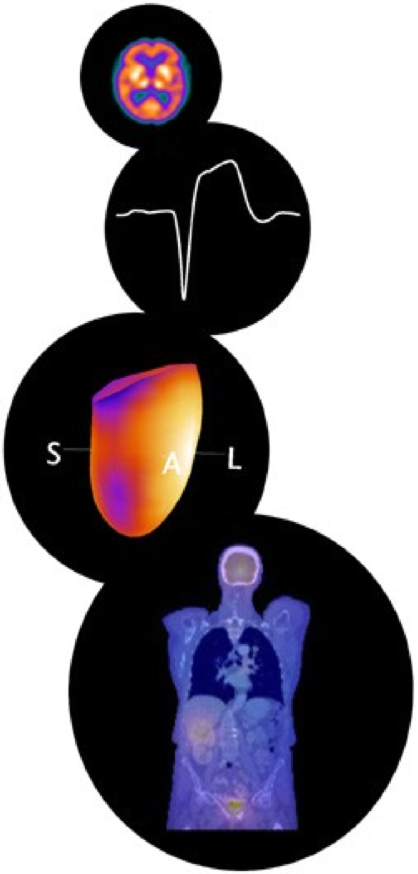


Medical Applications of Pattern Recognition

by

Neşe Yalabık

HIBIT'10, Antalya,April 2010

Outline

- **Part 1**:Introduction:Definitions and Terminology
- **Part 2**:Historical Background
- **Part 3**: PR Techniques used in Medicine and Application Examples

Part 1:Introduction:Definitions and Terminology

Definitions and Terminology

- **Medical Informatics :** Is an interdisciplinary scientific field of research that deals with the use of Information and Communication Technologies and Systems for clinical health care, for more accurate and faster service to people.
- **Pattern Recognition(PR):** Automated analysis of collected attributes of objects, events,etc. to classify them into categories.
- **Medical Pattern Recognition:** All PR Techniques in decision support and treatment of illnesses

Example Applications of Pattern Recognition

- Reading hand-written text to classify it into letters and words
- Analyzing fingerprints to find the owner
- Recognizing the faces of people to name them
- Finding buildings in a satellite image
- Naming a gun from its bullet mark(Ballistics)
- Identifying different objects on a conveyor belt
- Analyzing test results in decision support for any illness

Pattern Recognition and Classification: An Introduction

We human beings do pattern recognition everyday.

We “**recognize**” and **classify** many things, even if it is corrupted by **noise**, **distorted** and **variable**.

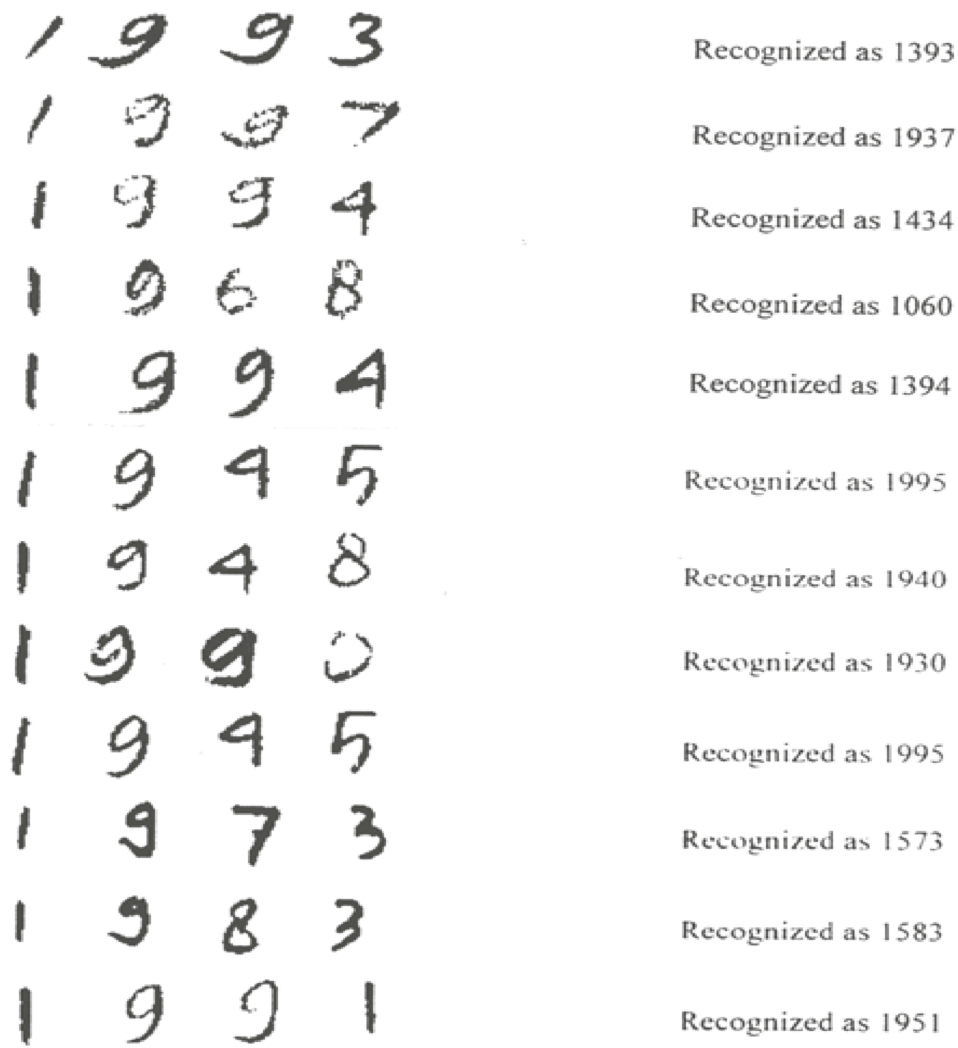
- Classification is the result of recognition: categorization, generalization
- A problem is a PR problem only if it involves ‘**statistical variation’**

How do we do it?

- Automatic pattern recognition has 50 years of history
- Many different approaches tried
- Limited success in many problems
- Successful only with restricted environments and limited categories.

Variation in PR Problems

- We see here that all 9's are different from each other and 9's and 4's can easily be mixed



Unlimited Recognition

Turns out that **unlimited recognition** is still a dream,

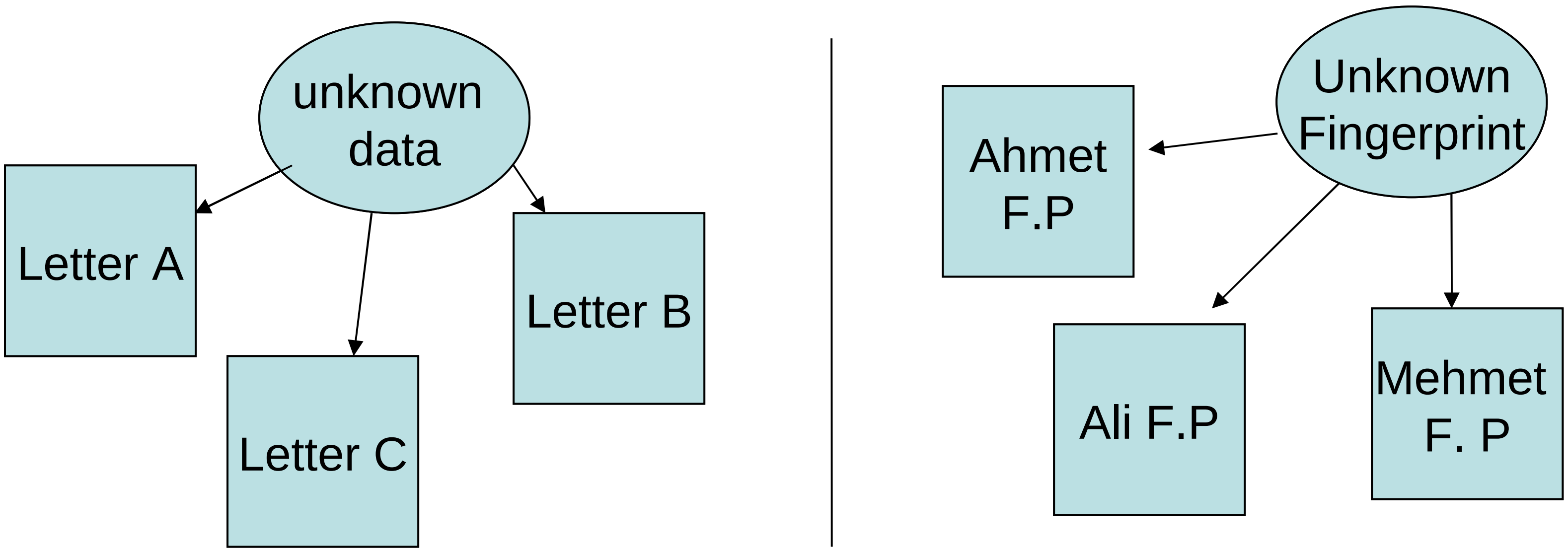
such as:

- Continuous speech recognition
- Cursive script
- Unlimited medical diagnosis
- Unlimited fingerprint recognition

Today **applications aim at limiting these to simpler problems**.

