Faculty: BioScienze e Tecnologie Agro-Alimentari e Ambientali
MASTER DEGREE IN FOOD SCIENCE AND TECHNOLOGY
I YEAR

Course:
EXPERIMENTAL DESIGN AND CHEMOMETRICS IN FOOD
(5 credits – 38 hours)

Teacher: Marcello Mascini

(mmascini@unite.it)

The Teacher is available to answer questions at the end of the lesson, or on request by mail

The course is split in 4 units

UNIT 1: statistical regression

Data, information, models, data types, analytical representation of data

Calibration and regression, Introduction to Statistics

Average & Variance

The Normal distribution, theory of measurement errors, the central limit theorem and the theorem of Gauss

Maximum likelihood, method of least squares, Generalization of the method of least squares

Polynomial regression, non-linear regression, the $\chi 2$ method, Validation of the model

UNIT 2: Design of Experiments

Basic design of experiments and analysis of the resulting data

Analysis of variance, blocking and nuisance variables

Factorial designs

Fractional factorial designs

Overview of other types of experimental designs (Plackett–Burman designs, D-optimal designs, Supersaturated designs, Asymmetrical designs)

Response surface methods and designs

Applications of designed experiments from various fields of food science

UNIT 3: Data Matrices and sensor arrays

Correlation

Multiple linear regression

Principal component analysis (PCA)

Principal component regression (PCR) and Partial least squares regression - (PLS)

UNIT 4: Elements of Pattern recognition

cluster analysis

Normalization

The space representation (PCA) Examples of PCA

Discriminant analysis (DA) PLS-DA

Examples of PLS-DA

UNIT 4: Elements of Pattern recognition

cluster analysis

Normalization

The space representation (PCA)
Examples of PCA

Discriminant analysis (DA) PLS-DA

Examples of PLS-DA

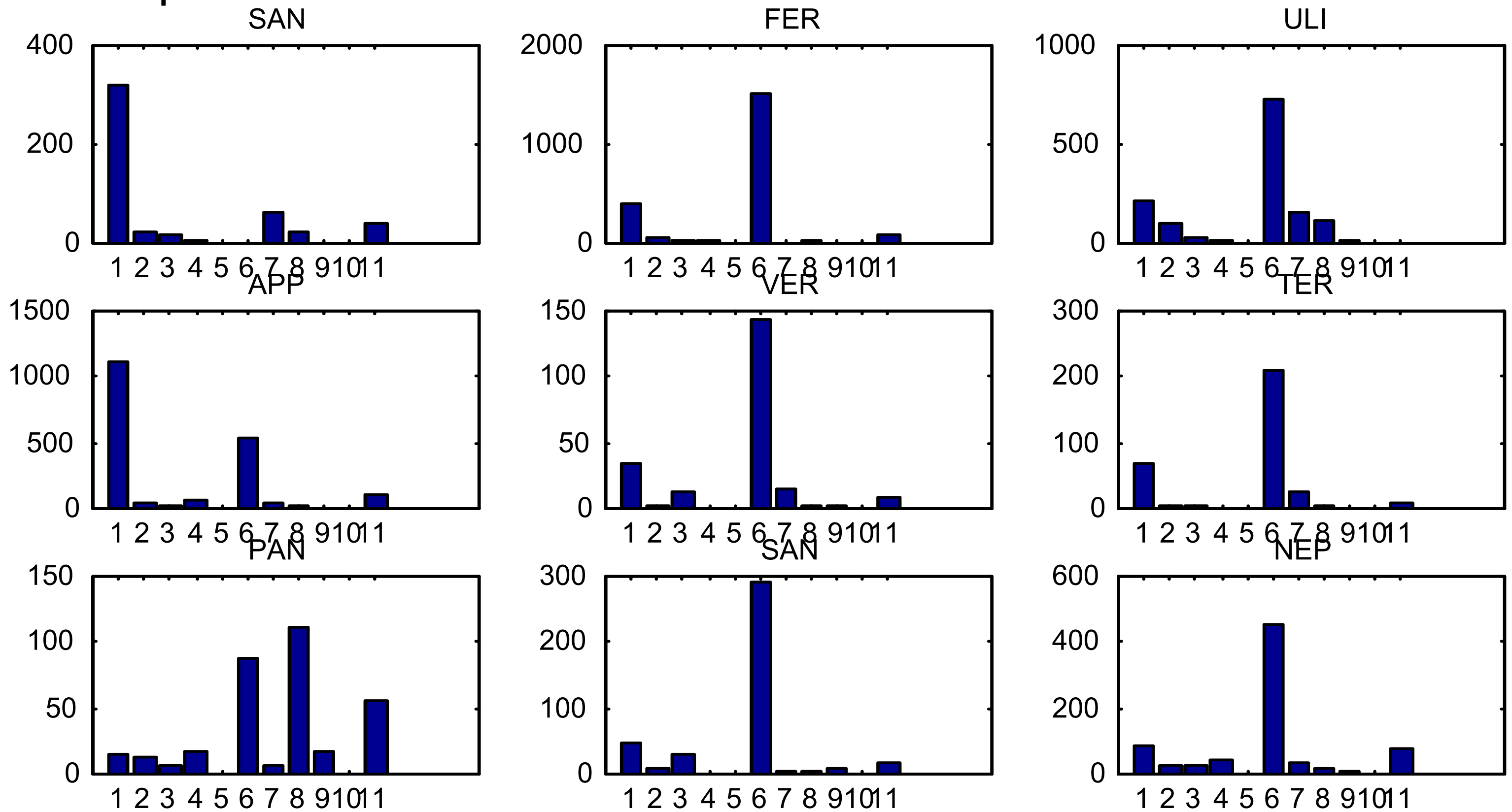
Definitions

- Pattern: set of features defining the properties of a complex object
- Class: a set of objects having some important properties in common
- Classifying: mathematical operation for which a sample, described by a number of features, is assigned to a particular class.
- The set of features is called Pattern, the classification operation is called Pattern Recognition.
- The relationship between the sample and the class is not explicit, it depends on the features chosen to describe the objects
- The pattern recognition problem is "interesting" when the individual features are not able to identify samples

- Fruits:
- Features: weight, shape, color, sugar, acid, ...

