Introduction to Machine Learning

Diego H. Peluffo-Ordóñez

dhpeluffo@yachaytech.edu.ec

www.diegopeluffo.com

www.sdas-group.com

May, 2018

CONTEXT
AI VS NI

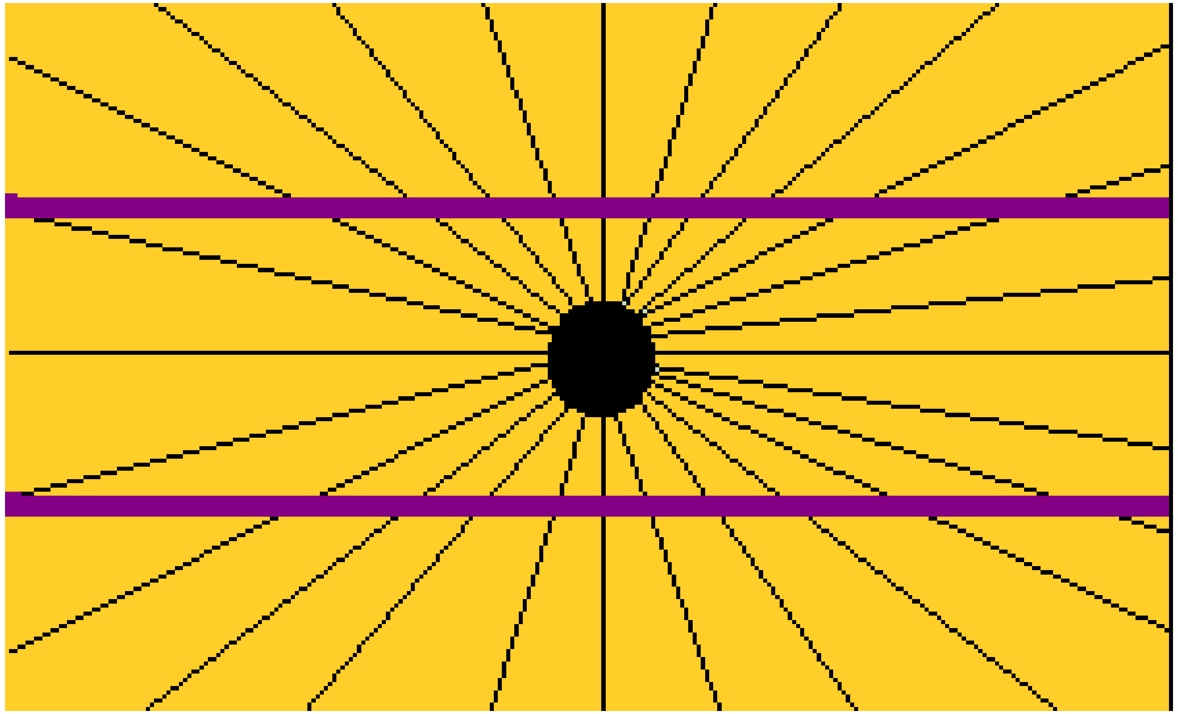
Artificial and natural intelligence

Easier for human

Perspective

Artificial and natural intelligence (2)

Are the purple lines inclined or angled?



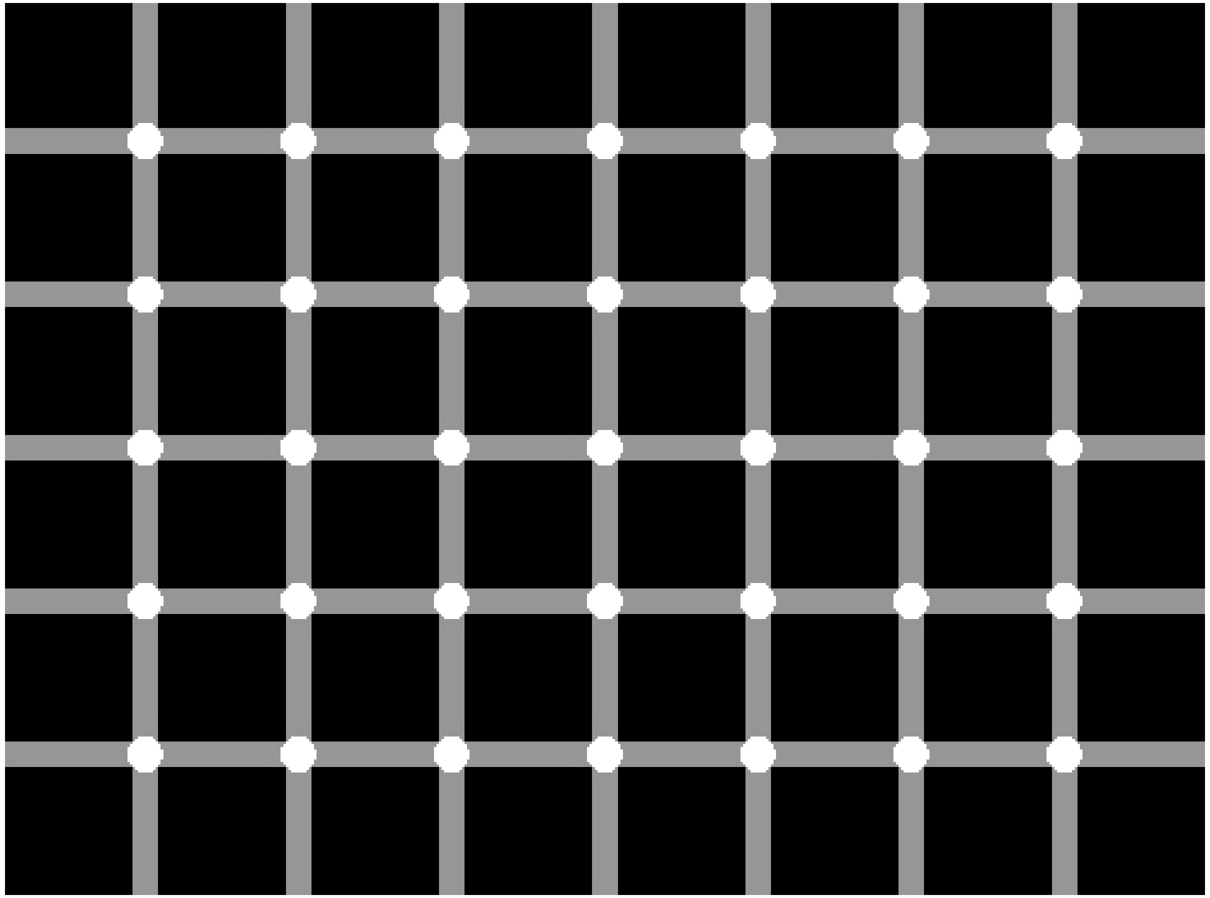
Easier for machine

Artificial and natural intelligence (3)

Easy for both

How many objects are in the picture?

Artificial and natural intelligence (4)



Easier for machine

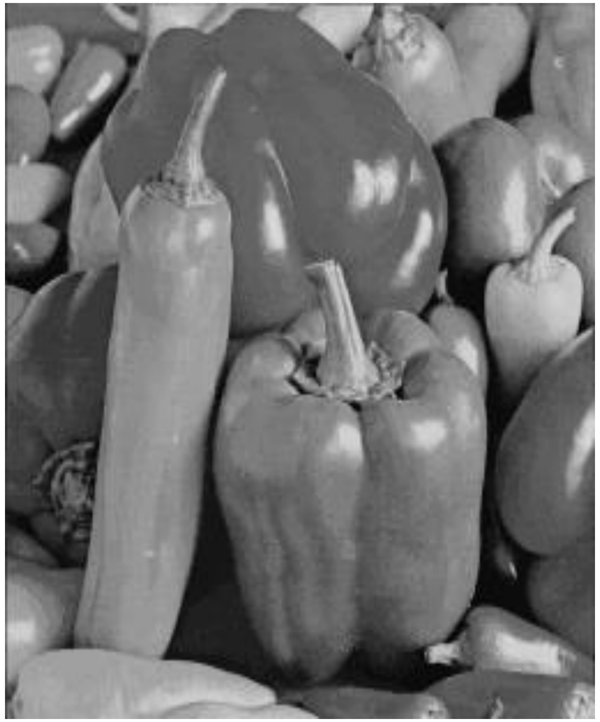
How many black points?

Artificial and natural intelligence (5)

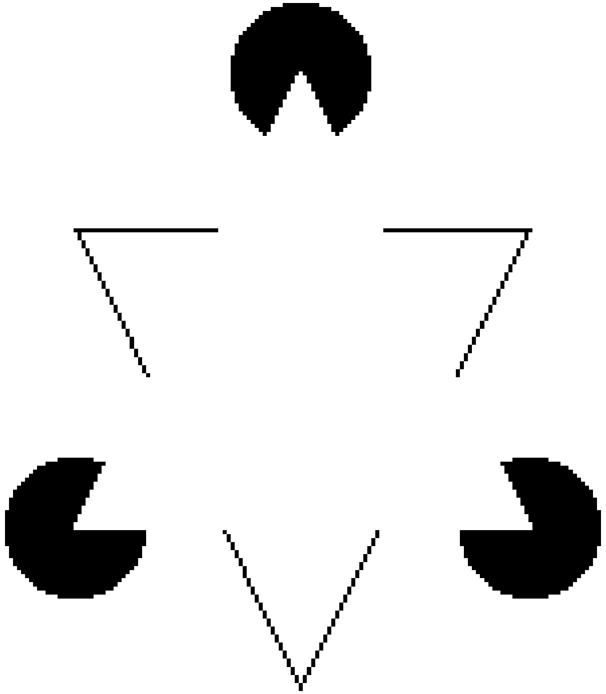
Perception is so important…

Artificial and natural intelligence (6)

How many faces?

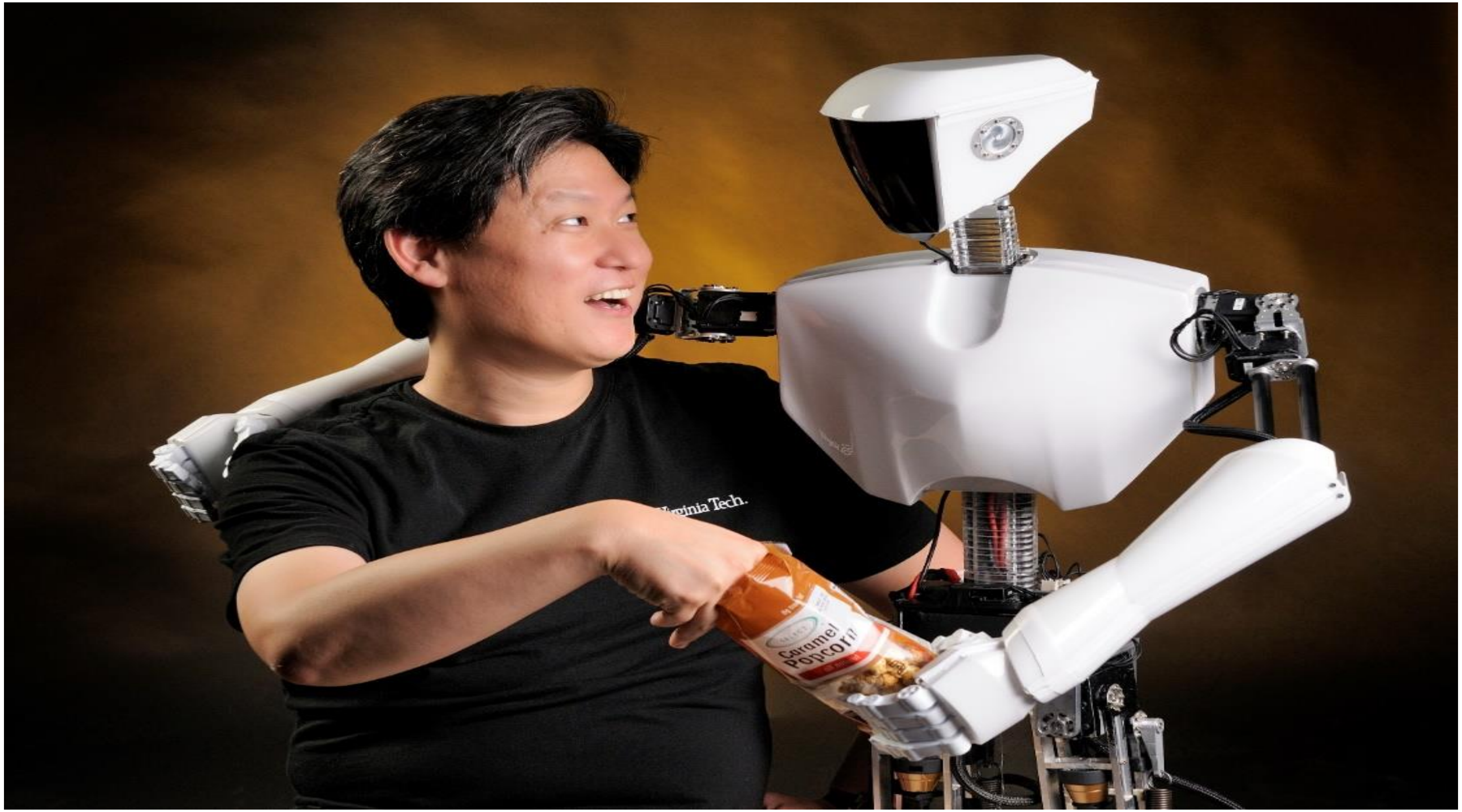


What is the picture about?



