

Pattern Recognition I

Basic PR Concepts

Michal Haindl

Faculty of Information Technology, KTI
Czech Technical University in Prague
Institute of Information Theory and Automation
Czech Academy of Sciences
Prague, Czechia



<http://ro.utia.cz/>



NI/MIE-ROZ 2023-2024

© M. Haindl

MI-ROZ - 01

1/72

PR PR-C PR PR Applications

Course Organization

- 2+2
- location FIT T9-346, exam UTIA CAS 471
- seminars - consulting hours (odd / even weeks) Ing. Radek Richtr, Ph.D.
- exam questions



© M. Haindl

MI-ROZ - 01

3/72

Outline - PR Basic Concepts

- 1 Organization
- 2 Content
- 3 Pattern Recognition
 - AI
 - PR Achievements
- 4 PR Applications
 - Literature
 - Notation
 - Notions
 - Representation
 - Classification



MI/MIE-ROZ 2023-2024

© M. Haindl

MI-ROZ - 01

2/72

PR PR-C PR PR Applications

Seminar work http://mosaic.utia.cas.cz/cvut2023_Z.html

Supervised melanoma classifier - FIT project

- individual implementation of parts of a classifier pipeline
- generic segmenter in Cygwin / (Linux)



© M. Haindl

MI-ROZ - 01

5/72

Supervised melanoma classifier - FIT project

- individual implementation of parts of a classifier pipeline
- generic segmenter in Cygwin / (Linux)
 - 1 image input
 - 2 melanoma detection
 - 3 textural features
 - 4 transformation
 - 5 segmenter
 - 6 multisegmenter
 - 7 postprocessing
 - 8 TM output
 - 9 verification



Institute of Information Theory and Automation



room 474



<http://www.utia.cas.cz/>



Supervised melanoma classifier - FIT project

- individual implementation of parts of a classifier pipeline
- generic segmenter in Cygwin / (Linux)
- programme the classifier (learning & classification) in C/C++
- individually improve on the selected classifier
- verify it on the PTSB server <http://mosaic.utia.cas.cz>
- write short report & presentation
- demonstration



Content

- Elements of pattern recognition.
- Basic PR concepts.
- Bayesian decision theory.
- Learning theory.
- Parametric & non-parametric classifiers.
- Context classifiers.
- Cluster analysis.
- Feature selection & extraction.
- Classification quality estimation.
- Structural & syntactic recognition.
- Neural networks.
- Genetic algorithms.
- Vector support machines.
- Applications.



Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization



Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model
object synthesis

- discriminative $p(\omega_i | X)$
- generative (descriptive) $p(X, \omega_i)$



honeysuckle



Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model



Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model
object synthesis



tomato harvesting



apple harvesting

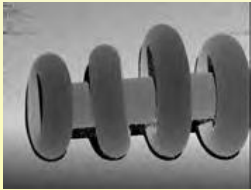


Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model
object synthesis

Analysis $\Rightarrow$

- supervised recognition,
- **unsupervised recognition**,
- inhomogeneity (edge) detection,
- defect detection,



Pattern Recognition & Modelling

Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model
object synthesis

Analysis

Modeling (Synthesis)

$\rightarrow$

- not identical (pixelwise)
- indiscernible (visually, aurally, tactually ...)



Pattern Recognition & Modelling

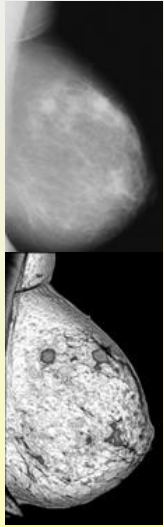
Pattern Recognition $\Rightarrow$ object categorization $\rightarrow$ statistical model
object synthesis

Analysis

Modeling (Synthesis)

$\Rightarrow$

- object multiplication / enlargement,
- reconstruction / recovery / restoration of unobservable data,
- **enhancement**,
- compression,
- artificial objects generation.



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data

agent based models, computer vision, robotics, gaming, automated reasoning, theorem proving, expert systems, natural language processing, machine learning, deep learning, neural nets, genetic algorithms, etc.



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

- Humans don't start out with skills; they start out with a broad ability to acquire new skills.
- Humans have innate ability to quickly learn (like what a smile means, what happens when you drop something ...).
- A self-driving car cannot intuit from common sense what will happen if a truck spills its load.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data

agent based models, computer vision, robotics, gaming, automated reasoning, theorem proving, expert systems, natural language processing, machine learning, deep learning, neural nets, genetic algorithms, etc.



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

- I AI a repertoire of fixed algorithms
- II AI limited recall to take future actions based on limited past observances
- III AI possess a social understanding of its surrounding and reacts accordingly
- IV AI conscious, self-aware capability, autonomous free will

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data

agent based models, computer vision, robotics, gaming, automated reasoning, theorem proving, expert systems, natural language processing, machine learning, deep learning, neural nets, genetic algorithms, etc.



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data

agent based models, computer vision, robotics, gaming, automated reasoning, theorem proving, expert systems, natural language processing, machine learning, deep learning, neural nets, genetic algorithms, etc.



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Notions & Synonyms

Pattern Recognition - objects categorization

Human Intelligence adaptability - power to generalize to never-seen-before situations.

Artificial Intelligence - a branch of computer science that is concerned with the automation of intelligent behavior (creativity & self-awareness - autonomous free will.)

Machine Learning - mathematical models parameters estimation

Data Mining - collecting, cleaning, processing, analyzing, and gaining useful insights from data

agent based models, computer vision, robotics, gaming, automated reasoning, theorem proving, expert systems, natural language processing, machine learning, deep learning, neural nets, genetic algorithms, etc.



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Pattern Recognition - Motivation

PR - identification of objects on the basis of their measurement patterns

human PR - sense stimuli are converted into meaningful perceptual experiences

process still largely unknown

- ① pattern is perceived by sense organs
- ② the same (similar) pattern have been previously perceived → meaningful perceptual experience
- ③ the past perception of the pattern must be remembered
- ④ some correspondence (equivalence) must be somehow established between past and present perception



Scene Understanding

The computer should be able to reason about all aspects of the scene and provide detailed pixel-level annotations describing the **semantics** and geometry of the scene.

