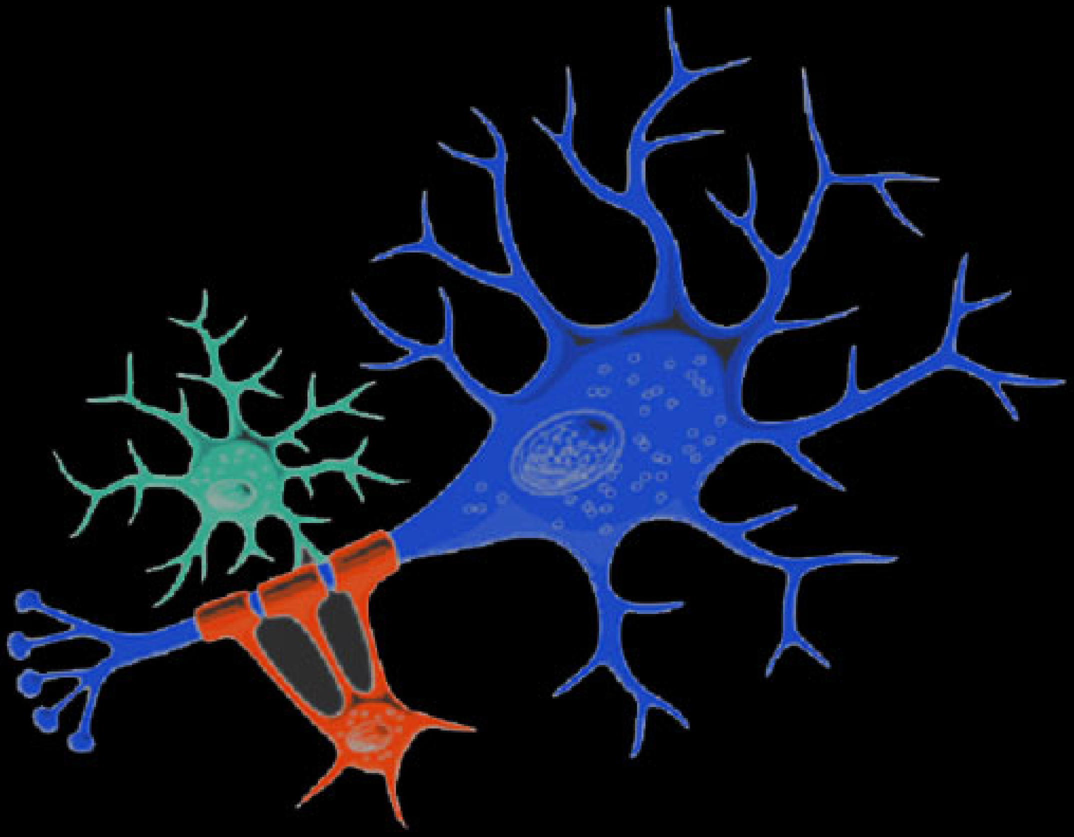
Pattern recognition methods for electronic tongue systems



Patrycja Ciosek

Electronic tongue

A SYSTEM FOR AUTOMATIC ANALYSIS AND CLASSIFICATION (RECOGNITION) OF LIQUID SAMPLES

AN ARRAY OF CHEMICAL SENSORS

PATTERN RECOGNITION SYSTEM

PROPER WORK OF SENSOR ARRAYS

The proper work of the sensor systems involves the application-oriented selection of:

- sensing materials and transducers,
- features, i.e. steady-state outputs, dynamic response descriptors or linear/nonlinear combinations of them,
- pattern recognition **(PARC)** tools.

PARC – why it is not so simple?

- **There is no single, optimal procedure** to analyze sensor array data (giving both reliable and low level of error results, which can be applied in any classification problem)
- **Strong local differences** in pattern space result in diversity of performance of various classifiers in pattern subspaces
- Fusion of linear and nonlinear processing closer simulates the **natural signal processing**

Learning by example

- All pattern recognition techniques are based on **learning by example** => having a set of feature patterns of known class (learning set), the classifier system is learned to give corresponding class membership responses. When satisfactory level of error for the object from learning set is obtained, class membership of **unknown vector “x”** can be estimated.

Overtraining of the model

- even if the error of learning samples classifications reaches 0, there is probability that samples with **not identical but similar** characteristics can be misclassified.
- This is due to the **overtraining of the model**.
- Too complex classifiers model noise in the data set, which results in **poor generalization ability** and incorrect classifications of the test data.

