Mathematical Foundations of Data Analysis (MFDA)

Boqiang Huang

huang@math.uni-koeln.de

Institute of Mathematics, University of Cologne, Germany

2019.04.02

MFDA

MFDA – I: Deterministic Data Analysis

Winter Semester 2018/2019

Deterministic Model: the output of the model is fully determined by the parameter values and the initial conditions.

MFDA – II: Statistical/Stochastic Data Analysis

Summer Semester 2019

Stochastic model: possess some inherent randomness. The same set of parameter values and initial condtions will lead to an ensemble of different outputs.

Lecture:	Tu. 16:00 – 17:30	Hörsaal 2.03
	Th. 14:00 – 15:30	Hörsaal 2.03
Exercise:	Th. 16:00 – 17:30	Seminarraum 3
Office Hrs:	Tu. 14:00 – 15:30	Gyrhofstr. 8b, Room 1.202 (Pavilion, 1st floor)

MFDA – II: Home Works, Projects and Exams

- Normal problem sets will be given in some weeks, each with 20 points
  - You may solve problems in groups. If so, all members should be clarified.
  - Write down the solutions by yourself.
  - Answer sheet should be handed in on next Tuesday.
  - Solutons will be discussed on Thursday.
  - You may be asked to explain your solutions on board.

MFDA – II: Home Works, Projects and Exams

- Normal problem sets will be given in some weeks, each with 20 points

- Two projects will be given in the semester, each with 40 points (Report 30 + Codes 10)
  - You are required to realize some algorithms, and apply them into well-designed applications.
  - Matlab will be the default programming language, but not only.
  - You may solve problems in groups. If so, all members should be clarified.
  - Write down your report by yourself. One group may release one code solution.
  - A whole month will be given to finish the project.
  - Please send your codes and report via E-Mail: huang@math.uni-koeln.de

MFDA – II: Home Works, Projects and Exams

- Normal problem sets will be given in some weeks, each with 20 points
- Two projects will be given in the semester, each with 40 points (Report 30 + Codes 10)
- Who collected at least 50% of total points can apply for the final oral exam
  - Exam week: the last week of the semester (before summer vacation)
  - Grading rule: Problem Sets 30% + Projects 30% + Final Exam 40%

MFDA – II: Home Works, Projects and Exams

- Normal problem sets will be given in some weeks, each with 20 points
- Two projects will be given in the semester, each with 40 points (Report 30 + Codes 10)
- Who collects at least 50% of total points can apply for the final oral exam
- Course webpage www.numana.uni-koeln.de/17864.html

MFDA – II: Main References

[1] A. Hyvaerinen, J. Karhunen, E. Oja, Independent component analysis, New York: John Wiley & Sons Inc., 2001.

[2] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning: with applications in R, Springer, 2013.

[3] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics, 2016.

[4] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT Press, 2016.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature, vol. 521, pp. 436-444, 2015.

[6] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time-series, The Handbook of Brain Theory and Neural Networks, vol. 3361, 1995.

[7] C. Goller, A. Kuechler, Learning task-dependent distributed representations by backpropagation through structure. IEEE Int. Conf. on Neural Networks, 1996.

[8] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.

MFDA – II: Contents

- Basics for Machine Learning
  - Probability and information theory
  - Numerical computation methods
  - Basic knowledge in machine learning

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
  - Simple linear regression
  - Multiple linear regression

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
  - Logistic regression
  - Linear discriminant analysis

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
  - Cross-validation
  - Bootstrap

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
  - Shrinkage methods
  - Dimension considerations

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Beyond Linearity
  - Regression splines
  - Generalized additive models

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Beyond Linearity
- Tree-Based Methods
  - Decision trees
  - Random forests and boosting

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Beyond Linearity
- Tree-Based Methods
- Support Vector Machines

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Beyond Linearity
- Tree-Based Methods
- Support Vector Machines
- Unsupervised Learning
  - Principal components analysis
  - Independent component analysis

MFDA – II: Contents

- Basics for Machine Learning
- Linear Regression
- Classification
- Resampling Methods
- Linear Model Selection and Regularization
- Beyond Linearity
- Tree-Based Methods
- Support Vector Machines
- Unsupervised Learning
- Typical Supervised Learning Methods

Pattern Recognition (in Artificial Intelligence)

- Everything can be the Pattern!!!

  depends on application fields, pre-knowledge, targets, expectations ... ...