- robot perception (autonomous driving, drone delivery, ...)
- surveillance and environmental monitoring
- content-based image search
- etc.



Scene Understanding

The computer should be able to reason about all aspects of the scene and provide detailed pixel-level annotations describing the **semantics** and geometry of the scene.

In the early years of computer-vision research an undergraduate student at MIT was asked to spend his summer getting a computer to describe what it saw in images obtained from a video camera.

Almost 60 years later researchers are still grappling with the same problem.



Scene Understanding

The computer should be able to reason about all aspects of the scene and provide detailed pixel-level annotations describing the **semantics** and geometry of the scene.

In the early years of computer-vision research an undergraduate student at MIT was asked to spend his summer getting a computer to describe what it saw in images obtained from a video camera.



Scene Understanding

- 2014 drone delivery available "Could it be four, five years? I think so." Amazon CEO Jeff Bezos
- 2015 "I almost view it (autonomous driving) as a solved problem," Tesla CEO Elon Musk
- 2018 Tesla's vehicles would "probably" be able to drive autonomously across the United States in "three to six months." Elon Musk
- 2019 The company expected to deliver packages by drone "within months." Amazon Worldwide Consumer CEO Jeff Wilke
- 2023 Tesla has yet to release software capable of handling the entire driving task (assisted driving only). Drone delivery still hasn't happened. The **human supervision of robots** is still necessary to prevent mistakes and help the robot out of confusing situations.



WHO Guidelines to AI in Healthcare (2022)

Critical regulatory factors concerning the use of AI in healthcare:

- 1 transparency and documentation (entire product lifecycle and tracking development processes),
- 2 risk management (intended use, continuous learning, human interventions, training models, cybersecurity),
- 3 validating data, clarity on the intended use of AI,
- 4 data quality,
- 5 privacy and data protection (GDPR - EU, HIPAA - USA),
- 6 collaboration (regulatory bodies, patients, healthcare professionals, industry).

AI requires an ecosystem of human skills and cannot replace human experience of **creativity and emotion**.



Pattern Recognition Course Goal

- **Understanding** - pattern recognition should not be viewed as a series of black boxes.
- **Trade-offs** - no single method will perform well in all possible applications.
- **Selection** - the most widely applicable PR methods.



Explainable AI

XAI:

- to correctly understand the decision making process (medical diagnostics, autonomous driving)
- to understand the key features that led the algorithm to make the respective decision and classification
- to make sure the classifier generalizes well



Pattern Recognition Course Goal

- **Understanding** - pattern recognition should not be viewed as a series of black boxes.
- **Trade-offs** - no single method will perform well in all possible applications.
- **Selection** - the most widely applicable PR methods.



Pattern Recognition Course Goal

- **Understanding** - pattern recognition should not be viewed as a series of black boxes.
- **Trade-offs** - no single method will perform well in all possible applications.
- **Selection** - the most widely applicable PR methods.



© M. Haindl

MI-ROZ - 01

19/72

PR PR-C PR PR Applications

AI PR Achievements

PR - Achievements

artificial PR - basic elements (sensors, A/D converters, memory, computer) are available
achievements

- OCR - mail sorting, documents, car template recognition



Machine Translation of Cuneiform

(Sumerian and Akkadian) Languages 3000 BC, 500,000 preserved cuneiform tablets - many untranslated (deciphered 1857), 40,000 British Museum - Electronic Babylonian Literature project

- visual inspection - woven materials, paper / steel mills, food, notes



© M. Haindl

MI-ROZ - 01

20/72

PR PR-C PR PR Applications

AI PR Achievements

PR - Achievements

artificial PR - basic elements (sensors, A/D converters, memory, computer) are available
achievements

- OCR - mail sorting, documents, car template recognition



show-through removal



Miroslav Gospel - 1186



Codex Gigas - early 13th century



- visual inspection - woven materials, paper / steel mills, food, notes

© M. Haindl

MI-ROZ - 01

20/72

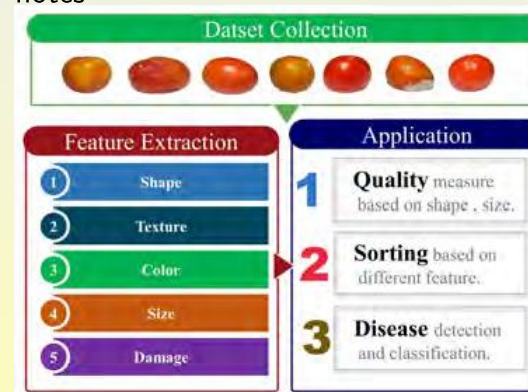
PR PR-C PR PR Applications

AI PR Achievements

PR - Achievements

achievements

- OCR - mail sorting, documents, car template recognition
- visual inspection - woven materials, paper / steel mills, food, notes



tomato quality-based grading

- driving assistance - road sign recognition, approaching cars, notes



© M. Haindl

MI-ROZ - 01

20/72

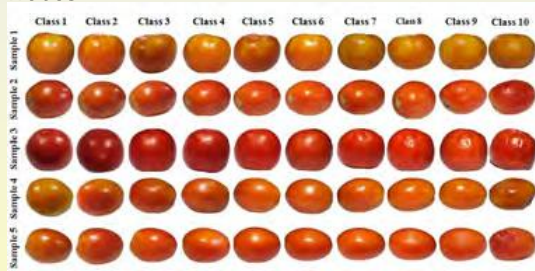
PR PR-C PR PR Applications

AI PR Achievements

PR - Achievements

achievements

- OCR - mail sorting, documents, car template recognition
- visual inspection - woven materials, paper / steel mills, food, notes



tomato quality-based grading

- driving assistance - road sign recognition, approaching cars, lane detection
- medical signal analysis - EKG, EEG, blood cells counting, mammography, dermatology, DNA
- speech recognition (spoken words, speaker identification, ChatGPT-4)