The independence of the learn and test set

- **The independence of the training and test** samples is often overlooked in practice
- The testing set should embrace samples with **similar but not identical** characteristics with the learning ones
- In the foodstuff analysis this can be achieved by choosing testing samples originating from **different manufacture lot/lots** than those used in the learning phase of the system
- The training data should exhibit **representative features** for the population of its class.

Data Analysis

- Data matrix
- Preprocessing
- PCA
- ANN

Data Analysis - **Data matrix**

- Pattern or **“chemical image”** of the analyzed sample denotes the **multidimensional vector x=(x1, x2, ..., xn)**
- **Its** components are features obtained from measurements and specified by the investigator
- His/her role is to choose particular variables obtained from the measurements, or their combinations, in order to enable suitable description of the object and its correct classification to one of the previously established classes
- Each sample is characterized by unique and typical set of data, forming **"fingerprint"** of an analyte in m-dimensional pattern space.

Data Analysis - **Data matrix**

Following sensor signals

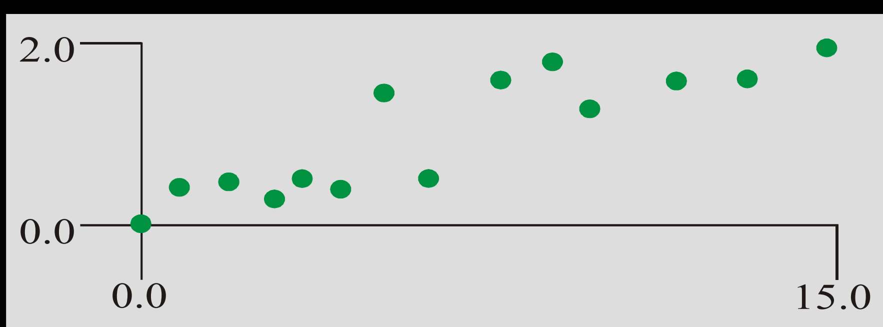
$$X = \begin{bmatrix} x_{11} \dots x_{1j} \dots x_{1m} \\ \dots\dots\dots\dots\dots\dots \\ x_{i1} \dots x_{ij} \dots x_{im} \\ \dots\dots\dots\dots\dots\dots \\ x_{n1} \dots x_{nj} \dots x_{nm} \end{bmatrix}$$