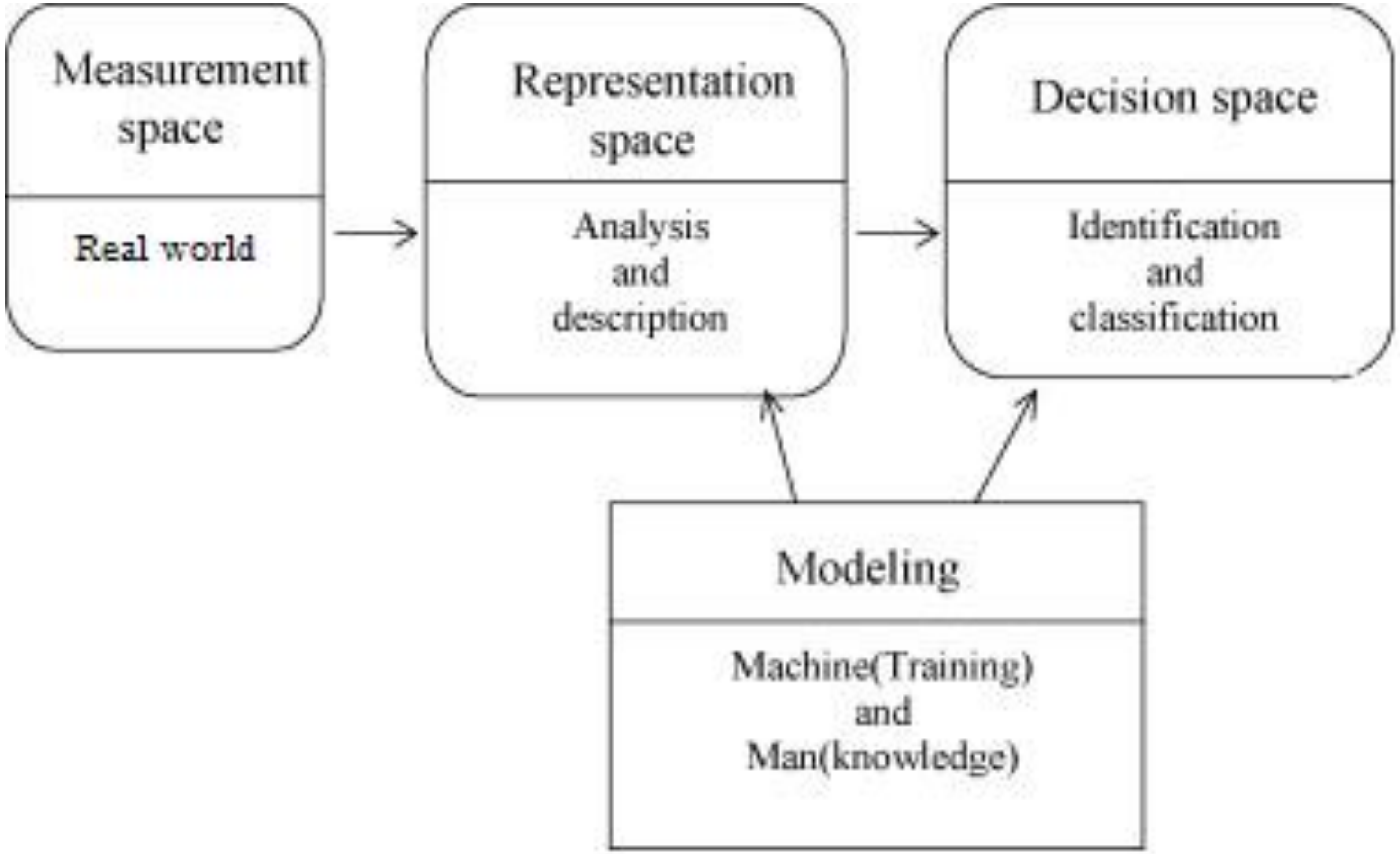
Is there a triangle?

Integration

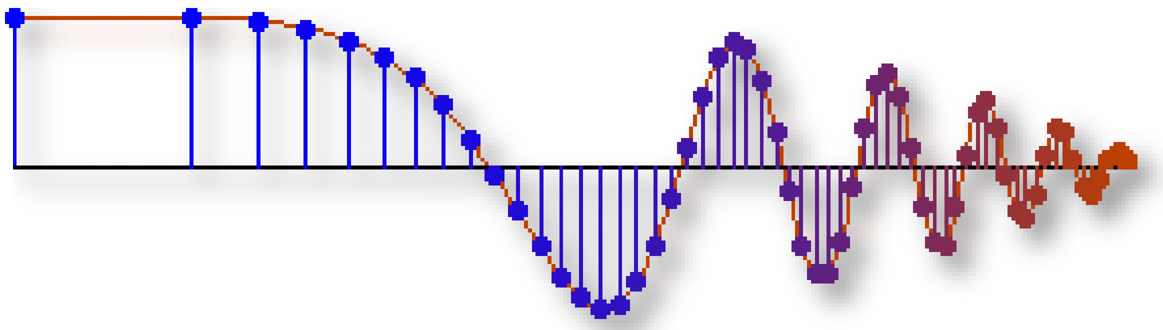


Machine Learning

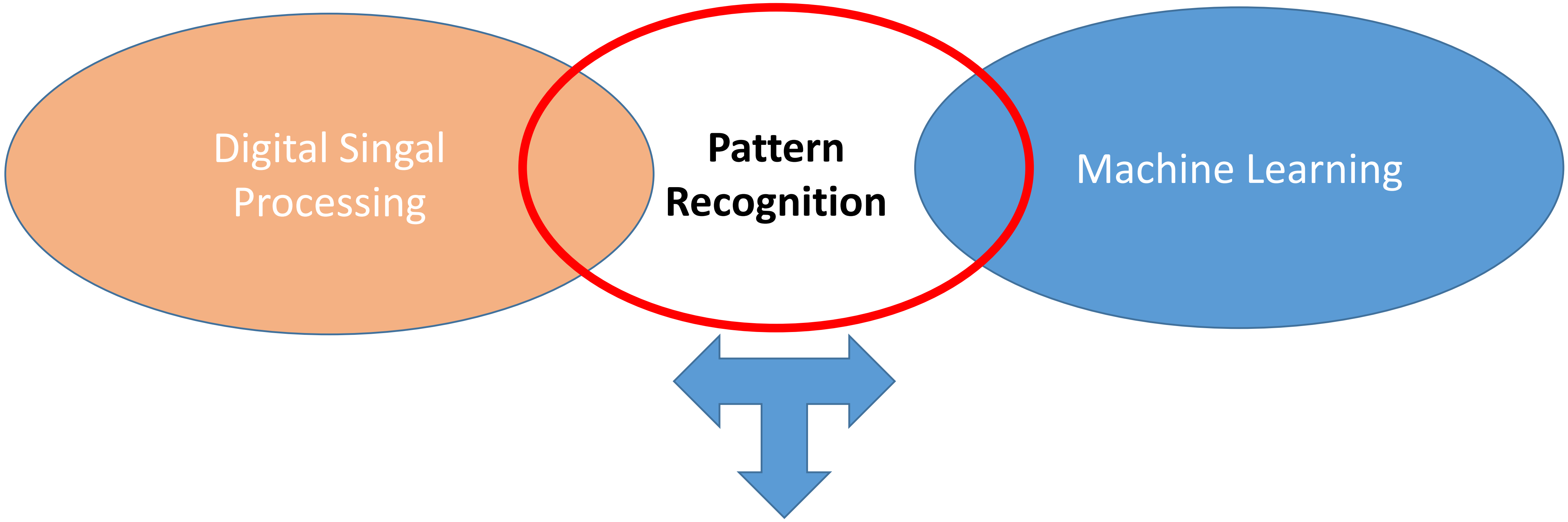
Machine Learning & Data Processing



Signal processing



Machine learning

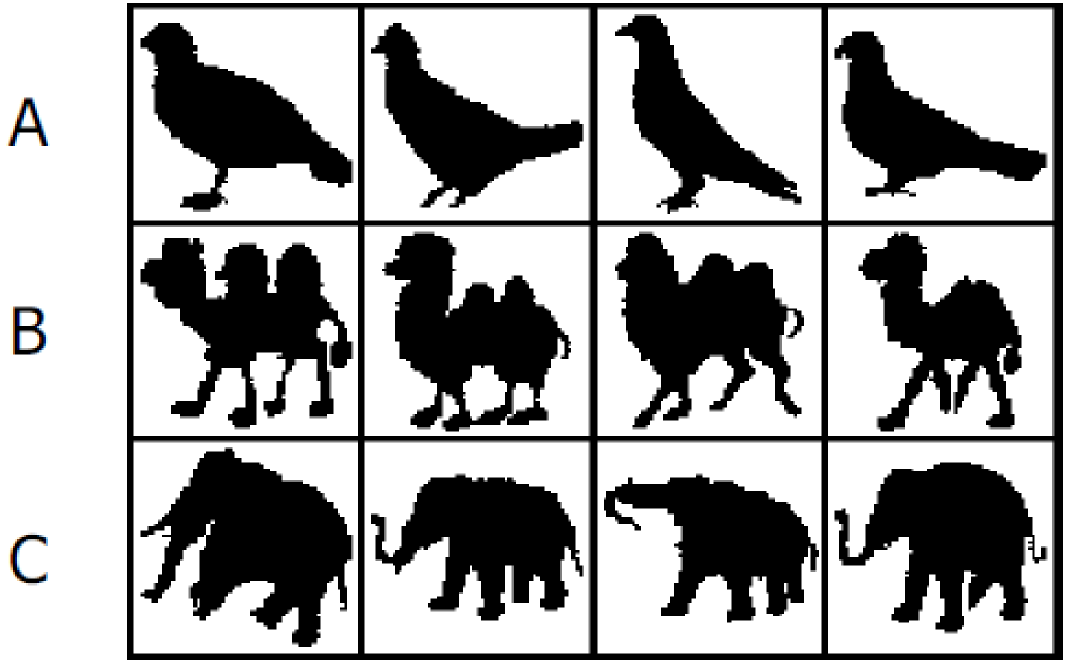

Digital Singal Processing
Pattern Recognition
Machine Learning

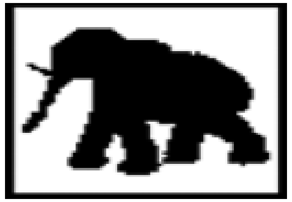
KNOWLEDGE

Pattern recognition

What is this? What occasion? Where are the faces? Who is who?

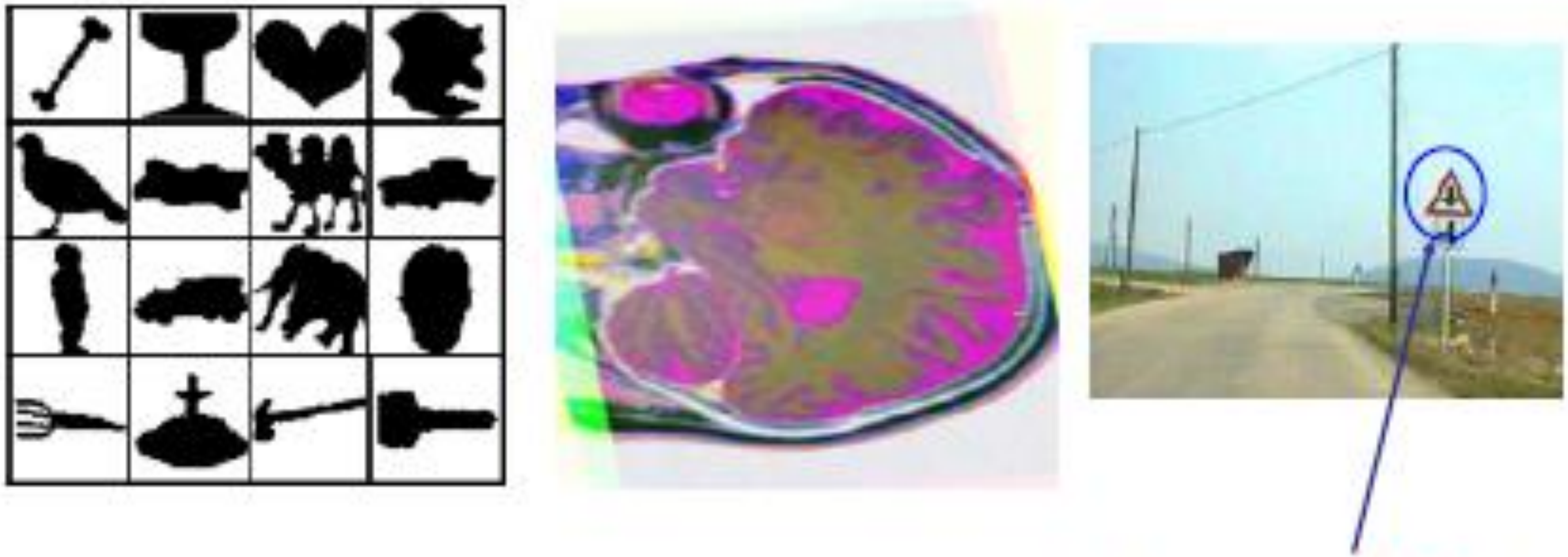
Pattern recognition





Which group?

Pattern recognition



To which class belongs an image

To which class (segment) belongs every pixel?

Where is an object of interest (detection); What is it (classification)?

