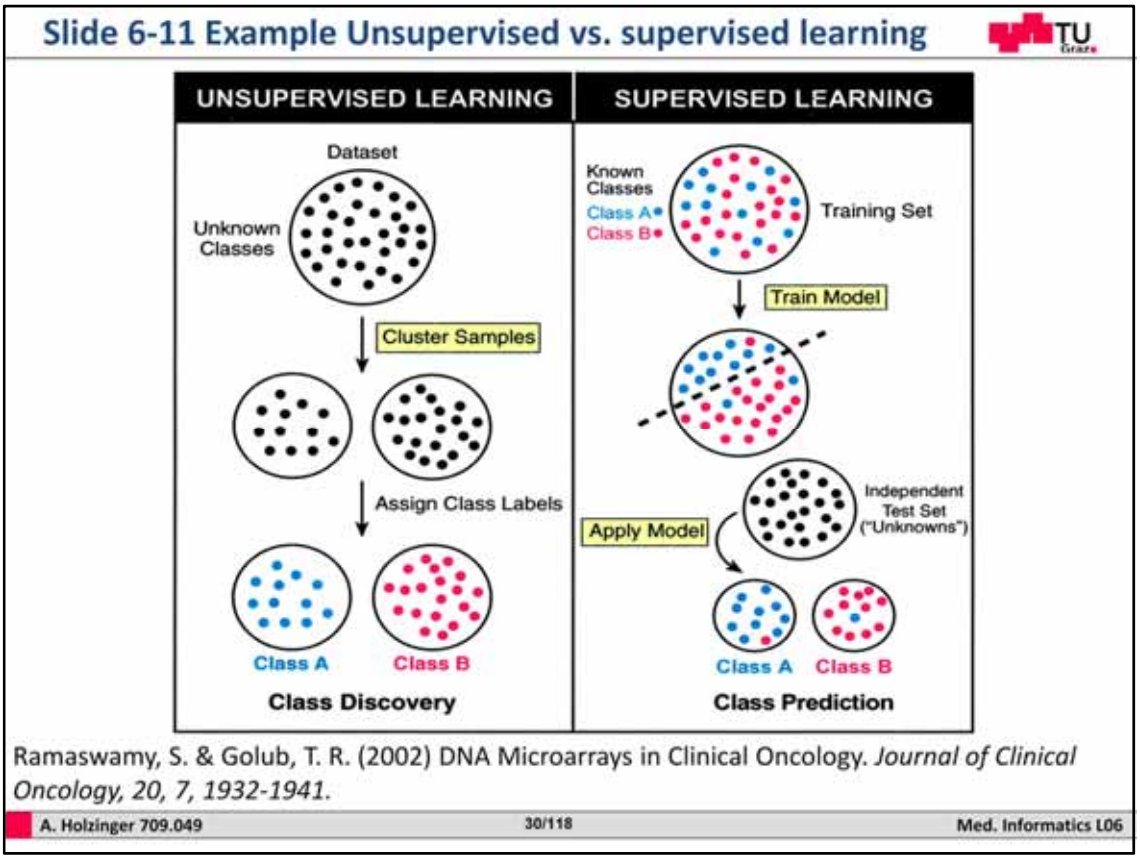
- Machine Learning is the development of algorithms which can **learn from data**
- Machine Learning has a pre-history in **statistical learning,** which is the application of statistical models and the assessment of **uncertainty**



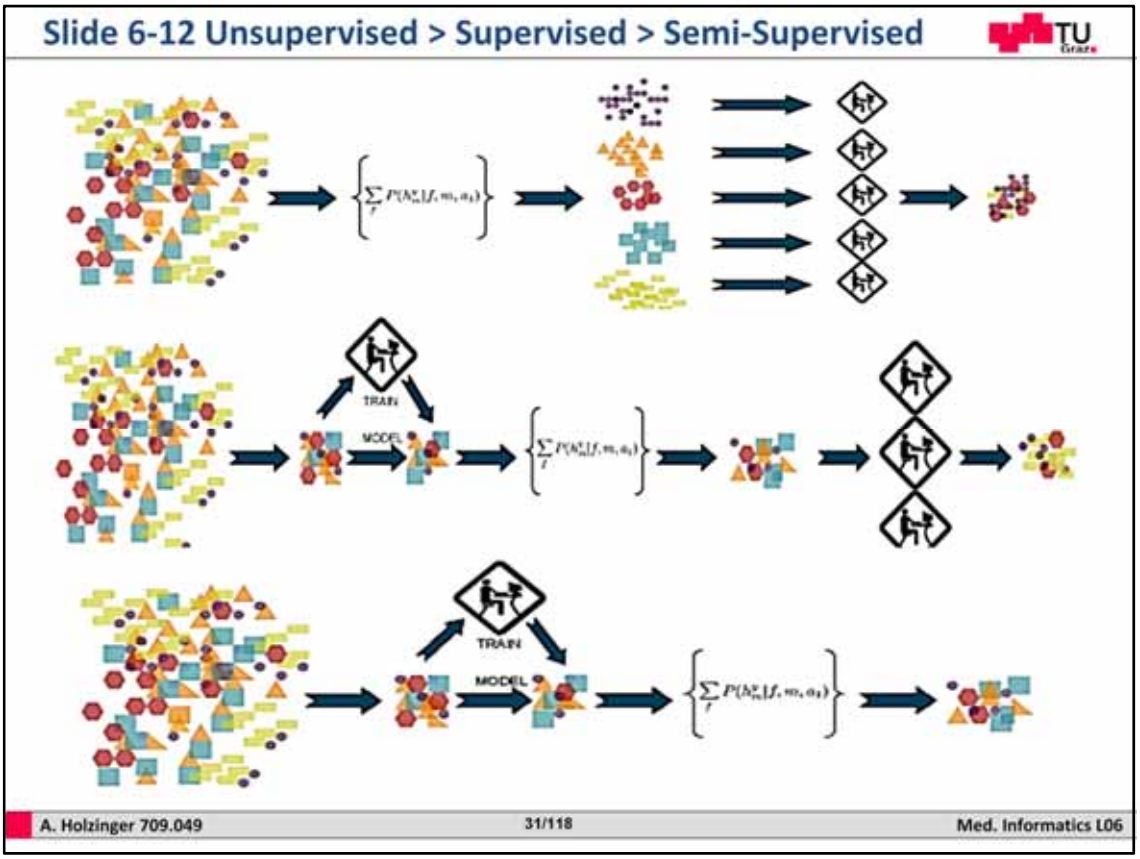
http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html
http://web.stanford.edu/~hastie/sldm.html