A more detailed definition of P.R.: The process of machine perception for an automatic labeling of an object or an event into one of the predefined categories.

Classifiers



Objective in PR

Minimize the average error (at least as good as a human being)

Minimize the risk: wrong decision could be more risky in some cases such as medical diagnosis

Why automize? Obvious reason: save from time and effort

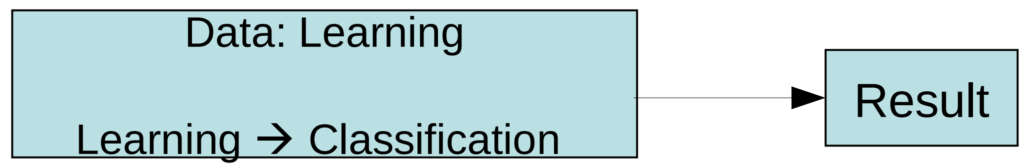
(Ex: consensus forms: enter 100 million records into electronic medium).

How do machines solve it: Many different approaches in history

- Template matching
- Use statistics, decision theory “statistical pattern recognition”
- Use “ neural networks” self learning systems
- Tree Classifiers
- Support Vector machines
- Multiclassifiers

Learning and Features

Whichever approach is used, there’s a classification process

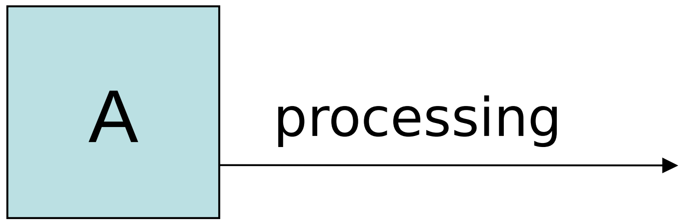


- “**Learning samples**” Large data sets to be used in training, or estimating parameters, etc.
- “**Result**” a decision on the category sample belongs.
- “**Test Samples**” used in testing the classifier performance.
- L.S and T.S may have an overlap.
- “**Data”** a raw data pre-processing feature set.
- “**Feature**” a discriminating, easily measurable characteristics of our data.

In all approaches, samples from different categories should give distant numerical values for features.

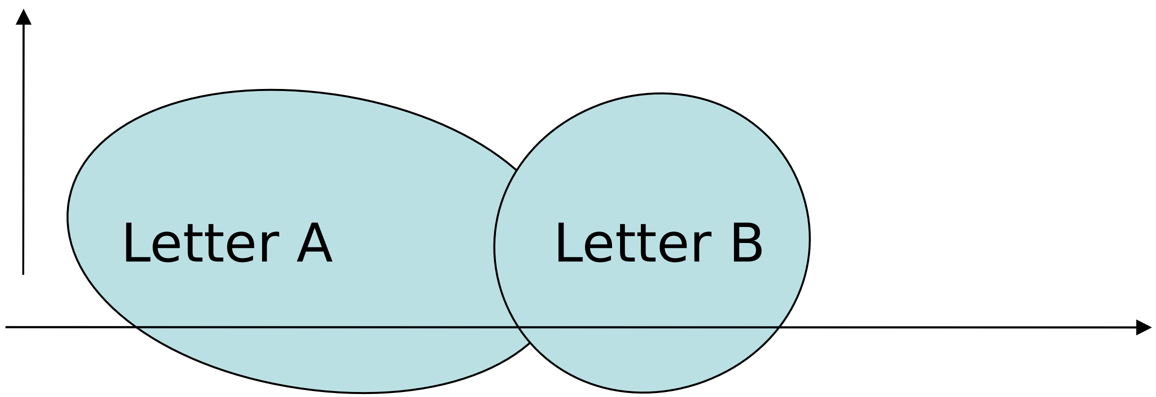
Ex. For letter A, a feature

2-d array



$$\left[M_0, M_1, \ldots, M_k \right]$$

M: moments invariants (center of growing obtained from the A feature vector! A model of the underlying system that generated it.



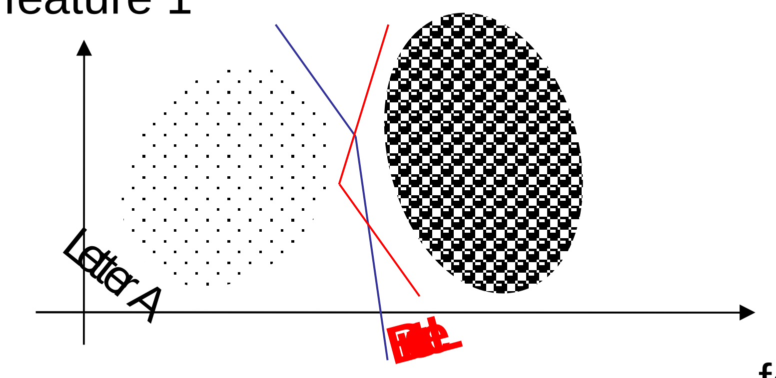
There is always an error probability in decision!

How many features should we use?

Not small, but not too large either.

(curse of dimensionality)

Classification



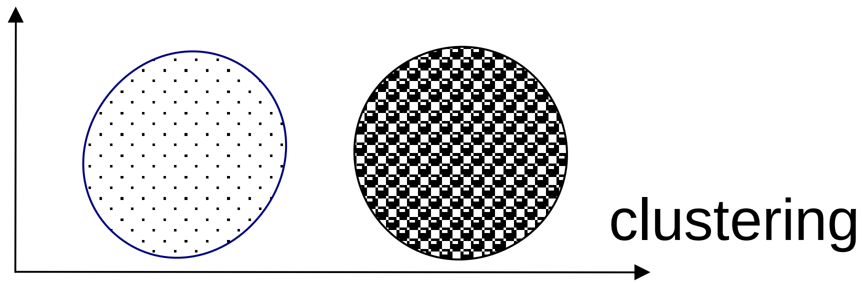
How do we separate A 's from B 's?

- From a decision boundary
- Classify the sample to the side it falls

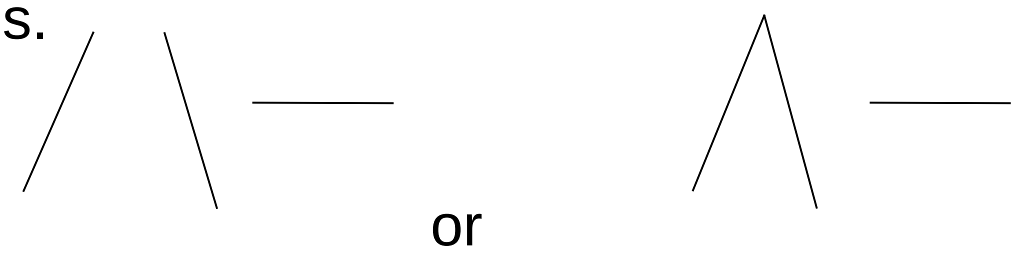
Many classification methods exist

- Parametric: Bayes Decision Theory, Parameterize as belonging to a probabilistic variable.
- Non-parametric: discriminant functions, nearest neighbor rule use only learning samples
- Tree classifiers

Given the learning data set, supervised learning, learn parameters of P.R.



If we do not have enough data, we incorporate “domain knowledge” for example, we already know that letter A is written by hand in form of 2 or 3 strokes.



So maybe recognizing strokes rather than the complete letters first is a better idea. Also consider the text.