Pattern of mineral water

- Ionic profile of mineral water



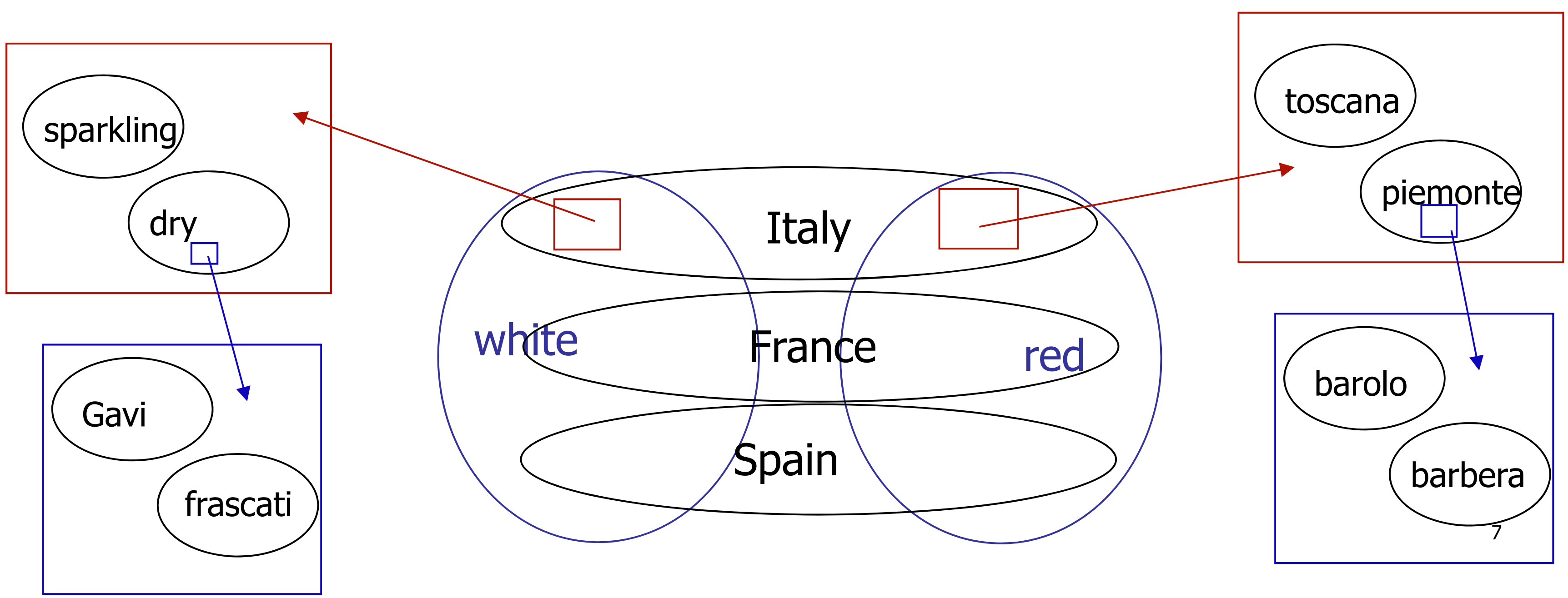
Ca	Na	Mg
K	NH4	HCO
SO4	Cl	NO3
F	SiO	

Pattern analysis

- The pattern recognition is a method to have information on the sample described by patterns
- Mathematically, a pattern is a vector that corresponds to the feature describing some aspects of a sample
- The features are linked to the type of information you want to be described
- The last operation of the pattern recognition is to assign a sample to a class (membership class)

Membership class

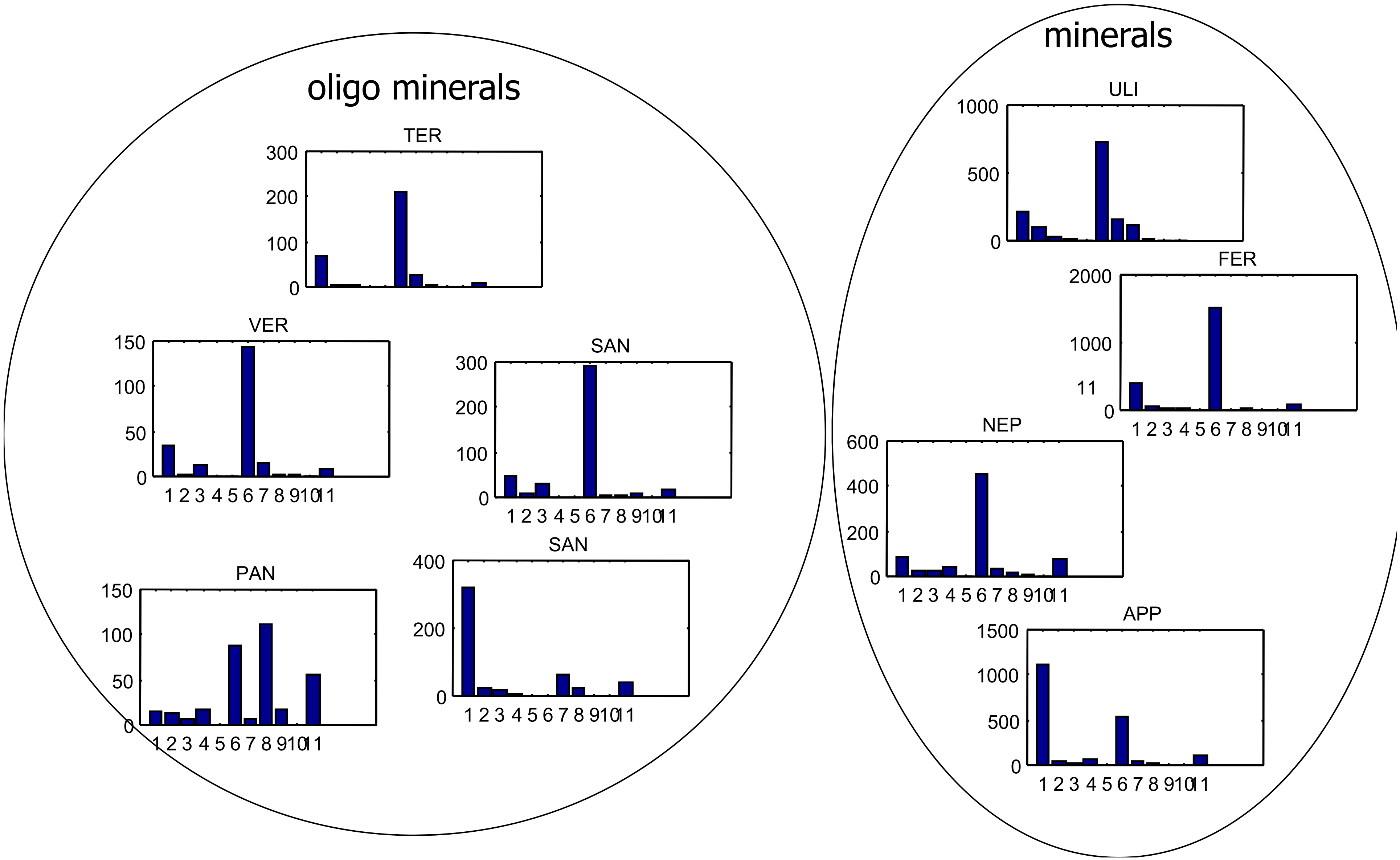
- The membership class is a theoretical set of elements sharing a global feature
- Items can be grouped according to different classification schemes depending on the global feature
- Example:
  - Wine classification



Patterns and membership class

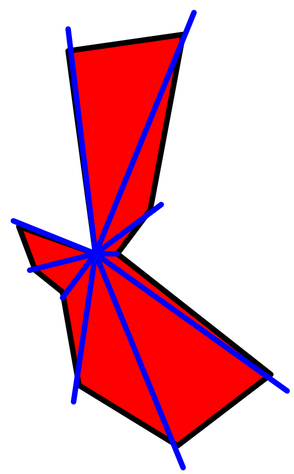
- If the features are appropriate for the classification then the elements belonging to a same class have similar patterns.
- The similarity between patterns can be detected with a visual inspection of suitable graph representation
- The simplest: column chart and radar-plot