- Pattern

  (I) Statistical properties

  e.g. probability density functions, correlation functions, information content ... ...

  statistics, stochastic process, information theory ... ...

  (II) Geometrical and scale features

  e.g. shapes, (multi-scale) structures, vector/gradient fields ... ...

  spatial-temporal process, transform-/decomposition- based characterization ... ...

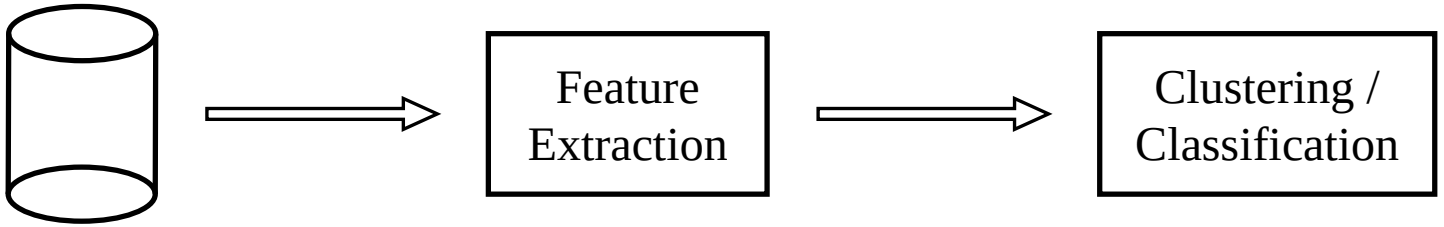
  (III) Recognize the pattern

  (I) + (II) $\xrightarrow[\text{machine learning}]{\text{human learning}}$ (III)

Pattern Recognition (in Artificial Intelligence)

- Everything can be the Pattern!!!

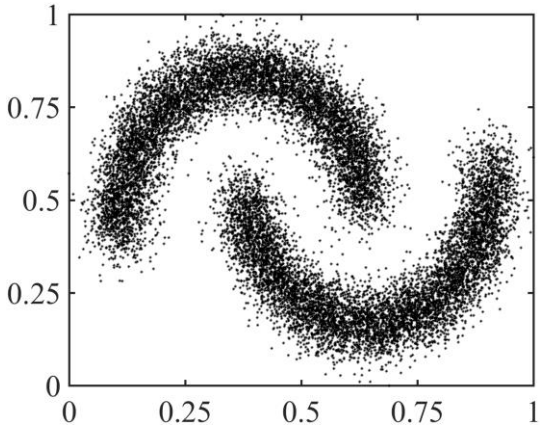
  depends on application fields, pre-knowledge, targets, expectations ... ...

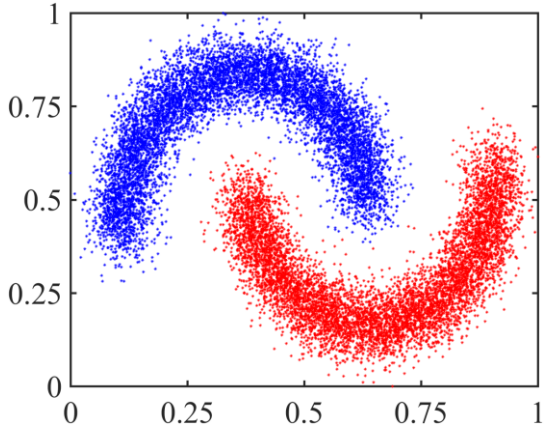


data
$f\colon \mathbb{R}^M \to \mathbb{R}^N$

feature space in $\mathbb{R}^S$

identify patterns

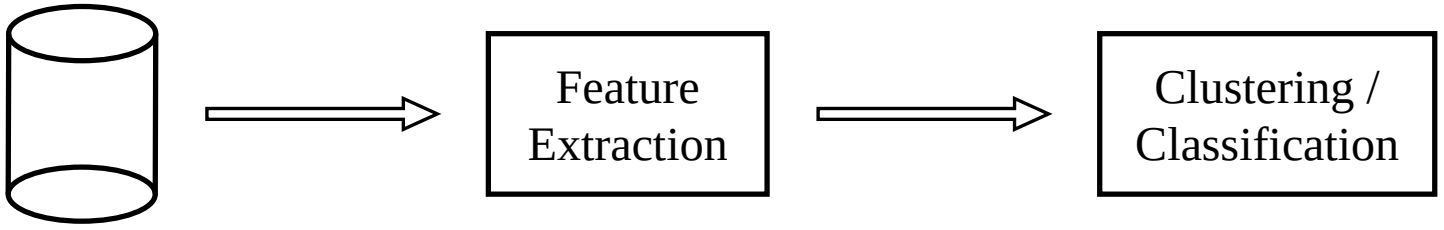




Pattern Recognition (in Artificial Intelligence)

- Everything can be the Pattern!!!

depends on application fields, pre-knowledge, targets, expectations ... ...