Statistical Approach to P.R

$X = [X_1, X_2, ..., X_d]$

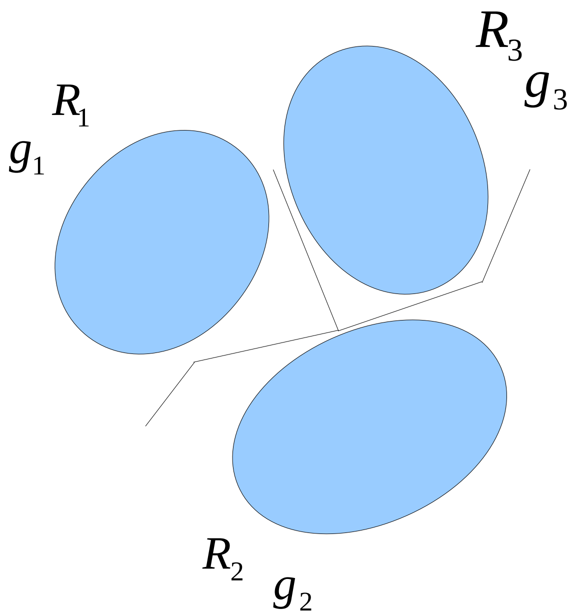
Dimension of the feature space: d

Set of different states of nature: c

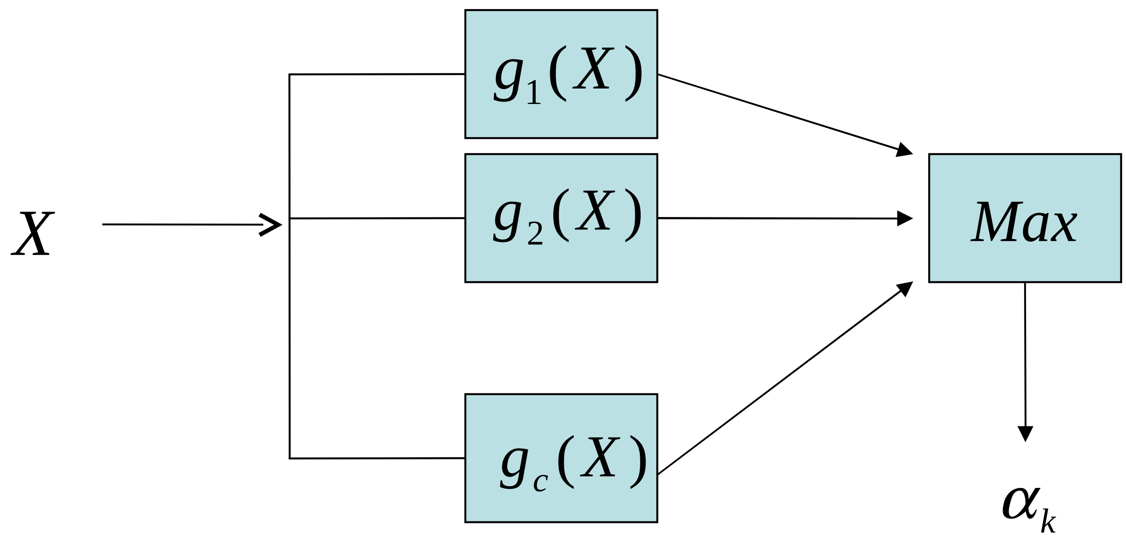
Categories: $\{\omega_1, \omega_2, ..., \omega_c\}$

find R_i $\quad R_i \cap R_j = \varphi \quad uR_i = R^d$

for R_i $\quad g_i(X) \geq g_j(X)$



A Pattern Classifier



So our aim now will be to define these functions $g_1, g_2, ..., g_c$ to *minimize* or *optimize* a criterion.

Pattern Recognition in Medical Decision Support

- 50 years ago, we tried to make systems that will 'diagnose' an illness without a physican
- Today, we make systems that we call '**decision support**' that only gives opinion to physician
- Interpreting all kinds of collected medical data, which is huge

Pattern Recognition in Medical Decision Support

- Examples:
- Interpreting 1-d data such as in ECG, EEG
- Interpreting 2-d data: detecting cells, tumors or any other abnormalities in any x-ray, MR, tomography etc.
- Sequence processing in genetic data
- Processing of any collected numerical data such as blood test results
- Processing any collected non-numeric data such as patient history, doctor interpretations and reports
- Using more than one of these together to use in decisions and treatment of an illness

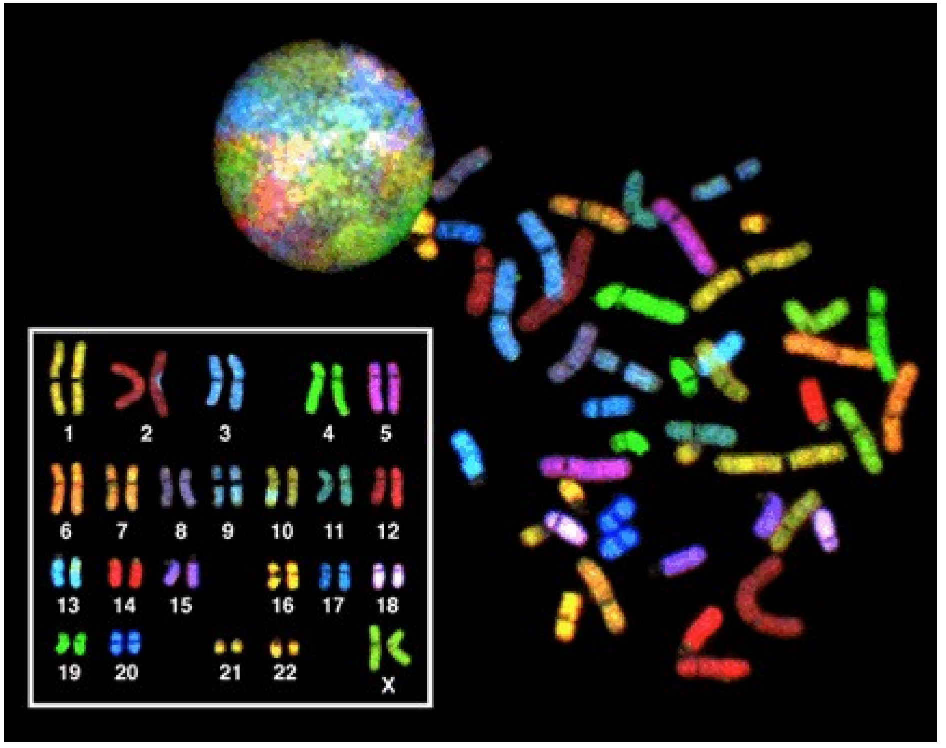
Part 2:Historical Background

Historical Background

- Earlier in 60's and 70’s of the 20th century where computers were thought to be able to solve any problems, it was thought that it was easy
- Enter the symptoms, diagnose the illness
- Unfortunately it did not work!
- As in all PR problems, you had to limit yourselves to very restricted problems

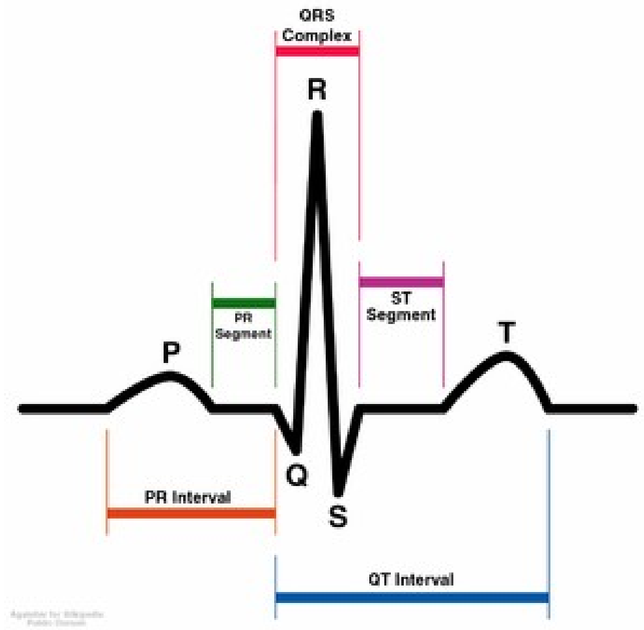
Chromosome Analysis

- **Karyotyping:** ordering and enumerating the chromosomes
- Detect the abnormalities in chromosome spreads to detect genetic deseases, cancer etc. still an unsolved problem.



ECG Analysis

- ECG and EEG analysis: First automated ECG interpreters available in '70's, improved later
- Today, many accurate machines available
- PQRST curve: abnormalities detected by measuring various features



Medical Diagnosis Decision Support

- In 80's and 90's, 'expert systems' were popular
- Most successful diagnostic application: **Mycin**
- was designed to diagnose infectious **blood diseases** and recommend **antibiotics** in Stanford University
- Used '**Expert Systems**' approach: 500 rules(if-then statements)
- a correct diagnosis rate of about 65%(better than most physicians),
- Legal issues : Who is responsible for the wrong diagnosis?
- Certainty factors in rules
- Never used in practice due to legal and ethical issues
- Also technical issues that are solved today

Example of a Decision Rule in MYCIN

```
RULE-507

IF:
   1. The infection which requires therapy is
      meningitis
   2. Organisms were not seen on the stain of the
      culture
   3. The type of the infection is bacterial
   4. The patient does not have a head injury defect
   5. The age of the patient is between 15 and 55 years
Then:
      The organisms that might be causing the infection
      are diplococcus-pneumoniae and neisseria-
      meningitidis
```

Medical Diagnosis Decision Support

- 90's and 2000's: Mycin-like system led to clinical 'decision support systems' or 'diagnostic Clinical Decision Support Systems' AI approach to PR
- Knowledge base, Inference Engine
- Non-knowledge based CDSS: Neural Networks, Bayesian Networks, Genetic Algorithms, Tree Classifiers, multiclassifiers etc.
- Shown to improve physician's performance in general

Part 3: PR Techniques used in Medicine and Application Examples

PR Techniques used in Clinical Medicine

Last 20 years many new approaches to PR, many successfully applied to medicine.