Column chart and membership class

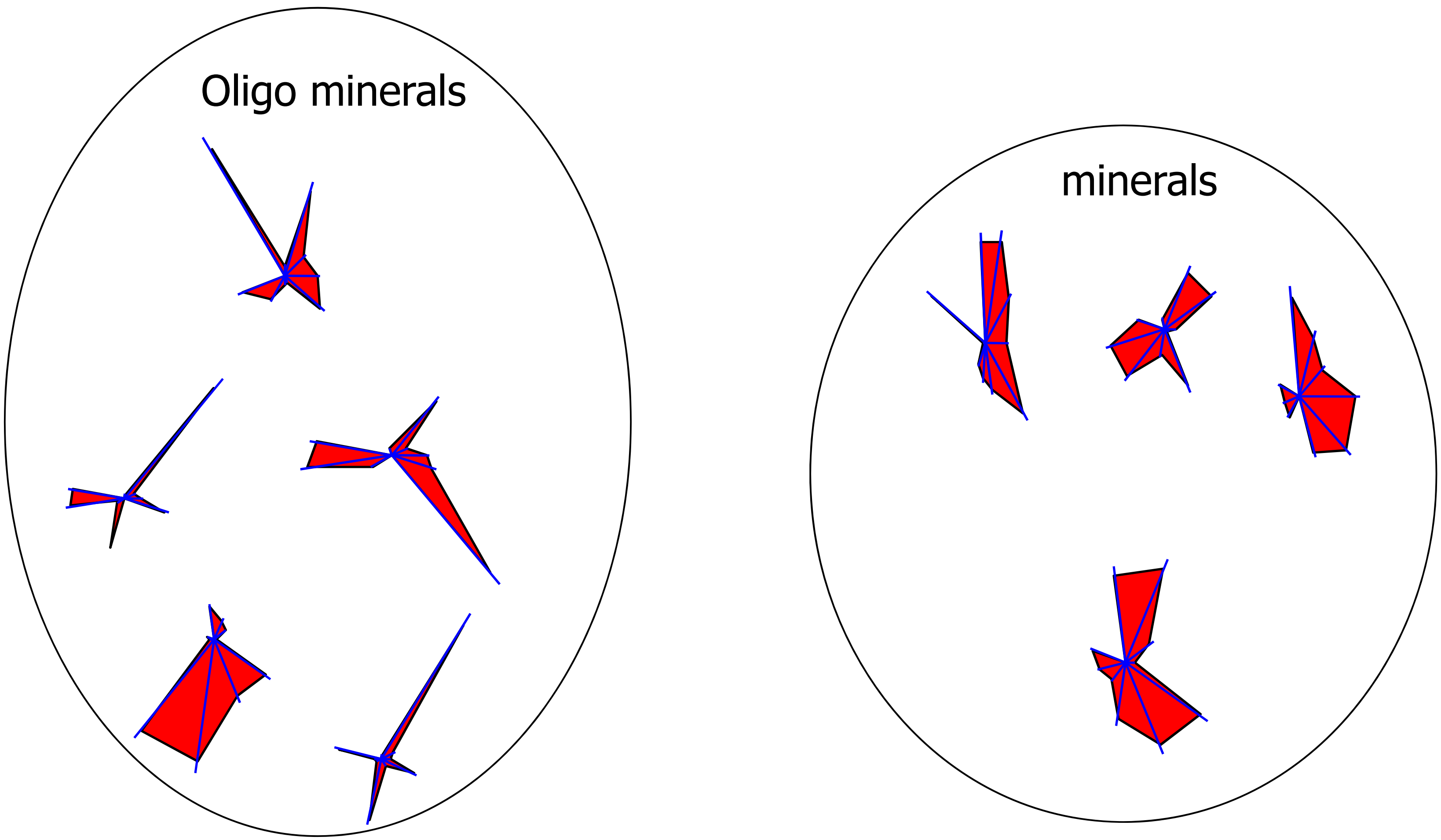


Radar plot

- A simple way to visualize multidimensional patterns
- the pattern variables are the same of the directions (axes)
- The axes are organized to form a circle
- Along each axis is plotted the value of the variable
- Joining the points on the axes you get a figure that forms the "profile" of the pattern
- often used in sensory analysis to define the sensory profile of foods.



Radar plot and class membership



distance Criteria

- Each pattern is a point in N-dimensional space
- The space is defined by the features that describe the pattern
- For each feature is assigned an axis
- All axes define an Orto-normal basis
- "class-membership" – distance relation
- Two close points (patterns) probably belong to the same class two far points belong to different classes

Classification criteria

- Criteria "unsupervised"
- Determining, on the basis of an a priori criterion, an internal classification scheme
- The criterion used is generally that of the distance

- Clusters of analysis
- hierarchical method used to form classes with more and more undefined.
- Exotic Methods
- Potential Method

Cluster Analysis

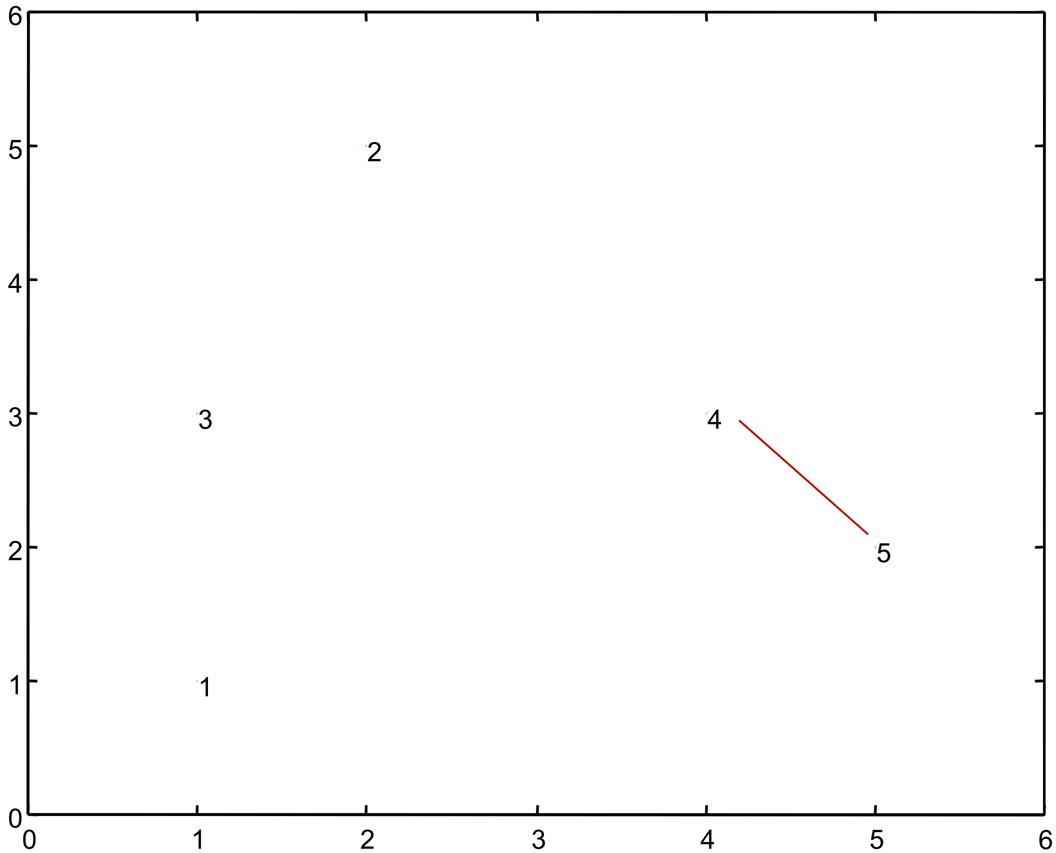
Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

- Cluster analysis groups in hierarchy set of points
- Depending on the distance the points are grouped together to form more and more large groups. Eventually all data can be grouped together
- The basic instrument of the cluster analysis is the distance matrix
- Given a set of Xi pattern, the distances dij is the following matrix:

$$d_{ij} = \left\| X_i - X_j \right\|$$

  - The matrix is symmetric

Distance matrix



	1	2	3	4	5
1	0	4.1231	2.0000	3.6056	4.1231
2	4.1231	0	2.2361	2.8284	4.2426
3	2.0000	2.2361	0	3.0000	4.1231
4	3.6056	2.8284	3.0000	0	**1.4142**
5	4.1231	4.2426	4.1231	1.4142	0