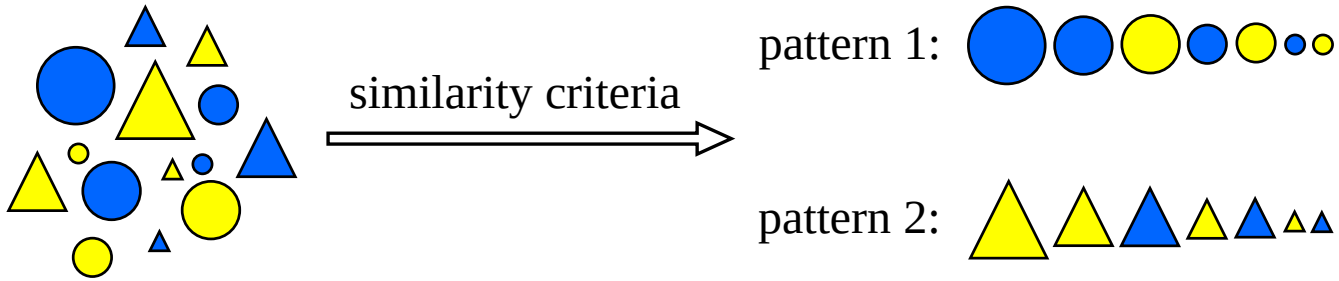


data
$f\colon \mathbb{R}^M \to \mathbb{R}^N$

feature space in $\mathbb{R}^S$

identify patterns

Example: unsupervised learning



- unsupervised learning

e.g. *k*-means clustering
Principle Component Analysis (PCA)
Independent Component Analysis (ICA)

- supervised learning

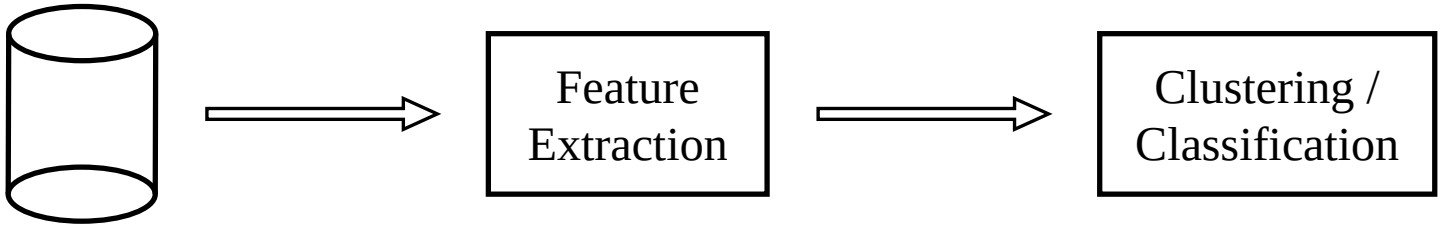
e.g. *k*-Nearest Neighbors (*k*NN)
Support Vector Machine (SVM)
Artificial Neural Network (ANN)

