Keywords of the 6th Lecture

- 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

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

Advance Organizer (2/2)

- Deep Learning = class of machine learning algorithms using layers of non-linear processing units for feature extraction (remember: features are key for learning and understanding); learning representations from data.
- Knowledge Extraction = is the creation of knowledge from structured (relational databases, XML) and unstructured (text, documents, images) sources.
- Multimedia = several data of different modalities are processed at the same time, i.e. encompassing audio data (sound, speech), image data (b/w and colour images), video data (time-aligned sequences of images), electronic ink (sequences of time aligned 2D and 3D co-ordinates of a stylus, pen, data gloves etc.)
- Principal Component Analysis (PCA) = statistical technique for finding patterns in high-dimensional data.
- Supervised learning = inferring a function from supervised training data on the basis of training data which consist of a set of training examples, the input objects (typically vectors) and a desired output value (also called the supervisory signal).
- Supervised learning algorithm = analyses the training data and produces an inferred function, called a classifier (if the output is discrete) or a regression function (if the output is continuous); the algorithm generalizes from the training data to unseen situations.
- Support vector machine (SVM) = concept for a set of related supervised learning methods to analyze data and recognize patterns, used for classification and regression analysis.
- Unsupervised learning = establishes clusters in data, where the class labels of training data is unknown.

Glossary

- ANN = artificial neural network
- ANN = Artificial Neural Network
- ANOVA = Analysis of Variance
- AUC = area under the curve
- CDT = Clinical Decision Tree
- DM = Data Mining
- KDD = Knowledge Discovery from Data(bases)
- MDM = Multimedia Data Mining
- MELD = model for end-stage liver disease
- MM = Multimedia
- NLP = Natural Language Processing
- ROC = receiver-operating characteristic
- SVM = Support Vector Machine
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;

Slide 6-1: Key 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 sensor nets, Multimedia, etc.);
  - d) Graph Data (e.g. Protein Network data, etc.);
  - e) Data restrictions (accessibility, privacy, safety, security, legal restrictions, fair use, etc.);
- 3) Context - Data Mining in a particular context
- 4) Interpretability
- 5) Computational Resources
- 6) Benchmarking against Gold-Standards
- 8) Embedded data mining

What is Knowledge?

World

Knowledge := a set of expectations

Adaptive Agent in a physical world

Implicit vs. Explicit Knowledge


What is the difference between Knowledge Discovery and Data Mining?
Slide 6-3: The classic differentiation between DM and KDD

- KDD = Knowledge Discovery and Data Mining
- DM = Data Mining


http://hci-kdd.org/

Slide 6-4 Interactive Knowledge Discovery and Data Mining

Slide 6-5 The Knowledge Discovery Process Chain

Privacy, Data Protection, Safety and Security

Data Mining
Data mining is the set of methods and techniques for exploring and analyzing (big) data sets; in an automatic or semi-automatic way, in order to find certain unknown or hidden rules, associations or tendencies; relevant essentials of the useful information while reducing the quantity of data; descriptive (or exploratory) techniques are designed to bring out information that is present but buried in a mass of data; predictive (or explanatory) techniques are designed to extrapolate new information based on the present information.

Data Mining, Knowledge Discovery, 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?”
Progress in ML is driven by the ongoing explosion in the availability of online data and at the same time low-cost computation.

What is a best practice example of Machine Learning

Machine Learning and Statistics are closely related

Unsupervised learning (e.g. clustering)
- The class labels of training data is unknown
- Given a set of measurements, observations, etc. with the aim of establishing the existence of clusters in the data;

Supervised learning (e.g. 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

Example: DNA Microarrays in Clinical Oncology

Unsupervised > Supervised > Semi-Supervised

Supervised Learning Process


Example for supervised learning: ANN


Neuron - Information flow through


Perceptron - Artificial Neural Network ANN


Classification Problem in Hyperplane - ANN


A. Holzinger LV 706.046 18.11.2015 11/16/2015

WS 2015 Medizinische Informatik 7


A. Holzinger LV 706.046 18.11.2015 11/16/2015

Biomedical Examples

Many applications, many other networks, for example:
- Hopfield networks
- Boltzmann machines
- Kohonen nets
- Unsupervised networks, ...

A. Holzinger LV 706.046 18.11.2015 11/16/2015


A. Holzinger LV 706.046 18.11.2015 11/16/2015


A. Holzinger LV 706.046 18.11.2015 11/16/2015
ANN application Example: Liver transplantation


Diagnostic accuracy of the ANN


Another Clinical Case Example


Important in Clinical practice -> prognosis!