- Neural Networks
- Bayesian Belief Networks
- Support Vector Machines
- Tree Classifiers
- Multiclassifiers. A combination of above

Neural Networks

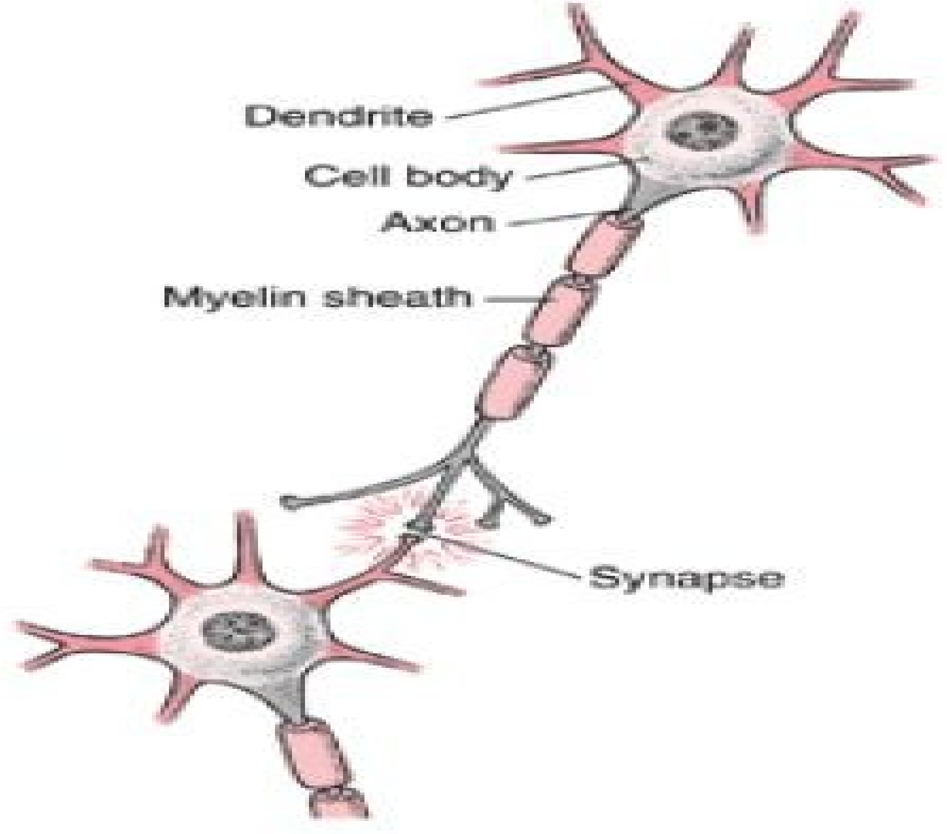
- Old approach. Perceptron in '50's by Rosenblatt
- Revived with new learning algorithms in 80's (Back Propagation)
- Used in many scientific problems

Biological vs. Artificial

Biological Neural Networks

A Neuron:

A nerve cell as a part of nervous system and the brain

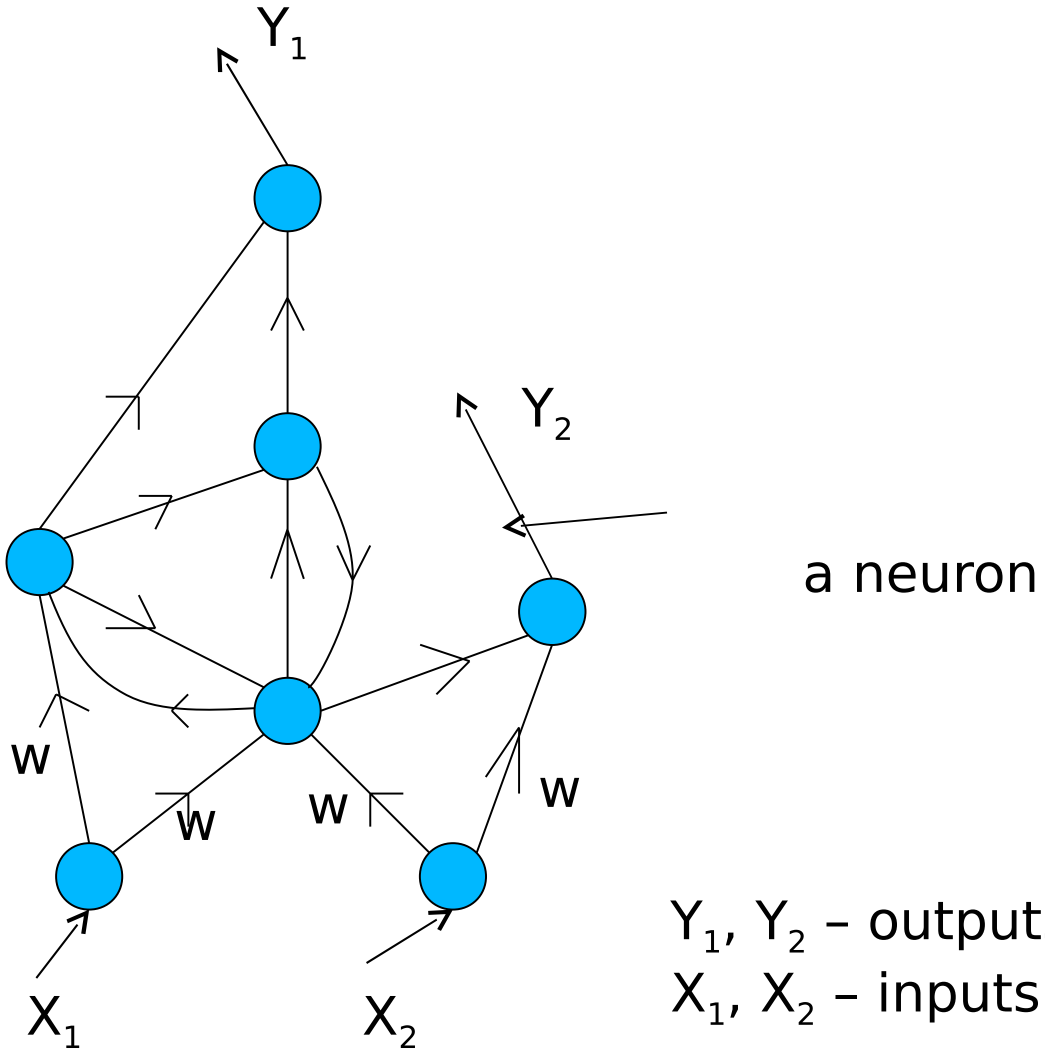


Biological vs. Artificial

- 10 billion neurons and a huge number of connections in human brain.
- thinking, reasoning, learning and recognition are performed by the information storage and transfer between neurons
- Each neuron “fires” sufficient amount of electric impulse is received from other neurons.
- The information is transferred through successive firings of many neurons through the network of neurons.

Artificial Neural Networks:

- An artificial NN, or ANN or (a connectionist model, a neuromorphic system) is meant to be
- A simple, computational model of the biological NN.
- A simulation of above model in solving problems in pattern recognition, optimization etc.



An Artificial Neural Net

Y_1, Y_2 – outputs
X_1, X_2 – inputs
w – neuron weights

Any application that involves

- Classification
- Optimization
- Clustering
- Scheduling
- Feature Extraction

may use ANN!

WHY ANN?

- Easy to implement
- Self learning ability
- When parallel architectures are used, very fast.
- Performance at least as good as other approaches, in principle they provide nonlinear discriminants, so solve any P.R. problem.