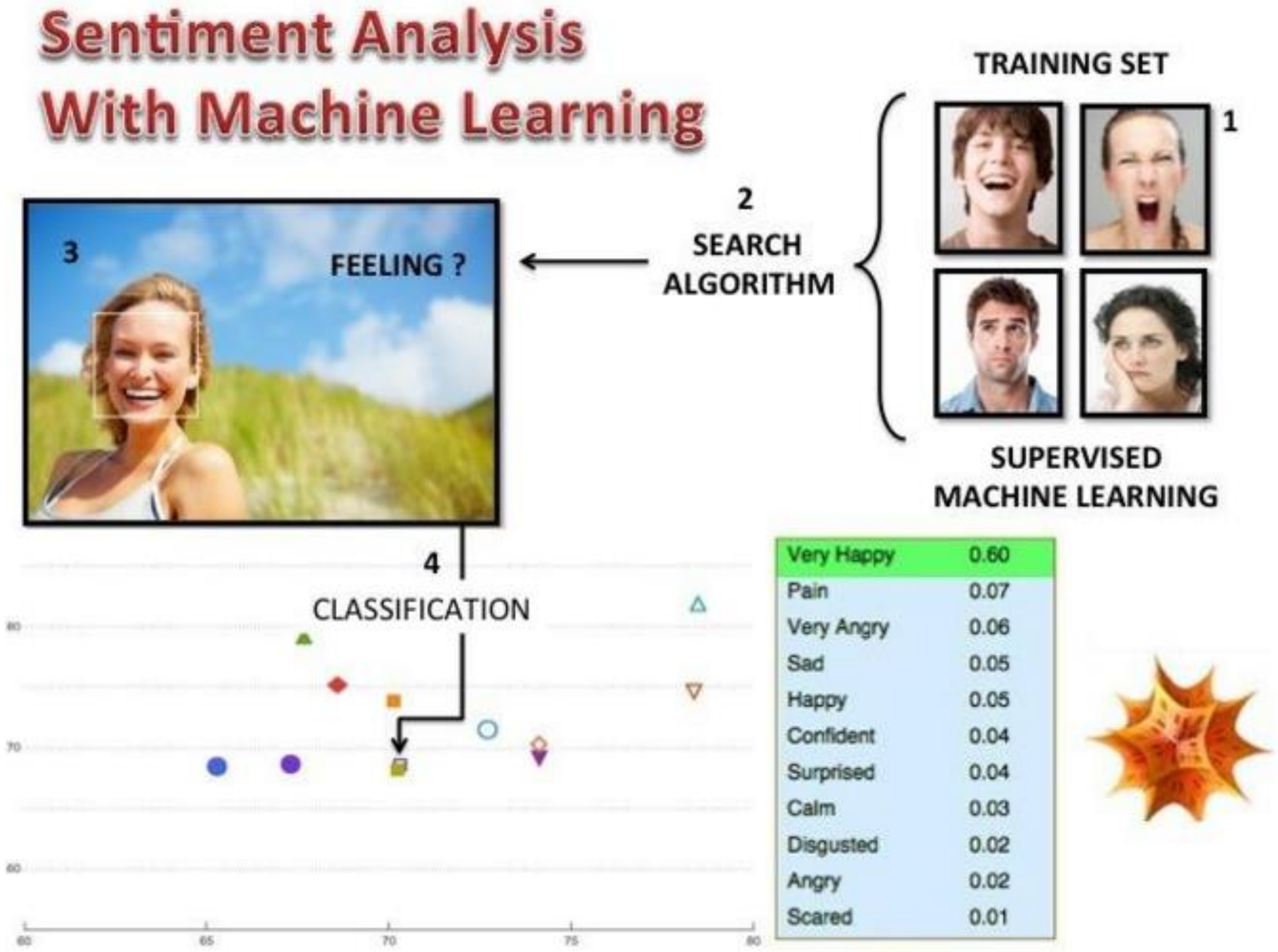
Pattern Recognition

is the **research area** that studies the operation and design of systems that **recognize patterns in data**.

It encloses: discriminant analysis, feature extraction, error estimation and cluster analysis.

Applications: image analysis, character recognition, speech analysis, man and machine diagnostics, person identification and industrial inspection.

Detection



Very Happy	0.60
Pain	0.07
Very Angry	0.06
Sad	0.05
Happy	0.05
Confident	0.04
Surprised	0.04
Calm	0.03
Disgusted	0.02
Angry	0.02
Scared	0.01

Matrix notation

Matrix

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\mathsf{T} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1{}^\mathsf{T} \\ \vdots \\ \boldsymbol{x}_i{}^\mathsf{T} \\ \vdots \\ \boldsymbol{x}_n{}^\mathsf{T} \end{bmatrix}; \boldsymbol{X} \in R^{n\text{x}p}$$

Vector

$$\boldsymbol{x}_i = \left(x_1, \ldots, x_p\right)^\mathsf{T} = [x_1, \ldots, x_p]; \boldsymbol{x}_i \in R^p$$

Data set representation

Data matrix

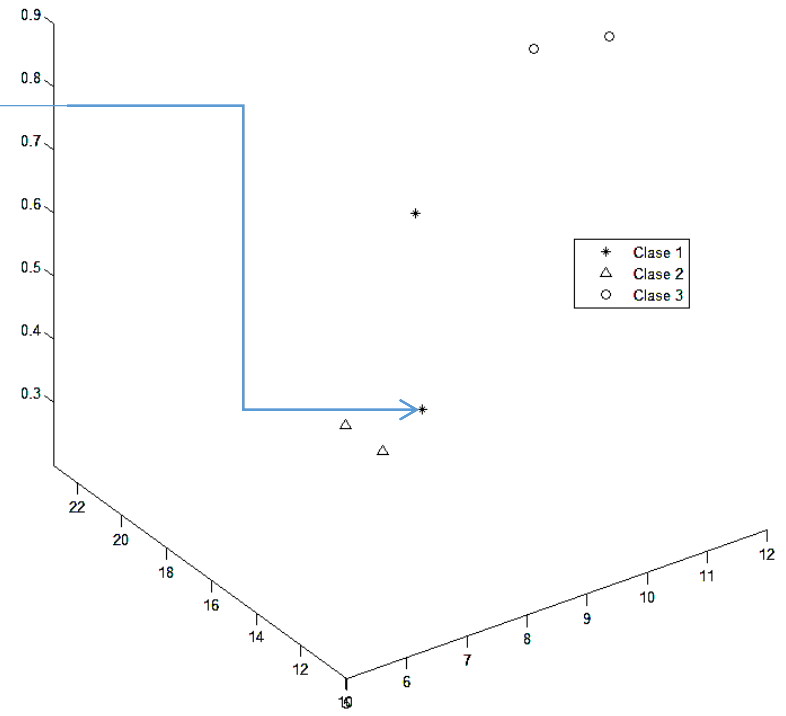
Label vector

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}; \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

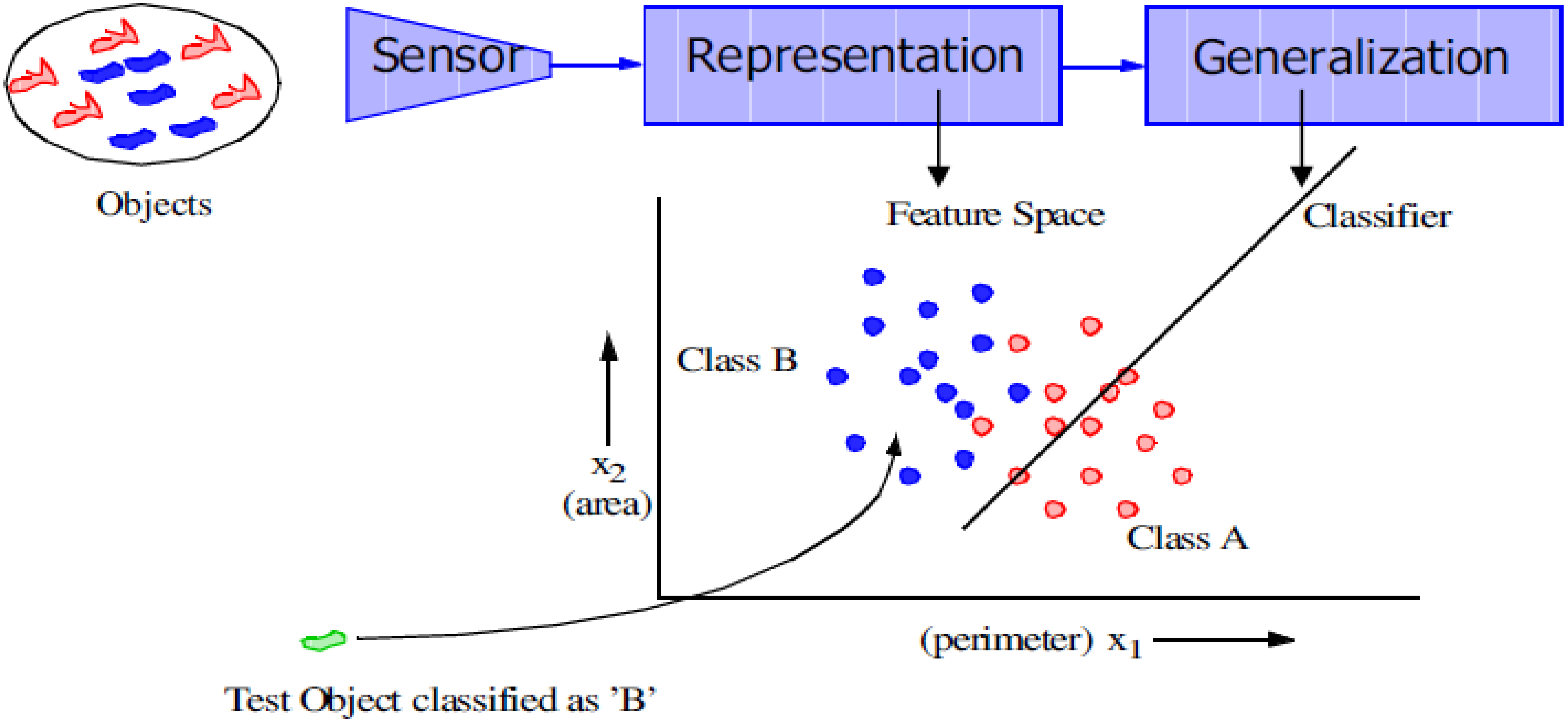
n samples, p features

Example:
3-dimensional 3-class data set

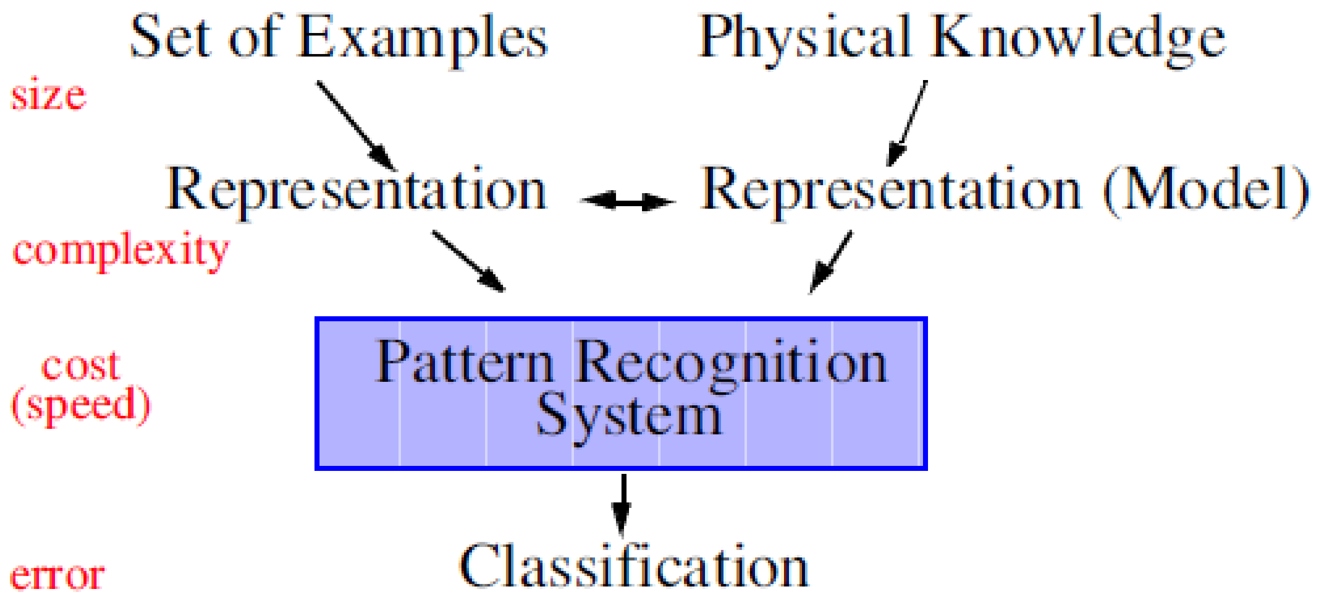
$$X = \begin{bmatrix} 10 & 20 & 0.2 \\ 11 & 23 & 0.4 \\ 5 & 10 & 0.6 \\ 6 & 11 & 0.5 \\ 10 & 15 & 0.9 \\ 12 & 17 & 0.8 \end{bmatrix}; \quad y = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 3 \end{bmatrix};$$



PR system


Sensor
Representation
Generalization
Objects
Feature Space
Classifier
Class B
Class A
x2
(area)
(perimeter) x1
Test Object classified as 'B'

Goals and requirements for PR



Minimize desired set of examples
Minimize amount of explicit knowledge
Minimize complexity of representation
Minimize cost of recognition
Minimize probability of classification error

PR books

Fukunaga, K., *Introduction to statistical pattern recognition*, second edition, Academic Press, 1990.

Kohonen, T., *Self-organizing maps*, Springer Series in Information Sciences, Volume 30, Berlin, 1995.

Ripley, B.D., *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.

Devroye, L., Gyorfi, L., and Lugosi, G., *A probabilistic theory of pattern recognition*, Springer, 1996.

Schurmann, J. *Pattern classification, a unified view of statistical and neural approaches*, Wiley, 1996

Gose, E., Johnsonbaugh, R., and Jost, S., *Pattern Recognition and Image Analysis*, Prentice-Hall, 1996.

Vapnik, V.N., *Statistical Learning Theory*, Wiley, New York, 1998.

Theodoridis, S. and Koutroumbas, K. *Pattern Recognition*, Academic Press, New York, 1999.

Duda, R.O., Hart, P.E., and Stork, D.G. Pattern Classification, 2d Edition Wiley, New York, 2001.

Hastie, T., Tibishirani, R., Friedman, J., *The Elements of Statistical Learning*, Springer, Berlin, 2001.

Webb, A. *Statistical Pattern Recognition* Wiley, 2002.