Unsupervised learning (e.g. clustering) means that the class labels of training data is unknown, i.e. given a set of measurements, observations, etc. we are aiming to establish the existence of clusters in the data;
Supervised learning (classification)
Supervision = the training data (observations, measurements, etc.) are accompanied by labels indicating the class of the observations;
New data is classified based on the training set.

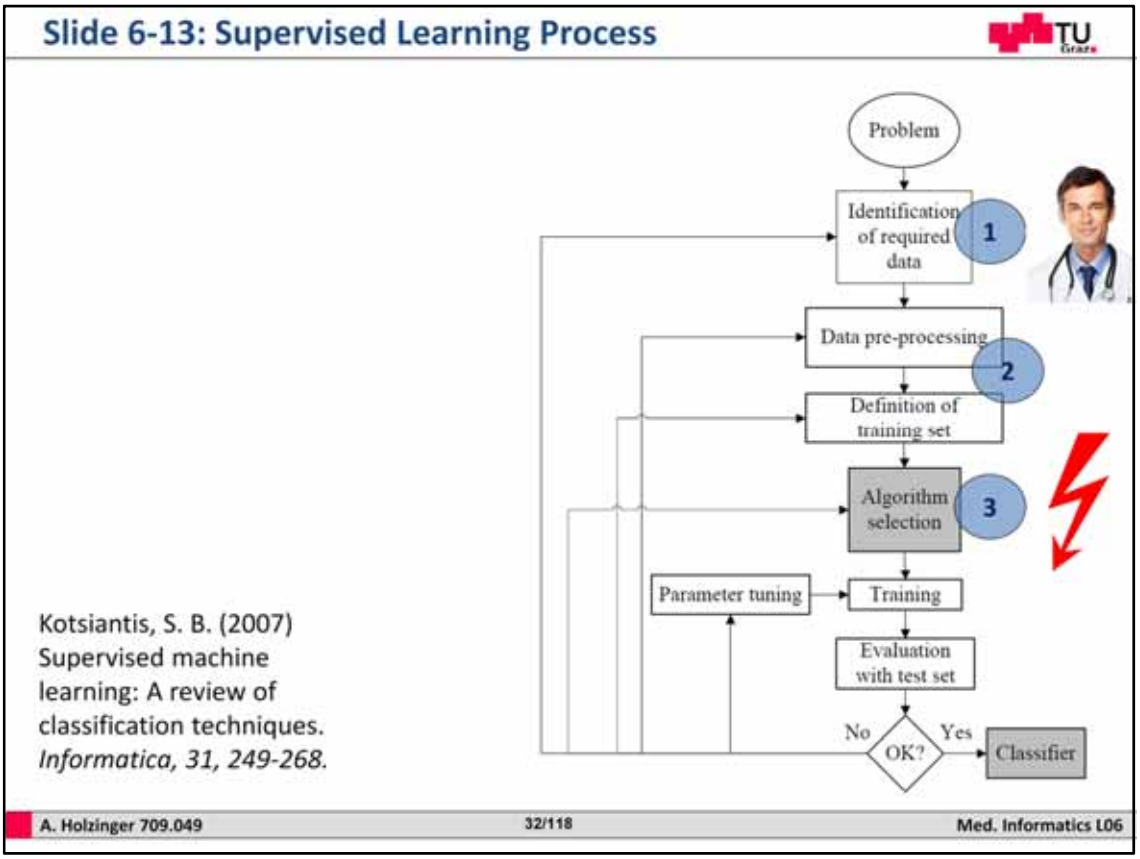


i) Unsupervised learning (clustering): multiple tumor samples are clustered into groups based on overall similarity of their gene expression profiles. This approach is useful for discovering previously unappreciated relationships.
ii) Supervised learning (classification): multiple tumor samples from different known classes are used to train a model capable of classifying unknown samples. This model is then applied to a test set for class label assignment.