PR - Achievements

achievements

- OCR - mail sorting, documents, car template recognition
- visual inspection - woven materials, paper / steel mills, food, notes
- driving assistance - road sign recognition, approaching cars, lane detection
- medical signal analysis - EKG, EEG, blood cells counting, mammography, dermatology, DNA
- speech recognition (spoken words, speaker identification, ChatGPT-4)
- security - fingerprint, face, DNA, iris, gait recognition, vein, spoofing detection



PR - Achievements

achievements

- OCR - mail sorting, documents, car template recognition
- visual inspection - woven materials, paper / steel mills, food, notes
- driving assistance - road sign recognition, approaching cars, lane detection
- medical signal analysis - EKG, EEG, blood cells counting, mammography, dermatology, DNA
- speech recognition (spoken words, speaker identification, ChatGPT-4)
- security - fingerprint, face, DNA, iris, gait recognition, vein, spoofing detection



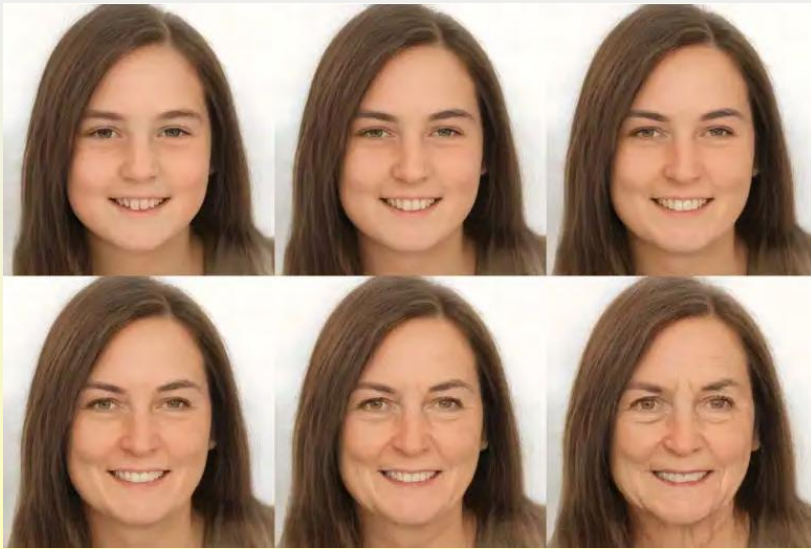
Face Recognition

photos have to be captured under controlled conditions:

- well-lit environment, normalized intensity and spectrum
- upright head holding
- looking into the camera
- no occlusions
- neutral expression
- no aging - similar age with the stored database photo
- age, race or sex variable accuracy - black females are 10 times more likely to be misidentified than white males (NIST)



Face Recognition



Face recognition - unreliable after 5 years from captured images



History

- ♣ PR problem formulation
C.K. Chow, "An optimum character recognition system using decision functions", IRE Trans. Elec. Comp., vol. 6, Dec. 1957.
- ♣ discriminant analysis
R.A. Fisher, "The use of multiple measurements in taxonomic problems", Ann. Eugenics, vol. 7, 1936.
- ♣ statistical decision theory
A. Wald, "Statistical Decision Functions", Wiley, 1950.
- ♣ 1960 Frank Rosenblatt at Cornell Univ. builds a computer - the Perceptron - that can learn by trial and error through a neural network
- ♣ feature selection
G. Sebestyen, "Decision Making Processes in Pattern Recognition", Macmillan, 1962



Face Recognition

- Janus 2014 - 2020 funded by Intelligence Advanced Research Projects Agency (IARPA), unconstrained face recognition (in the wild)
- Horus facial recognition system (doubled accuracy), FBI & Pentagon 640 M face photos worldwide



History

- ♣ 1962 L. Roberts MIT, the first computer vision system
- ♣ 1970 Shakey from SRI International is the first robot to use AI to navigate
- ♣ 1975 HWIM (hear what I mean) Harpy, Hearsay - the first large vocabulary (1000 words) speech system



IAPR International Association for Pattern Recognition
<http://www.iapr.org>

Books:

- Chen, C.H. Pau, L.F. Wang, P.S.P. ed. Handbook of PR & Comp. Vision, World Scientific, 93, ISBN 981-02-1136-8.
- Devijver, P.A., Kittler, J. Pattern Recognition: A Statistical Approach, 1982
- Duda, R.O., Hart, P.E., Stork, D.G. Pattern Classification, Wiley, 2000, ISBN 978-0-471-05669-0.
- Fu, K.S., Syntactic Methods in Pattern Recognition, Academic Press, 1974, ISBN 0-12-269560-7.
- Fukunaga, K. Introduction to Statistical Pattern Recognition, Academic Press, 1972



Pattern Classification



Pattern Classification,
 2nd Edition, J. Wiley
 R. O. Duda, P. E. Hart, D. G. Stork
 ISBN: 978-0-471-05669-0, 680 pages, November 2000

- Bayesian Decision Theory.
- Maximum-Likelihood and Bayesian Parameter Estimation.
- Nonparametric Techniques.
- Linear Discriminant Functions.
- Multilayer Neural Networks.
- Stochastic Methods.
- Nonmetric Methods.
- Algorithm-Independent Machine Learning.
- Unsupervised Learning and Clustering.