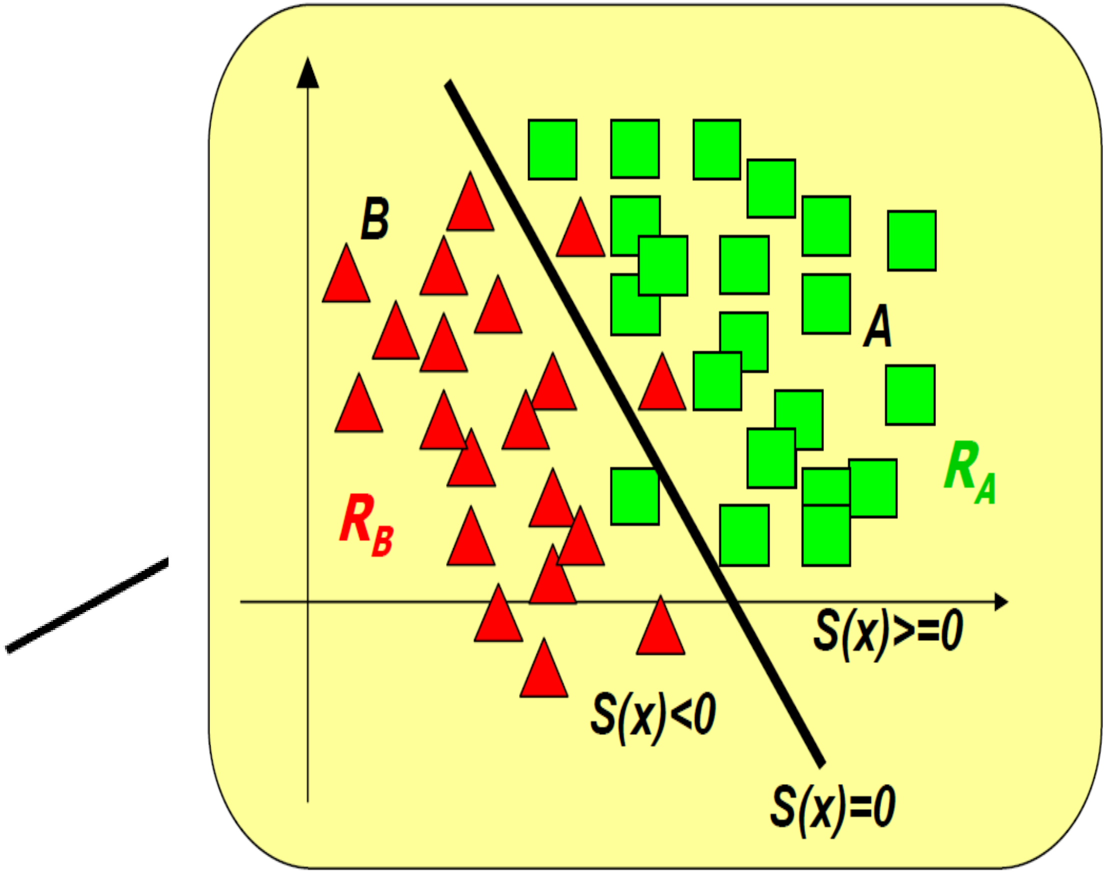
F. van der Heiden, R.P.W. Duin, D. de Ridder, and D.M.J. Tax, *Classification, Parameter Estimation, State Estimation: An Engineering Approach Using MatLab*, Wiley, New York, 2004, 1-440.

Bishop, C.M., *Pattern Recognition and Machine Learning*, Springer 2006

Statsoft, Electronic Textbook, http://www.statsoft.com/textbook/stathome.html

Data classification

$\mathbf{x} = [x_1, x_2]^T$



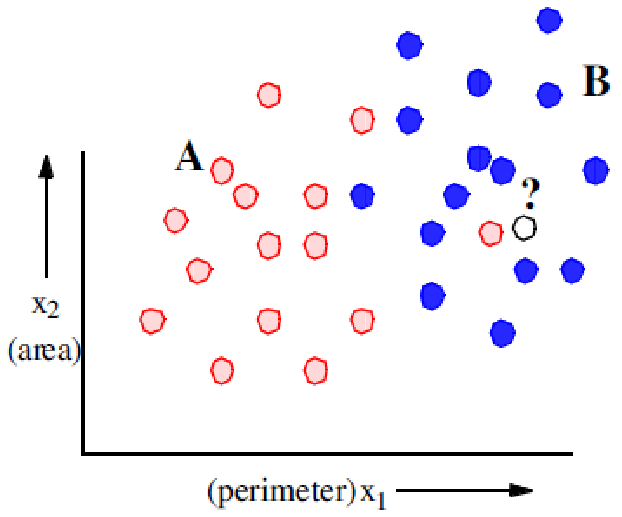
Data classification

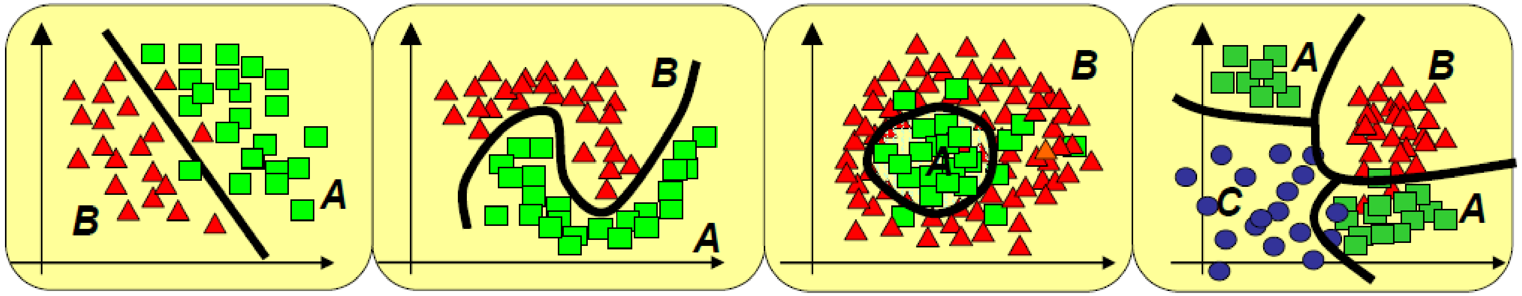
Not always it is that simple…

? to be classified as

A – because it is most close to an object A

B – because the local density of B is larger.

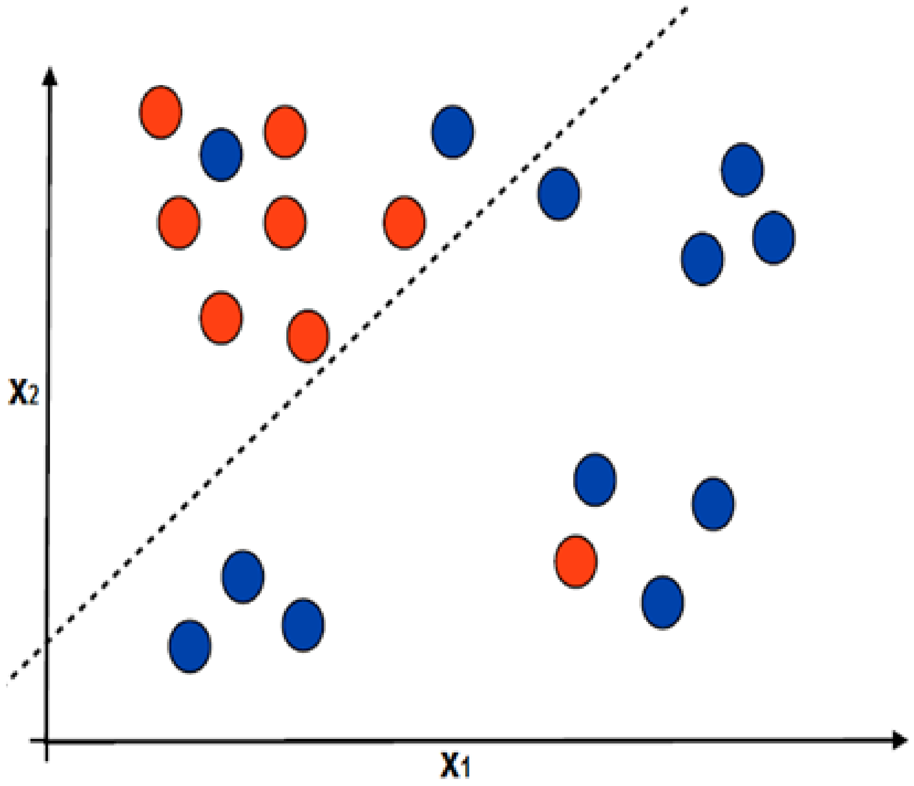




Unsupervised vs Supervised

Supervised approach

- Requires a **labeled training data** set.
- Reaches **high accuracy** when model and training size is enough.
- Is able to **generalize** from input data.
- Outputs **automatic classification**.

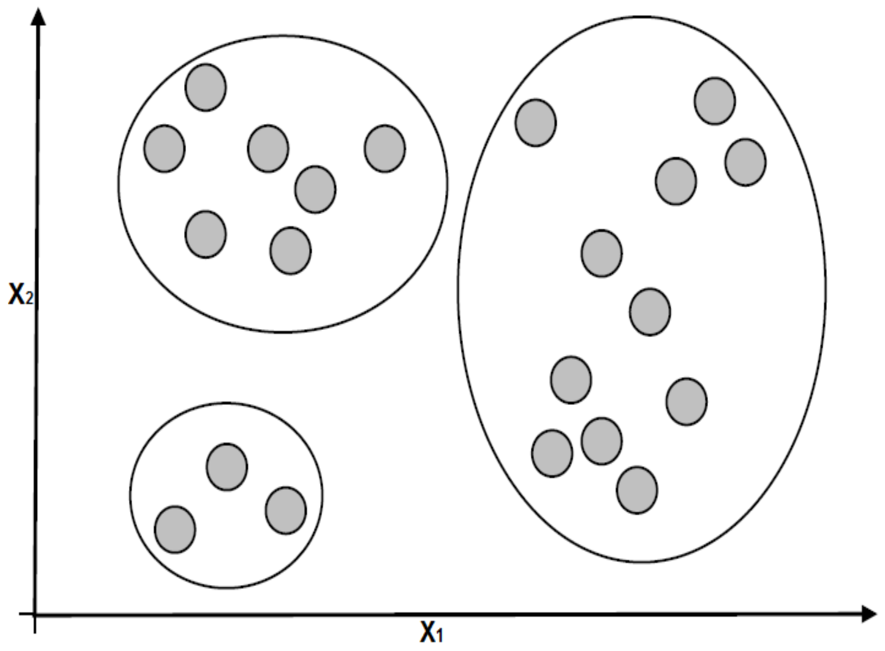


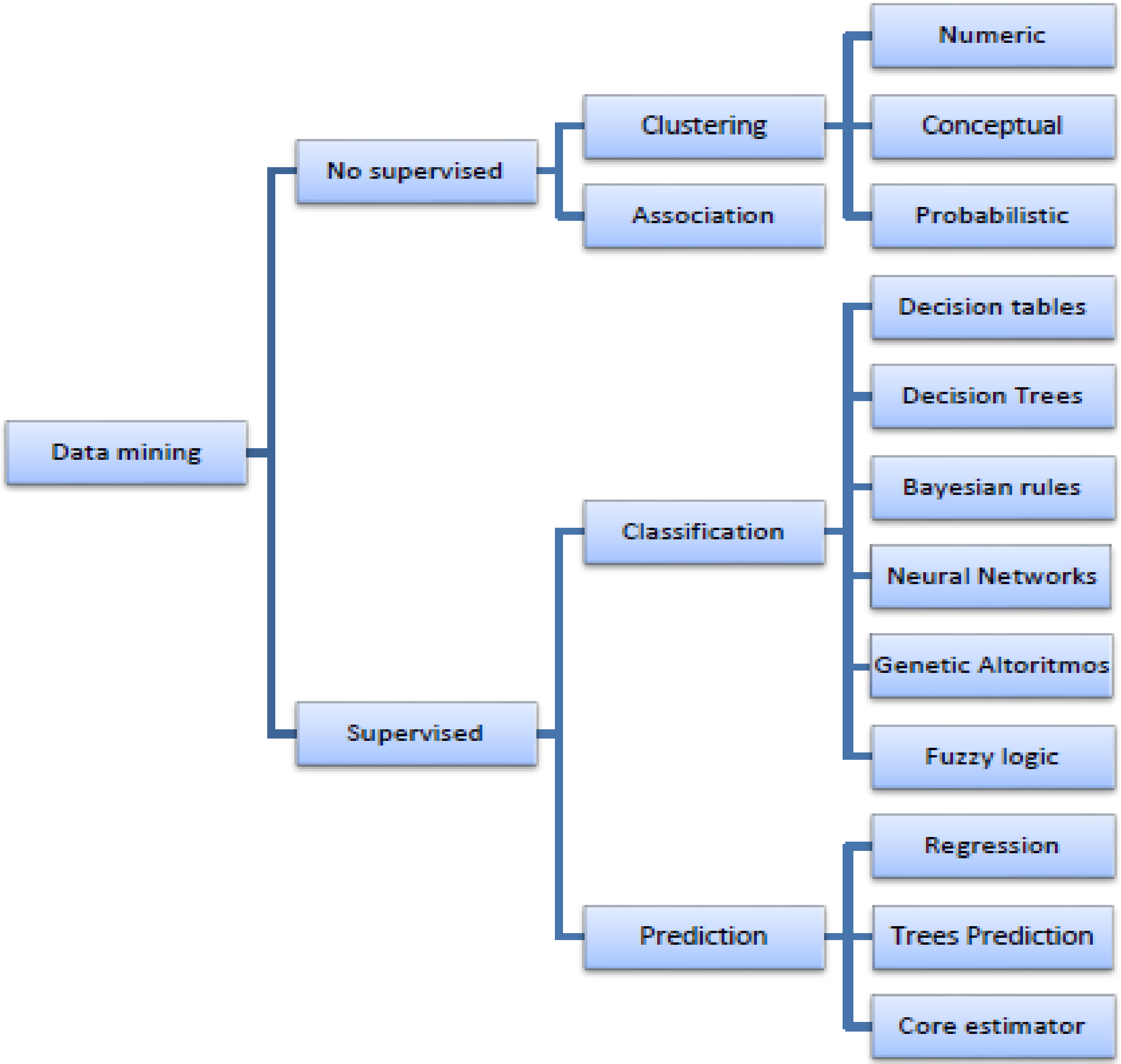
K-NN, Bayes Classifier, SVM

Unsupervised vs Supervised

Unsupervised approach

- Requires a **grouping** criterion (No labeled data are required).
- Generally reaches **lower accuracy** than supervised approaches.
- Has **no** a natural **generalization ability**.
- Outputs **clusters holding similar objects**.