Multilayer Perceptron

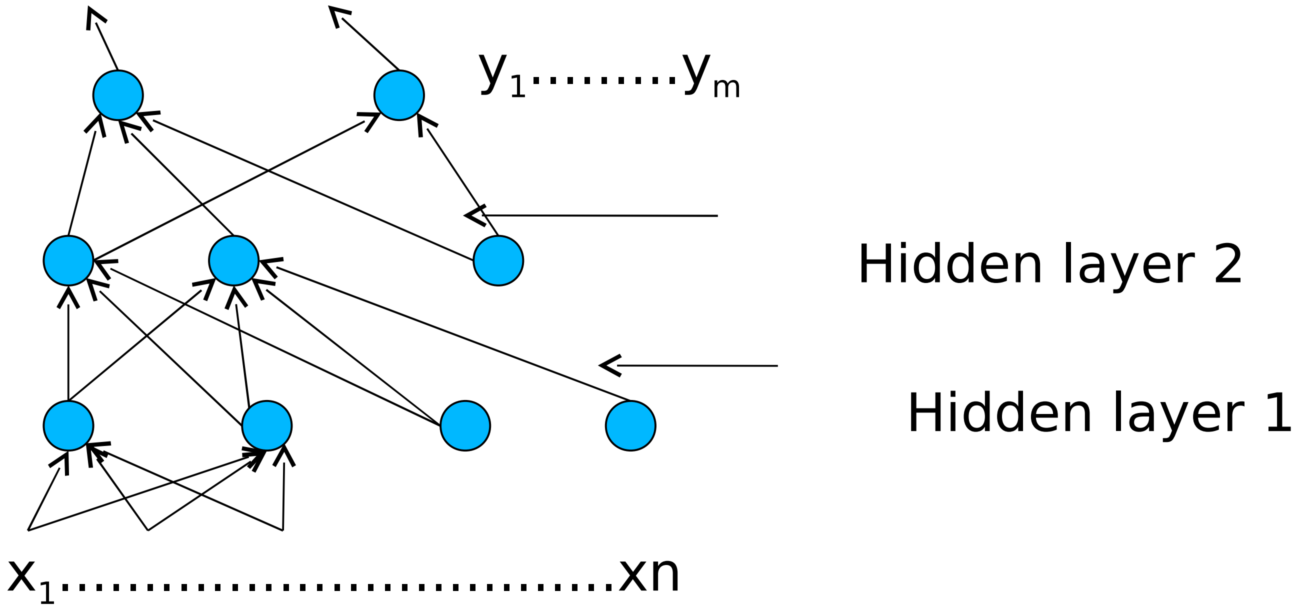


Figure: Fully Connected Multilayer Perceptron

Multilayer Perceptron

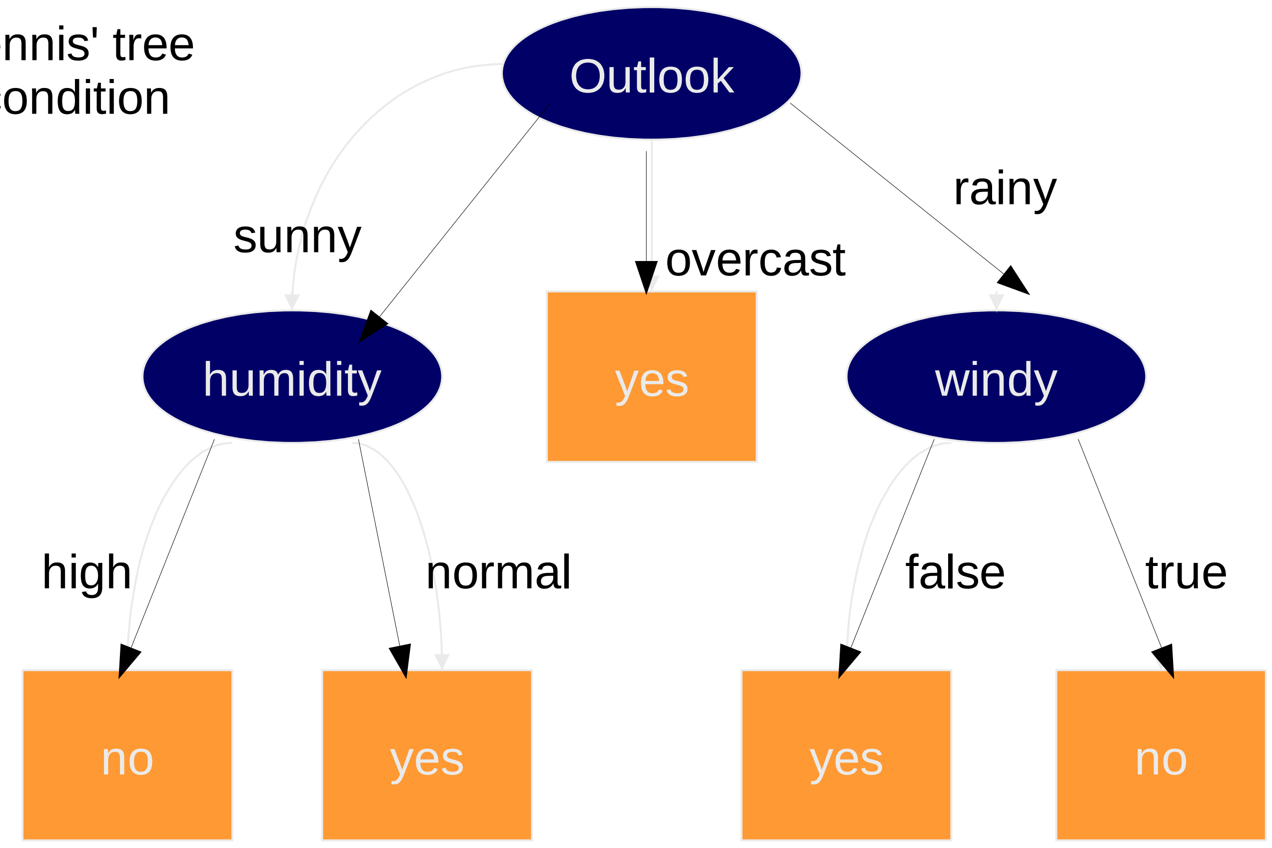
- It was shown that a MLP with 2 hidden layers can solve any decision boundaries.
- **Back-propagation learning algorithm:** iteratively update the weights to obtain required input-output pairs.
- **Inputs:** Features, **Outputs:** one output/class.
- Successfully used in many bio-medical decision making problems

Tree Classifiers

- Consider the feature vector X= (x1, x2, x3....xn)
- A tree classifier considers features one by one instead of as a whole and measures them one by one, following the leaves of a tree. The features are usually binary valued .
- An optimum tree can be constructed using learning samples.
- Leaves of the tree correspond to the classes.
- Example will be seen in the following .

Decision Tree Example

The decision 'to play tennis' tree
According to weather condition



Decision tree for the weather data.

Example study

‘**OAGAIT**’: A Decision Support System for Grading Knee Osteoarthritis using Gait Data'

N. Köktaş, N. Yalabık, G. Yavuzer,P. Dunn, V. Atalay

A Tübitak Project , 2006-2008 and a Ph.D. Thesis

METU Computer Engineering Dept. and Ankara University Gait Laboratories

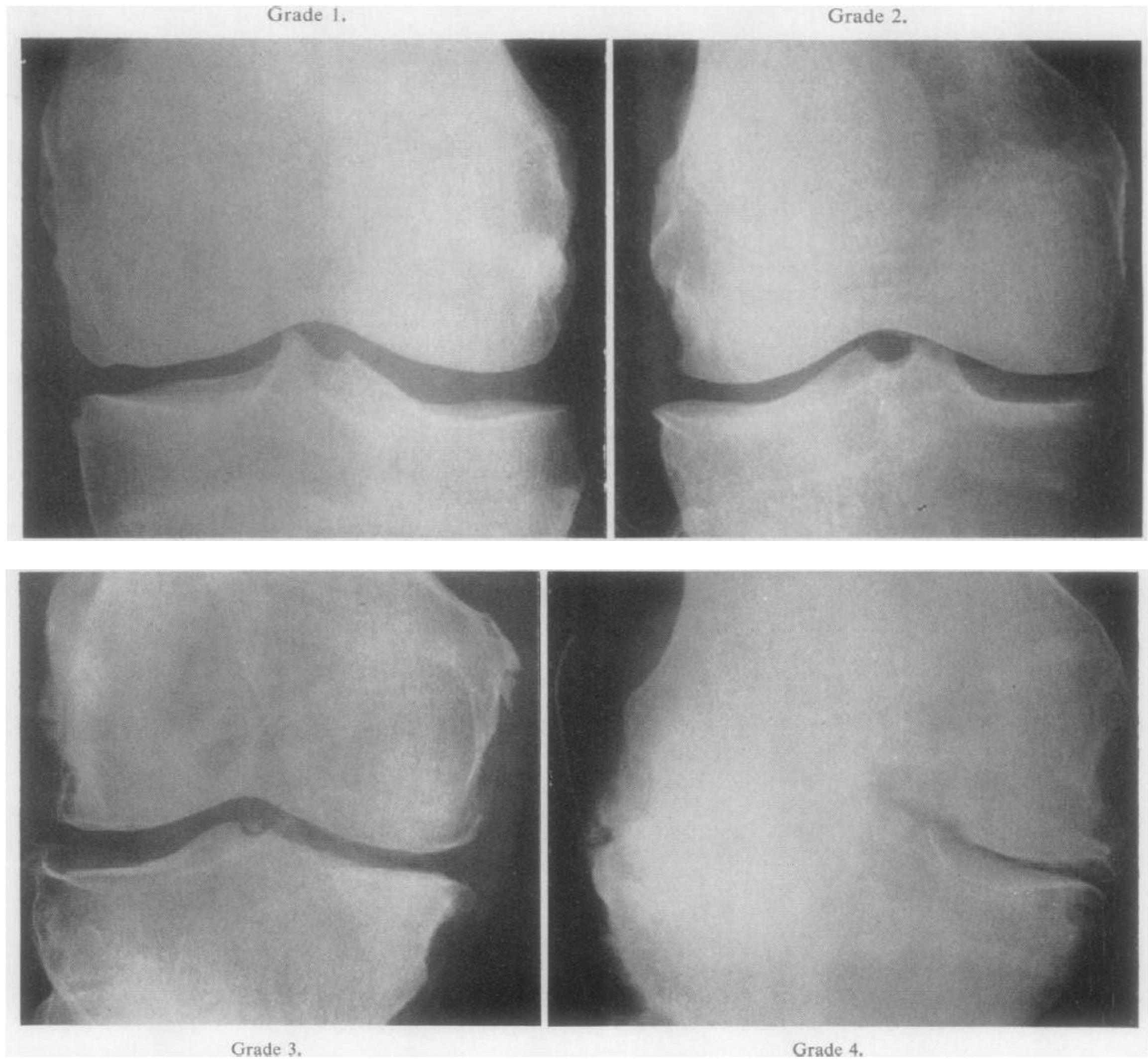
Gait Analysis

- What is gait analysis?
  - process of collecting and analyzing quantitative information about walking patterns of people
- Where is it used?
  - human identification
  - clinical applications
- Why is it important?
  - for diagnosis, developing treatment plans and tracking the progression of diseases

Osteoarthritis (OA)

- OA is a disorder that affects joint cartilage and surrounding tissue
- Shows itself by pain, stiffness and loss of function of knee
- Kellgren-Lawrence method is used for radiological assessment
  - *Grade 0:* Normal
  - *Grade 1:* Doubtful narrowing of joint space and possible outgrowth of the bone
  - *Grade 2:* Definite outgrowth of the bone and possible narrowing of joint space
  - *Grade 3:* Moderate multiple outgrowths, definite narrowing of joints space, some hardening and possible deformity of bone contour;
  - *Grade 4:* Large outgrowths, marked narrowing of joint space, severe hardening and definite deformity of bone contour.