- semi-supervised learning

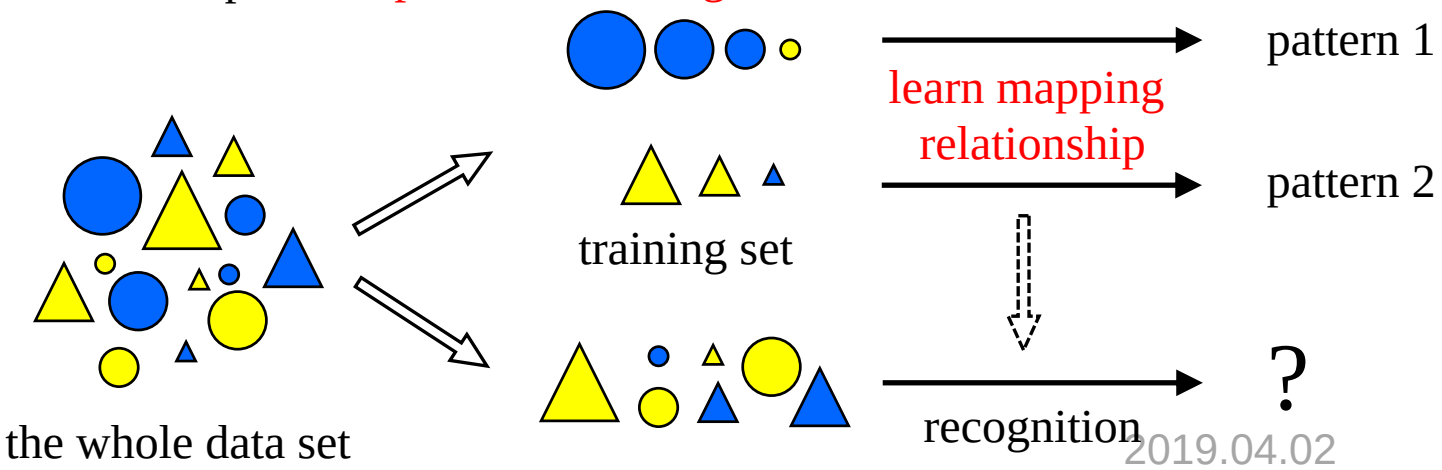
Pattern Recognition (in Artificial Intelligence)

- Everything can be the Pattern!!!

depends on application fields, pre-knowledge, targets, expectations ... ...



Example: supervised learning



- unsupervised learning

e.g. *k*-means clustering
Principle Component Analysis (PCA)
Independent Component Analysis (ICA)

- supervised learning

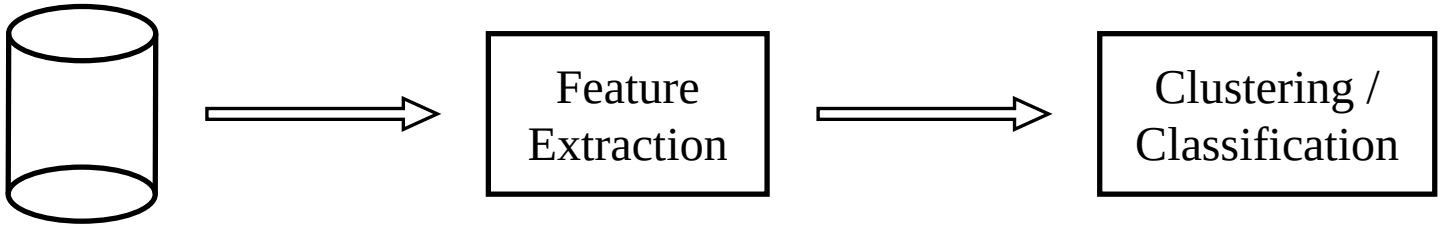
e.g. *k*-Nearest Neighbors (*k*NN)
Support Vector Machine (SVM)
Artificial Neural Network (ANN)

- semi-supervised learning

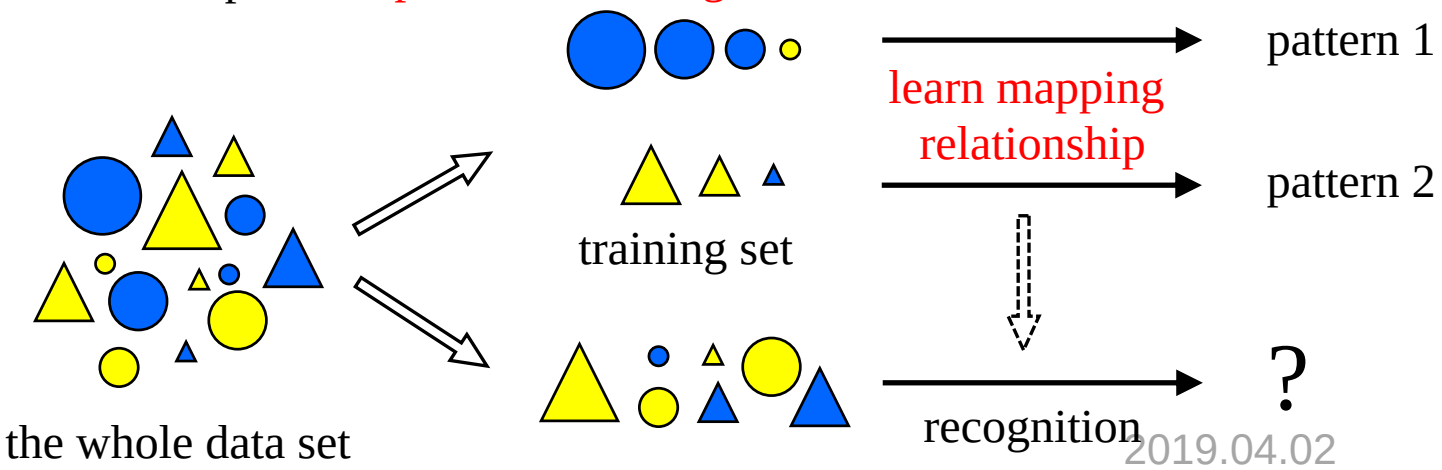
Pattern Recognition (in Artificial Intelligence)