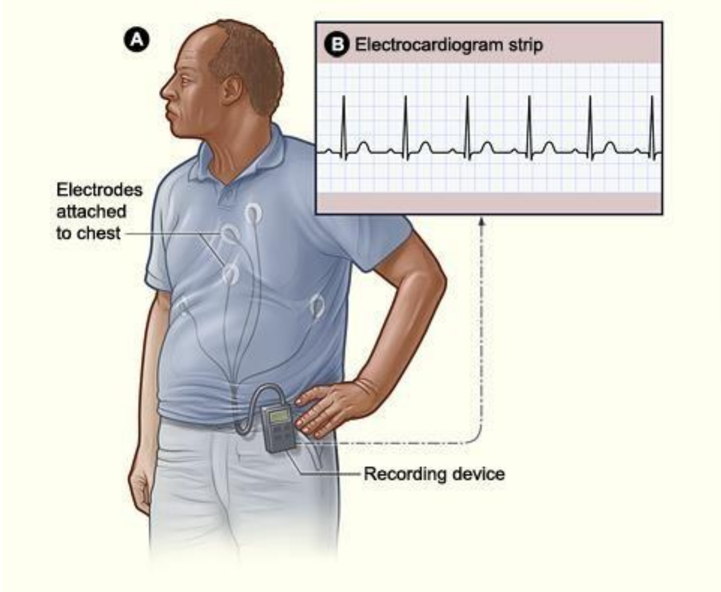



Data mining
No supervised
Supervised
Clustering
Association
Classification
Prediction
Numeric
Conceptual
Probabilistic
Decision tables
Decision Trees
Bayesian rules
Neural Networks
Genetic Altoritmos
Fuzzy logic
Regression
Trees Prediction
Core estimator