- Anderberg M.R. Cluster Analysis for Applications, Academic Press, 1973.
- MacLachlan, G. J. Discriminant analysis and statistical pattern recognition, Wiley 1992.
- Mitchel M. An Introduction to Genetic Algorithms, Bradford Book, 1996.
- Muller, B. J. Reinhardt Neural Networks - An Introduction, Springer, 1991.
- Haykin, S. Neural Networks A Comprehensive Foundation, Macmillan College Publishing Company, 1994, ISBN 0-02-352761-7.
- Sankar K.P. ed. Genetic Alg. for Pattern Rec., CRC Press, 1996.
- Young, T. Y. Handbook of Pattern Recognition and Image Processing, Academic Press, 1986, ISBN 0-12-774560-2.
- Kotek, Z., Chalupa, V., Brůha, I., Jelínek, J. Adaptivní a učící



Pattern Recognition



Pattern Recognition
 4th Edition, Academic Press
 S. Theodoridis and K. Koutroumbas
 984 pages, 2009

- 1 Introduction
- 2 Classifiers based on Bayes Decision
- 3 Linear Classifiers
- 4 Nonlinear Classifiers
- 5 Feature Selection
- 6 Feature Generation I: Data Transformation and Dimensionality Reduction
- 7 Feature Generation II
- 8 Template Matching

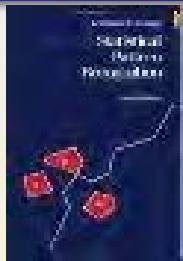


Pattern Recognition

- 9 Context Dependant Clarification
- 10 System Evaluation
- 11 Clustering: Basic Concepts
- 12 Clustering Algorithms: Algorithms L Sequential
- 13 Clustering Algorithms II: Hierarchical
- 14 Clustering Algorithms III: Based on Function Optimization
- 15 Clustering Algorithms IV: Clustering
- 16 Cluster Validity



Statistical Pattern Recognition

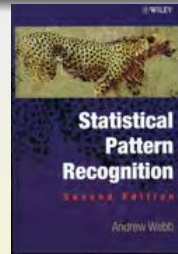


Statistical Pattern Recognition
2nd Edition, A. Wiley
K. Fukunaga
496 pages, 2004

- 1 Introduction
- 2 Random vectors and their properties
- 3 Hypothesis testing
- 4 Parametric classifiers
- 5 Parameter estimation
- 6 Nonparametric density estimation
- 7 Nonparametric classification and error estimation
- 8 Successive parameter estimation
- 9 Feature extraction and linear mapping for signal representation
- 10 Feature extraction and linear mapping for classification



Statistical Pattern Recognition

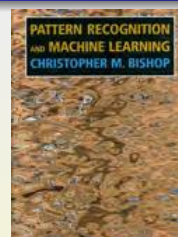


Statistical Pattern Recognition
3th Edition, J. Wiley
A. Webb
496 pages, 2004

- 1 Introduction to statistical pattern recognition
- 2 Density estimation - parametric
- 3 Density estimation - nonparametric
- 4 Linear discriminant analysis
- 5 Nonlinear discriminant analysis -kernel methods
- 6 Nonlinear discriminant analysis - projection methods
- 7 Tree-based methods
- 8 Performance
- 9 Feature selection and extraction
- 10 Clustering



Pattern Recognition and Machine Learning

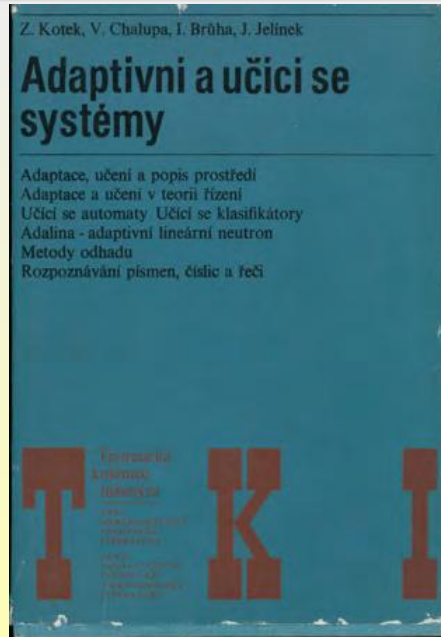
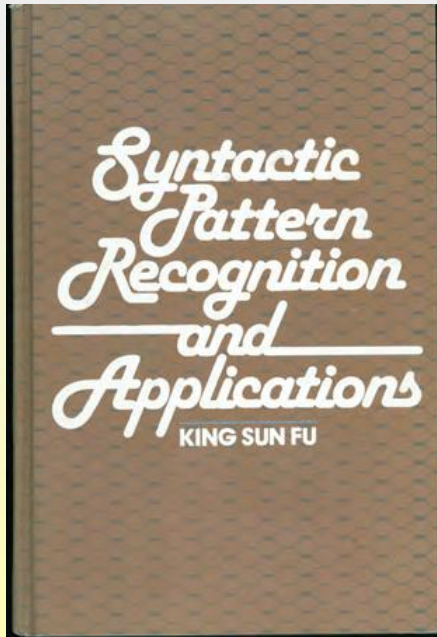


Pattern Recognition and Machine Learning
1st Edition, Springer
C.M. Bishop
738 pages, 2006

- 1 Introduction
- 2 Probability distribution
- 3 Linear models for regression
- 4 Linear models for for classification
- 5 Neural networks
- 6 Kernel methods
- 7 Sparse kernel machines
- 8 Graphical models
- 9 Mixture models and EM
- 10 Approximate inference
- 11 Sampling methods
- 12 Continuous latent variables
- 13 Sequential data
- 14 Combining models



Pattern Recognition



© M. Haindl

MI-ROZ - 01

43/72

PR PR-C PR PR Applications

Literature Notation Notions Representation Classification

Literature

Journals:

- IEEE T. on Pattern Analysis and Machine Intelligence
- IEEE T. on Information Theory
- IEEE T. on Image Processing
- IEEE T. on Medical Imaging
- IEEE T. on Geoscience and Remote Sensing
- Pattern Recognition
- Pattern Recognition Letters
- Computer Vision, Graphics and Image Processing
- International Journal of Pattern Recognition and Artificial Intelligence