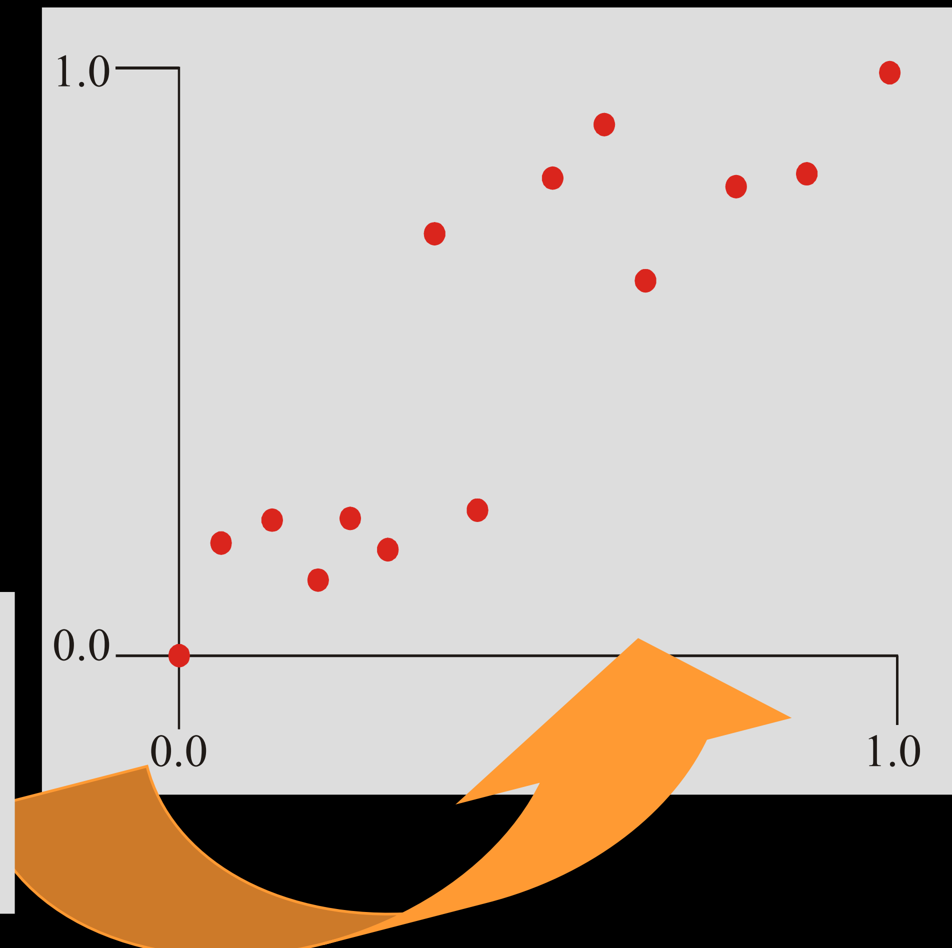
Pattern vector

Following samples

Data Analysis - **Preprocessing**

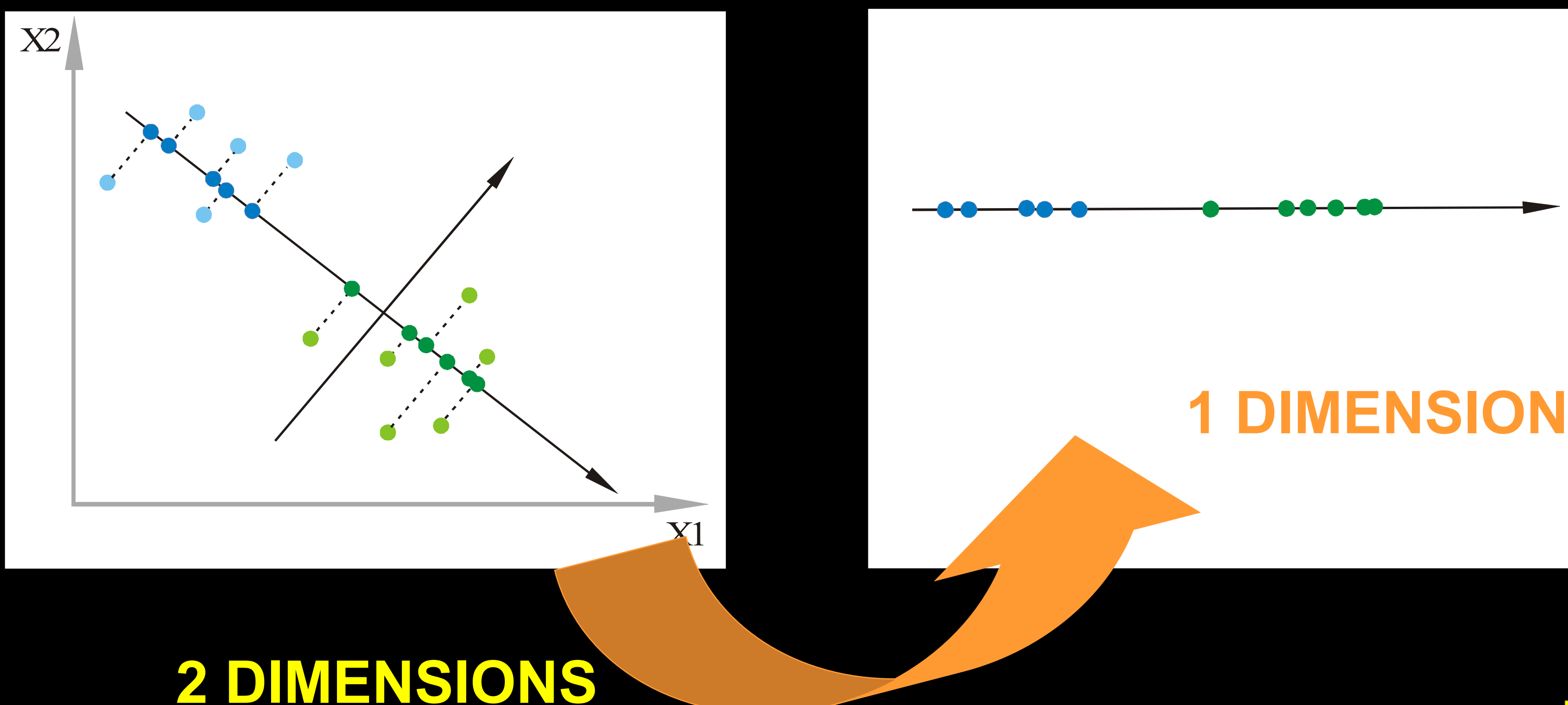
- Specific transformations
- Meancentering
- Autoscaling
- ....