- Everything can be the Pattern!!!

depends on application fields, pre-knowledge, targets, expectations ... ...



Example: supervised learning



- unsupervised learning

e.g. *k*-means clustering
Principle Component Analysis (PCA)
Independent Component Analysis (ICA)

- supervised learning

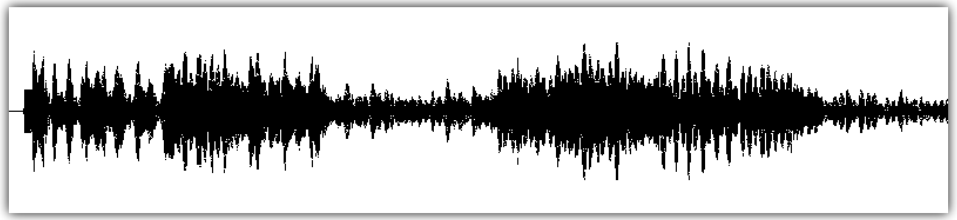
e.g. *k*-Nearest Neighbors (*k*NN)
Support Vector Machine (SVM)
Artificial Neural Network (ANN)

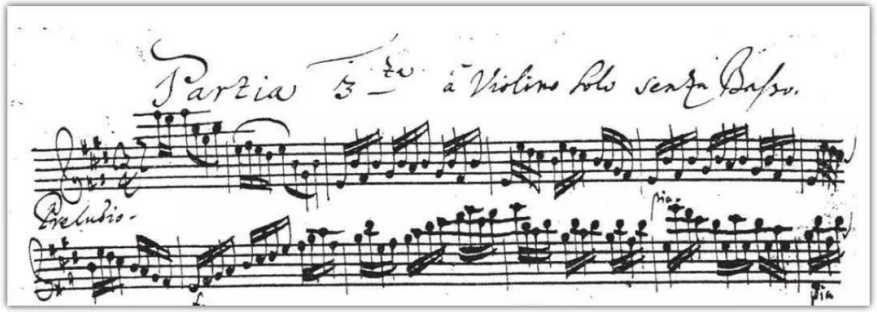
- semi-supervised learning

Pattern Recognition (in Artificial Intelligence)

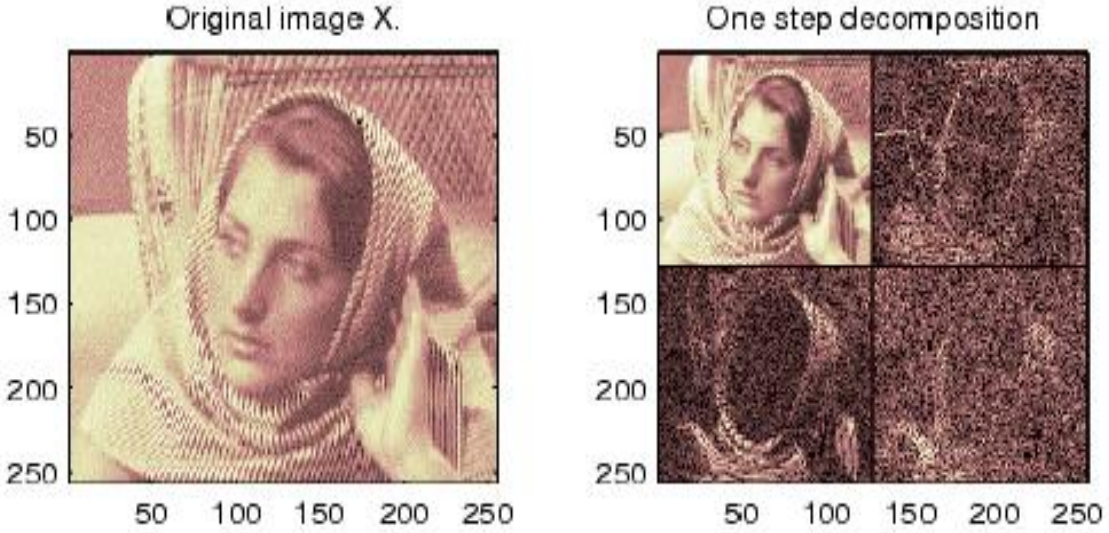
- All the problems come from how to understand the world and represent the world