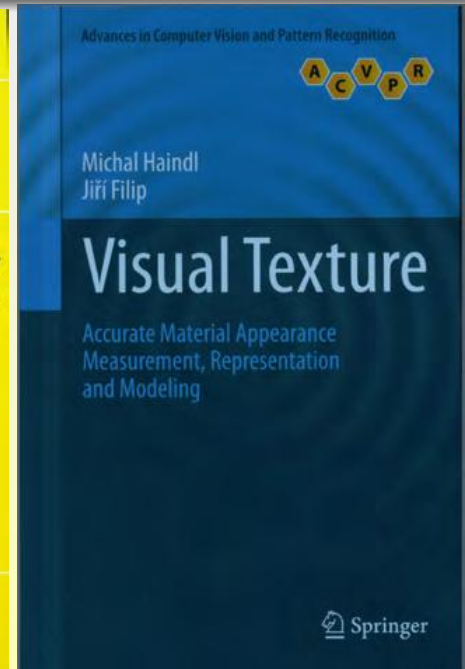
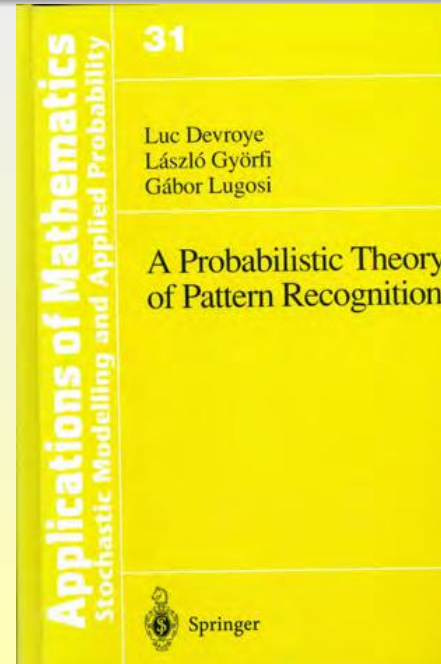


© M. Haindl

MI-ROZ - 01

45/72

Pattern Recognition



© M. Haindl

MI-ROZ - 01

44/72

PR PR-C PR PR Applications

Literature Notation Notions Representation Classification

Notation

- $\mathcal{S}$ pattern space
- X feature vector $X = [x_1, \dots, x_I]$
- $I = \dim\{X\}$ number of features
- $\mathcal{X}$ feature space
- K number of classes
- ω_i class indicator
- $\Omega = \{\omega_1, \dots, \omega_K\}$
- $g(X)$ discriminant function
- H decision boundary
- $n_i = \text{card}\{T_i\}$
- T_i training set for class i
- $\tilde{T}_i$ test set for class i
- μ_i mean value
- θ_i $p(X|\omega_i)$ parameters
- Σ_i covariance matrix



© M. Haindl

MI-ROZ - 01

46/72

Pattern Recognition Notions

♠ set of patterns $S = \{p_1, p_2, \dots\}$ (pattern space)

♠ pattern recognition

$$S = \bigcup_{k=1}^K S_k$$

$$S_i \cap S_j = \emptyset \quad i \neq j$$

♠ classification - the assignment of an object (pattern) to one of several prespecified categories $\Omega = \{\omega_1, \dots, \omega_K\}$

- Repeated observation of the same pattern should produce the same class.
- Two different pattern should give rise to two different classes.
- A slight distortion of a pattern should produce a small displacement of its representation.



Notions 3

♠ discriminant function - a scalar function $g(X)$, whose domain is usually measurement space and whose range is usually real numbers selection of

$$g(X) : \forall X \in \omega_i \quad g_i(X) > g_j(X), \quad j = 1, \dots, K; j \neq i$$

assign X into class ω_j if

$$g_j(X) = \max_k \{g_k(X)\}$$



Notions 2

♠ classifier - a set of decision rules that partition a feature space into disjoint subspaces (sorts patterns into categories or classes)

- sequential classifier - K class problem solves as $K - 1$ two-class problems
- hierarchical classifier - decision rule in a tree form, each terminal node contains the assigned class
- parallel classifier - K class problem solves as $K(K - 1)/2$ -class problems

♠ separable classes $\Leftrightarrow$ their class region do not overlap

♠ decision rule - assigns one class on the basis of the unit features



Notions 3

♠ discriminant function - a scalar function $g(X)$, whose domain is usually measurement space and whose range is usually real numbers selection of

$$g(X) : \forall X \in \omega_i \quad g_i(X) > g_j(X), \quad j = 1, \dots, K; j \neq i$$

assign X into class ω_j if

$$g_j(X) = \max_k \{g_k(X)\}$$

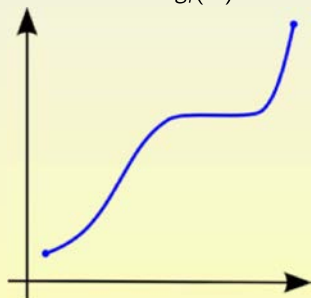


Discriminant Function

- ♠ discriminant function is not unique
(multiply or add $C > 0$, replace by $f(g_i(X))$)

$f()$ a monotonically increasing function

$$\frac{\partial f(g_i(X))}{\partial g_i(X)} > 0$$



$$f(g_i(X)) : \log(g_i(X)), \sqrt{g_i(X)}, a^{g_i(X)} \quad a > 1,$$



Discriminant Function

- ♠ discriminant function is not unique
- ♠ linear discriminant function

$$g(X) = a_0 + \sum_{j=1}^l a_j x_j$$



Discriminant Function

- ♠ discriminant function is not unique
e.g., identical minimum-error rate discriminant functions

$$g_j(X) = P(\omega_j|X)$$

$$g_j(X) = \frac{p(X|\omega_j)P(\omega_j)}{\sum_{i=1}^K p(X|\omega_i)P(\omega_i)}$$

$$g_j(X) = p(X|\omega_j)P(\omega_j)$$

$$g_j(X) = \log p(X|\omega_j) + \log P(\omega_j)$$



Notions 4

- ♠ decision boundary between ω_i, ω_j
 $H = \{X \in M : g_i(X) = g_j(X)\}$
classification is not uniquely defined
- ♠ hyperplane decision boundary - decision boundary for linear discriminant function (linear decision rule)