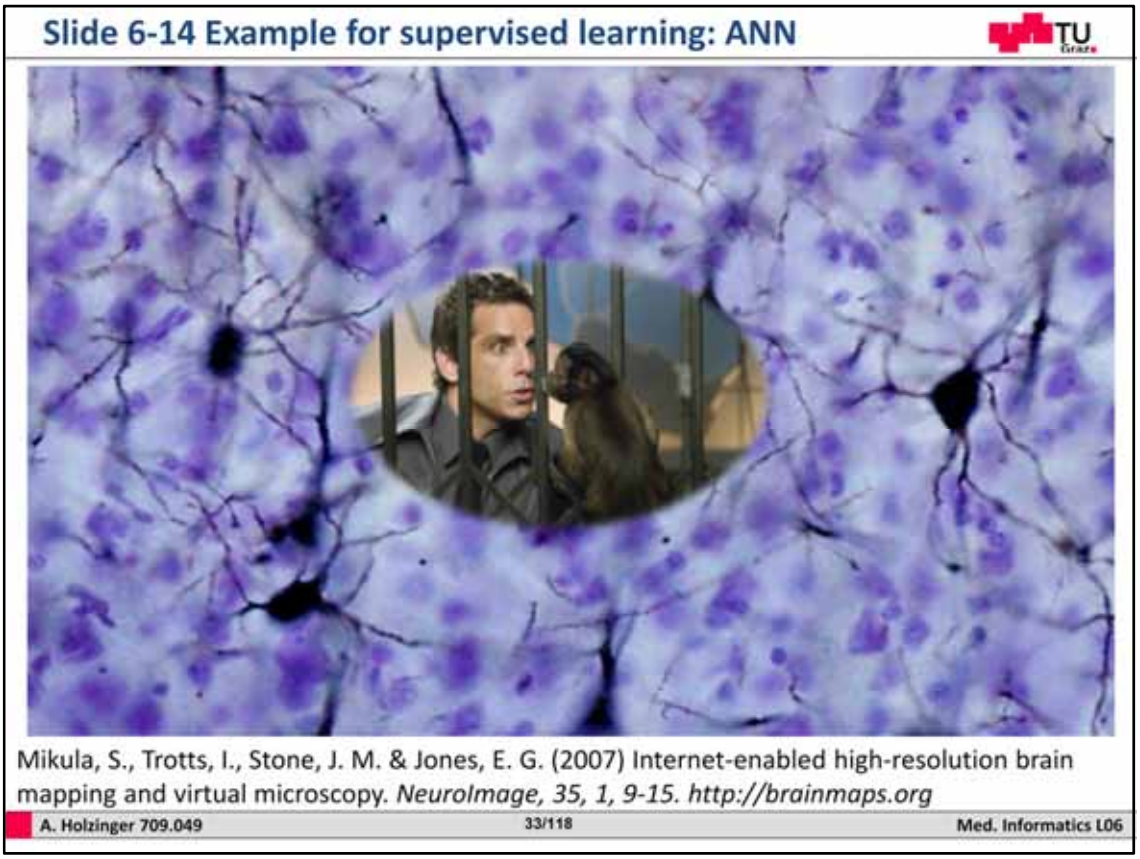


a) Unsupervised Learning: An clustering algorithm (e.g. Latent semantic indexing, k- means, principal component analysis etc.) is applied on the raw data and produces automated clusters, which must then be manually reviewed before the data can be used. Algorithms to unsupervised learning include: k-means, mixture models, hierarchical clustering, hidden Markov models, blind signal separation using feature extraction techniques for dimensionality reduction (e.g., principal component analysis (PCA), independent component analysis (ICA) etc.) (Bandyopadhyay & Saha, 2013).
b) Supervised learning: Sample data sets are manually selected and used as a model for training the algorithm. The product of the trained algorithm must then be manually verified before the data can be used.
c) Semi-supervised learning: This begins in the same manner as supervised learning, but the manual review of the data happens before the algorithm application.

See more at: http://hudsonlegalblog.com/e-discovery/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth.html#sthash.tG9Xh6oe.dpuf



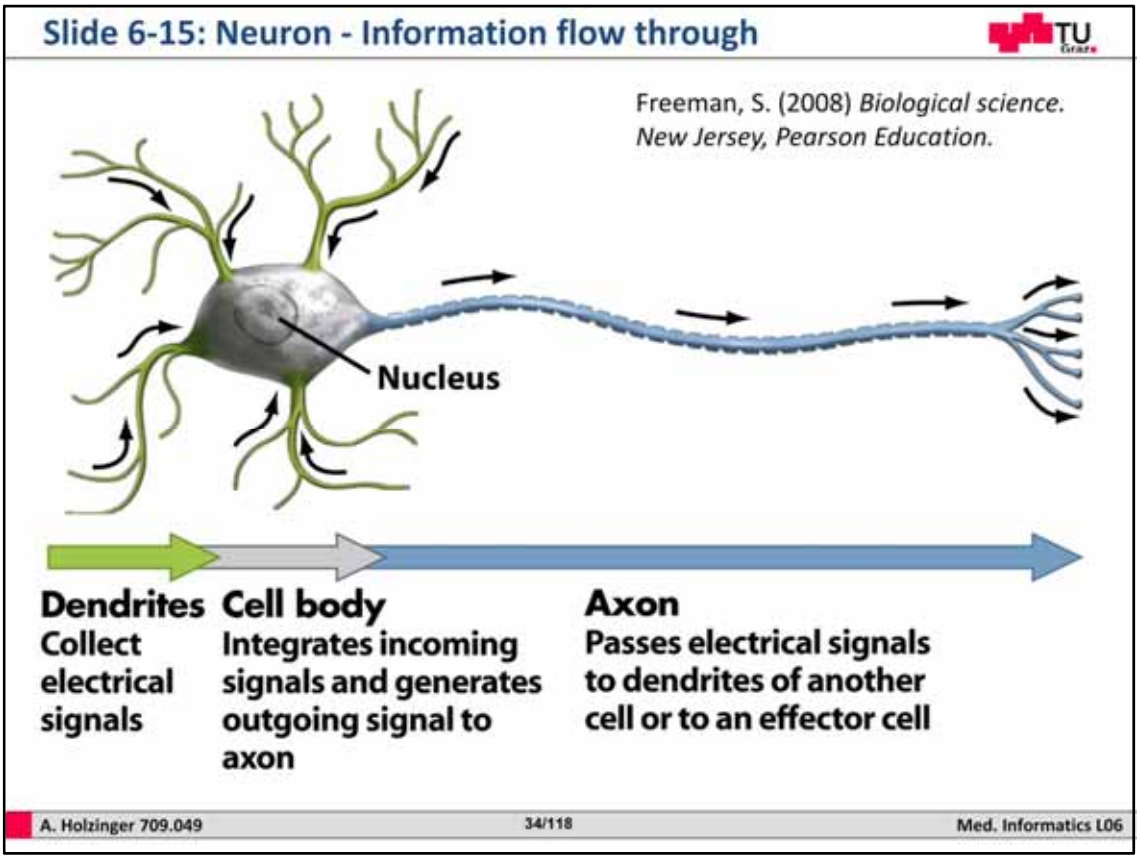
This Slide shows a more detailed view on the supervised learning principle in three steps according to (Kotsiantis, 2007):
Step 1: Identification of required data: the first step is collecting the dataset by help of a medical expert. If no expert is available, then the simplest method is "brute-force," i.e. measuring everything, in the hope that the right (relevant!) features can be isolated. However, a dataset collected by the "brute-force" method is not directly suitable for induction. It contains in most cases noise, artifacts and missing feature values, hence requires much pre-processing effort.
Step 2: Data pre-processing: the second step is the data preparation to handle missing data. Instance selection is not only used to handle noise but to cope with the infeasibility of learning from very large datasets. Instance selection in these datasets is an optimization problem that attempts to maintain the mining quality while minimizing the sample size (Liu, Motoda & Yu, 2004). It reduces data and enables a data mining algorithm to function and work effectively with very large datasets. There is a variety of procedures for sampling instances from a large dataset and feature subset selection is the process of identifying and removing as many irrelevant and redundant features as possible.
Step 3: Algorithm selection: Once preliminary testing is judged to be satisfactory, the classifier (mapping from unlabeled instances to classes) is available for routine use. The classifier's evaluation is most often based on prediction accuracy (the percentage of correct prediction divided by the total number of predictions). There are at least three techniques which are used to calculate a classifier's accuracy. One technique is to split the training set by using two-thirds for training and the other third for estimating performance. In another technique, known as cross-validation, the training set is divided into mutually exclusive and equal-sized subsets and for each subset the classifier is trained on the union of all the other subsets. The average of the error rate of each subset is therefore an estimate of the error rate of the classifier.