Model-based Clinical Decision Making Strategy


Note: Probable Information p (x)!
Slide 6-28: Bayesian Network (BN) - Definition

- is a probabilistic model, consisting of two parts:
- 1) a dependency structure and
- 2) local probability models.

\[ p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | Pa(x_i)) \]

Where \( Pa(x_i) \) are the parents of \( x_i \).

BN inherently model the uncertainty in the data. They are a successful marriage between probability theory and graph theory, allowing to model a multidimensional probability distribution in a sparse way by searching independence relations in the data. Furthermore this model allows different strategies to integrate two data sources.


Slide 6-30 Breast cancer – big picture – state of 1999


Slide 6-31: 10 years later: Integration of microarray data

- Integrating microarray data from multiple studies to increase sample size;
- = approach to the development of more robust prognostic tests


Slide 6-32 Example: BN with four binary variables

Gene 2
Gene1 on 0.8
Gene1 off 0.2

Gene2 on 0.3 0.6
Gene1 on 0.3 0.6
Prognosis on 0.7 0.4
Prognosis off 0.5


Slide 6-33 Concept Markov-Blanket

First the structure is learned using a search strategy.
- Since the number of possible structures increases super exponentially with the number of variables,
- the well-known greedy search algorithm K2 can be used in combination with the Bayesian Dirichlet (BD) scoring metric:
  \[ p(S | D) \propto p(S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + N_{ijk})} \]
  \[ N_{ijk} \] — number of cases in the data set \( D \) having variable \( i \) in state \( k \) associated with the \( j \)-th instantiation of its parents in current structure \( S \).
  \( n \) is the total number of variables.

Next, \( N_{ij} \) is calculated by summing over all states of a variable:
  \[ N_{ij} = \sum_{k=1}^{r_i} N_{ijk} \cdot N'_{ij} \] and \( N'_{ij} \) have similar meanings but refer to prior knowledge for the parameters.
- When no knowledge is available they are estimated using \( N_{ijk} = N/(r_i q_i) \)
  with \( N \) the equivalent sample size, \( r_i \) the number of states of variable \( i \) and \( q_i \) the number of instantiations of the parents of variable \( i \).
- \( \Gamma(\cdot) \) corresponds to the gamma distribution.
- Finally \( p(S) \) is the prior probability of the structure.
  \[ p(S) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} p(t_{ij} \rightarrow x_i) \prod_{k=1}^{r_i} p(m_{ijk}) \]
  with \( p_t \) the number of parents of variable \( x_i \) and \( r_i \) all the variables that are not a parent of \( x_i \).
- Next, \( p(a \rightarrow b) \) is the probability that there is an edge from \( a \) to \( b \) while \( p(ab) \) is the inverse, i.e. the probability that there is no edge from \( a \) to \( b \).

Estimating the parameters of the local probability models corresponding with the dependency structure.
- CPTs are used to model these local probability models.
- For each variable and instantiation of its parents there exists a CPT that consists of a set of parameters.
- Each set of parameters was given a uniform Dirichlet prior:
  \[ p(\theta_{ij}) = \text{Dir}(\theta_{ij} | N'_{ijk}, \ldots, N'_{ijr_i}) \]

Note: With \( \theta_{ij} \) a parameter set where \( j \) refers to the variable and \( i \) to the \( j \)-th instantiation of the parents in the current structure. \( \theta_{ij} \) contains a probability for every value of the variable \( x_i \) given the current instantiation of the parents. \( Dir \) corresponds to the Dirichlet distribution with \( (N'_{ij1}, \ldots, N'_{ijr_i}) \) as parameters of the Dirichlet distribution. Parameter learning then consists of updating these Dirichlet priors with data. This is straightforward because the multinomial distribution that is used to model the data, and the Dirichlet distribution that models the prior, are conjugate distributions. This results in a Dirichlet posterior over the parameter set:
  \[ p(\theta_{ij} | D, S) = \text{Dir}(\theta_{ij} | N'_{ij1} + N_{ij1}, \ldots, N'_{ijr_i} + N_{ijr_i}) \]
  with \( N_{ijk} \) defined as before.

Uses a nonlinear mapping to transform the original data (input space) into a higher dimension (feature space)
- = classification method for both linear and nonlinear data;
- Within the new dimension, it searches for the linear optimal separating hyperplane (i.e., “decision boundary”);
- By nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated with a hyperplane;
- The SVM finds this hyperplane by using support vectors (these are the “essential” training tuples) and margins (defined by the support vectors);

- SVM
  - Deterministic algorithm
  - Nice generalization properties
  - Hard to learn – learned in batch mode using quadratic programming techniques
  - Using kernels can learn very complex functions
- ANN
  - Non-deterministic algorithm
  - Generalizes well but doesn’t have strong mathematical foundation
  - Can easily be learned in incremental fashion
  - To learn complex functions — use multilayer perceptron (non-trivial)

Are there alternatives to such network approaches?
Clinical use: SVM are more accurate than ANN.