$$(a_{i,0} - a_{j,0}) + \sum_{k=1}^l (a_{i,k} - a_{j,k})x_k = 0$$

- ♠ non-linear decision boundary

- ♠ piecewise linear

$$g_j(X) = \max_{i=1, \dots, n_j}^i g_j(X)$$



Notions 5

♠ quadratic discriminant function

$$g(X) = a_0 + \sum_{i=1}^I \sum_{j=i}^I a_{ij} x_i x_j + \sum_{j=1}^I a_j x_j$$

♠ general discriminant function - two equivalent options

- ① non-linear discriminant function e.g.
 $g(X) = a_1 \ln(x_1) + a_2 x_2^3$ (+ linearisation)
- ② non-linear mapping $\Phi : S^n \rightarrow S^m$ combined with linear classifier

$$\Phi(X) = [\Phi_1(X), \dots, \Phi_m]$$

$$g_j(X) = \sum_{k=1}^m a_{j,k} \Phi_k(X)$$



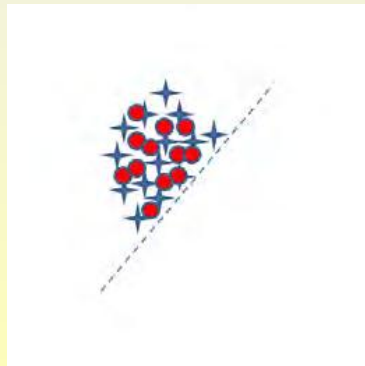
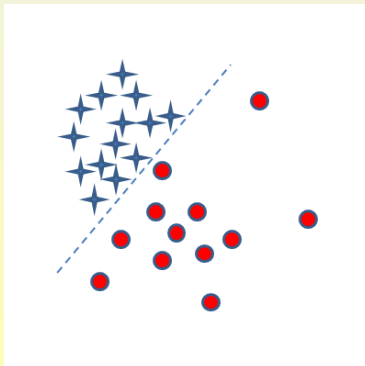
Notions 7

♠ feature vector - functions of the initial measurement variables of an object (pattern), or some subset of the initial measurement pattern variables, input for classifier

$$X = [x_1, \dots, x_I]$$

good features

bad features

♠ feature space $\{X\}$

Notions 6

♠ dichotomy decision rule - $K = 2$

$$g(X) = g_1(X) - g_2(X)$$

- ♠ supervised classification - $T = \cup_{k=1}^K T_k$ training set partition known
- ♠ unsupervised classification - T training set partition unknown
 - number of classes K known
 - number of classes K unknown



Notions 7

♠ feature vector - functions of the initial measurement variables of an object (pattern), or some subset of the initial measurement pattern variables, input for classifier

$$X = [x_1, \dots, x_I]$$

♠ feature space $\{X\}$

- ♠ feature selection - determination of the most discriminative pattern measurements (features),
- ♠ feature extraction - a mapping from the original I dimensional measurement space into the $\tilde{I}$ dimensional feature space $\tilde{I} < I$



Notions 8

- ♠ ground truth - known classification of some patterns
- ♠ training (supervised) set - $T = \{(X_j, \omega_j)\}$
- ♠ test set - $\tilde{T} = \{(X_j, \omega_j)\}$
- ♠ a classifier is learning if its iterative training procedure increases the classification performance accuracy after each few iterations
- ♠ parametric learning - discriminant function (conditional density $p(X|\omega_i)$) is assumed to be known except for some unknown parameters e.g. multivariate normal density

$$p(X|\omega_i) = \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i)\right\}$$



Notions 10

- ♣ statistical PR - the measurement patterns have the form of a random vector realization, each component is a measurement of a particular quality, property, or characteristic of the unit
- ♣ syntactic PR - the measurement patterns have the form of sentences from the language of a phrase structure grammar
- ♣ structural PR - the measured unit is encoded in terms of its parts, their mutual relationships and their properties



Notions 9

- ♠ nonparametric learning - no special functional form of the conditional probability distribution $p(X|\omega_i)$ is assumed (distribution-free), e.g. the nearest-neighbour rule, LS fit to an empirical cumulative distribution,
- ♠ supervised learning - learning from labelled training set $T = \{(X_j, \omega_j)\}$
- ♠ unsupervised learning (learning without a teacher) - learning from unlabeled training set $T = \{(X_j)\}$
- ♠ semi-supervised learning $T = \{(X_j, \omega_j), (X_k)\}$
- ♠ adaptive learning - learning in changing environment



Representation

Environment

- ♠ system - an object with inputs and outputs
 - dynamic system

$$u(t) \rightarrow \square \rightarrow y(t)$$

$$\dot{x}(t) = F_1(x(t), u(t)) \quad \text{differential state eq.}$$

$$y(t) = F_2(x(t), u(t)) \quad \text{alg. output eq.}$$

- automaton - a system with countable number of states
- ♠ pattern - an object where I/O are meaningless
- ♠ identification - a mathematical model building
- ♠ identifier - e.g. an adaptive / learning system
- ♠ system control - appropriate input signals to guarantee required output
- ♠ controller - (regulator)
- ♠ pattern recognition $pattern \rightarrow \square^{repr} \cdot X \rightarrow \square^{class.} \rightarrow \omega_i$



Representation

object representation

- featured - numerical $X = [1.2, 23.34, 12.10^8, 0.231]$
- syntactic (structural) - symbolic $X = [\clubsuit, \heartsuit, \diamondsuit, \spadesuit, \text{X}, \blacksquare]$