Audio wave of Bach‘s Partita No.3





2-D Discrete
Wavelet Transform



Pattern Recognition (in Artificial Intelligence)

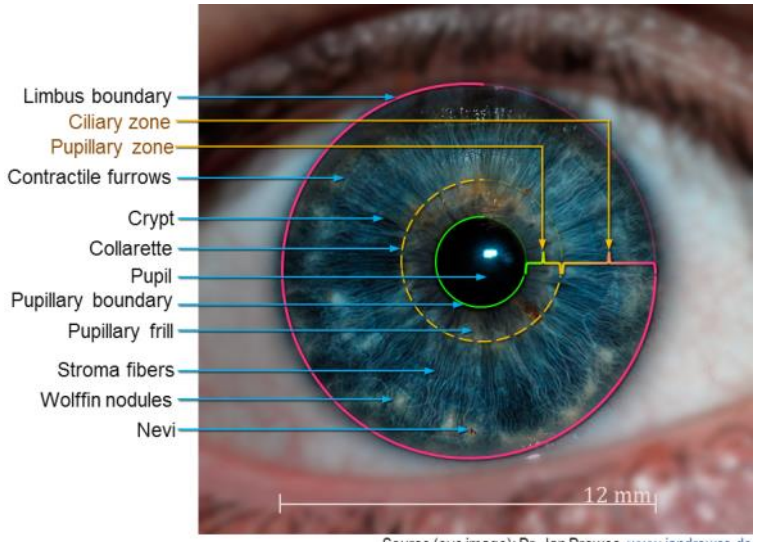
- All the problems come from how to understand the world and represent the world

❖ Understanding

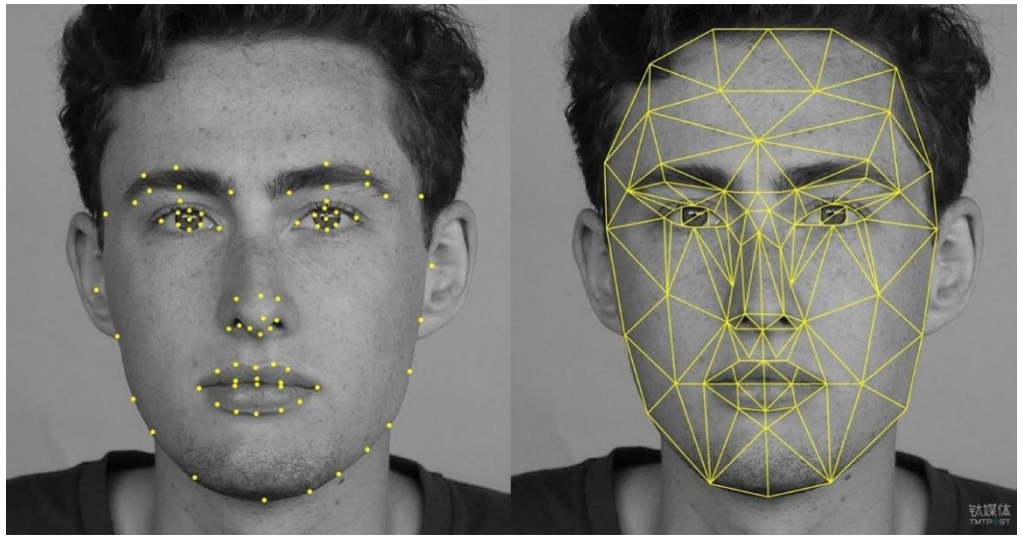
Feature Extraction, Classifiction, Learning, Artificial Intelligence …

Audio Recognition, Image Recognition, Video Review, Information Fusion …

www.studyread.com



Source (eye image): Dr. Jan Drewes. www.jandrewes.de


www.hackernoon.com, @Cole Murray