

# Mining and Processing Biomedical Data

Dr. rer. nat. Krisztian Buza

adiunkt naukowy

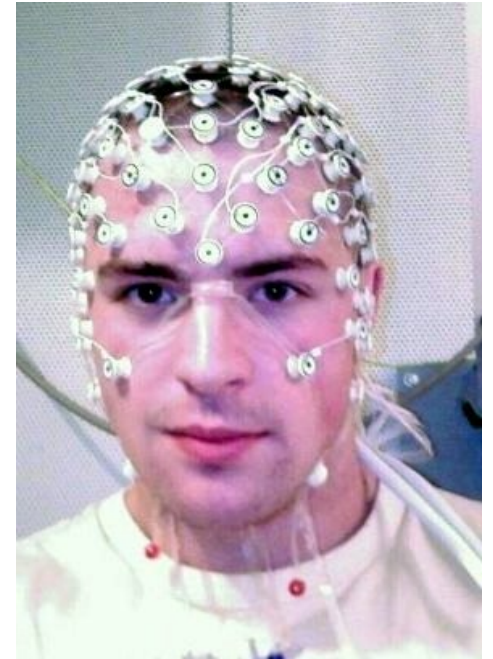
Faculty of Mathematics, Informatics and Mechanics

University of Warsaw, Poland

[chrisbuza@yahoo.com](mailto:chrisbuza@yahoo.com)

# Applications related to EEG signals („brain waves“)

- Medical diagnosis
- EEG-based tools for paralyzed patients
  - EEG-controlled spelling device
  - EEG-controlled web browser
- Study sleepiness in long-distance track driving
- ... and many more ...
- Data processing and mining techniques are relevant to such applications



Public domain image from Wikipedia

**What is Data Mining?**

# Data Mining and Machine Learning

- Data Mining: a collection of data analysis techniques that aim to find hidden patterns, hidden knowledge in huge datasets
- Machine Learning: computer „learns“ from the data, identifies patterns, regularities; for example:
  - Correlations between variables (temperature and the amount of ice cream that is sold)
  - Identification of groups of patients / objects
  - Characterization of the differences between groups, for example in form of rules that describe how to decide to which group an object belongs

E.B. Celikkaya, C.R. Shelton, D. Kale, R.C. Wetzel, R.G. Khemani:

## ***Non-invasive Blood Gas Estimation for Pediatric Mechanical Ventilation***

Machine Learning for Clinical Data Analysis and Healthcare,  
NIPS Workshop 2013

- “ventilator management for children with Acute Lung Injury or Acute Respiratory Distress Syndrome”
- “ventilator machine is used to support breathing in a pediatric intensive care unit”
- adapt ventilation settings according to individual needs, because this “can lead to faster and safer weaning of the patient”
- Parameters of the ventilation that have to be set: ventilator rate, tidal volume, peak inspiratory pressure

E.B. Celikkaya, C.R. Shelton, D. Kale, R.C. Wetzel, R.G. Khemani:

# ***Non-invasive Blood Gas Estimation for Pediatric Mechanical Ventilation***

Machine Learning for Clinical Data Analysis and Healthcare,  
NIPS Workshop 2013

- Parameters can be set based on:
  - “how much oxygen is delivered to the lungs”
    - non-invasive measurement = „simple“ to measure
  - “how much carbon dioxide is cleared in the lungs”
    - can be determined based on arterial pH and the partial pressure of carbon dioxide
    - require invasive measurements = „hard“ to measure

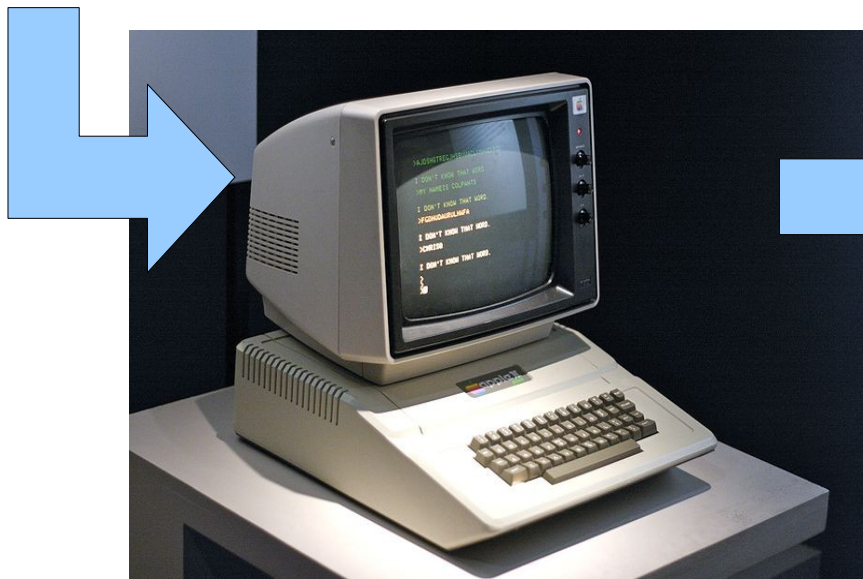
E.B. Celikkaya, C.R. Shelton, D. Kale, R.C. Wetzel, R.G. Khemani:

## ***Non-invasive Blood Gas Estimation for Pediatric Mechanical Ventilation***

- Concentration of carbon dioxide exhaled from the lungs can be measured non-invasively (i.e., „simply“) and it is closely related to the partial pressure of carbon dioxide
- But: we do not know the exact formula
- The partial pressure of carbon dioxide may depend on other things too
- Similar situation with arterial pH
- Use a computer software to find **how** the partial pressure of carbon dioxide and arterial pH depend on other things that can be measured „simply“ (such as the concentration of carbon dioxide exhaled from the lungs)

A, B – parameters that are hard to measure  
C, D, E – parameters that are simple to measure

	A	B	C	D	E
Day 1	2.5	4.3	32	6.2	7.5
Day 2	-	-	36	5.9	7.2
Day 3	-	-	33	6.1	6.9
Day 4	4.8	6.1	37	5.8	7.7
Day 5	-	-	39	5.9	7.5



$$A \approx 0.5 * C - (D+E)$$
$$B \approx 0.6 * C - (D+E) + 0.9$$

Image from Wikipedia is used on this slide. Licence info:  
[http://commons.wikimedia.org/wiki/File:Apple\\_II.jpg](http://commons.wikimedia.org/wiki/File:Apple_II.jpg)



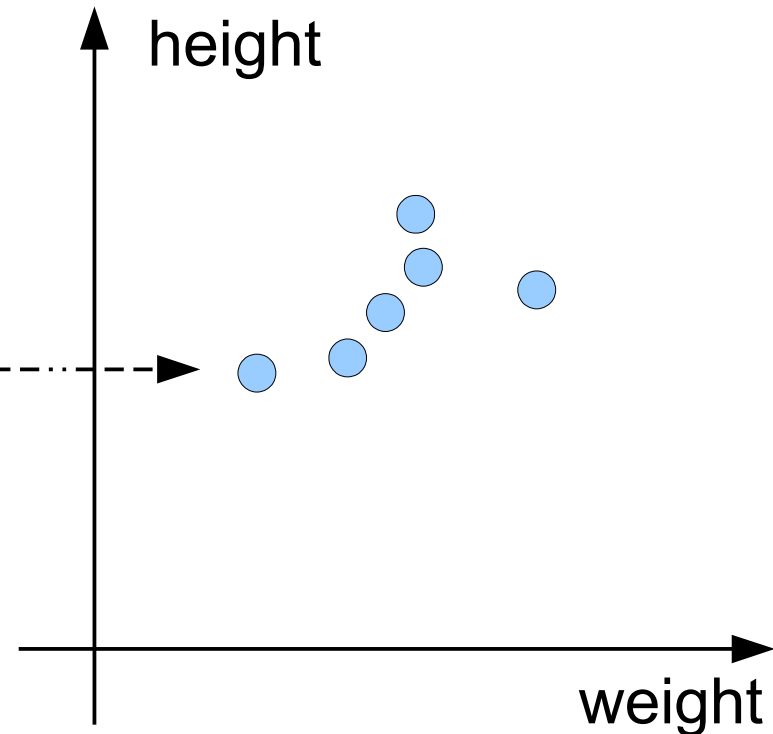
# Data Mining Techniques

# Basic concepts

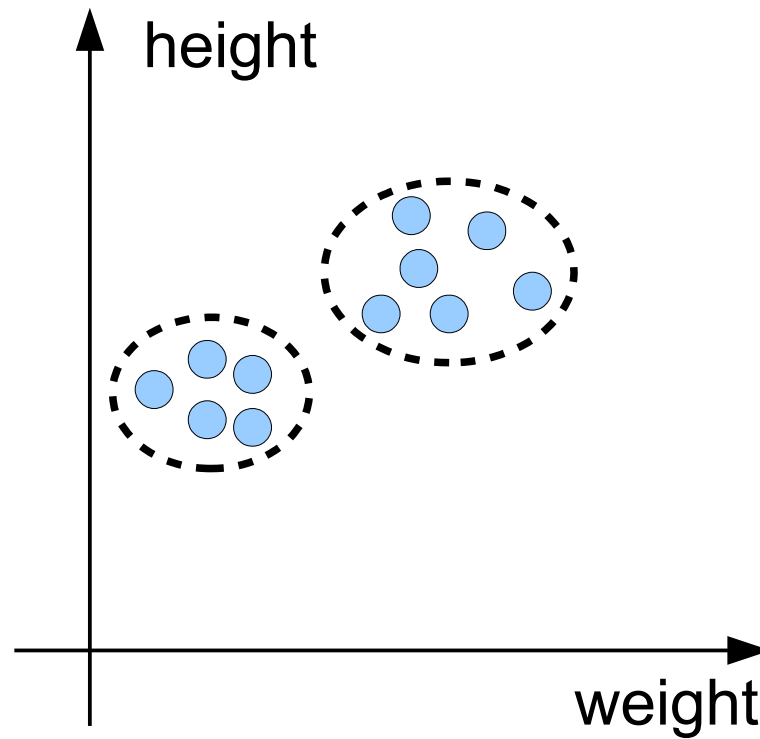
Attributes / Features

Instances / Objects

	Weight (kg)	Height (cm)
Patient 1	85	182
Patient 2	103	178
Patient 3	45	165
Patient 4	72	168
Patient 5	80	175
Patient 6	83	191
...	...	...