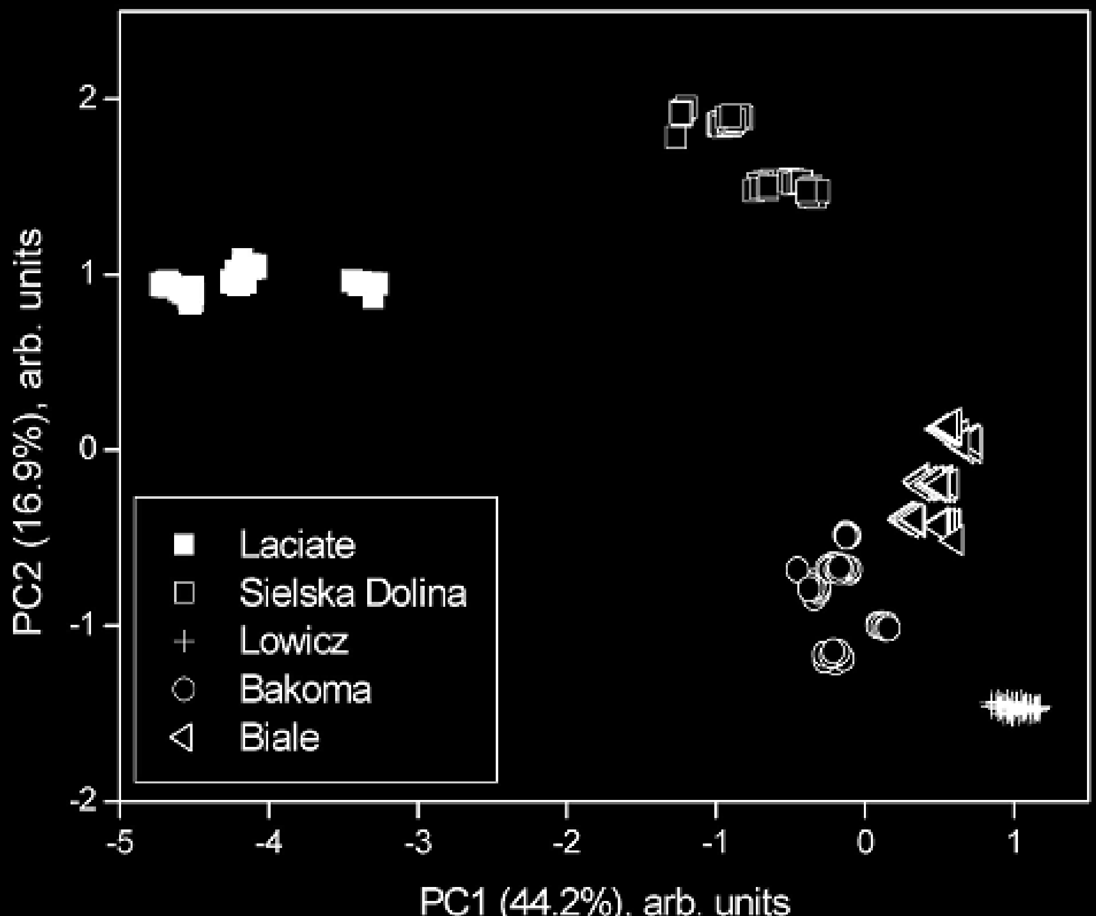


Data Analysis –
Principal Components Analysis (PCA)

linear feature-extraction technique finding most influential, new directions in the pattern space, explaining as much of the variance in the data set as possible.