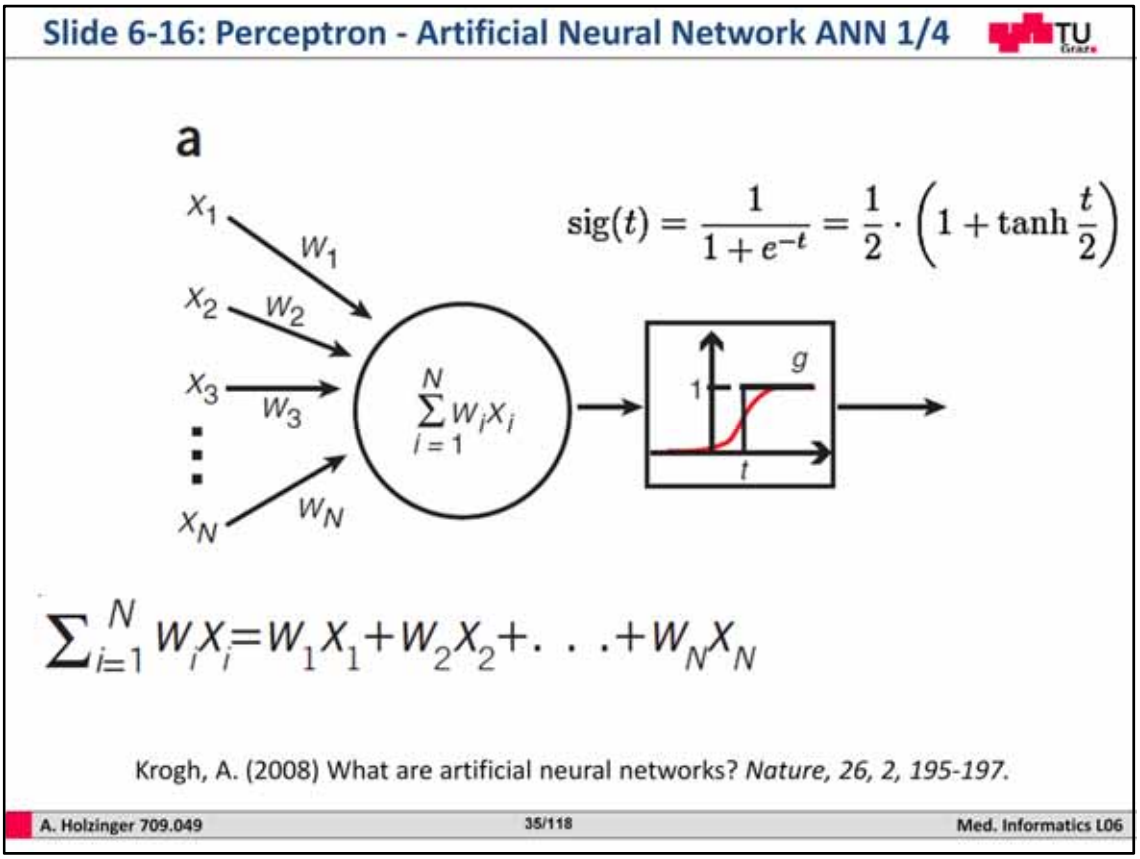
In this slide we see a neural network on the example of Macaca fascicularis (monkey) and we can see the structure of the nerve cells, see brainmap.org for more examples; (Mikula et al., 2007).
Artificial Neural networks (ANN) mimic the human neural network and literally "learn" from examples to find patterns in data or to classify data. The advantage is that it is not necessary to have any specific model in mind when running the analysis. ANN finds interaction effects (e.g. effects from the combination of age and gender) which must be explicitly specified in regression. The disadvantage is that it is harder to interpret the resultant model with its layers of weights and arcane transformations. Neural networks are therefore useful in predicting a target variable when the data are highly non-linear with interactions, but they are not very useful when these relationships in the data need to be explained (Vikram & Upadhayaya, 2011).
Note: The human brain can be described as a highly parallelized biological neural network, an interconnected network of neurons transmitting patterns of electrical signals, where the dendrites receive input signals and, based on those inputs, fire an output signal via an axon (see →Slide 6-15).
Remark: When it comes to tasks other than number crunching, the human brain possesses numerous advantages over a digital (Von-Neumann) computer. For example, humans can quickly recognize a face, even when seen from the side in bad lighting in a room full of other objects. We can easily understand natural speech, even that of an unknown person in a noisy room. Despite years of focused research, computers are far from performing well at this level. Moreover, the brain is also remarkably robust; it does not stop working just because a few cells die. The computations of the brain are done by a highly interconnected network of neurons, which communicate by sending electric pulses through the neural wiring consisting of axons, synapses and dendrites (see →Slide 6-15). McCulloch & Pitts (1943) modeled a neuron as a switch which receives input from other neurons and, depending on the total weighted input, is either activated or remains inactive .