Applications

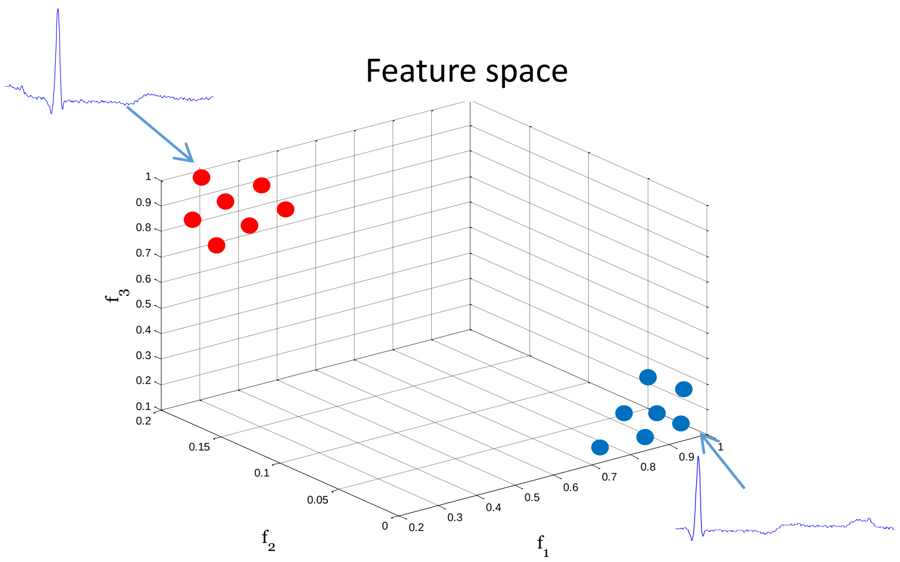
Biomedical signals - ECG

- Holter Recordings

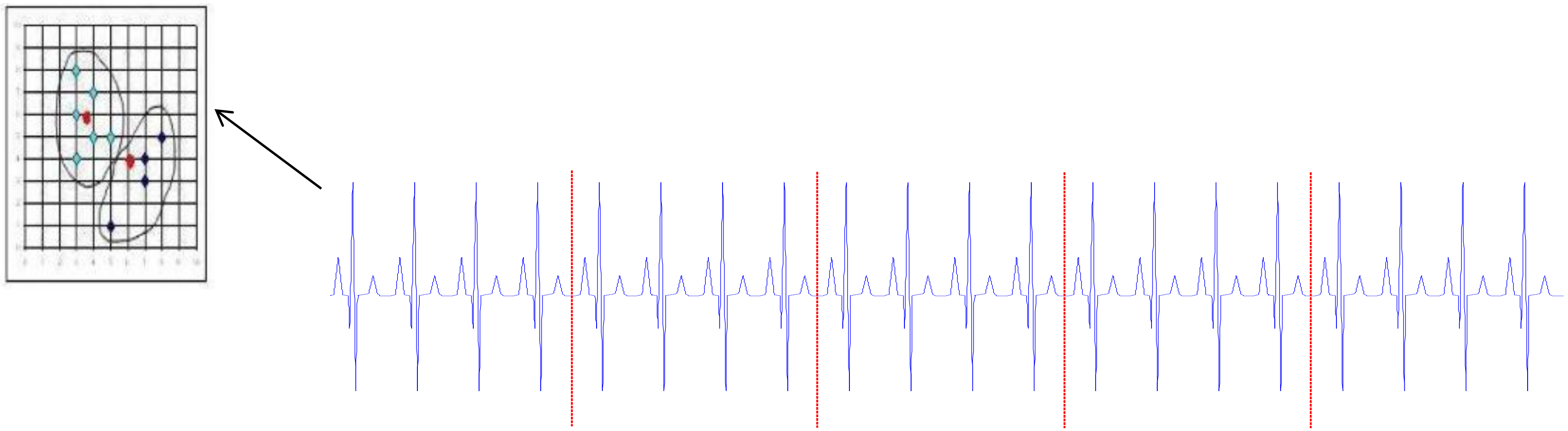


- Long term monitoring
- Holter analysis
- Holter monitoring

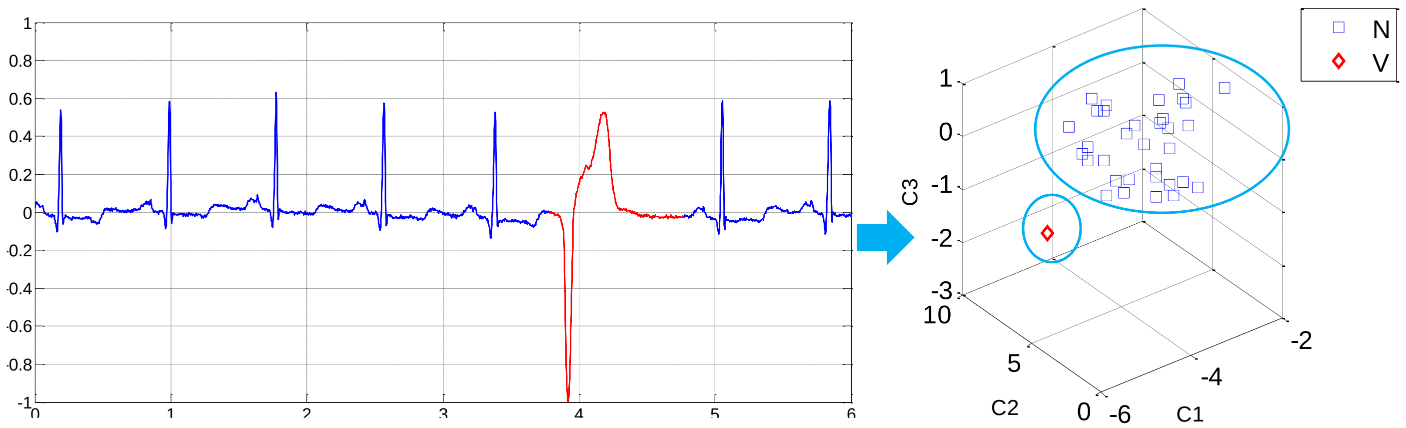
Biomedical signals - ECG



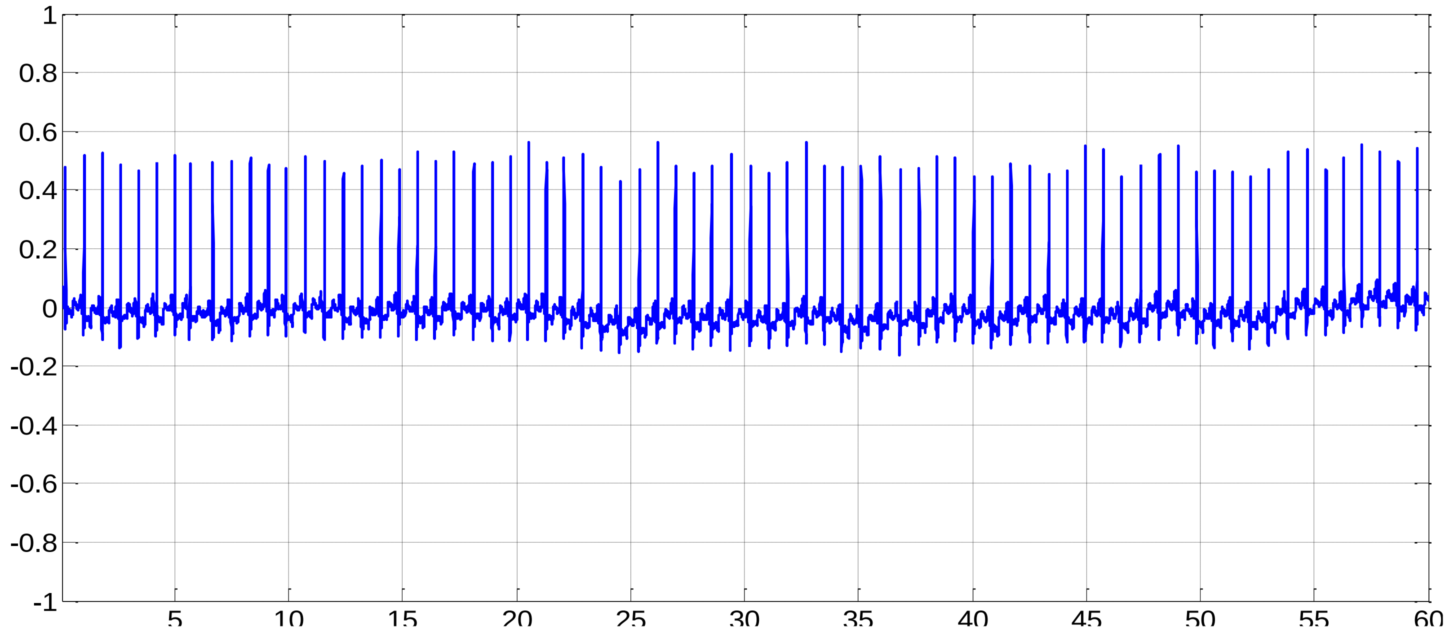
Biomedical signals - ECG



Biomedical signals - ECG

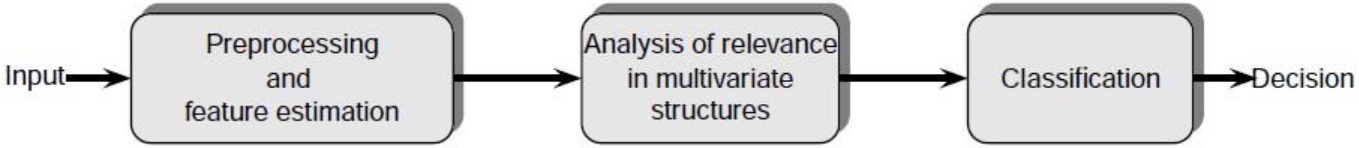


Biomedical signals - ECG



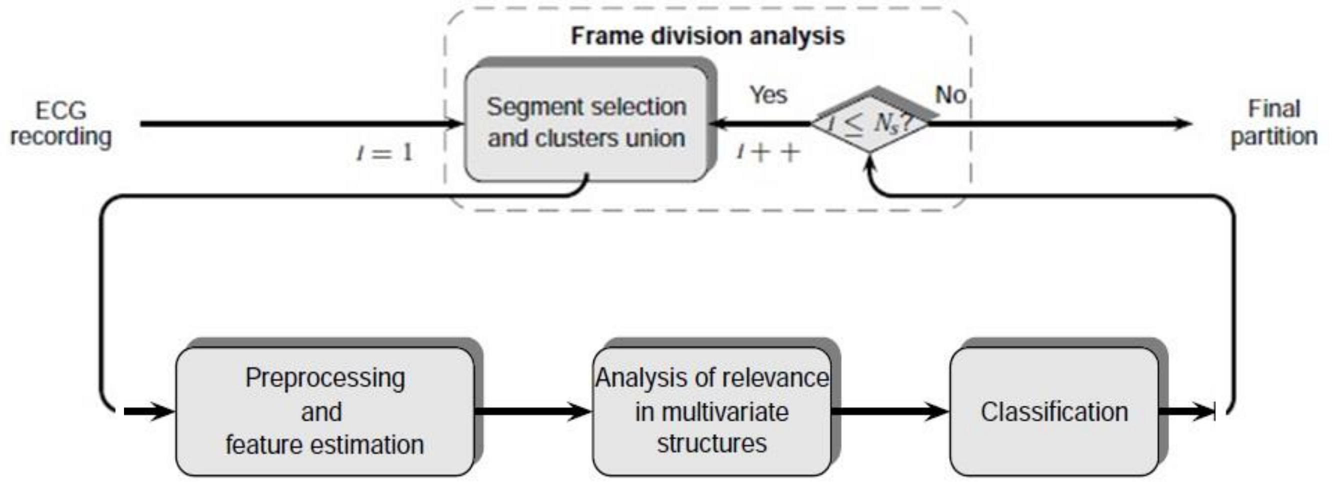
Biomedical signals - ECG

General scheme



Biomedical signals - ECG

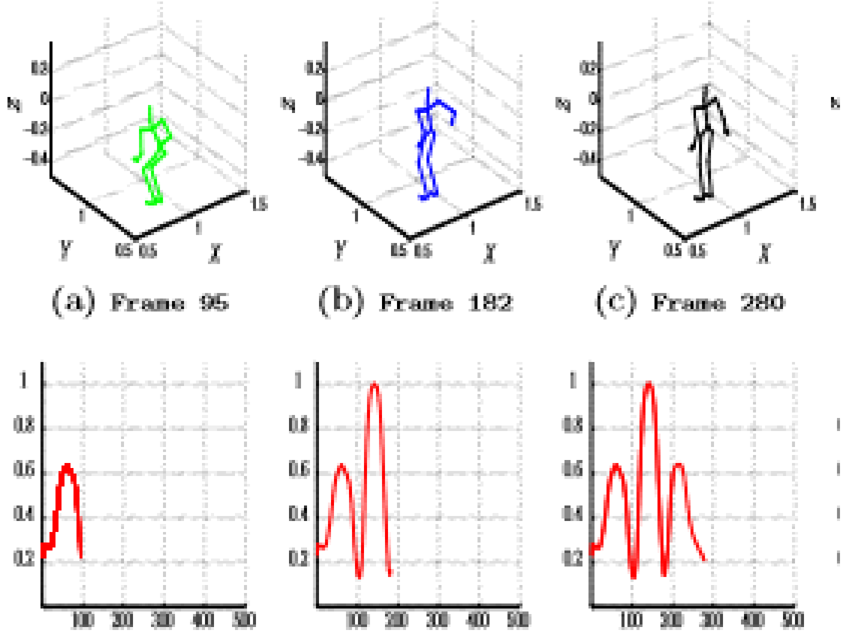
Segment-based scheme

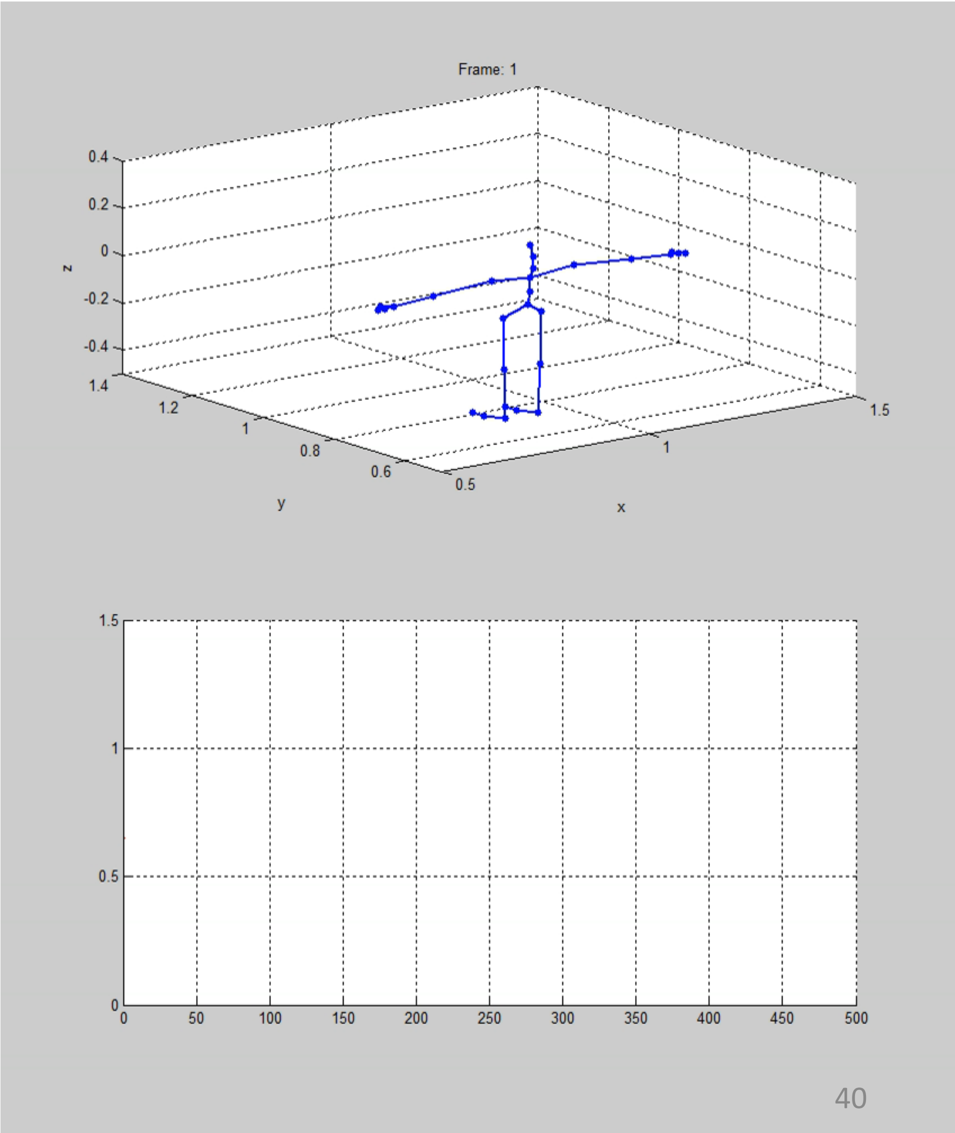


Biomedical signals - ECG

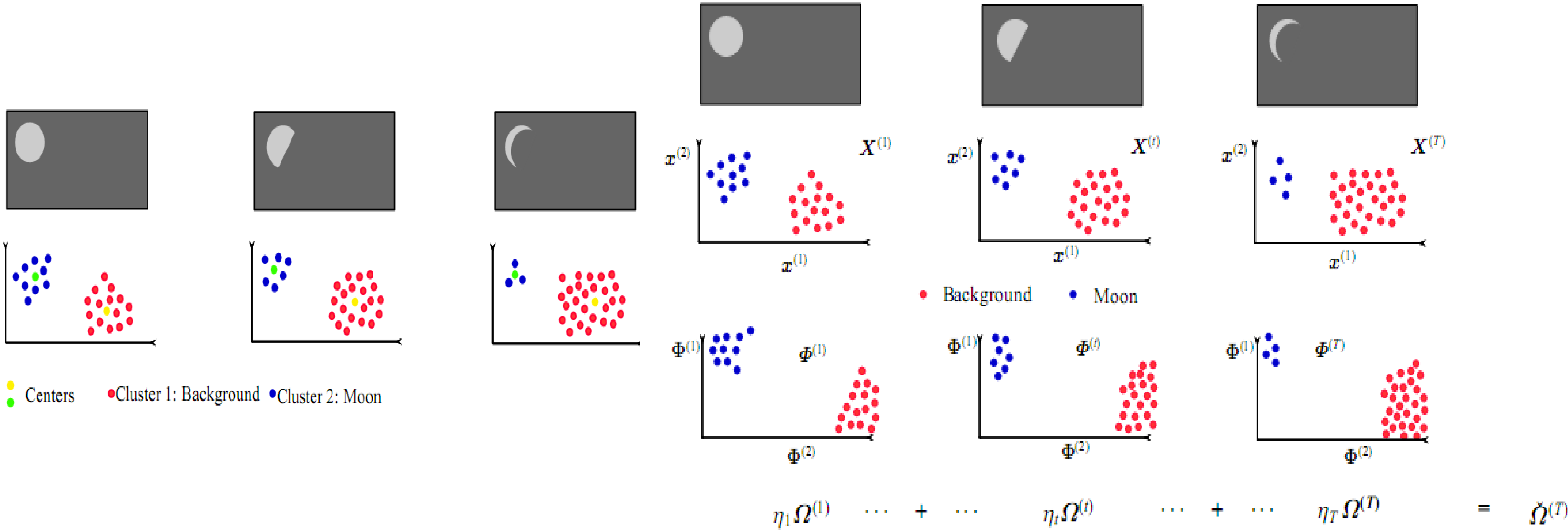
Biomedical signals - ECG

Dynamic data analysis – Tracking

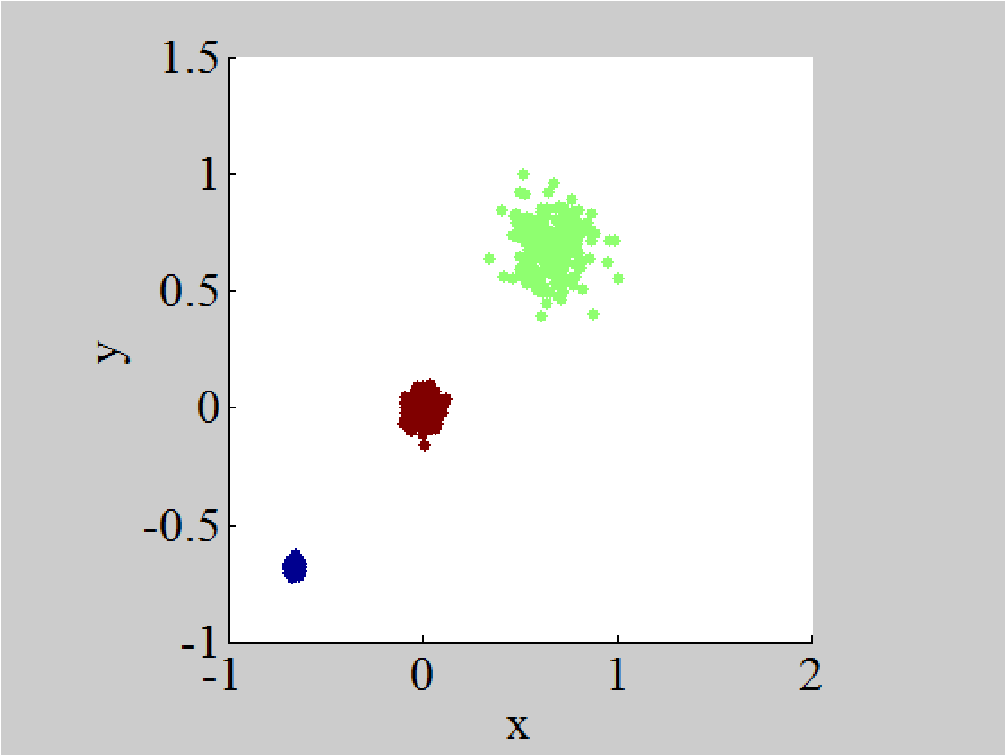




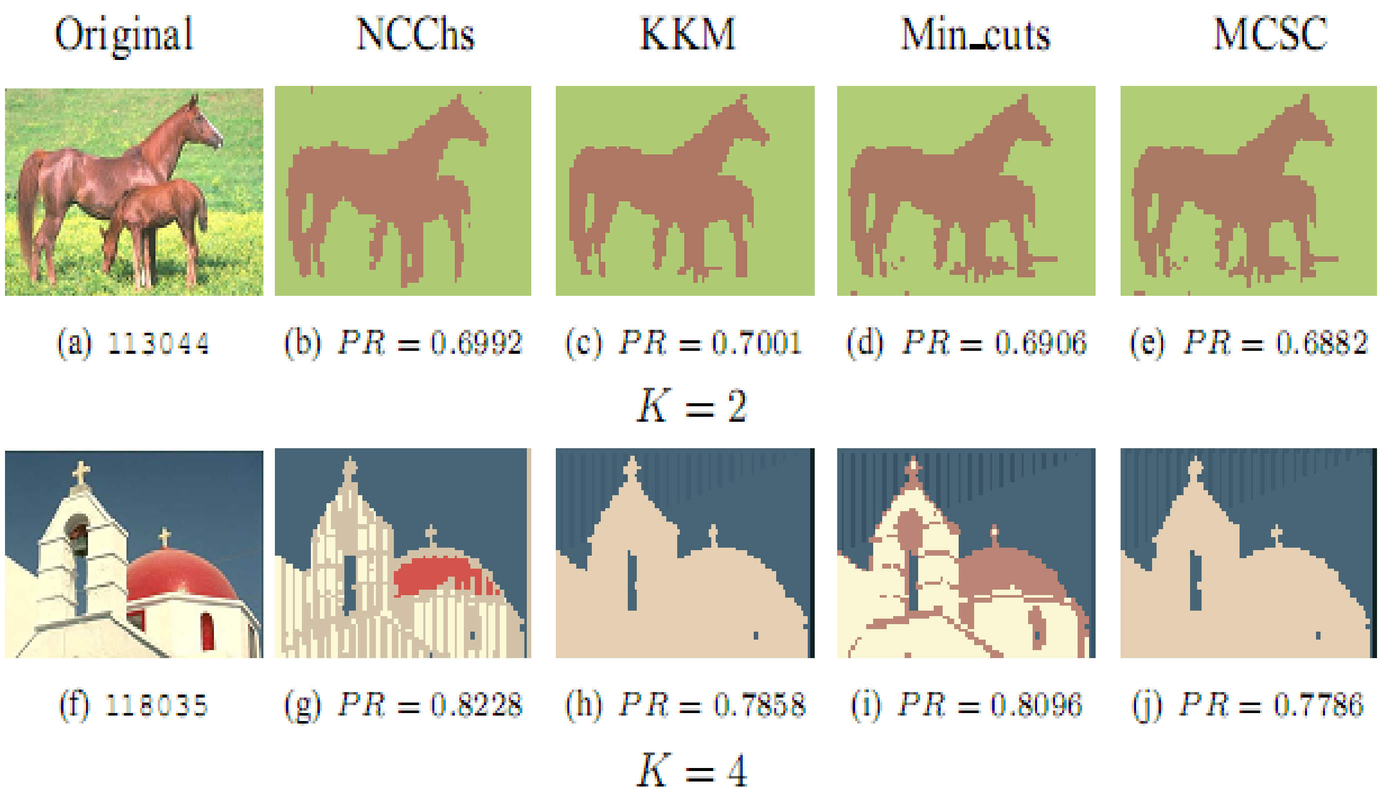
Dynamic data analysis – Hidden objects

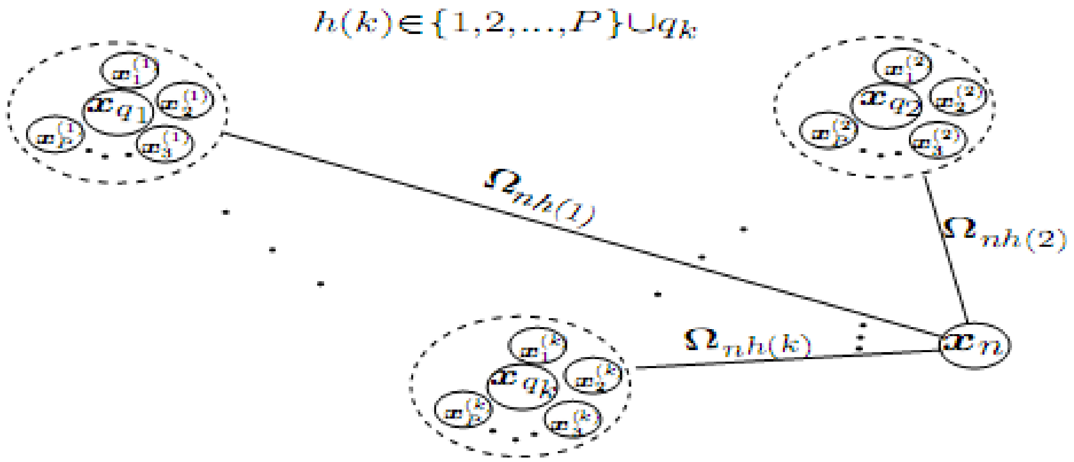