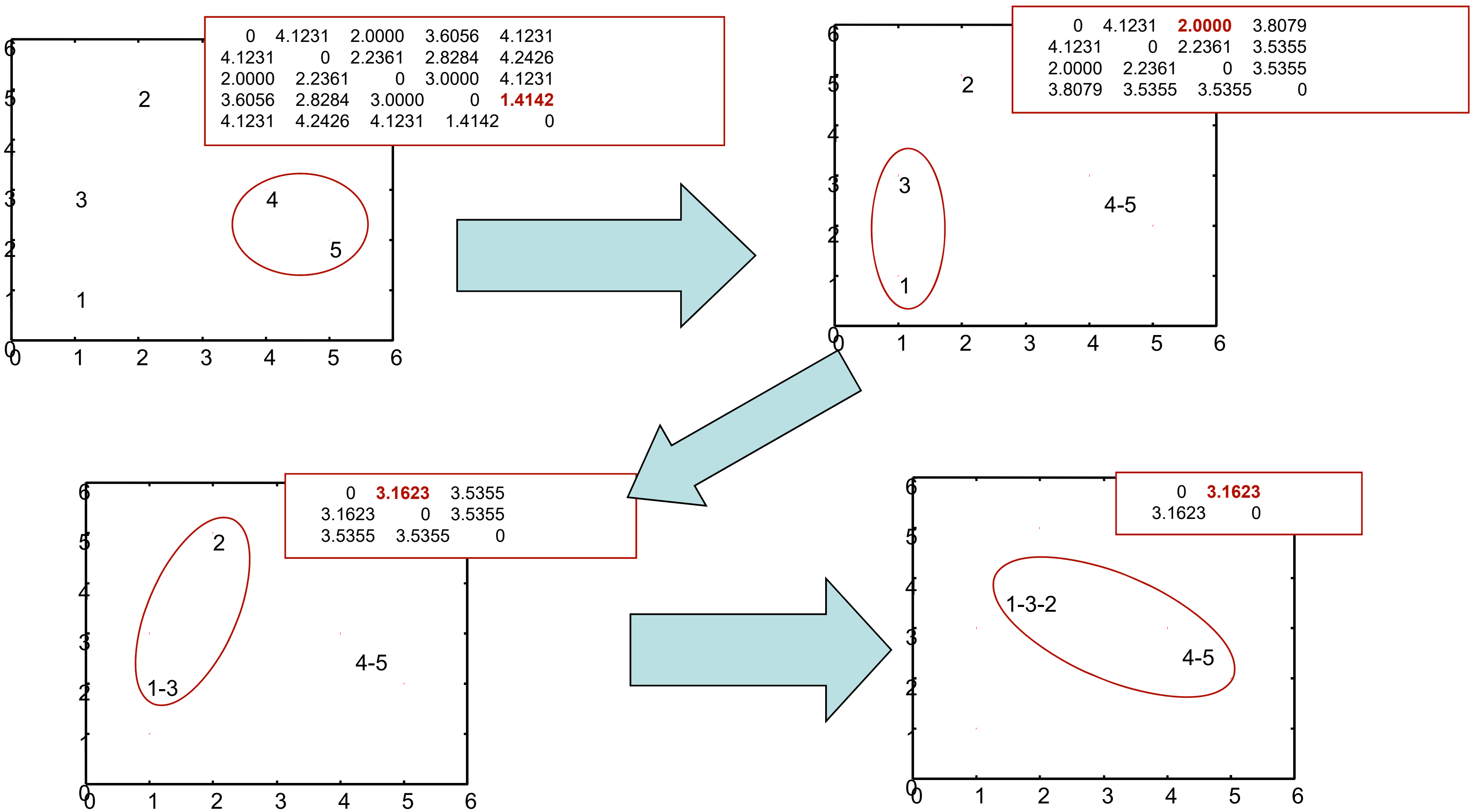
- Points 4 and 5 are the closest
- The pair 4-5 is isolated from 1-2-3
- Probably there are two sets of data:
- 1-2-3
- 4-5
- The cluster analysis makes the analysis rational and it allows to operate on N-dimensional space

Cluster analysis

- The hierarchical cluster analysis is an iterative process by making a dendrogram that defines the classes depending on the distance.

- Step i
- distance matrix calculation
- Detecting points with smaller distance
- Formation of clusters by combining Points
- Replacement of the cluster with the average point
- Refining the procedure until there is only one point

Example



0	4.1231	2.0000	3.6056	4.1231
4.1231	0	2.2361	2.8284	4.2426
2.0000	2.2361	0	3.0000	4.1231
3.6056	2.8284	3.0000	0	**1.4142**
4.1231	4.2426	4.1231	1.4142	0

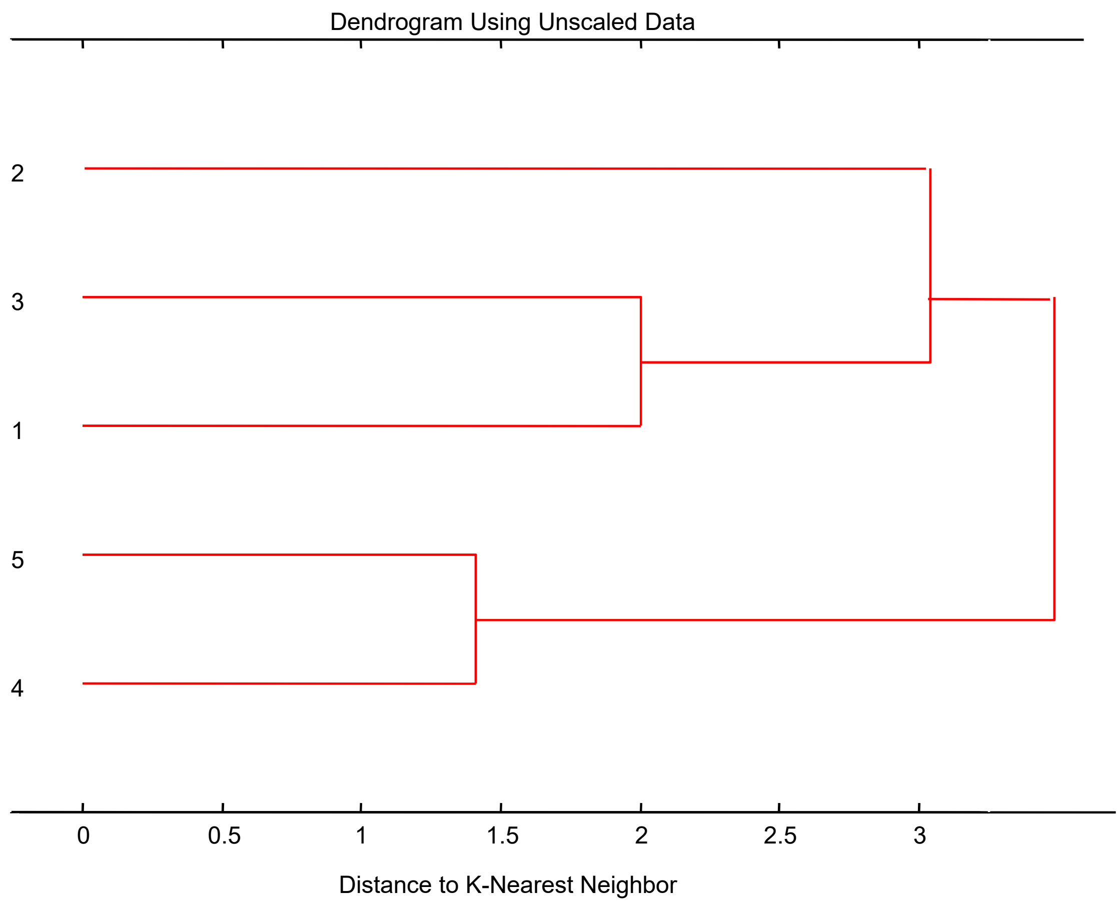
0	4.1231	**2.0000**	3.8079
4.1231	0	2.2361	3.5355
2.0000	2.2361	0	3.5355
3.8079	3.5355	3.5355	0