Macaca fascicularis is a primarily arboreal macaque native to Southeast Asia. It is also called the Cynomolgus Monkey, The Crab-eating Macaque, and the Long-tailed Macaque.
http://www.wolfram.com/products/applications/neuralnetworks/

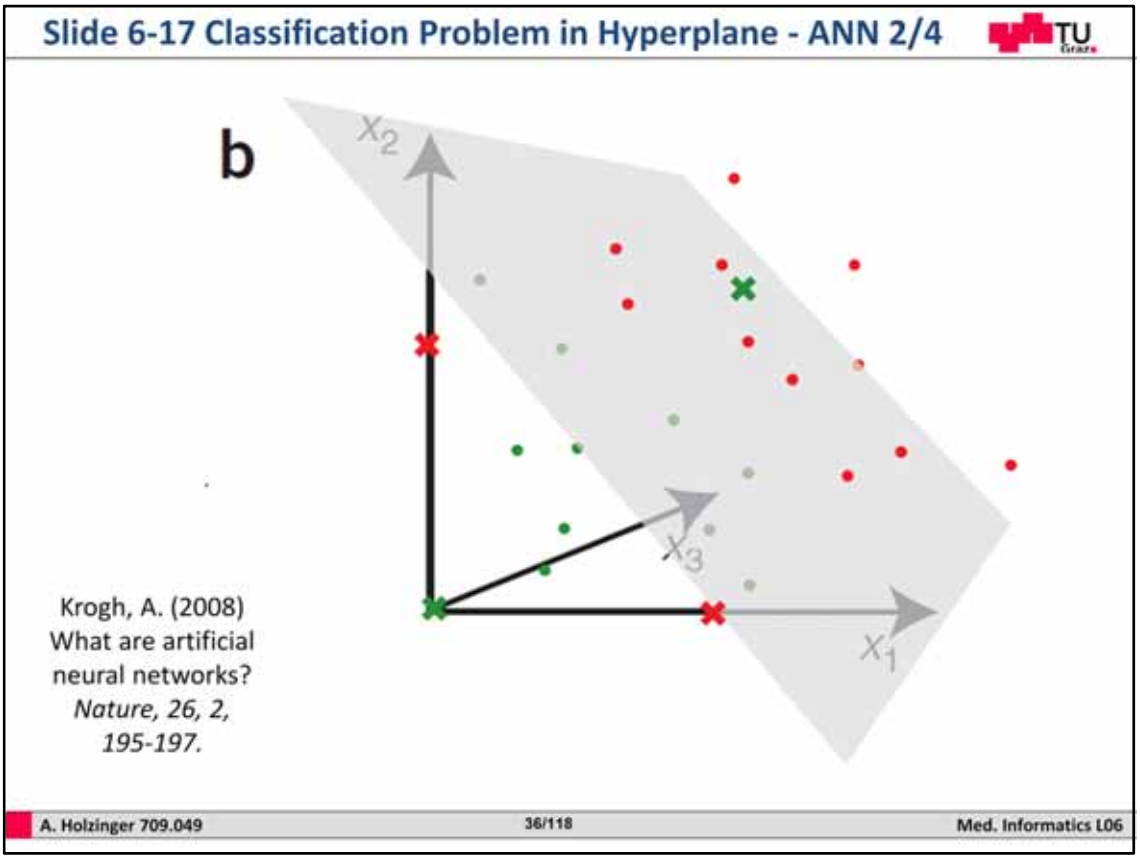
Recommended Reading:
http://natureofcode.com/book/
How can we capture the unpredictable evolutionary and emergent properties of nature in software? How can understanding the mathematical principles behind our physical world help us to create digital worlds? This book focuses on the programming strategies and techniques behind computer simulations of natural systems using Processing.
http://natureofcode.com/book/chapter-10-neural-networks/



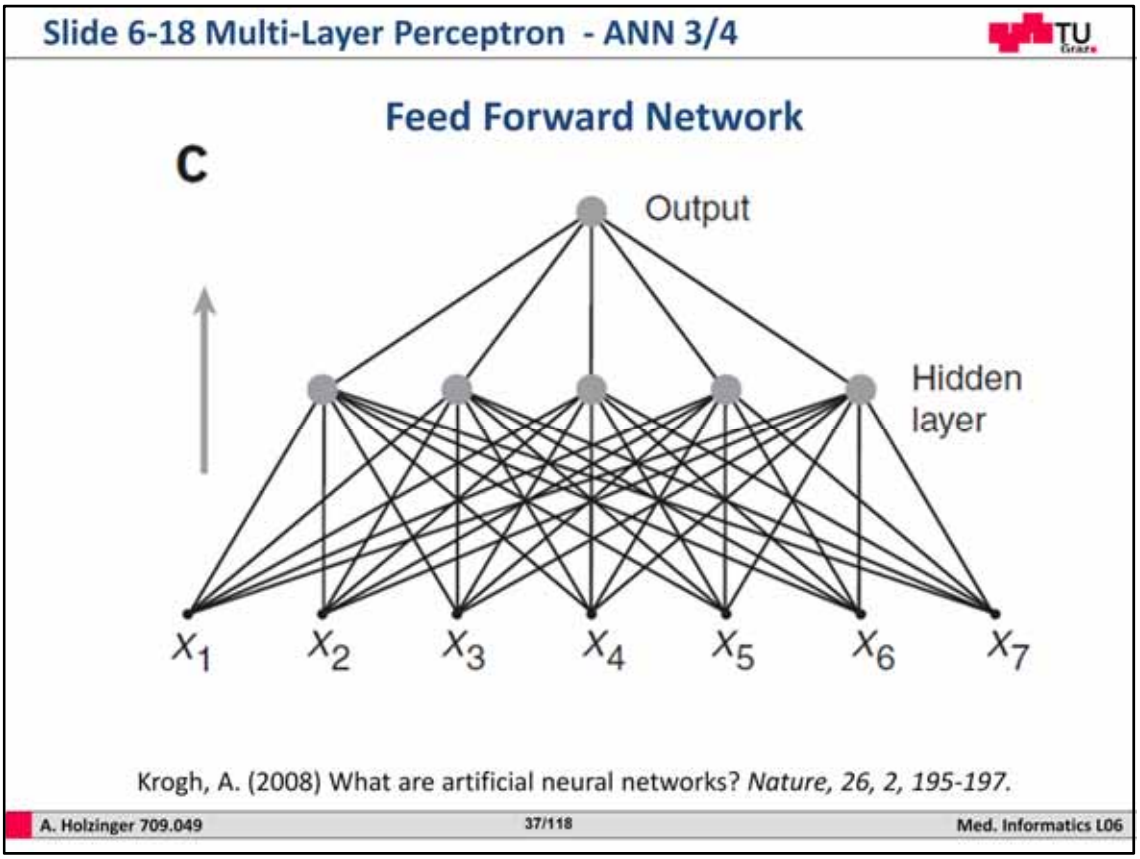
Here we see the information flow within a neuron: the dendrites collect electrical signals, the cell body integrates the incoming signals and generates outgoing signals to the axon, the axon passes electrical signals to dendrites of another cell or to an effector cell (Freeman, 2008).
The weight, by which an input from another cell is multiplied, corresponds to the strength of a synapse (the neural contacts between nerve cells). These weights can be both positive (excitatory) and negative (inhibitory).