XR image showing OA of the knee joint



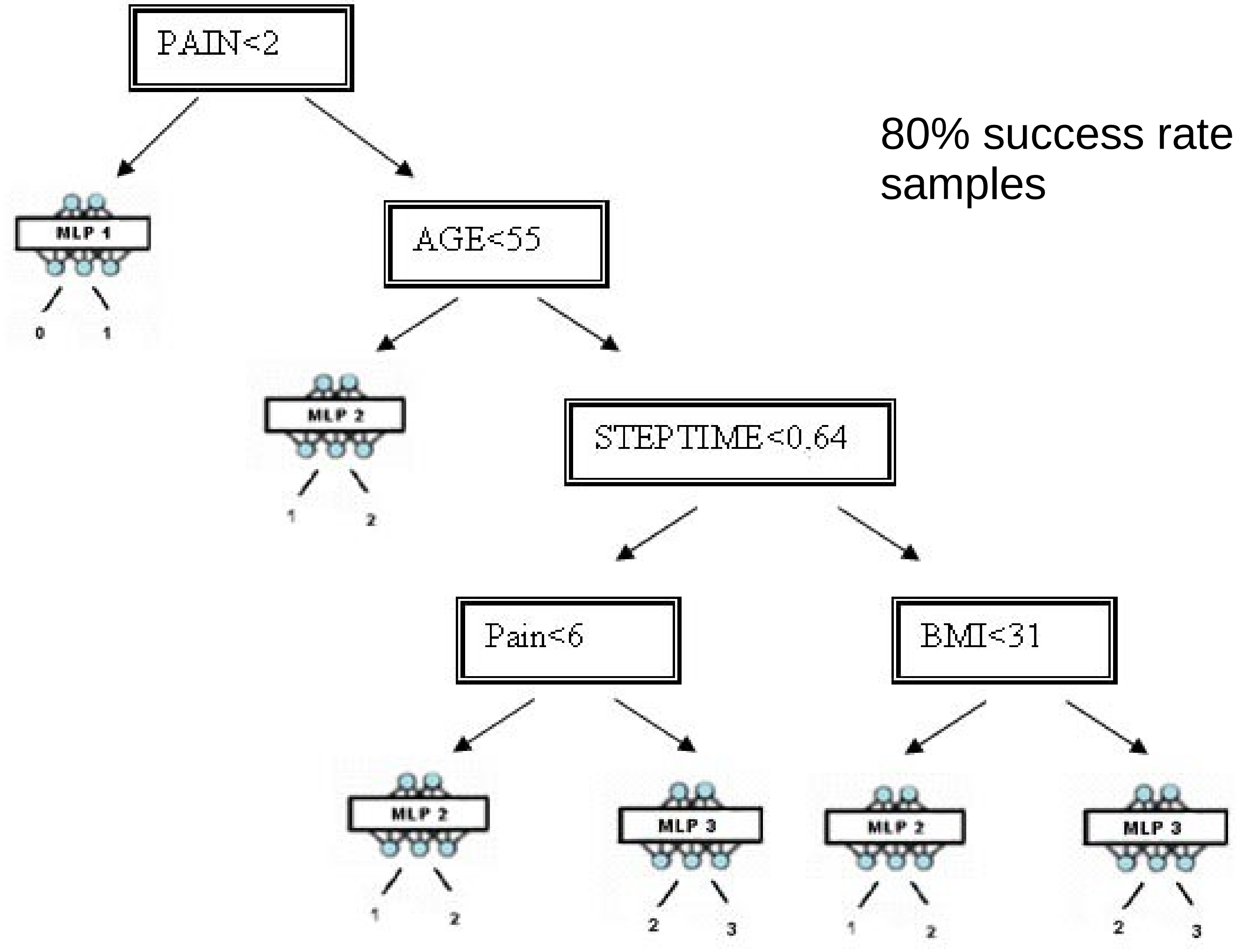
Gait Classification

- The aim is to support the physicians' decision making
- Most popular PR algorithms for gait classification are NNs, SVMs, FFT, PCA etc.
- Gait Laboratories in hospitals in Turkey are becoming very popular
- There are 5 gait laboratories only in Ankara
- The increasing amounts of collected data need to be analyzed intelligently
- MD.s are seeking help of computer scientists for developing tools

Properties of Gait Data

- Three sets of data is gathered in gait laboratory
  - History and symptoms of the patients
    - ***A = {age, BMI, pain, stiffness, history, period, sex}***
  - Time-distance parameters of the gait
    - ***B = {Cadence, Walking Speed, Stride Time, Step Time, Single Support, Double Support, Stride Length, Step Length}***
  - Temporal changes of the joint angles (kinetic and kinematic gait variables)
    - ***C = {PTilt, PObliq, PRot…… APRot}***

Implementation and results



80% success rate with 100 test samples

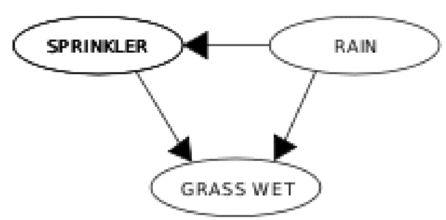
Bayesian Networks(BN)

- **A Bayesian Belief Network**: a knowledge-based **graphical representation** that shows a set of variables and their probabilistic relationships between diseases and symptoms. They are based on conditional probabilities, the probability of an event given the occurrence of another event, such as the interpretation of diagnostic tests. **In the context of CDSS, the Bayesian network can be used to compute the probabilities of the presence of the possible diseases given their symptoms.**
- Some of the advantages of Bayesian Network include the knowledge and conclusions of experts in the form of probabilitiesas an **assistance in decision making.**

A Simple Bayes Net

- Below net shows the probabilities between the case of grass being wet and sprinkler and rain conditions.
- Using the net, we can find the probability of rain if the grass is wet.

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Example Study

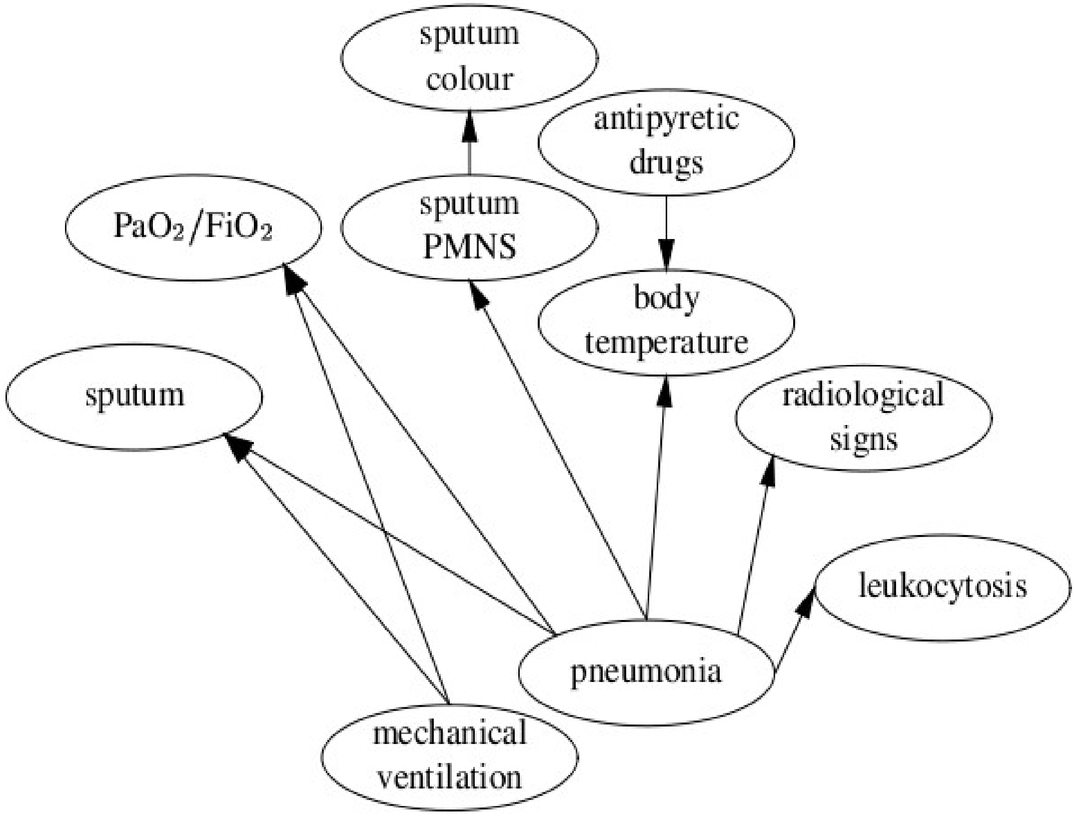
'Bayesian Networks in Medicine: a Model-based Approach to Medical Decision Making'

Peter Lucas,K-P. Adlassnig (ed.), Proceedings of the EUNITE workshop on Intelligent Systems in patient Care, Vienna, Oct. 2001, pp. 73-97)