0	**3.1623**	3.5355
3.1623	0	3.5355
3.5355	3.5355	0

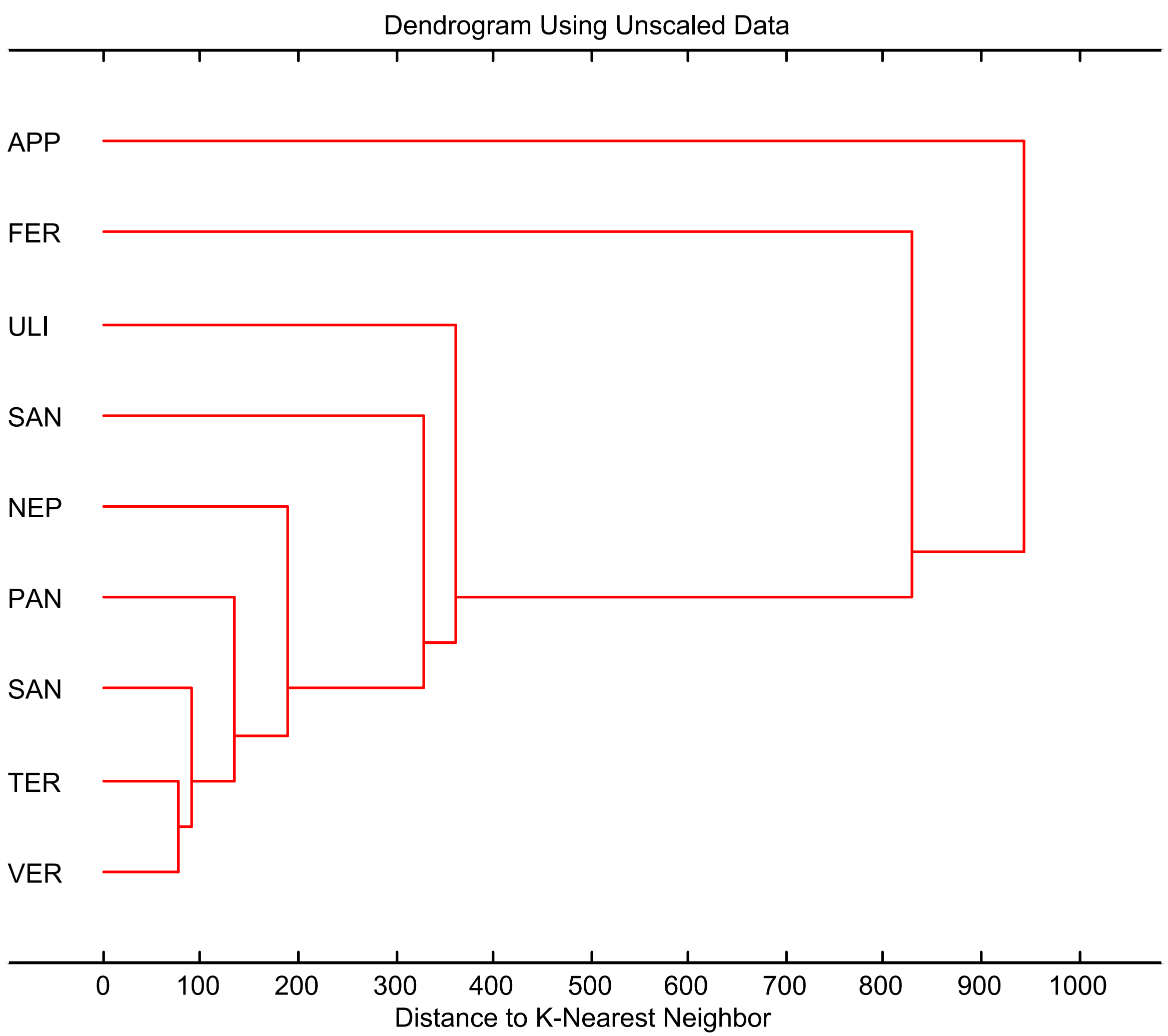
0	**3.1623**
3.1623	0

Dendrogram

- The result of the cluster analysis is depicted in a tree diagram (dendrogram) where all the points are joined to form clusters, as a function of distance.