In the 1960s, it was shown that networks of such model neurons have properties similar to the brain: they can perform sophisticated pattern recognition, and they can function even if some of the neurons are destroyed. Simple networks of such model neurons are called a perceptron (Rosenblatt, 1958).
However, such models can solve only a very limited class of linearly separable problems. Nonetheless, the error back-propagation method, which can make fairly complex networks of simple neurons learn from examples, showed that these networks could solve problems that were not linearly separable. In this slide we see such a model neuron perceptron: It receives input from a number of other units (or external sources), weighs each input and adds them up. If the total input is above a threshold, the output of the unit is one; otherwise it is zero. Therefore, the output changes from 0 to 1 iff the total weighted sum of inputs is equal to the threshold. The points in the input space satisfying this condition define a so called hyperplane. In 2D such a hyperplane is a line, whereas in 3D it is a normal plane. Data points on one side of the hyperplane are classified as 0 and those on the other side as 1.
So, we have a typical classification problem, which can be solved by a threshold unit if the two classes can be separated by a hyperplane (Krogh, 2008) – as can be seen in →Slide 6-17.

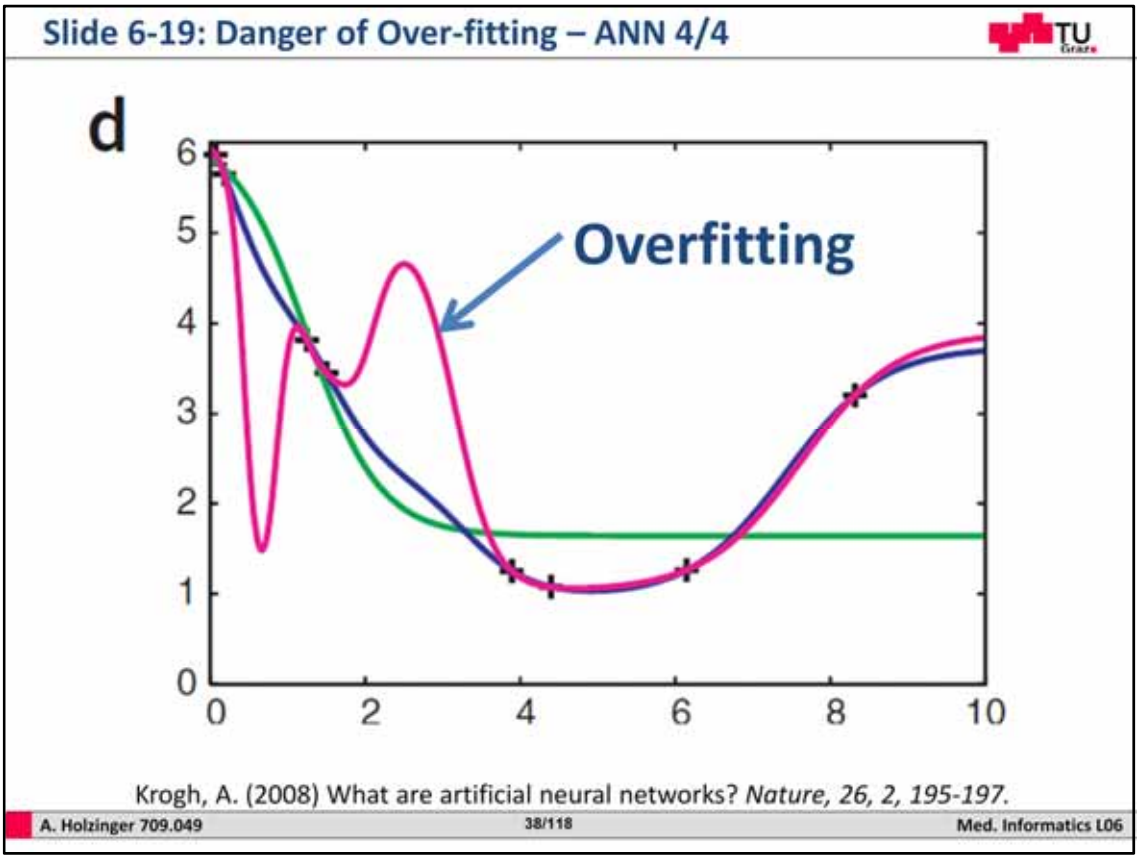


Such problems are called linearly separable and can be seen in this slide on the example of a classification problem in R^3.
If this classification problem is separable, we still need a way to set the weights and the threshold, such that the threshold unit correctly solves the classification problem. This can be done in an iterative manner by presenting examples with known classifications: This process is called learning (or training), which can be implemented by various algorithms. During this learning process, the hyperplane moves around until it finds its correct position in space.
This is nicely illustrated by a set of Java Applets which can be found here: http://lcn.epfl.ch/tutorial/english/index.html