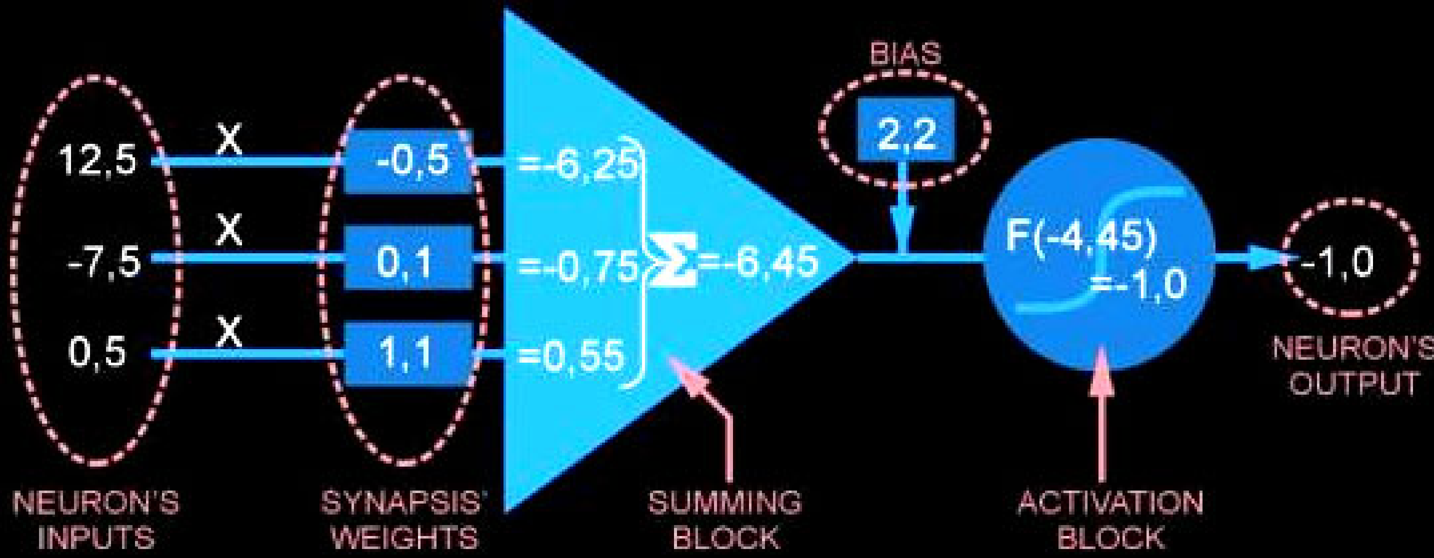
Data Analysis –
Principal Components Analysis (PCA)



PCA plot of milk measurements

(electronic tongue developed at Warsaw University of Technology)

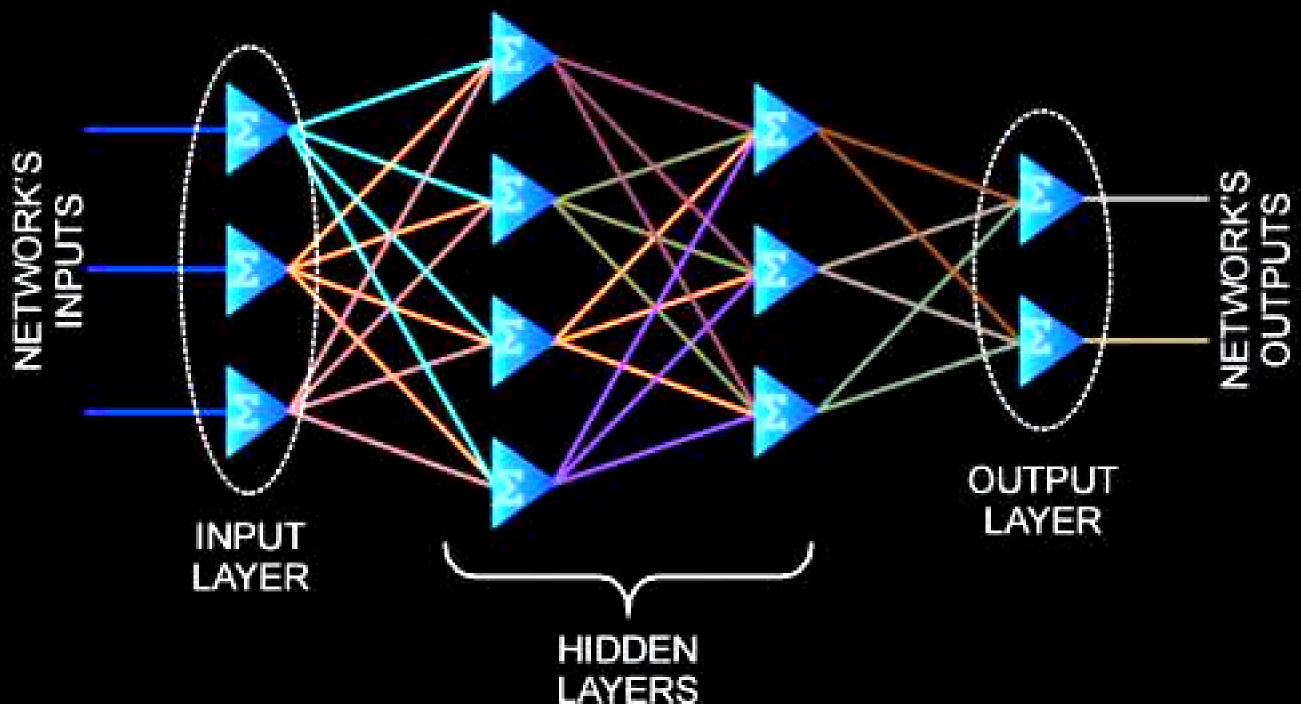
Data Analysis –
Artificial Neural Network (ANN)