Dynamic data analysis – Overlapping

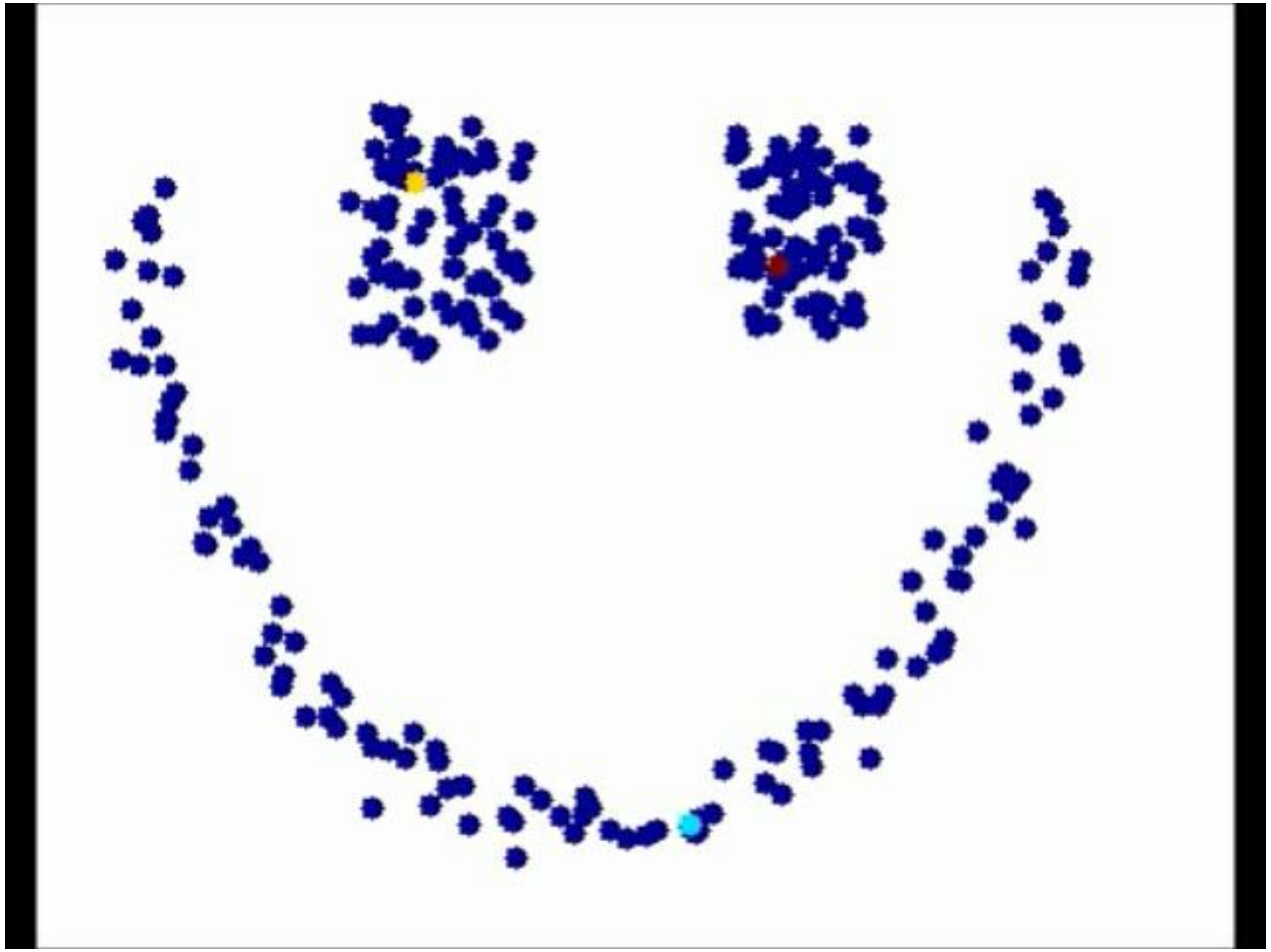


Complex data clustering



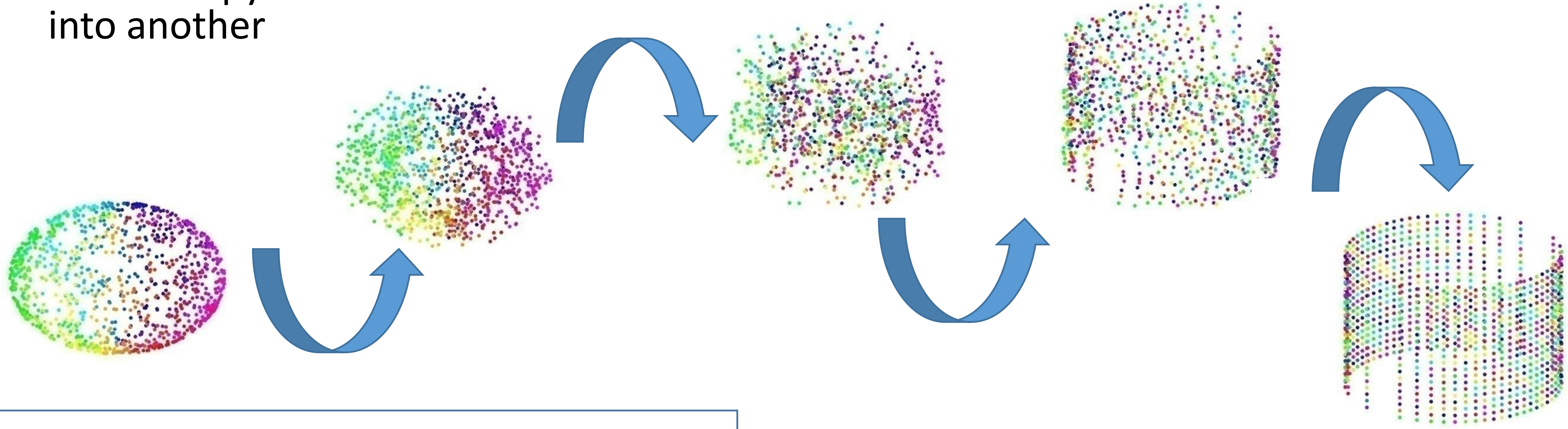


Complex data clustering



Homotopy - Deforming objects

A homotopy between two functions is a continuous transformation from one function into another



Homotopy

$$h: \chi \times [0,1] \to \psi$$

$$f1, f2\ \lambda \mapsto h(f1, f2, \lambda)$$

$$h(f1, f2, 0) = f1$$

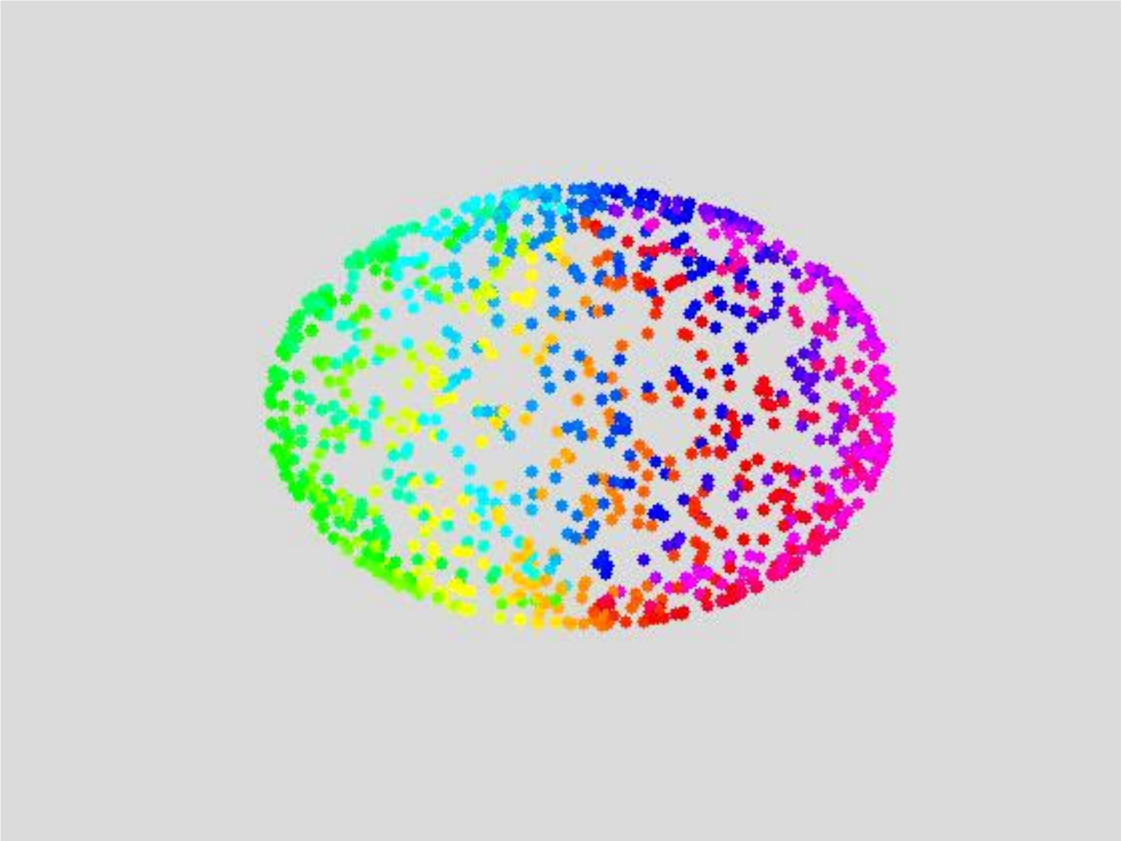
$$h(f1, f2, 1) = f2$$

(Nash et al., 2013)

A feasible way to write a **homotopy** mapping is through a linear combination in which coefficients are given in terms of **λ**

$$h: C_1(\lambda) f_2 + C_2(\lambda) f_1$$

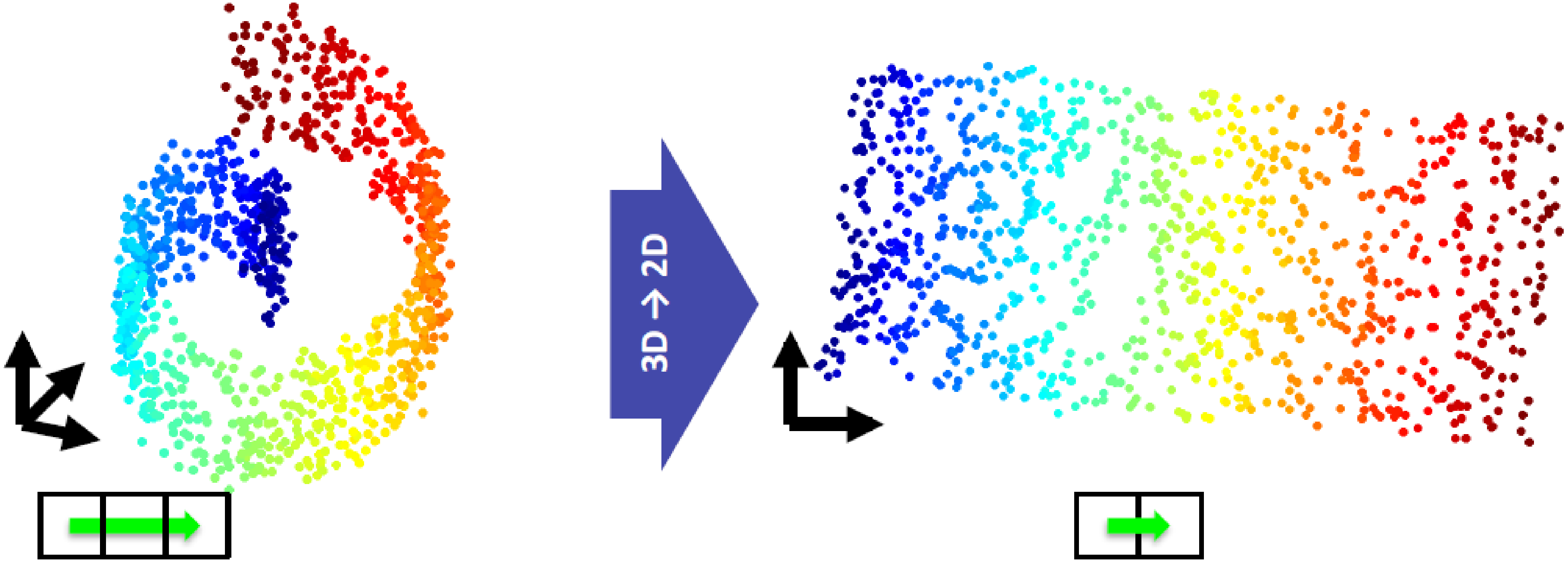
Homotopy - Deforming objects



Generalized exponential approach **(GExpH)** m=2 (video)

Dimensionality reduction - Structure

Aims at representing **high-dimensional (HD)** data in **low-dimensional (LD)** spaces, while **preserving structure**