Medical Example: Let us think of two classes of cancer, only one responds to a certain treatment. As there is no simple biomarker to discriminate the two, we can use gene expression measurements of tumor samples to classify them. Assuming to measure gene expression values for 20 different genes in 50 tumors of class 0 (nonresponsive) and 50 of class 1 (responsive). On the basis of these data, we train a threshold unit that takes an array of 20 gene expression values as input and gives 0 or 1 as output for the two classes, respectively. If the data are linearly separable, the threshold unit will classify the training data correctly. But many classification problems are not linearly separable, so we must introduce more hyperplanes; that is, by introducing more than one threshold unit. This is usually done by adding an extra (hidden) layer of threshold units each of which does a partial classification of the input and sends its output to a final layer, which assembles the partial classifications to the final classification, the so-called multi-layer perceptron, see →Slide 6-18.

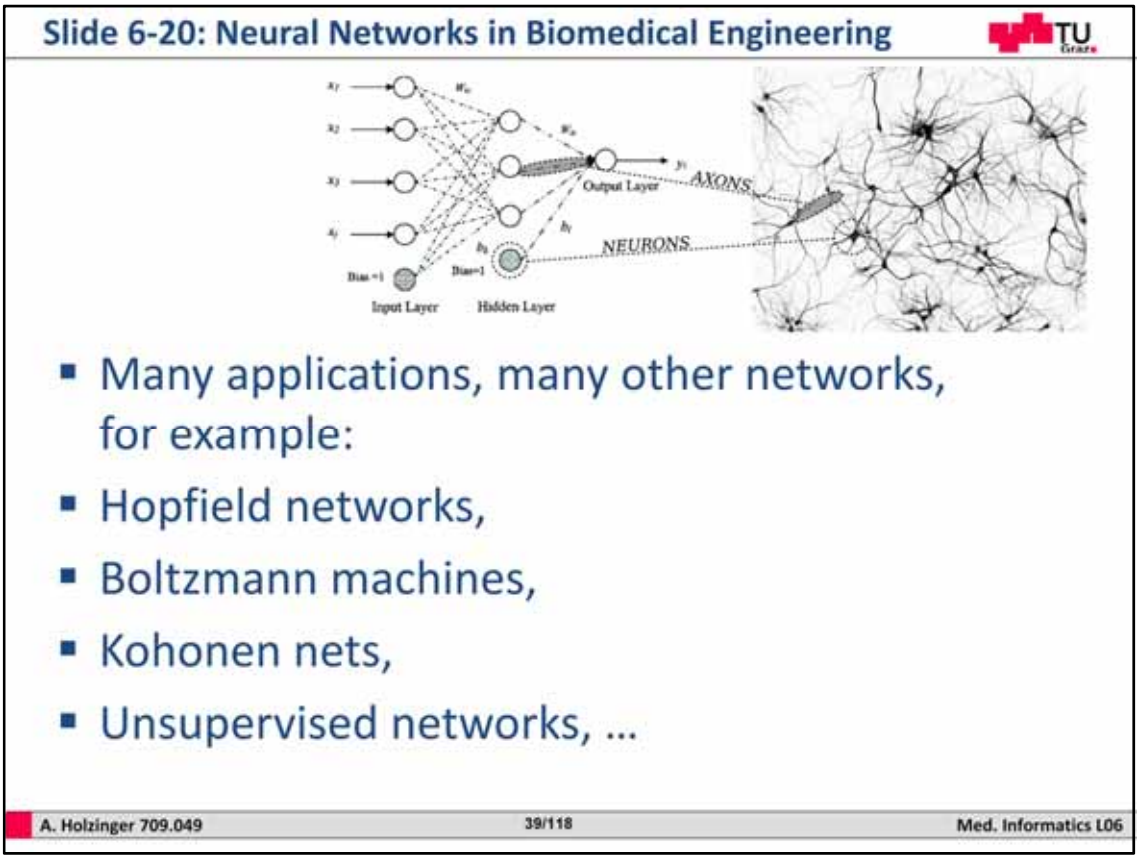


In this slide we see a multi-layer perceptron aka feed-forward network. These can be used for regression problems, which require continuous outputs, as opposed to binary outputs (0 and 1). By replacing the step function with a continuous function, output is a real number. A so-called sigmoid function is used, i.e. a soft version of the threshold function. The sigmoid function can also be used for classification problems by interpreting an output below 0.5 as class 0 and an output above 0.5 as class 1; often it makes sense to interpret the output as the probability of class 1.
In our example we can have a situation where class 1 is characterized by either a highly expressed gene 1 and a silent gene 2 or a silent gene 1 and a highly expressed gene 2; if neither or both of the genes are expressed, it is a class 0 tumor. This corresponds to the logical exclusive or function and it is the canonical example of a nonlinearly separable function. In this case, it would be necessary to use a multi-layer network to classify the tumors (Krogh, 2008).