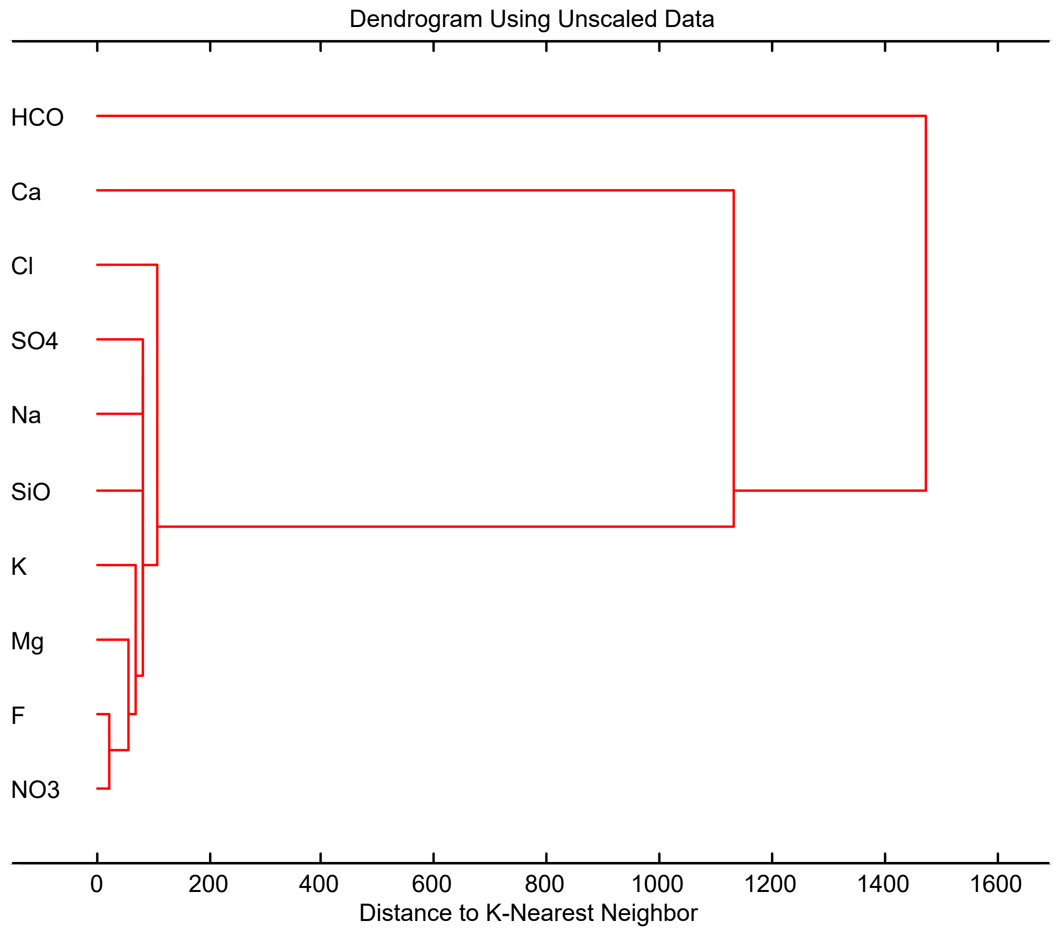


Cluster analysis mineral waters

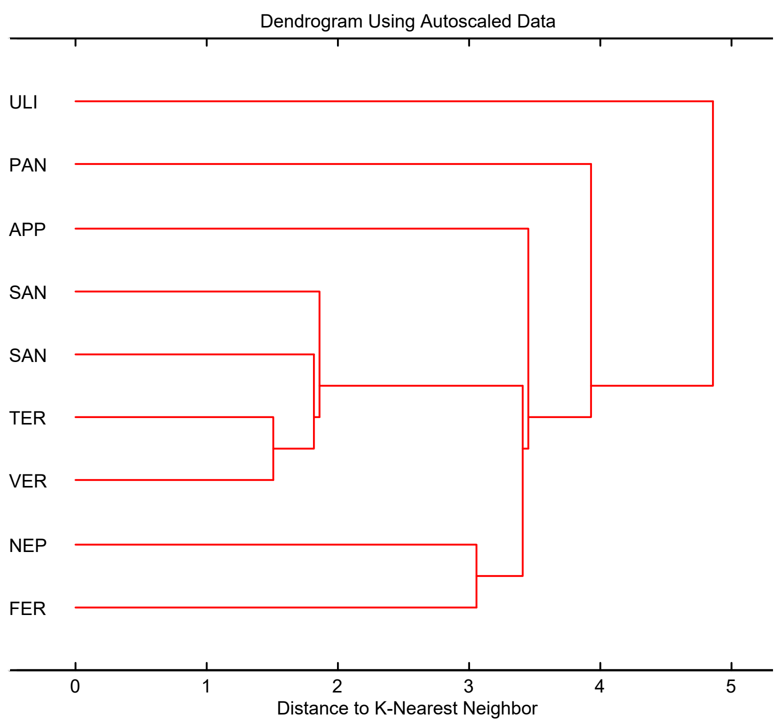


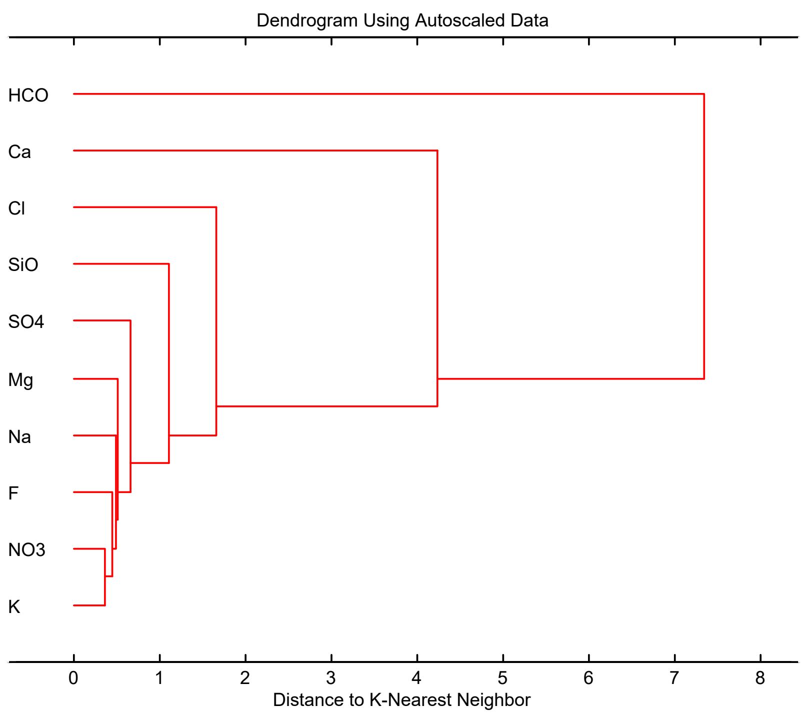
"inverse" Pattern recognition

- To study the role of the features we can study the problem in a transposed manner where the features become samples and samples the features.
- Example: cluster analysis of mineral waters



Normalization of the tree diagram for mineral waters



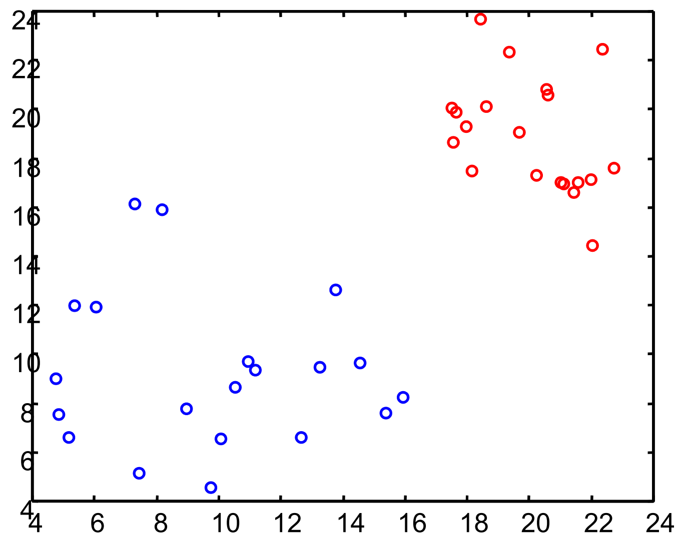


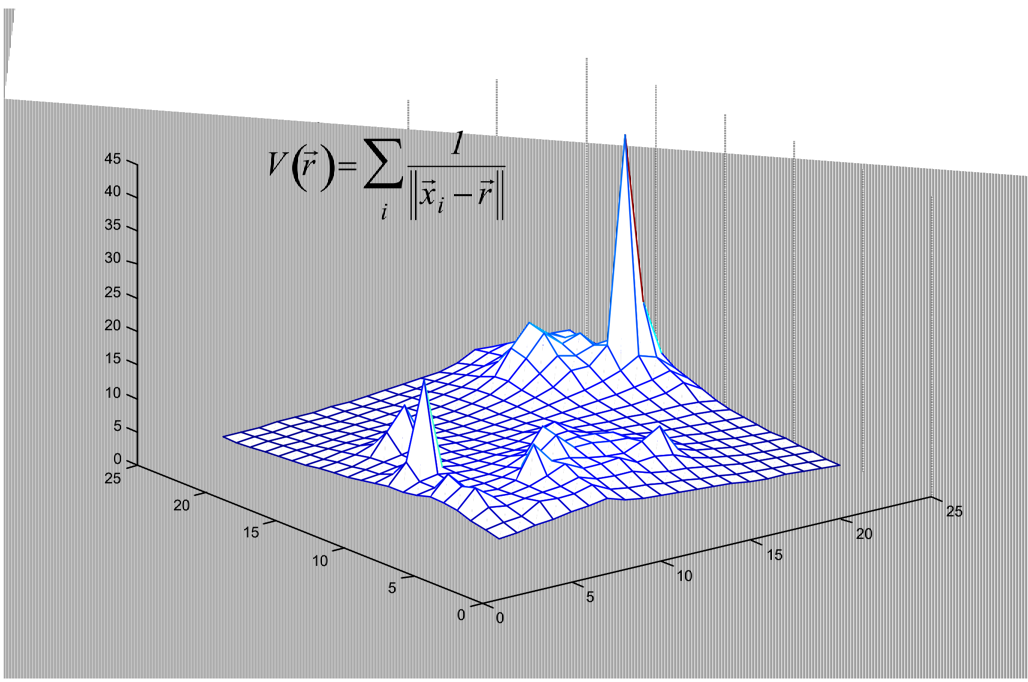
Gravitational clustering

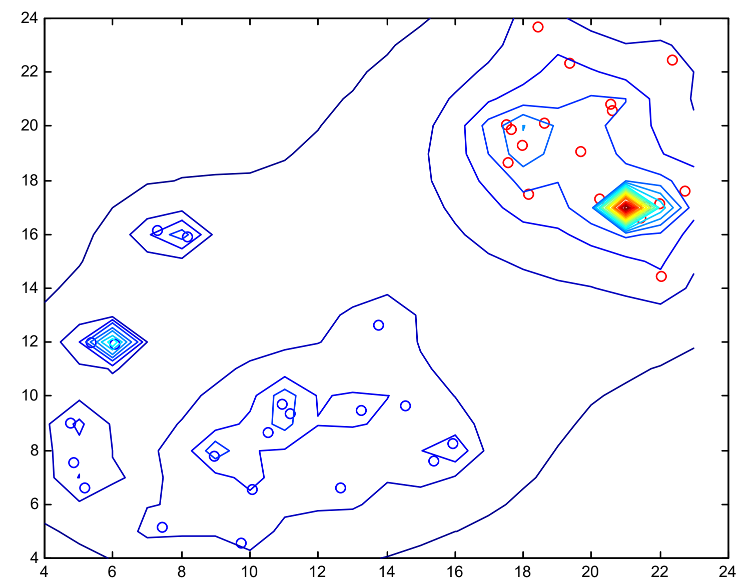
- An interesting "exotic" method of "unsupervised" classification is based on the analogy with the force fields.
- Suppose that each point possesses a mass M (equal for all). This mass will generate a potential V point in the space. Where The points are grouped (in a class) will generate a higher potential, and then studying the evolution of potential, and in particular Its top it is possible to identify the spatial regions of maximum densification (classes).
- The analogy with the masses is just an example you can use any potential function.
- If gravitational analogy, N data points xi the potential at the point r is given by:

$$V(\vec{r}) = \sum_i \frac{1}{\left\| \vec{x}_i - \vec{r} \right\|}$$

Gravitational clustering

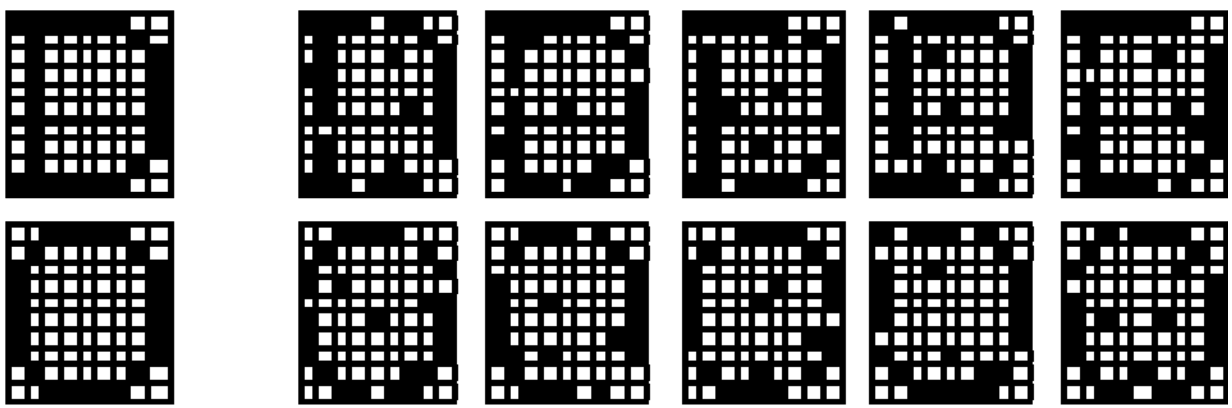






Template Matching

- This method is efficient when each class contains only a pattern. The measured patterns are affected by additive noise (no translation, rotation, deformation of pattern)
- Typical example: Image Recognition
- For each class, the pattern with low noise is taken as the class template
- Two methods to assign a pattern to a class:
- Count the number of agreements: maximum correlation
- Count the number of disagreements: lowest error