Single neuron

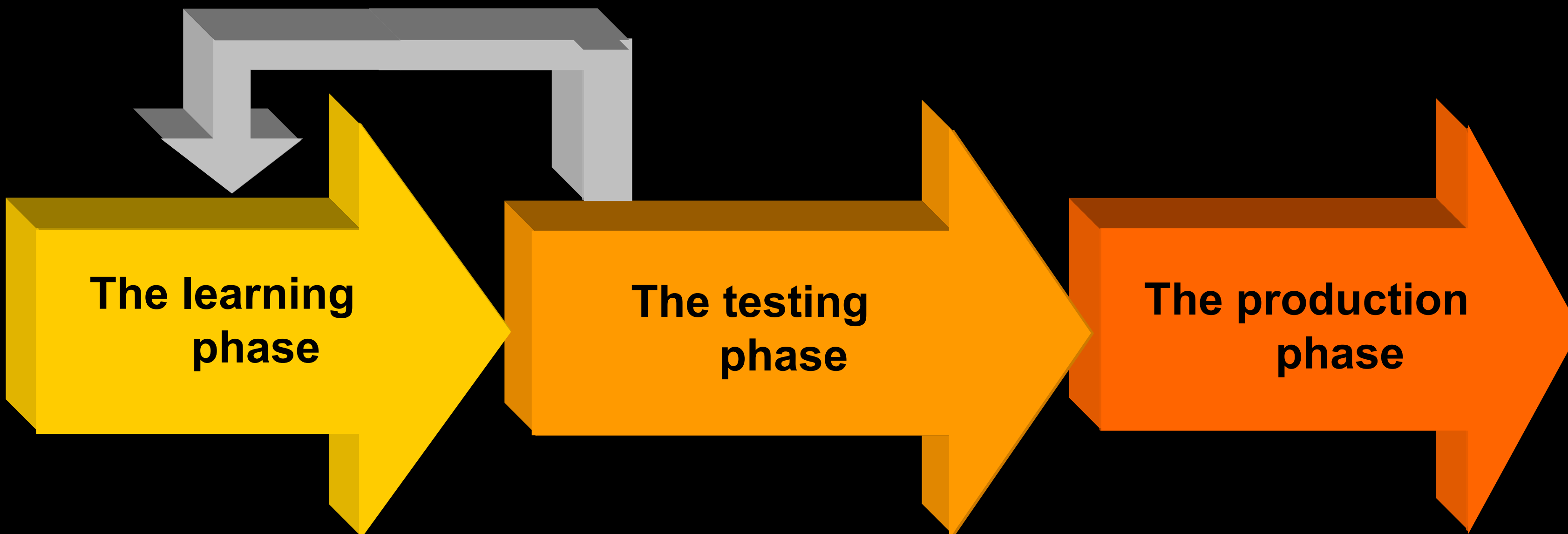
Data Analysis –
Artificial Neural Network (ANN)

The most common PARC tool used for electronic tongue and nose applications!!!



Patrycja Ciosek

Data Analysis –
Artificial Neural Network (ANN)



- network is forced to provide desired outputs corresponding to a determined input
- adjusting the synapses' weights

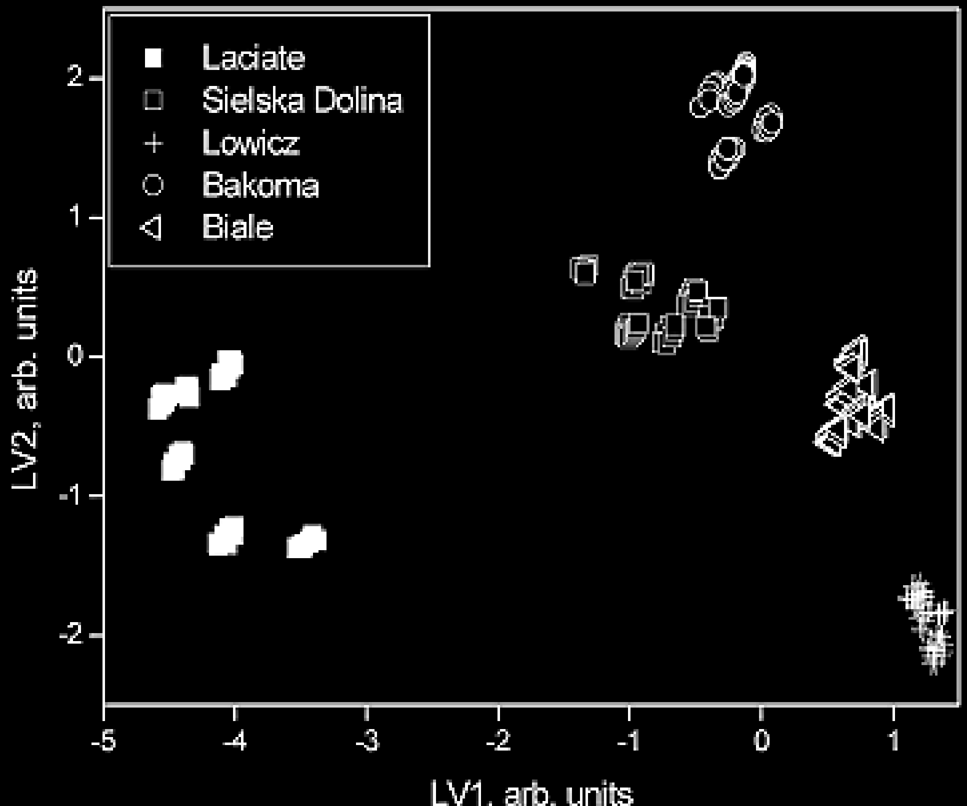
✓verification of the generalization capability of network