The 10 top data mining algorithms:

- C4.5
- k-means
- Apriori
- EM
- PageRank
- Adaptive Boost
- k-Nearest Neighbor
- CART
- Naive Bayes
- SVM

Selection of Semantic Methods:

- Latent Semantic Analysis (LSA)
- Probabilistic latent semantic analysis (PLSA)
- Latent Dirichlet allocation (LDA)
- Hierarchical Latent Dirichlet Allocation (HDLA)
- Semantic Vector Space Model (SVSM)
- Latent semantic mapping (LSM)
- Principal component analysis (PCA)

Future Outlook:

- ... is a grand challenge in future computing
- ... most of information in the hospital is unstructured and based on natural language
- ... masses of information is not easily processable by humans
- ... legacy approaches have all failed; “searching” not the right approach - Search is a way to gather information - but not to answer questions

A new approach is needed, leveraging content analytics and natural language processing [1]

Slide 6-45: Watson – a Workload Optimized System

- 90 IBM Power 750 servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB per sec on-chip bandwidth
- 10 GB Ethernet network
- 15 Terabytes of memory
- 10 Terabytes of disk, clustered
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts of unstructured information with UIMA & Hadoop open source components
- Linux provides a scalable, open platform, optimized to exploit POWER7 performance
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts of unstructured information with UIMA & Hadoop open source components
- Linux provides a scalable, open platform, optimized to exploit POWER7 performance

Slide 6-46: Example: IBM Watson for Healthcare

Much open work to do ...

"Natural language understanding remains a tremendously difficult challenge, and while Watson demonstrated a powerful approach, we have only scratched the surface," says David Ferrucci.

Future Outlook: Image data mining open research issues

<table>
<thead>
<tr>
<th>Task</th>
<th>Approach/Application</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing</td>
<td>Pixel, region and grid level</td>
<td>Global vs. local level features</td>
</tr>
<tr>
<td>Feature extraction and transformation</td>
<td>Color histograms [35], color moments [35], color sets [35], shape descriptors, texture descriptors, edges</td>
<td>Sensitive to the parameters e.g., number of bins, bin boundaries, selection of feature vectors (Filter etc.)</td>
</tr>
<tr>
<td>Image classification</td>
<td>GMM [10], SVM [41], Bayesion classifier [40] to classify images</td>
<td>Large training data needed for GMM, SVM kernel functions, etc. Unkown number of clusters</td>
</tr>
<tr>
<td>Image clustering</td>
<td>K-means in preproceesing stage to identify patterns [61]</td>
<td></td>
</tr>
<tr>
<td>Image association</td>
<td>A priori based association between structures and functions of human brain [91]</td>
<td>Solvability issue in terms of number of candidate patterns generated</td>
</tr>
</tbody>
</table>

**Future Outlook: Video data mining open research issues**

<table>
<thead>
<tr>
<th>Task</th>
<th>Approach/Application</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing</td>
<td>Shot level, frame level, scene level, clip level</td>
<td>Depends on video structure and application type</td>
</tr>
<tr>
<td>Feature extraction and transformation</td>
<td>Image features at frame level as well as motion descriptors, camera metadata like motion [161], date, place, time etc.</td>
<td>Motion calculation is computationally expensive</td>
</tr>
<tr>
<td>Video classification</td>
<td>Human body motion recognition [18], goal detection [16], gestures etc.</td>
<td>Not enough training data to learn events of interest, domain knowledge dependence</td>
</tr>
<tr>
<td>Video clustering</td>
<td>Clustering of shots [152, 153], segments [109] for video, indexing etc.</td>
<td>Scalability for large video clusters is an issue</td>
</tr>
<tr>
<td>Video association</td>
<td>A graph based association finding sequence of events in movies [127]</td>
<td>Finding semantic event boundaries and temporal distance thresholds</td>
</tr>
</tbody>
</table>