templates

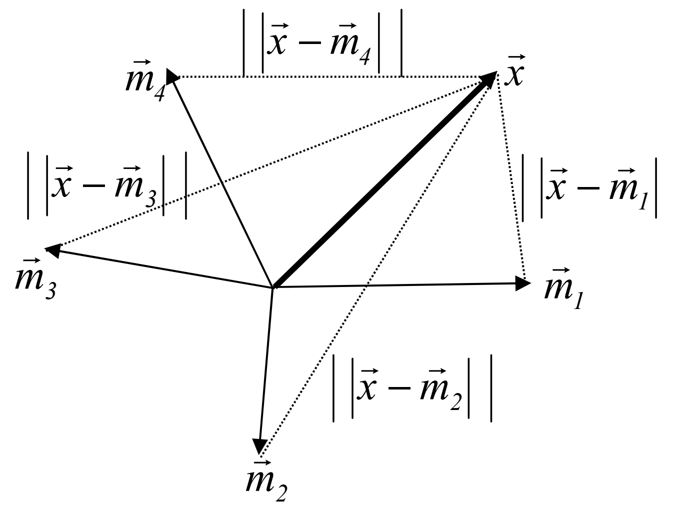
Measured Patterns

K-nearest neighbour

- mathematical calculation of template matching

- The templates are represented by vectors **mj**
- For class **K**, the distance between a pattern (**x**) and the corresponding template (**m**) is :

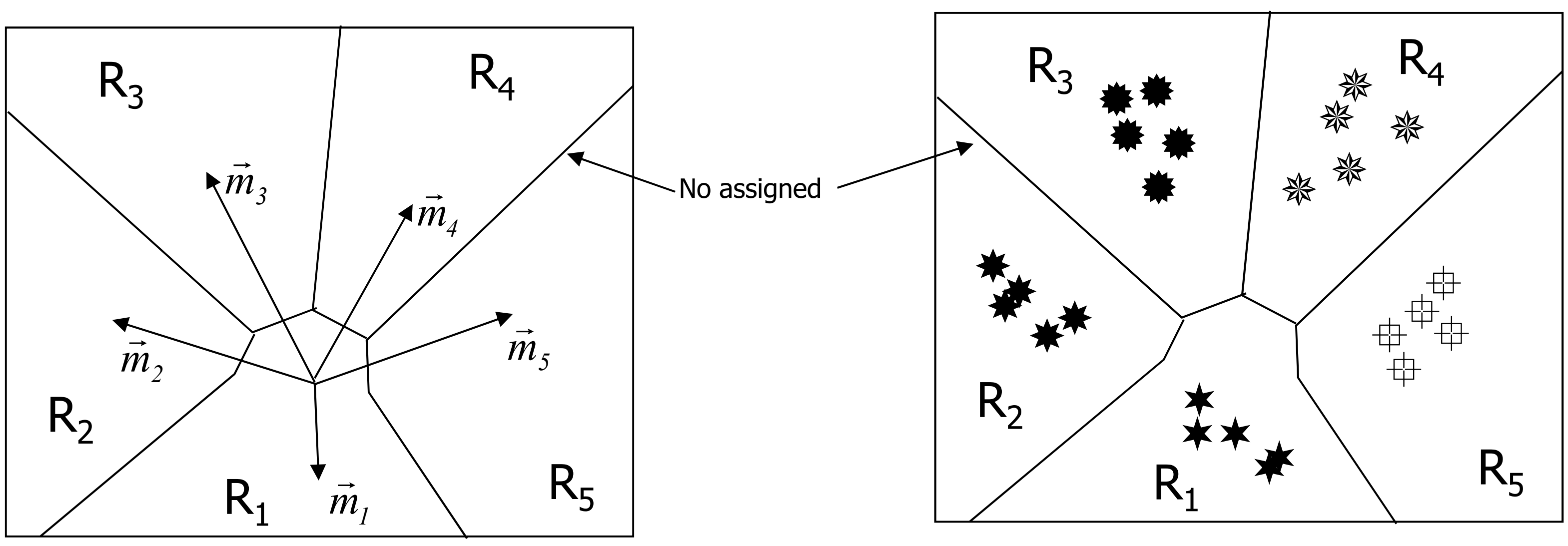
$$\varepsilon_k = \left\| \vec{x} - \vec{m}_k \right\|$$



- The pattern **x** is assigned to the class to which the distance from the corresponding template is minimum
- The transaction requires the definition of a measurement system, a rule for the distances calculation

Cluster analysis contours

- The discriminant functions divides he space into class regions
- The contours are points at the same distance from two or more template
- The Linear functions produce polygonal contours.

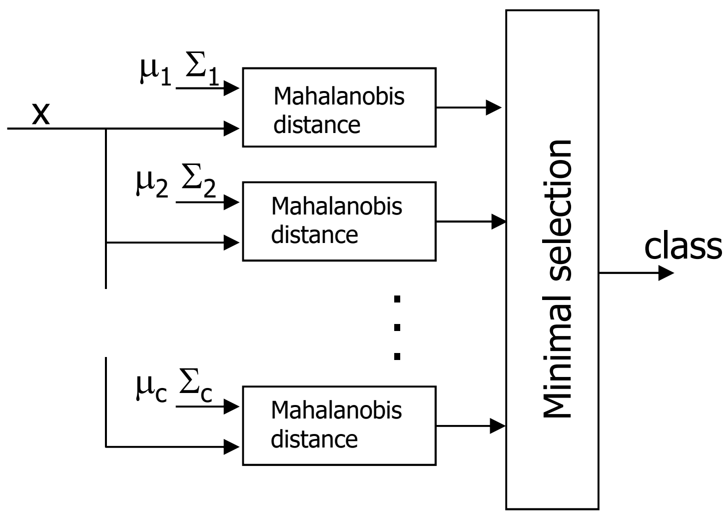


Mahalanobis distance

- For a Gaussian distribution, the iso-probability points are given by the following quadratic form:

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

- The probability value is called the Mahalanobis distance, or statistical distance,
- the probability that a vector **x** belonging to a Gaussian distribution is defined by μ and Σ



- By evaluating the Mahalanobis distance of a pattern is possible to assign a pattern to each class to which the Mahalanobis distance is smaller, that is, towards which the probability is greater

Nonparametric classification

- If the probability distribution of the patterns is not known or is not Gaussian, you should use a non-parametric statistical classification that is independent of the classes and in which you search for a combination of variables that allows the identification of classes.
- This operation is similar to the discriminant analysis
- The calculation is similar to multiple linear regression, using the linear relation between the matrix X (set of patterns of samples measured) and a Y matrix (numerical coding of class membership)
- to solve the problem it is necessary to find the regression matrix B

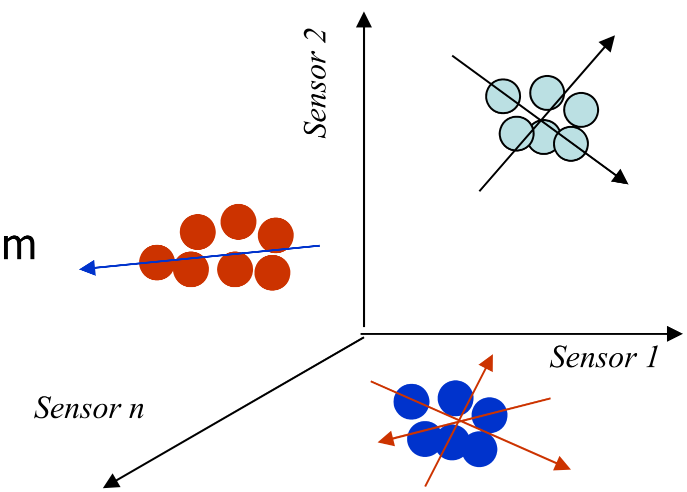
$$Y = X \cdot B^T \Rightarrow B^T = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y$$

“one-of-many”

- The Y-matrix is constructed with a number of columns equal to the number of classes in the problem
  Each Y column then identifies a class
  Given a pattern Xi the corresponding Yi line is made by setting to zero all the elements except the one corresponding to the class of Xi which is set to 1
  When identifying the regression model, we will have a finite accuracy, so the pattern is assigned to the class whose corresponding value is larger.