Bayesian Networks in Medicine

- ' The BN formalism offers a natural way to represent the uncertainties involved in medicine when dealing with diagnosis, treatment selection, planning, and prediction of prognosis '
- 'A BN model that was developed to assist clinicians in the diagnosis and selection of antibiotic treatment for patients with pneumonia'
- Domain expert knowledge is used in developing BN
- Results show a close match between expert opinion and BN

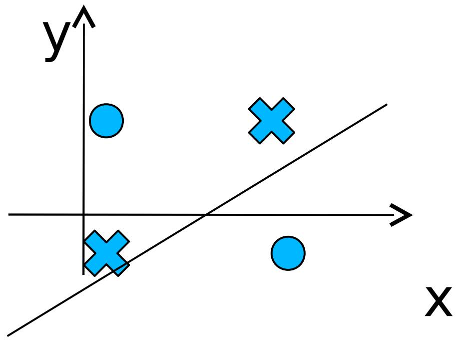
- A BN for pnomonia



Support Vector Machines(SVM)

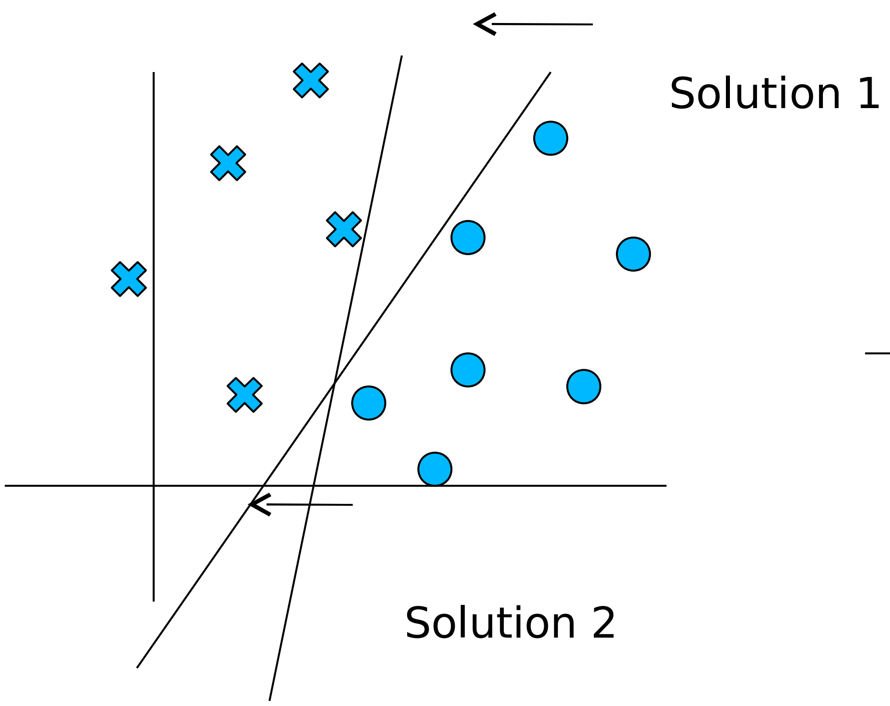
- **Support Vector Machines are extensions of Linear Discriminant Functions**
- Linear Discriminant Functions have linear decision boundaries and found using learning samples only
- **Linear separability:** All learning samples are correctly classified by a linear decision boundary
- Not possible for many cases
- **An SVM: An optimum linear discriminant function where linear separability is provided by a feature space extension to a higher dimension**

Linear Separability

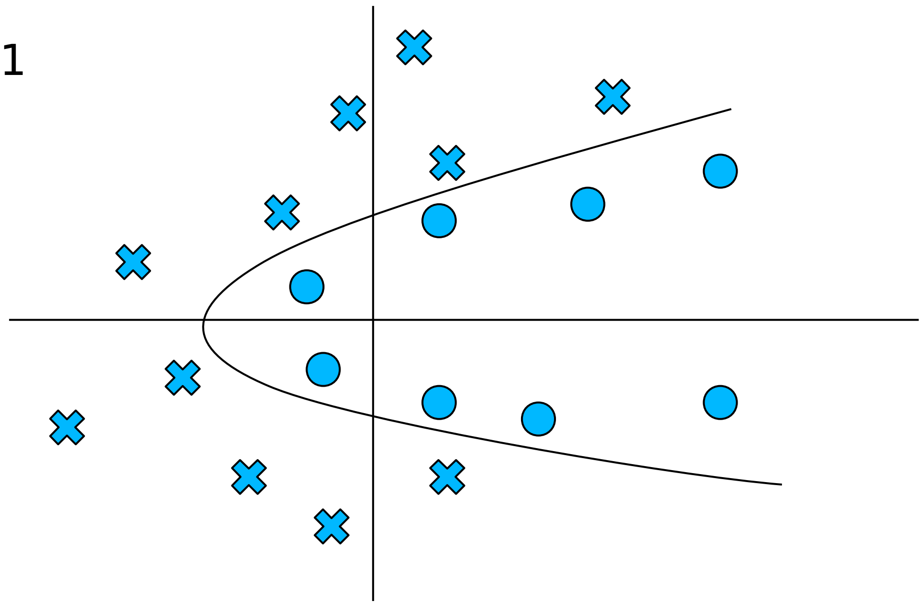


XOR Problem
Not linearly separable

Linearly seperable



not seperable



Many or no solutions possible

Here we see that by carrying the samples to a higher dimension results with separability which was not the case in lower dimension.

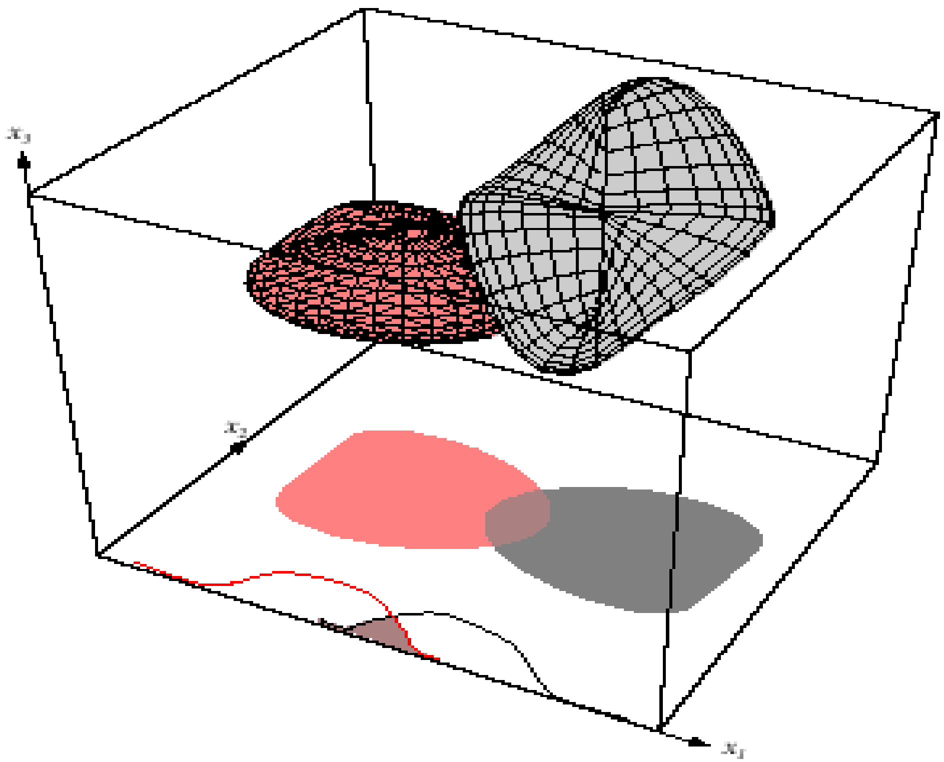
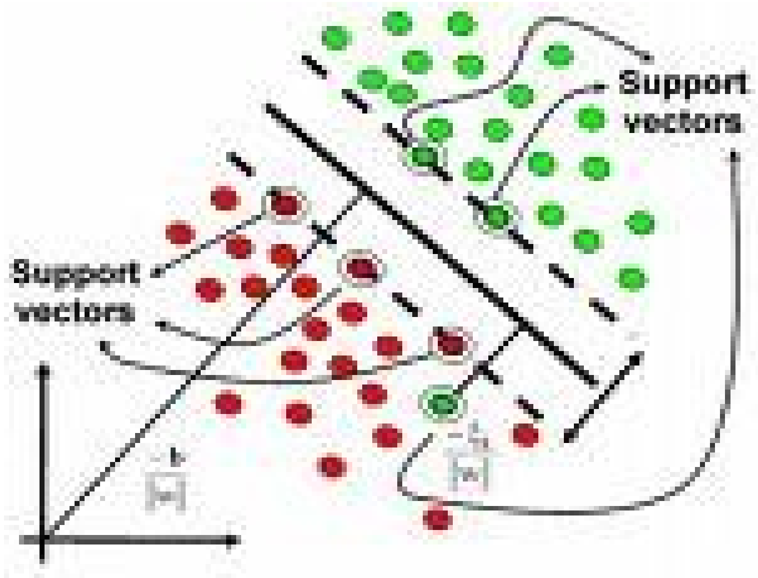