---

**Future Outlook: Audio data mining open research issues**

<table>
<thead>
<tr>
<th>Task</th>
<th>Approach/Application</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing</td>
<td>Phonetic level, word level, window of fixed size</td>
<td>Domain dependence e.g., phonemes level if foreign terms, segmenting out silence, noise etc.</td>
</tr>
<tr>
<td>Feature extraction and transformation</td>
<td>Plate rate, zero crossing rate, Mel frequency cepstral coefficients [19], bandwidth, spectral centroid, frequency spectrum, frequency sparsity, etc.</td>
<td>Sensitive to the parameters e.g., number of bins, frequency resolution, smoothing</td>
</tr>
<tr>
<td>Audio classification</td>
<td>HMM [24], GMM [20], SVM [46], Bayesian classifier to do segmentation and classification of speech, music etc.</td>
<td>Model based approaches have problem to work well in real time as more number of iterations and lot of data are needed for training.</td>
</tr>
<tr>
<td>Audio clustering</td>
<td>Clustering speaker gender/speech segment of same speaker [22]</td>
<td>Large clusters are sometimes biased</td>
</tr>
</tbody>
</table>


---

**Future Outlook: Text data mining open research issues**

<table>
<thead>
<tr>
<th>Task</th>
<th>Approach/Application</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text mining</td>
<td>Part of speech tagging, stemming, stop word removal, text chunking etc.</td>
<td>Resulting ambiguity is a main problem, single word different meaning in different context</td>
</tr>
<tr>
<td>Feature extraction and transformation</td>
<td>Identification of important keywords using TF, IDF [177] etc.</td>
<td>Different sizes and contexts of the documents</td>
</tr>
<tr>
<td>Text classification</td>
<td>K-nearest neighbor classification, decision trees, naïve bayes classifier [51]</td>
<td>Final decisions depend on relatively few terms</td>
</tr>
<tr>
<td>Text clustering</td>
<td>Bi section k means clustering, co-clustering [14]</td>
<td>Finding good distance measures</td>
</tr>
</tbody>
</table>

Bhatt & Kankanahalli (2011)

---

**Future Outlook: Multimodal data mining open research issues**

<table>
<thead>
<tr>
<th>Task</th>
<th>Approach/Application</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing</td>
<td>Trust multiple streams separately [98] or consider multiple streams as one [161]</td>
<td>Multimodal data stream synchronisation</td>
</tr>
<tr>
<td>Feature transformation</td>
<td>Metadata fusion [146]</td>
<td>External knowledge fusion</td>
</tr>
<tr>
<td>Multimodal classification</td>
<td>Pre-filtering heuristic rule based [16], SVM classifier based [14], sub-space based [129], fusion: late [99], early [116] and multi [85, 152]</td>
<td>Class imbalance, multimodal classifier fusion</td>
</tr>
<tr>
<td>Multimodal clustering</td>
<td>Mixed media graph [100], clustering [103, 140], EEMI [17]</td>
<td>Cross modal correlation discovery</td>
</tr>
<tr>
<td>Multimodal association</td>
<td>Generalized sequence pattern mining [90, 126]</td>
<td>Automatic identification of temporal structures</td>
</tr>
</tbody>
</table>


---

**Sample Questions (1)**

- What is the main goal in knowledge discovery?
- Describe the classical process of KDD!
- How do we define a data object?
- What are the most important data mining tasks?
- What is the difference between supervised and unsupervised learning?
- What is the difference between classification and numeric prediction?
- Why is cluster analysis in medicine important?
- How is data mining defined?
- Describe the taxonomy of data mining paradigms!
- What is a neural network and a neuron?
- What is an artificial neural network?
- How does an ANN work?

- Provide an example on the use of ANN for medical decision making!
- What can you infer from a Receiver Operating Characteristic (ROC)?
- How can you rate the diagnostic accuracy of an ANN?
- What is model based clinical decision making?
- What is very important in clinical patient management?
- What is a Bayesian Network (BN Definition)?
- Why do we need a Markov-Blanket?
- What is the principal function of a Support Vector Machine?
- How would you describe the differences between ANN and SVM?
- Why is text mining in the medical domain practice so difficult?
- Just name some important semantic methods for NLP!
- What is the typical system architecture of a NLP System?
Some Useful Links (1)

- http://www.support-vector-machines.org
- http://www.neurosolutions.com/ (End-user friendly Neural Network Development Environment)
- http://www.kernel-machines.org
- http://www.kernel-machines.org
- http://www.youtube.com/watch?v=eDYOH9q2QDA
- http://leenissen.dk/fann/wp/ (Fast Artificial Neural Network Library)
- http://www.support-vector-machines.org
- http://www.kernel-machines.org