✓network is capable of providing outputs corresponding to any input

Partial Least Squares (PLS)

- PLS is a supervised one-step method
- besides ANN is also very often utilized in sensor array data processing
- PLS determines a set of latent variables, corresponding to principal components in PCA but explaining as much as possible of the covariance between the data set and class affiliation matrix.
- Practical and common way to determine complexity of the PLS model, i.e. the number of latent variables constructing the data set, is to estimate their predictive power by cross validation (CV).

Partial Least Squares (PLS)



PLS plot of milk measurements

(electronic tongue developed at Warsaw University of Technology)

Soft Independent Modeling of Class Analogy (SIMCA)

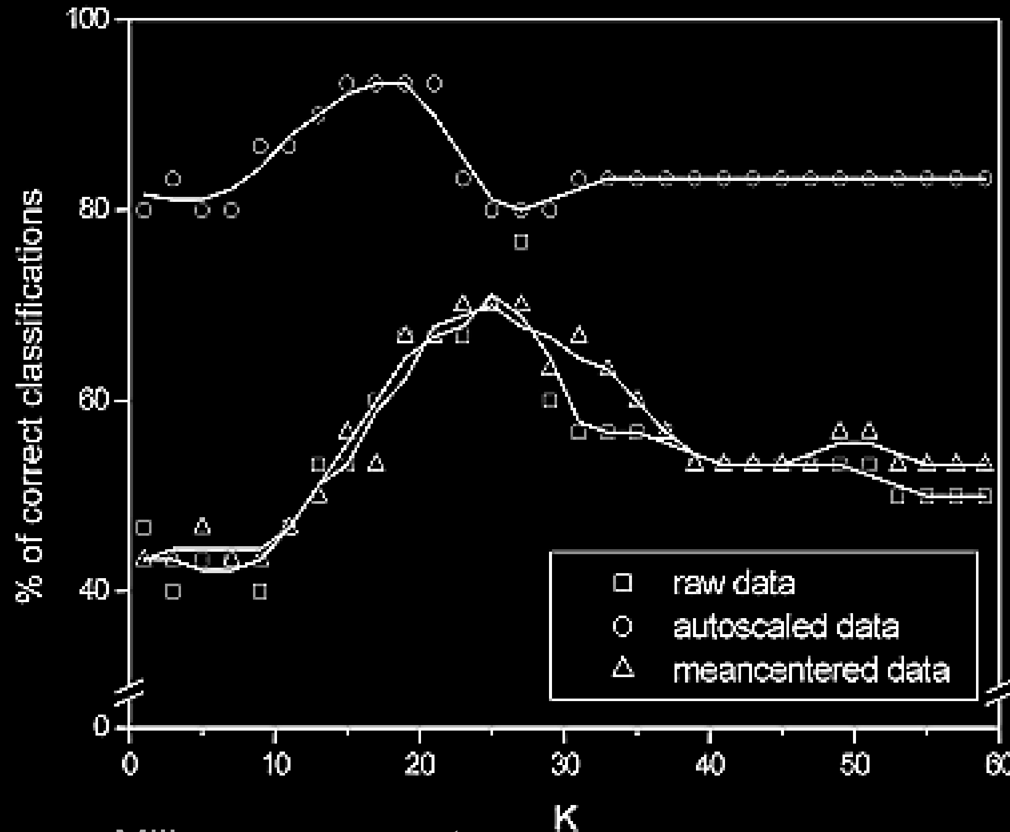
- PCA-based tool with supervised classification possibility
- It considers each class separately i.e. for each class PCA is performed to produce so called disjoint class models
- These models are used to predict test set class memberships
- A test sample is assigned to the class established in the model, when it falls within its class boundaries
- The selection of number of PCs forming each class model strongly influences the classification results