FIGURE 3.3. Two three-dimensional distributions have nonoverlapping densities, and thus in three dimensions the Bayes error vanishes. When projected to a subspace—here, the two-dimensional $x_1 - x_2$ subspace or a one-dimensional x_1 subspace—there can be greater overlap of the projected distributions, and hence greater Bayes error. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- SVM carries the feature space to a higher dimension by processing it with a nonlinear function called **'Kernel Function'**
- Then, finds an optimum boundary by making it equally spaced from samples from different classes using samples called '**Support Vectors**'



SVM in Medical Decision Making

- A newer tool than others in medical decision making as well as other applications
- Concluded to outperform other approaches in many studies as compared to NN, BN and others
- Even though it can be used for any problem, especially found to be successful in breast cancer studies

Example Study

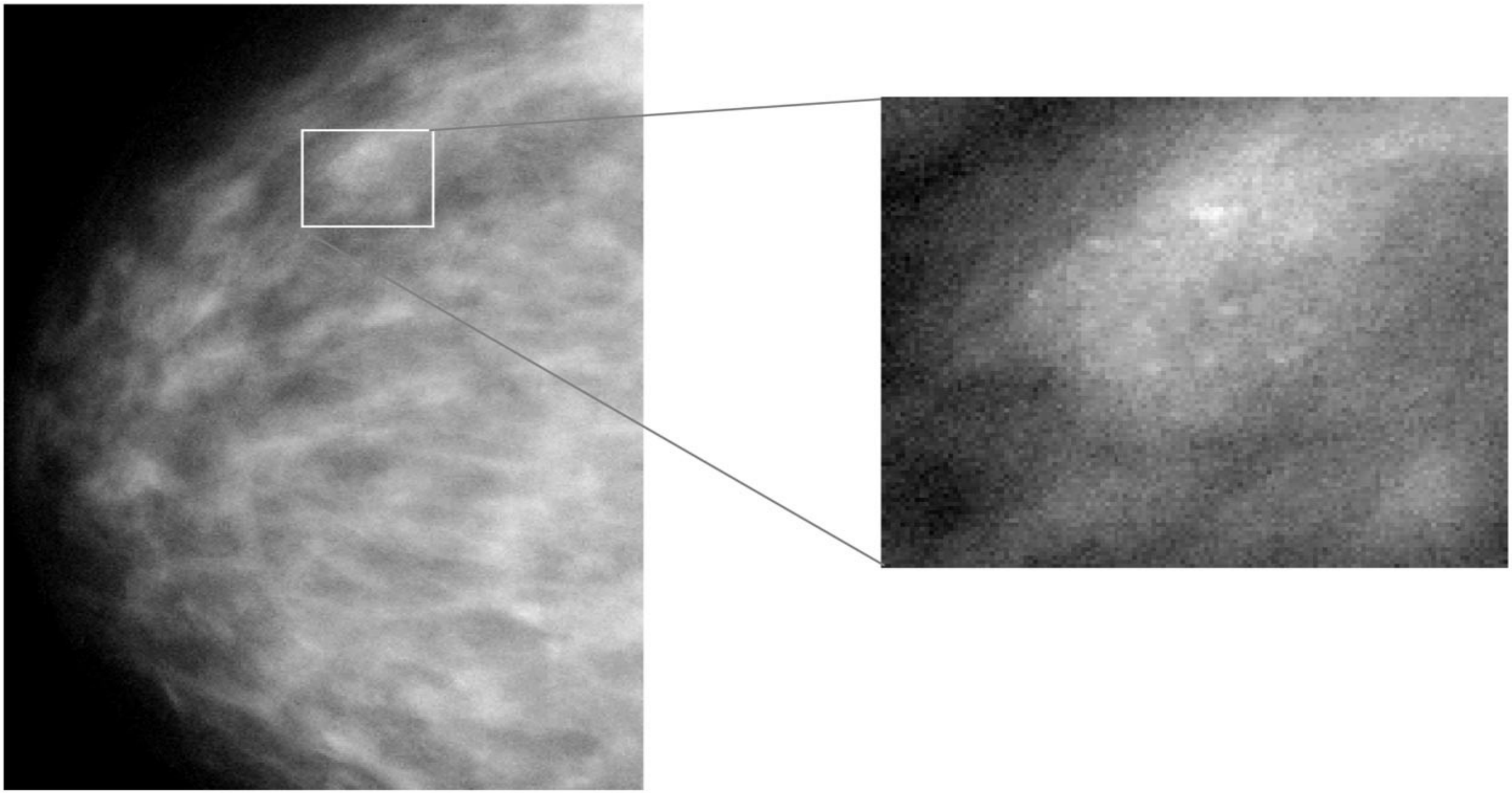
'A Support Vector Machine Approach for Detection of Microcalcifications'

Issam El-Naqa et al

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 21, NO. 12, DECEMBER 2002

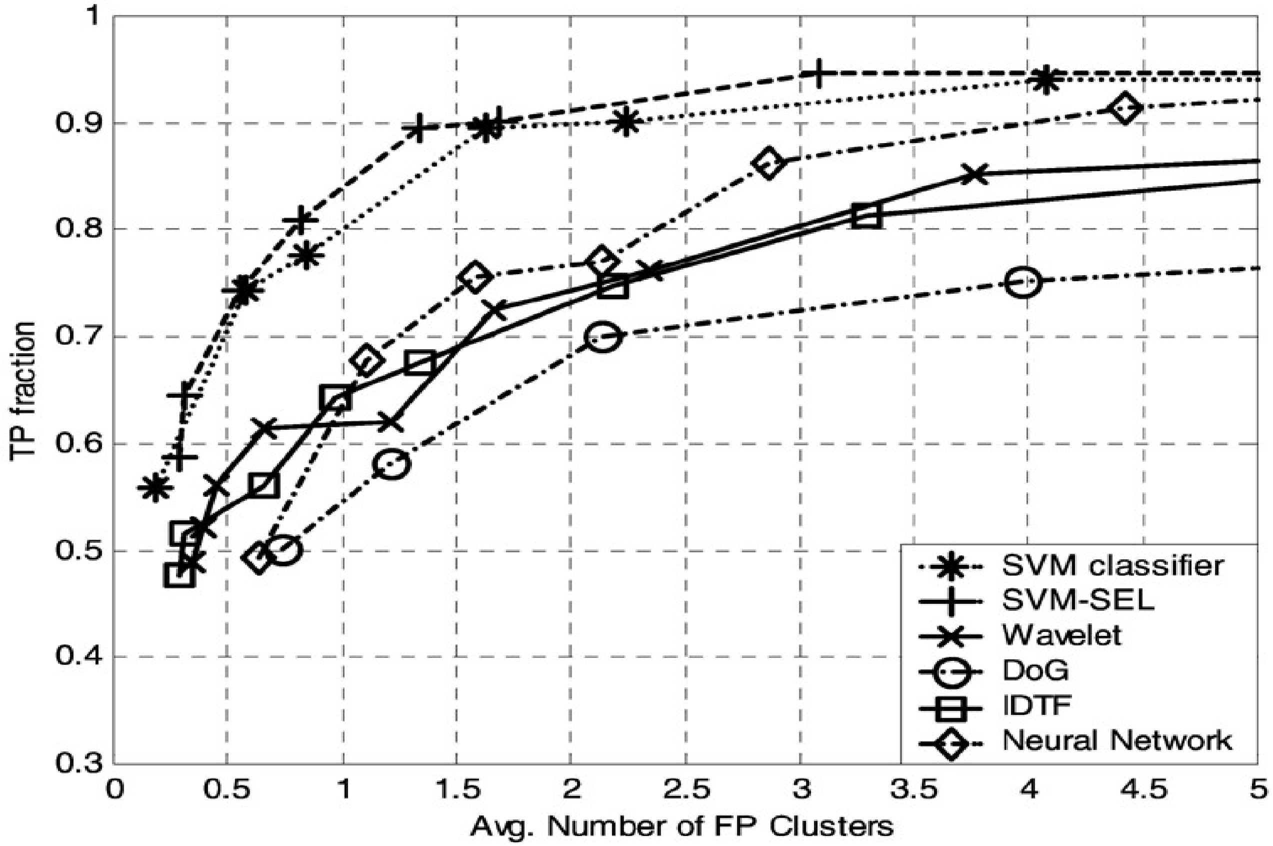
- Finds microcalcifications, that are pre-cancerous cycsts in breasts, from digital mammographs using SVM and compares it with other approaches

Microcalcifications in mammogram



Performance Comparison using a FROC curve

- Higher the curve is, better the performance



Conclusions

- We discussed many methods to automatically label illnesses, medical images and plots
- Recent methods are usually used as a part of a Decision Support System
- Ethical and legal issues prevent the development of fully automatic systems
- **Today, Pattern Recognition methods are accepted as useful tools in the service of M.D.'s as consultants in clinical decision making.**

References

- MIN720 Pattern Classification in Biomedical Applications' Course Lecture Notes, METU Informatics Institute, METU , 2010
- 'Pattern Classification' Duda, Hart, Stork, Wiley 2001
- Wikipedia Free Encyclopedia - www.wikipedia.com
- Other references in their respective pages