Featured Representation

feature vector (given object measurements, an object image)

$$X = [x_1, \dots, x_l]^T$$

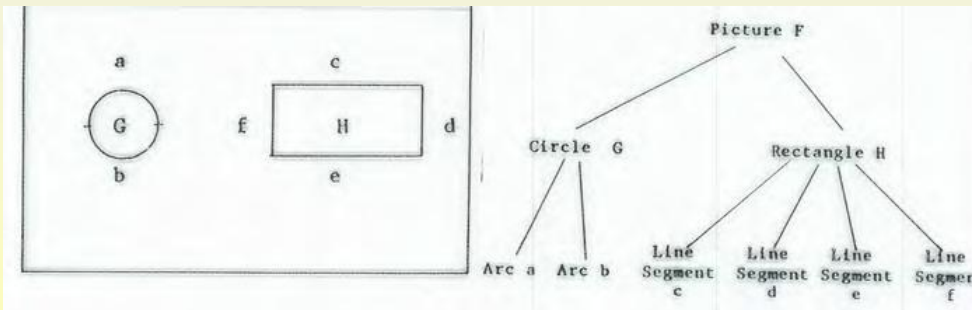
$$X \in \mathbb{R}^l$$



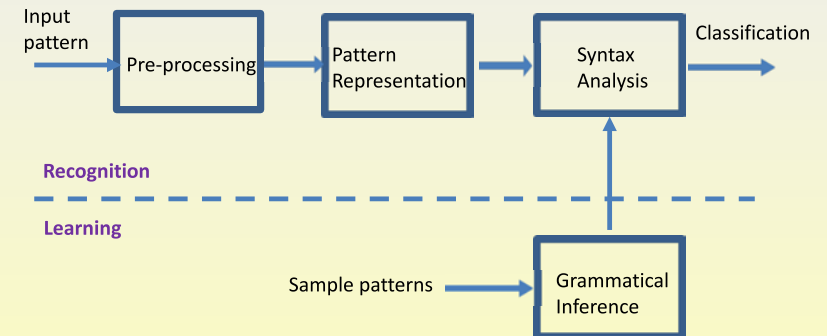
Syntactic Representation

primitive - an elementary subpattern, cannot be further divided

relational graph - consists from several related hierarchical trees

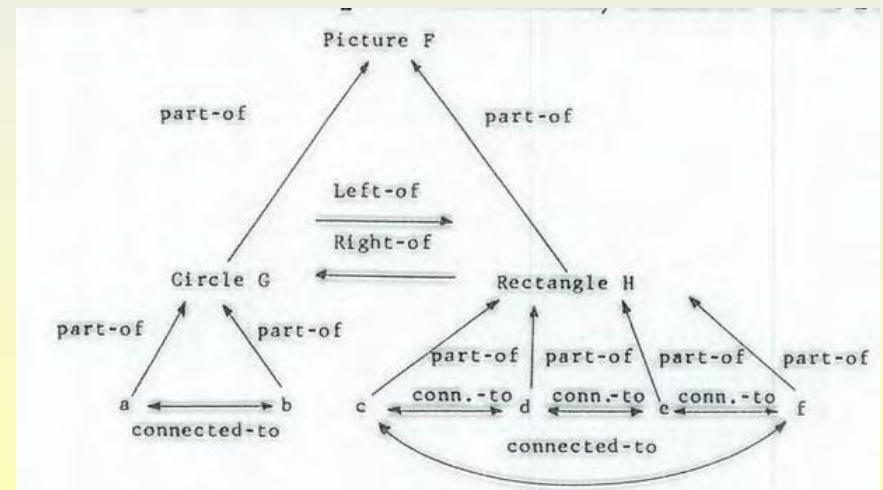


Syntactic pattern recognition system



Syntactic Representation 2

tree - hierarchical pattern representation, subset of a relational graph



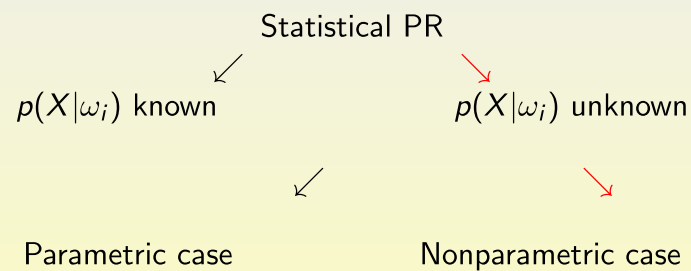
Syntactic Representation 3

formal language - analogous representation

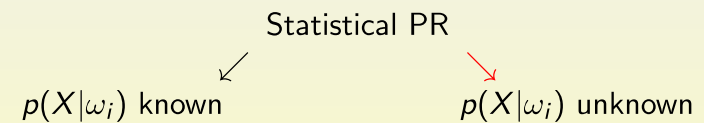
- terminal symbol - an elementary primitive property
- word (terminal symbol chain) - pattern representation
- formal language - set of patterns, not unique for a given pattern set
- substitution rules - rules how to generate words from terminal symbols
- grammar - substitution rules + set of terminal symbols, usually recursive
- $\omega_i \longleftrightarrow$ grammar
- syntactic analysis (pattern recognition) - the assignment of an object (word X) to one of several prespecified grammars



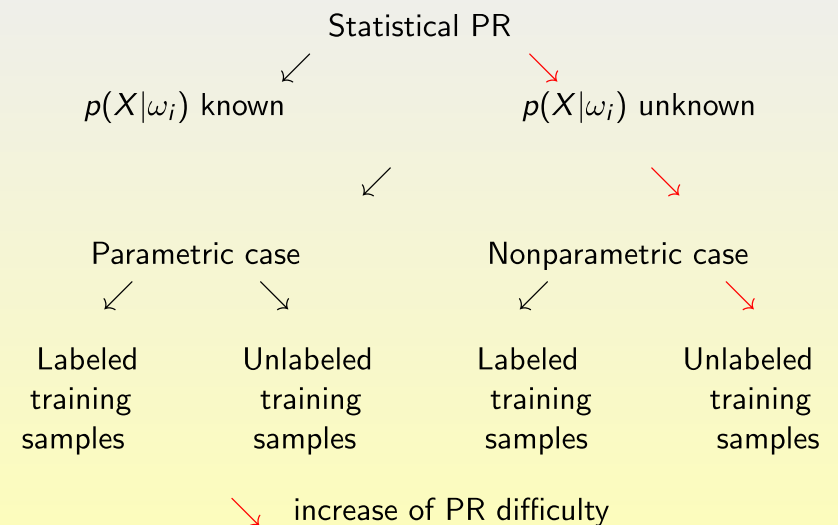
Statistical PR



Statistical PR

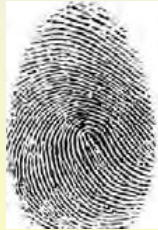


Statistical PR



Classification

- 1 data capturing
- 2 preprocessing (data normalisation)
- 3 training & test sets selection
- 4 feature selection
- 5 classification
- 6 postprocessing - thematic map filtering
- 7 probability of error estimation
- 8 interpretation