K-Nearest Neighbors (KNN)

- **Minimal distance classifier** => based on the rule, that similar patterns are assigned as belonging to one class
- KNN algorithm provides classification of pattern vectors based only on calculated distance between points in the pattern space
- **None training** is necessary
- the only user specified parameters are **distance metric** (usually Euclidean distance) and the **value of K** (equals to the number of closest neighbors among which the majority indicates the class affinity of the test object)
- The 1-NN rule, which assigns patterns to the class of the nearest training pattern is the most straightforward method => it provides reliable and application independent estimation of classification ability with the particular feature sets => it can be used as a benchmark and reference for the other classifiers.

KNN

- Usually increasing the value of K involves also the increase in the classification performance, because of various spread of data belonging to various classes
- After reaching certain point (optimal value of K) the performance decreases assimptotically.



Milk measurements
(electronic tongue developed at Warsaw University of Technology)

Learning Vector Quantization (LVQ)

- **Minimal distance classifier** => based on the rule, that similar patterns are assigned as belonging to one class
- neural networks with a **competitive layer**
- can be regarded as a **supervised version** of Self-Organizing Maps
- main attribute is the **distance between input vectors**
  - Responses are approximated by quantized references, so-called codebook vectors
  - Several codebook vectors are assigned to each class of objects
  - New pattern is classified to the class to which the nearest codebook vector belongs
  - First layer, so-called competitive, learns to classify input vectors and the second one, so-called linear, transforms the competitive layer's classes into target classifications defined by the user

Learning Vector Quantization

- The main advantages of LVQ nets are

  faster training

  high ability to generalize features in the data set with robustness

- they can approximate almost every classification problem
- they are **not limited in the number of dimensions** (in contrast to KNN)
- Model accuracy is **strongly dependent on the the data distribution** and learning parameters used
- LVQ **may give more reliable results** with the sensor array data, that commonly used pattern recognition tools (such as ANN)

Comparison of PARC tools

	PARC tool				
	KNN	PLS	SIMCA	ANN	LVQ
Learning phase	no	yes	yes	yes	yes
Speed of performance	slow	fast	fast	fast	fast
Linearity	linear	linear	linear	nonlinear	nonlinear
Overfitting sensitivity	-	yes	sometimes	easily trapped in local minima	sometimes*
	Type of preprocessing				
User specified parameters of the model	K distance metric***	number of LVs	number of PCs for each class	net architecture, learning rate, momentum coefficient, number of epochs	net architecture, learning rate

References:

P. Ciosek, W. Wroblewski, The analysis of sensor array data with various pattern recognition techniques, Sens. Actuators, B, in press

W. Gopel, Chemical imaging: I. Concepts and visions for electronic and bioelectronic noses, Sens. Actuators, B, 52 (1998) 125-142.

R. E. Shaffer, S. L. Rose-Pehrsson, R. A. McGill, A comparison study of chemical sensor array pattern recognition algorithms, Anal. Chim. Acta, 384 (1999) 305-317.

A. K. Jain, R. P. W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Transactions on pattern analysis and machine intelligence, 22 (2000) 4-37.

P. Ciosek, Z. Brzózka, W. Wróblewski, E. Martinelli, C. Di Natale, A. D'Amico, Direct and two stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue – effect of supervised feature extraction, Talanta, in press.

A. Ortega, S. Marco, T. Sundic, J. Samitier, New pattern recognition systems designed for electronic noses, Sens. Actuators, B, 69 (2000) 302-307.

D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a textbook, Elsevier, Amsterdam, 1998, p. 44-45, 385-412.

S. Osowski, Neural Networks for information processing, Oficyna Wydawnicza Politechniki Warszawskiej, Warsaw, 2000, p. 20-40 (in Polish).

A. Webb, Statistical Pattern Recognition, 2nd Ed., Wiley, New York, 2002, p. 390.

M. Bicego, G. Tessari, G. Tecchilli, M. Bettinelli, A comperative analysis of basic pattern recognition techniques for the development of small size electronic nose, Sens. Actuators, B, 85 (2002) 137-144.

F. Camastra, A. Vinciarelli, Cursive character recognition by learning vector quantization, Pettern Recognition Letters, 22 (2001) 625-629.