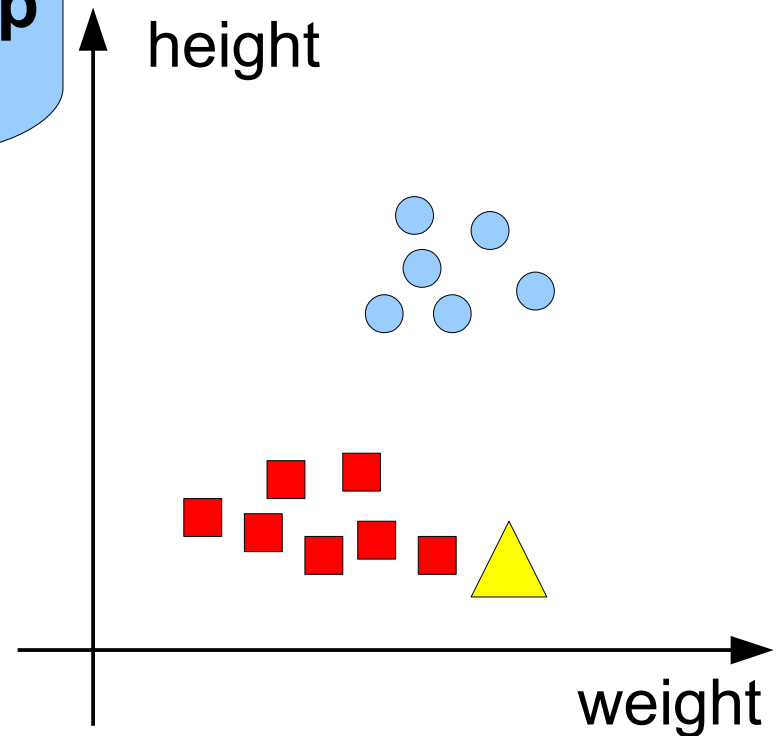


# Clustering

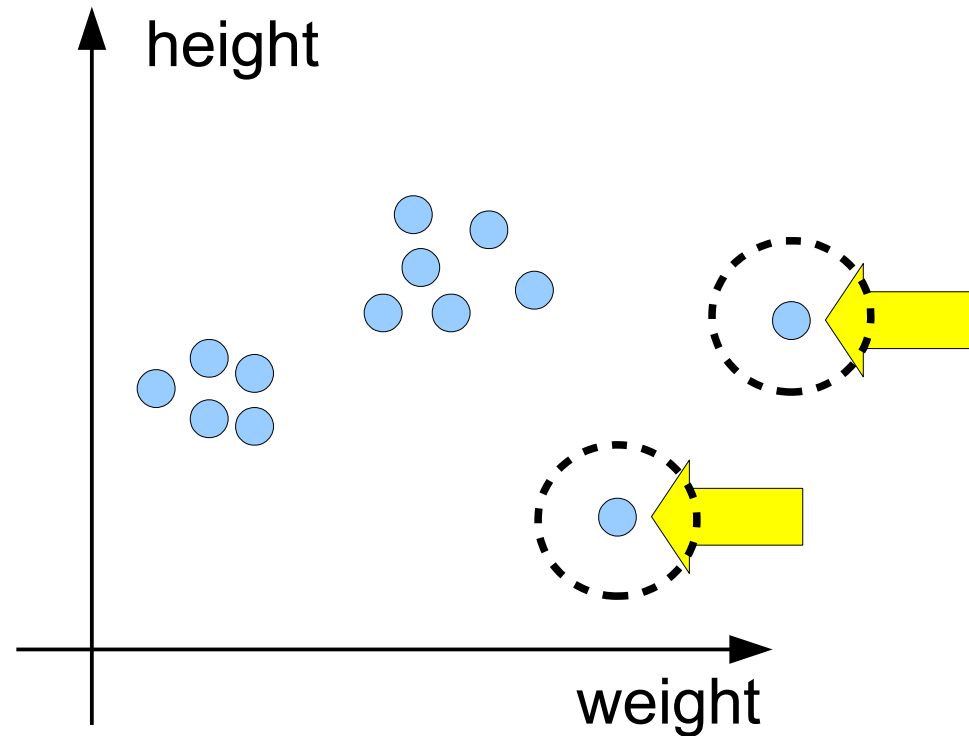


# Classification

**The yellow triangle  
most probably  
belongs to the group  
of red rectangles**



# Outlier detection (anomaly detection)



# Data Mining Techniques

- Classification
- Clustering
- Anomaly detection / Outlier detection
- Instance selection / Prototype selection
- Frequent pattern mining / Frequent itemset mining
- Regression
- Matrix completion (matrix factorization)