Supervised Classification

- statistical - based on decision theory, assumed knowledge of cpdf $p(X|\omega_j)$
- deterministic - no assumptions about cpdf, discrimination function deterministic

steps:

- 1 determination of the number of classes of interest K
- 2 training set selection and its partition into single class subsets $T = \cup_{k=1}^K T_k$
- 3 determination of a classifier
- 4 feature selection
- 5 estimation of classifier parameters from training set data
- 6 classification of new patterns



Supervised Classification

- statistical - based on decision theory, assumed knowledge of cpdf $p(X|\omega_j)$
- deterministic - no assumptions about cpdf, discrimination function deterministic

steps:

- 1 determination of the number of classes of interest K
- 2 training set selection and its partition into single class subsets $T = \cup_{k=1}^K T_k$
- 3 determination of a classifier
- 4 feature selection
- 5 estimation of classifier parameters from training set data
- 6 classification of new patterns



Deterministic Classification

- thresholding
- geometric classification

problem - determination of decision hypersurfaces

- ⊕ knowledge of $p(X|\omega_j)$ not required
- ⊕ easy implementation
- ⊖ determination of decision hypersurfaces, feasible solution only for hyperplanes



Geometric Classification

linear discriminant function, w_0 the threshold weight
 $(-w_0 < W^T X \Rightarrow S_1)$

$$g(X) = w_0 + W^T X = a^T y$$

$K = 2$ $g(X) > 0 \Rightarrow X \in S_1$ otherwise $X \in S_2$
 $g(X) = 0$ defines the decision surface, $y = [1, x_1, \dots, x_n]^T$



Geometric Classification

linear discriminant function, w_0 the threshold weight
 $(-w_0 < W^T X \Rightarrow S_1)$

$$g(X) = w_0 + W^T X = a^T y$$

$K = 2$ $g(X) > 0 \Rightarrow X \in S_1$ otherwise $X \in S_2$

minimization of the criterion function:

$$J(a, b) = \frac{1}{2} \|Ya - b\|^2$$

$$Y = [1^1 y_1, \dots, -2^1 y_1, \dots] \quad m \times (n+1)$$

subject to the constraint $b > 0$



Geometric Classification

linear discriminant function, w_0 the threshold weight
 $(-w_0 < W^T X \Rightarrow S_1)$

$$g(X) = w_0 + W^T X = a^T y$$

$K = 2$ $g(X) > 0 \Rightarrow X \in S_1$ otherwise $X \in S_2$

$$\{1^1 y_j : 1^1 y_j \in T_1\}, \quad \{2^2 y_j : 2^2 y_j \in T_2\}$$

learning - to find $a : a^T y_j > 0 \quad \forall 1^1 y_j, 2^2 y_j \quad (a^T (-2^2 y_j) > 0)$
 not unique solution, a margin has to be introduced to avoid
 convergence to a limit point on the boundary, i.e., $a^T y_j \geq b > 0$



Gradient Descent Procedure

$$\nabla_a J = Y^T (Ya - b)$$

$$\nabla_b J = -(Ya - b) \quad \text{for a given } b \quad a = (Y^T Y)^{-1} Y^T b$$

Ho - Kashyap Algorithm (determination of b)

- 1 $b(0) > 0$ otherwise arbitrary
- 2 $a(i) = (Y^T Y)^{-1} Y^T b(i)$
- 3 $b(i+1) = b(i) + \rho [e(i) + |e(i)|]$

$$e(i) = -\nabla_b J(i) = Ya(i) - b(i) \quad 0 < \rho < 1$$

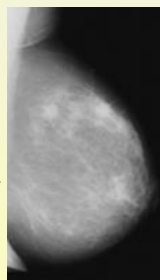
⊕ convergence in a finite number of steps for linearly separable training samples



Unsupervised Classification

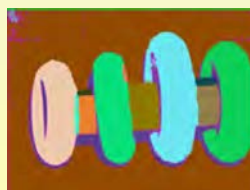
training set partition unknown

- statistical - estimation of unknown quantities in mixture density $p(X) = \sum_{k=1}^K p(\omega_k)p(X|\omega_k)$
from T some of $\{K, p(\omega_k)p(X|\omega_k), k = 1, \dots, K\}$ may be unknown
 ⊖ no general solution known
 for known $\{K, p(\omega_k)\}$ and the form of $p(X|\omega_k)$ solution exists
- deterministic - no assumptions about pdf, discrimination function deterministic based on similarity measures
 - clustering



Unsupervised Classification

- (determination of the number of classes of interest K)
- training set selection, its partition into single class subsets $T = \cup_{k=1}^K T_k$ unknown
- determination of a classifier
- feature selection
- estimation of classifier parameters from training set data
- classification of new patterns



Unsupervised Classification

- (determination of the number of classes of interest K)
- training set selection, its partition into single class subsets $T = \cup_{k=1}^K T_k$ unknown
- determination of a classifier
- feature selection
- estimation of classifier parameters from training set data
- classification of new patterns



Information

<http://mosaic.utia.cas.cz/>
<http://ro.utia.cz>