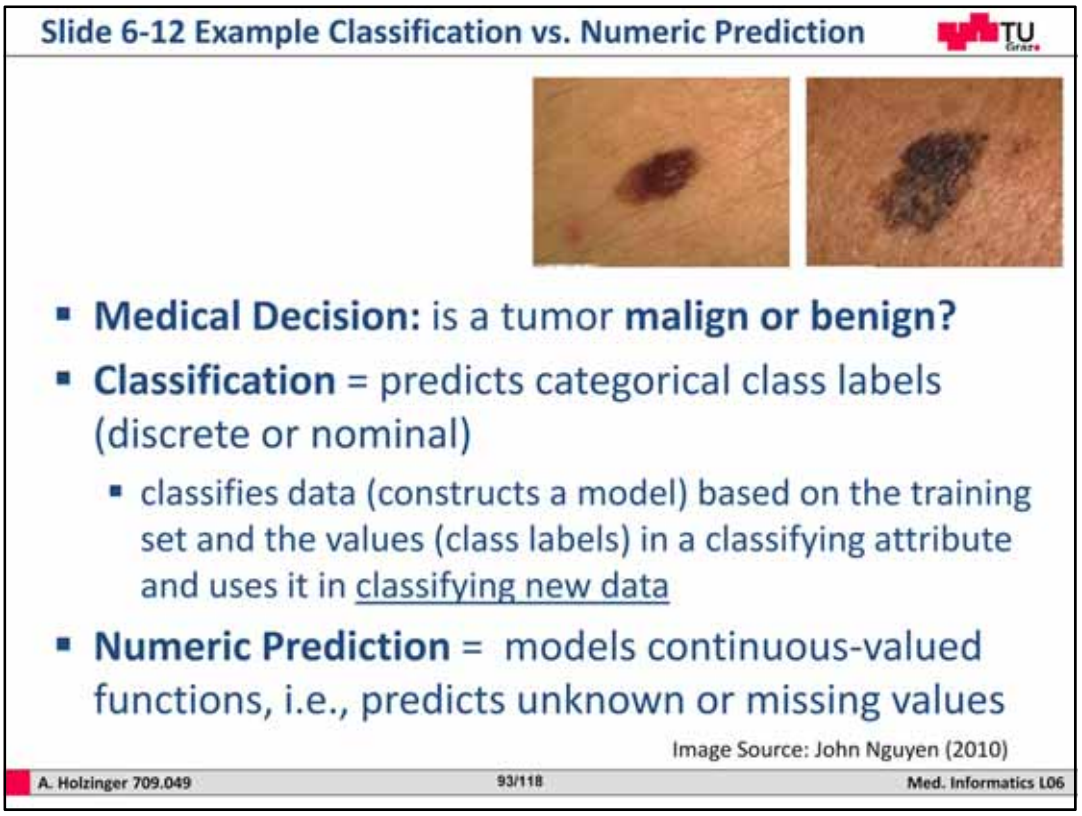
Hinton, G. E., Osindero, S. & Teh, Y. W. 2006. A fast learning algorithm for deep belief nets. Neural Computation, 18, (7), 1527-1554.



Over-fitting occurs when the network has too many parameters to be learned from the number of examples available, i.e. when a few points are fitted with a function with too many free parameters. In the slide we clearly see that the original 8 data points shown by the"+ symbols" are located on a parabola (apart from some noise). They were used to train three different neural networks. The networks all take an x value as input (one input) and are trained with a y value as desired output. As expected, a network with just one hidden unit (green) does not do a very good job. A network with 10 hidden units (blue) approximates the underlying function remarkably well. The last network with 20 hidden units (purple) over-fits the data: the training data points are learned perfectly, but for some of the intermediate regions the network is overly "creative" (Krogh, 2008).

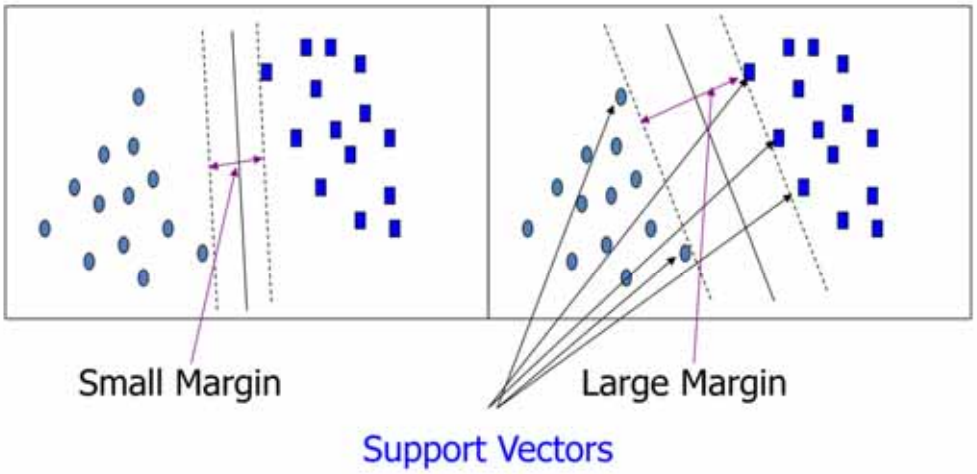


Neural networks have been applied to many interesting problems in medicine and biology and there are many other types of neural networks, for example:
Hopfield networks,
Boltzmann machines,
Kohonen nets, and
Unsupervised networks.
Hopfield networks are recurrent ANNs serving as content-addressable memory systems with binary threshold nodes (Hopfield, 1987), (Hopfield, 1982). Recurrent networks mean that the graph may contain (directed) cycles. Actually, Hopfield is the simplest form, which originated as physical models to describe magnetism.
Boltzmann machines are the stochastic, generative counterpart of Hopfield nets. They were one of the first examples of a neural network capable of learning internal representations, and are able to represent and to solve difficult combinatory problems (Barra et al., 2012), (Leitner et al., 2006).
Kohonen networks are self-organizing maps and a computational method for the visualization and analysis of high-dimensional data. The basic idea of (Kohonen, 1982) was, that topologically correct maps of structured distributions of signals can be formed in a one or two-dimensional array of processing units which did not have this structure initially. This principle is a generalization of the formation of direct topographic projections between two laminar structures known as retinotectal mapping.



Visual comparison of a benign lesion, dysplastic nevus, (left) and a malignant, lentigo maligna, melanoma (right). Both look very similar and can often be confused for one another.
http://bme240.eng.uci.edu/students/10s/nguyenjq/index.htm

Support Vectors



Han, J., Kamber, M. & Pei, J. (2011) Data Mining: Concepts and Techniques (Third Edition). The Morgan Kaufmann Series in Data Management Systems. San Francisco, Morgan Kaufmann Publishers.