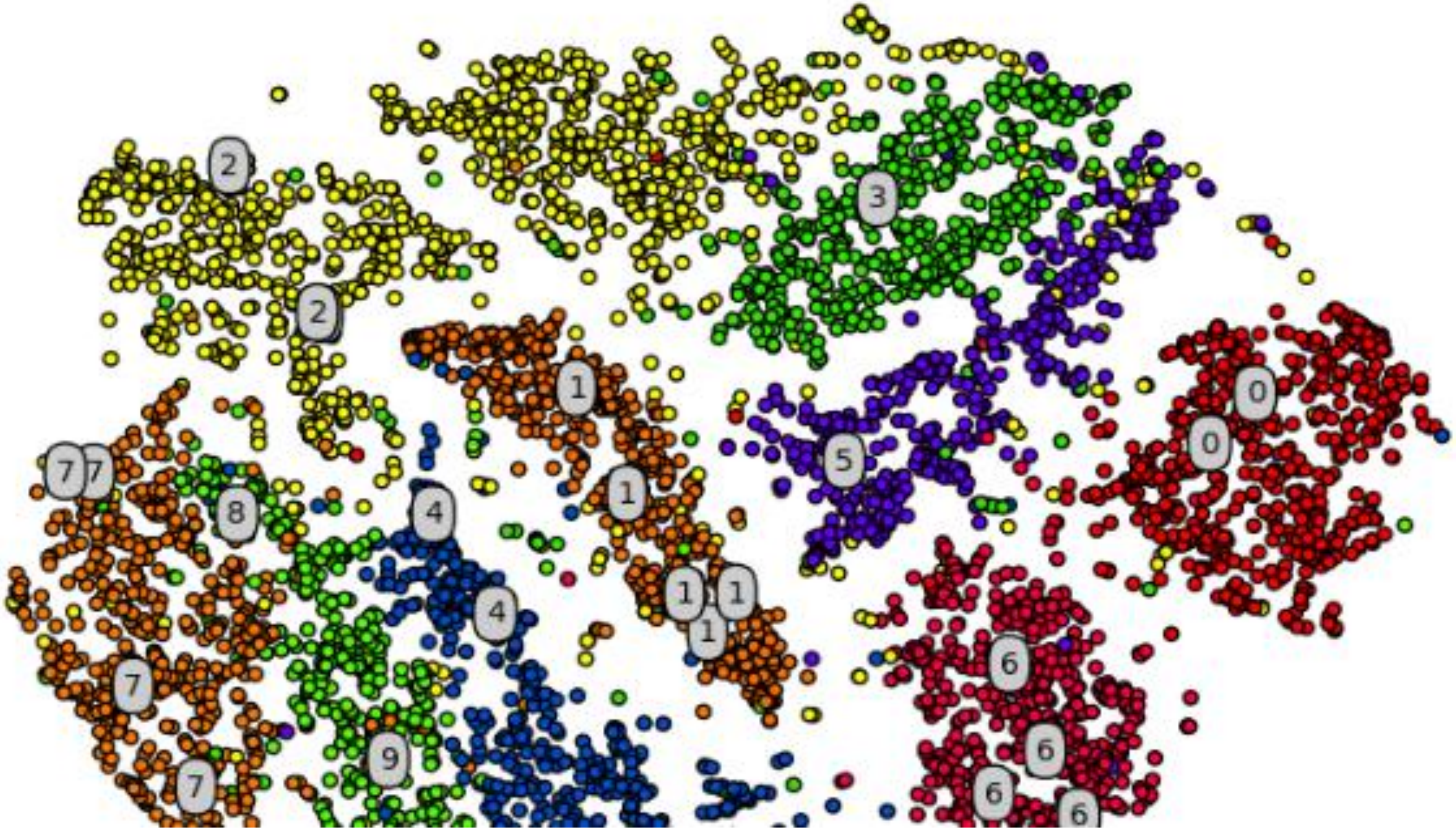
$$\boldsymbol{Y} = [\boldsymbol{y}_i]_{1 \leq i \leq N}$$

$$\boldsymbol{y}_i \in \mathbb{R}^D$$

$$\boldsymbol{X} = [\boldsymbol{x}_i]_{1 \leq i \leq N}$$

$$\boldsymbol{x}_i \in \mathbb{R}^d \quad d < D$$

Dimensionality reduction - Separability



Different perspectives for kernel spectral clustering: A theoretical study

Diego H. Peluffo O., Paul D. Rosero M., Carlos H. Pupiales Y., Luis E. Suárez Z., Edgar D. Jaramillo V., Edgar A. Maya O., Jaime R. Michilena C., and Carlos A. Vásquez A.
Universidad Técnica del Norte
Avenida 17 de Julio, 5-21
E-mail:
{dhpeluffo,pdrosero,chpupiales,lesuarez,edjaramillo,eamaya,jrmichilena,cavasquez}@utn.edu.ec

***Abstract*—Spectral clustering is a suitable technique to deal vith problems involving unlabeled clusters and having a complex tructure, being kernel-based approaches the most recommended ones. This work aims at demonstrating the relationship between ı widely-recommended method, so-named kernel spectral clustering (KSC) and other well-known approaches, namely normalized cut clustering and kernel k-means. Such demonstrations are done by following a primal-dual scheme. Also, we mathematically and experimentally prove the usability of using LS-SVM formulations with a model. Experiments are conducted to assess the clustering performance of KSC and the other considered methods on image**

show clearly the links with the other considered methods. Also, we mathematically and experimentally prove the usability of using LS-SVM formulations with a model. Additionally, in order to assess the clustering performance, we explore the benefit of each considered method on image segmentation. In this connection, images extracted from the free access Berkeley Segmentation Data Set [13] are used.

The remaining of this paper is structured as follows: Section II outlines a brief explanation of KSC, posing a primal-

Dimensionality reduction for interactive data visualization via a Geo-Desic approach

Jose A. Salazar-Castro,
Diego Peña-Unigarro

Universidad de Nariño
Pasto, Colombia
Universidad Nacional de Colombia Sede Manizales
Manizales, Colombia
alejo26st@udenar.edu.co,
diferpun@gmail.com

Diego H. Peluffo-Ordóñez,
Paul D. Rosero-Montalvo,
H. Mauricio Domínguez-Limaico

Universidad Técnica del Norte
Ibarra, Ecuador
Universidad Cooperativa de Colombia
Pasto, Colombia
dhpeluffo@utn.edu.ec,
pdrosero@utn.edu.ec,
hmdominguez@utn.edu.ec

Juan C. Alvarado-Pérez,
Roberto Therón

Corporación Universitaria Autónoma de Nariño
Universidad de Salamanca
Salamanca, España
jcalvarado@usal.es, theron@usal.es

Abstract—**This work presents a dimensionality reduction (DR) framework that enables users to perform either the selection or mixture of DR methods by means of an interactive model, here named Geo-Desic approach. Such a model consists of linear combination of kernel-based representations of DR methods, wherein the corresponding coefficients are related to coordinated latitude and longitude inside of the world map. By incorporating the Geo-Desic approach within an interface, the combination may be made easily and intuitively by users –even non-expert ones– fulfilling their criteria and needs, by just**

criteria considered from a priori information and design parameters and optimization preset criteria. However, in the visualization, the goal is not to reduce the data set to its intrinsic dimension, but a viewable (2 or 3), preserving the information as possible.

This paper presents an attempt to link the field of dimensionality reduction with the information–visualization. DR can be improved by importing some properties of the

Sitting-Pose Detection Using Data Classification and Dimensionality Reduction

Santiago Nuñez-Godoy, Vanessa Alvear-Puertas, Staling Realpe-Godoy, Edwin Pujota-Cuascota
Henry Farinango-Endara, Ivan Navarrete-Insuasti, Franklin Vaca-Chapi, Paul Rosero-Montalvo, Diego H. Peluffo
Universidad Técnica del Norte
Avenida 17 de Julio 5-21
E-mail:

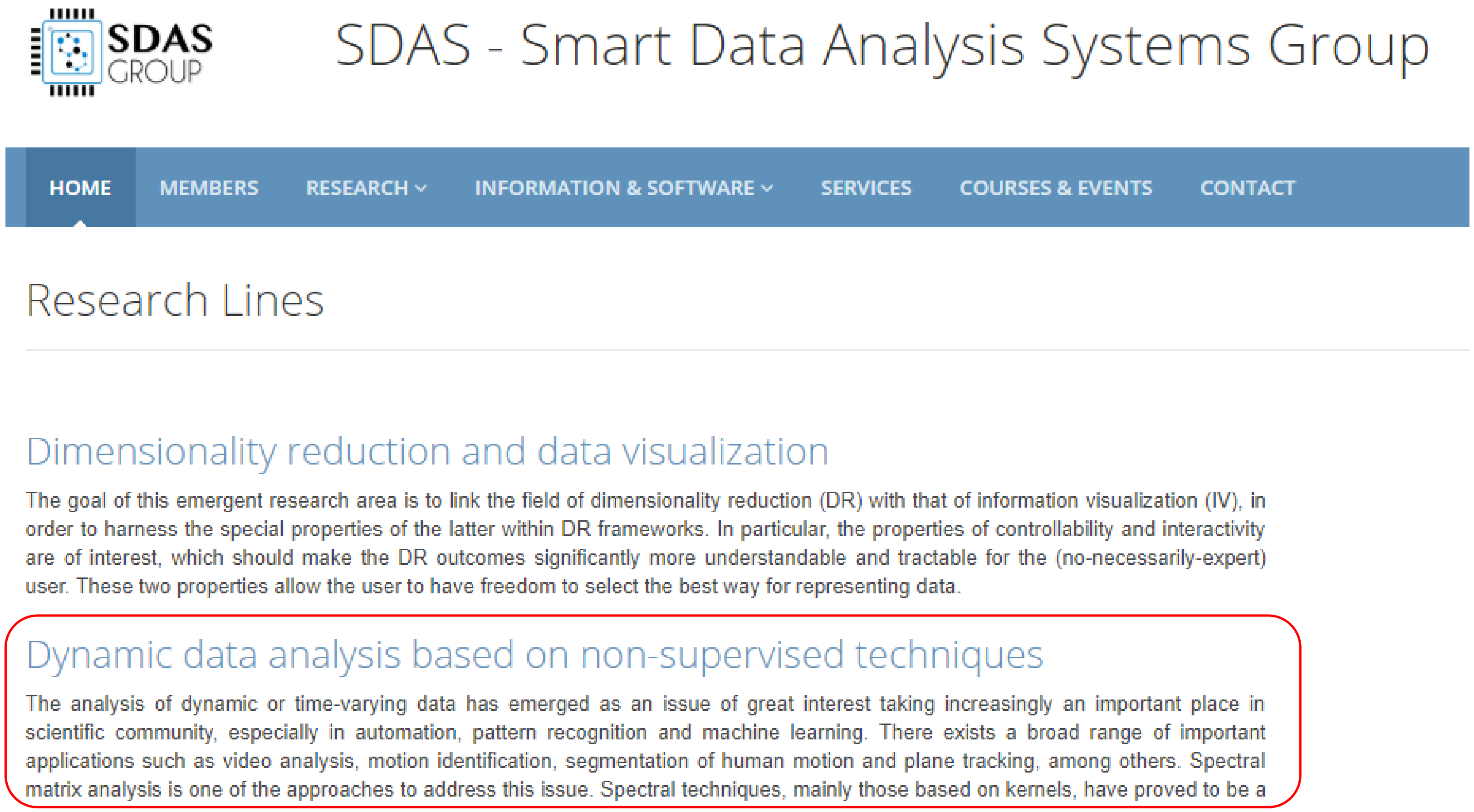
{bsnuniezg, vealvearp, swrealpeg, edpujota, hpfarinangoe, ienavarrete, fvacac, pdrosero,dhpeluffo}@utn.edu.ec

***Abstract*— The research area of sitting-pose analysis allows for preventing a range of physical health problems mainly physical. Despite that different systems have been proposed for sitting-pose detection, some open issues are still to be dealt with, such as: Cost, computational load, accuracy, portability, and among others. In this work, we present an alternative approach based on a sensor network to acquire the position-related variables and machine learning techniques, namely dimensionality reduction (DR) and classification. Since the information acquired by sensors is high-dimensional and therefore it might not be saved into embedded system memory, a DR stage based on principal component analysis (PCA) is performed. Subsequently, the automatic posed detection**

In this work, we present an alternative approach based on a sensor network to acquire the position-related variables and machine learning techniques, namely dimensionality reduction (DR) and classification. To design our detection system, we start with a comfortable chair that helps a user to readily reach a good posture so that his/her body keeps the natural state wherein the spin position and the space between the vertebrae do not change and the muscles of the waist and back relax. Otherwise, when changing to an incorrect position, intervertebral discs are narrowed and ligaments are stressed causing pain and discomfort [7].

More information at…

http://sdas-group.com/research-lines/



SDAS GROUP

SDAS - Smart Data Analysis Systems Group

HOME | MEMBERS | RESEARCH | INFORMATION & SOFTWARE | SERVICES | COURSES & EVENTS | CONTACT

Research Lines

Dimensionality reduction and data visualization

The goal of this emergent research area is to link the field of dimensionality reduction (DR) with that of information visualization (IV), in order to harness the special properties of the latter within DR frameworks. In particular, the properties of controllability and interactivity are of interest, which should make the DR outcomes significantly more understandable and tractable for the (no-necessarily-expert) user. These two properties allow the user to have freedom to select the best way for representing data.

Dynamic data analysis based on non-supervised techniques

The analysis of dynamic or time-varying data has emerged as an issue of great interest taking increasingly an important place in scientific community, especially in automation, pattern recognition and machine learning. There exists a broad range of important applications such as video analysis, motion identification, segmentation of human motion and plane tracking, among others. Spectral matrix analysis is one of the approaches to address this issue. Spectral techniques, mainly those based on kernels, have proved to be a

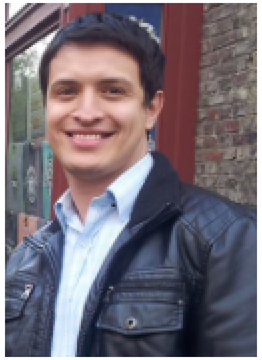
Final remarks

- Several applications
- Open issues and big challenges
- Not that expensive
- Nice people

To learn more…

http://diegopeluffo.com/

Home | Publications | Teaching | Research Areas | Projects | Thesis advisory | Software | Links of interest



Diego Hernán Peluffo-Ordóñez
Personal website

Professional Interests

- Applied maths
- Data clustering
- Dimensionality reduction
- Data visualization
- Unsupervised analysis and kernel methods
- Time-varying data analysis
- Biosignals
- Electromedical science

Conference papers | Journals | Book Chapters | Theses

- **2016**
  - **On the Relationship Between Dimensionality Reduction and Spectral Clustering from a Kernel Viewpoint**
    13th International Conference on Distributed Computing and Artificial Intelligence, Advances in Intelligent Systems and Computing. DCAI 2016
    D. H. Peluffo-Ordóñez, M. A. Becerra, A. E. Castro-Ospina, X. Blanco-Valencia, J. C. Alvarado-Pérez, R. Therón, A. Anaya-Isaza
    See abstract See full paper
- **2015**
  - **Low-cost prototype for biofeedback applications**
    Segundo encuentro nacional de semilleros de investigación en Ingeniería Electrónica 2015
    S. Cano, A. Ruano, J. Revelo, D. Peluffo
    See abstract See full paper
  - **Generalized Bonhoeffer-van der Pol oscillator for modelling cardiac pulse: Preliminary results**
    IEEE 2nd Colombian Conference on Automatic Control. CCAC 2015
    Peluffo-Ordonez, D.H.; Rodriguez-Sotelo, J.L.; Revelo-Fuelagan, E.J.; Ospina-Aguirre, C.; Olivard-Tost, G.
    See abstract See full paper

Thank you for your attention

Any question? Here or later at… dpeluffo@yachatech.edu.ec -
www.diegopeluffo.com
www.sdas-group.com

http://sdas-group.com/gallery/
https://www.dropbox.com/s/bhbnaecke1belnl/VideoTutorial.mp4?dl=0