Appendix: WEKA Tool – The knowledge flow interface

Appendix: Criteria for selecting a data mining system

- Scalability
  - Row (or database size) scalability
  - Column (or dimension) scalability
  - Curse of dimensionality: it is much more challenging to make a system column scalable that row scalable
- Availability of Visualization tools
  - Visualization categories: data visualization, mining result visualization, mining process visualization, and visual data mining
- Data mining query language and graphical user interface
- Easy-to-use and high-quality graphical user interfaces are essential for end-user guided, highly interactive data mining systems

Appendix: Examples of professional data mining systems

- Microsoft SQLServer 2005 (Integrated DB and OLAP with mining Support OLEDB for DM standard)
- SAS Enterprise Miner (A variety of statistical analysis tools, Data warehouse tools and multiple data mining algorithms)
- IBM Intelligent Miner (A wide range of data mining algorithms; Scalable mining algorithms; Toolkits: neural network algorithms, statistical methods, data preparation, and data visualization tools; Tight integration with IBM’s DB2 relational database)
- SGI MineSet (advanced statistics and visualization tools)
- Clementine (SPSS – integrated data mining development environment – multiple data mining algorithms)

Appendix: Example Bio2RDF Architecture

Appendix: Sample Architecture for a visual retrieval system

Some Data Mining wording first

- Data sets are made up of data objects.
- Each data object represents an entity.
- Data objects are described by attributes.
- We want to mine samples (aka examples, instances, data points, objects, tuples) out of databases
- Database rows $\rightarrow$ data objects; columns $\rightarrow$ attributes

Schedule

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

Knowledge Discovery is Making Data Useful

- Masses of data –
- Graph-based data may contain information about design principles and/or the evolutionary history of complex systems
- As in Paleontology: Discovery from past data

Big Data Players

Slide 6-12 Example Classification vs. Numeric Prediction

- **Medical Decision**: is a tumor malign or benign?
- **Classification** = predicts categorical class labels (discrete or nominal)
  - classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data
- **Numeric Prediction** = models continuous-valued functions, i.e., predicts unknown or missing values

Image Source: John Nguyen (2010)

Slide 6-13 Which one is malign – which one is benign?

Slide 6-14 Example: Benign vs. Malign – molecular level

Slide 6-15 Cluster analysis
H. Clustering Explorer: melanoma gene expression


ANN Demo: Learning Process

http://www.youtube.com/watch?v=d9ta0fHdKso

Slide 6-24 Artificial Neural Network ANN

- The n-dimensional input vector \( x \) is mapped into variable \( y \) by means of the scalar product and a nonlinear function mapping

\[
\sum w_k (Learning \ Rate) \rightarrow f(\sum) \rightarrow y
\]

Input vector \( x \)  weight vector \( w \)  weighted sum  Activation function


Artificial Neural Network ANN

http://www.youtube.com/watch?v=0Str0Rdkxxo

Typical ANN architecture


The Tool of the Liver Transplantation Example

http://www.neurosolutions.com

Support Vectors

- Small Margin
- Large Margin

Support Vectors


Support Vector Machine Demonstration

Video removed due to file size

SVM with a polynomial kernel – Visualization by Udi Aharoni

Remember: The Curse of Dimensionality

- The complexity of trained classifier is characterized by the # of support vectors — not the dimensionality of the data
- The support vectors are the essential or critical training examples — they lie closest to the decision boundary
- If all other training examples are removed and the training is repeated, the same separating hyperplane would be found
- The number of support vectors found can be used to compute an (upper) bound on the expected error rate of the SVM classifier, which is independent of the data dimensionality
- Thus, an SVM with a small number of support vectors can have good generalization, even when the dimensionality of the data is high


SVM advantages in high dimensional data

Slide 6-42: Latent Dirichlet allocation (LDA)

**Slide 6-43: Principal component analysis (PCA)**

![PCA Diagram](http://www.nlpca.org/pca_principal_component_analysis.html)

**History of using NLP in biomedicine and molecular biology**


**Example: Medical Text Mining process (1)**


**Example: Medical Text Mining process (2)**

Uramoto, N. et al. (2004)

**Example: Medical Text Mining architecture (3)**

Uramoto et al. (2004)
Example: Medical Text Mining results (4)

Uramoto et al. (2004)

NLP system architecture for text mining in biology


Summary: Multimedia Data Mining Overview


Slide 6-5 The Knowledge Discovery Process Chain

HIO, Interactive Visualization, Analytics, Decision Support
Topological Data Mining - Pattern Discovery
Sampling, Cleansing, Preprocessing, Mapping
Data Integration, Data Fusion, Pre-selection
Privacy, Data Protection, Data Security, Data Safety