Soft Independent Modeling of Class Analogy (SIMCA)

- For each class of samples, a PCA model is constructed. The PC bases are different for each class and the number of meaningful components is also different
- Each class defines a proper hypervolume
- Unknown samples are identified applying them to each model and looking for the matching one.
- A probability of membership is obtained.



A case of neural network paradigm

Self Organizing Map

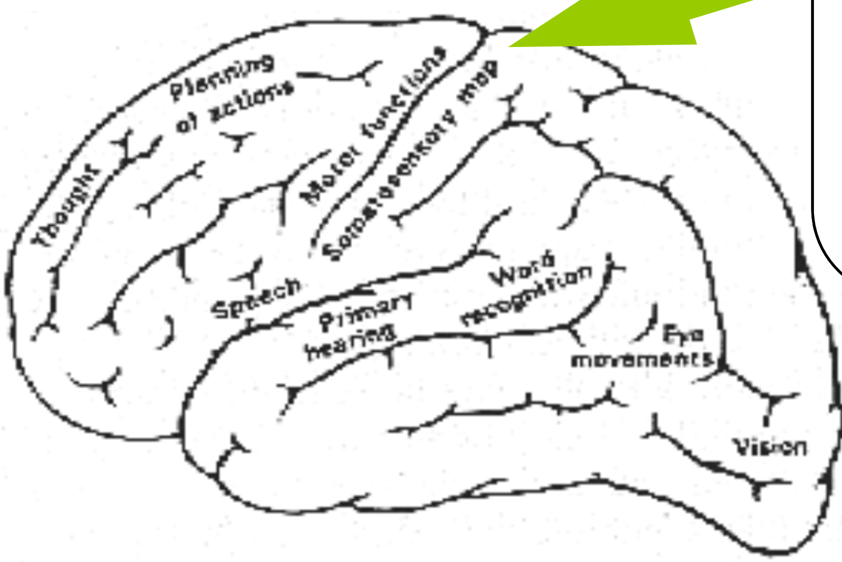
Linear Vector Quantization

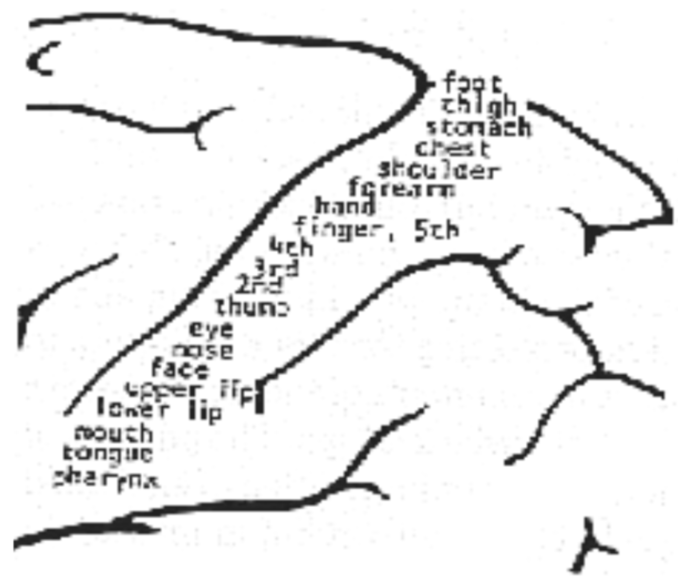
Self Organizing Map

- it is a neural network based on strong biological similarities . It aims at mimicking the functionality of the cerebral cortex
  - sensorial map
- Principal features:
  - it learns from the experience; unsupervised:; adaptive
- It provides a powerful tool to map a phenomena (represented as a multidimensional system) into a bidimensional grid discovering any intrinsic classification property
- The map is formed by a bidimensional grid of neurons (discrete space)
- Each neuron is identified with a codebook vector belonging to the sensor space and representing the link between the SOM and the input space
- Learning algorithm (*Kohonen algoritm*) is structured in two step
  - Response
  - Adaptation

SOM: the Biological Paradigm

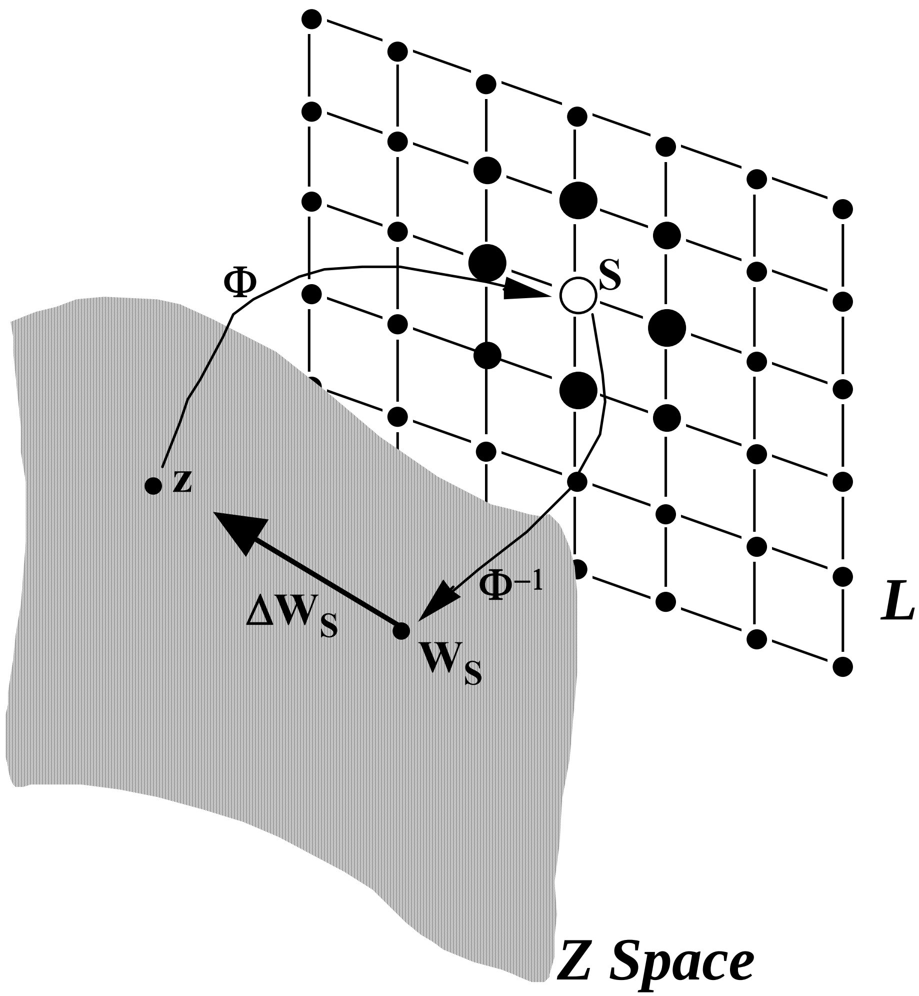
Brain areas





The somatotopic map

SOM: the Learning Algorithm



1. Response

$$\|\boldsymbol{z}-\boldsymbol{w_s}\| \le \|\boldsymbol{z}-\boldsymbol{w_r}\|$$

2. Adaptation

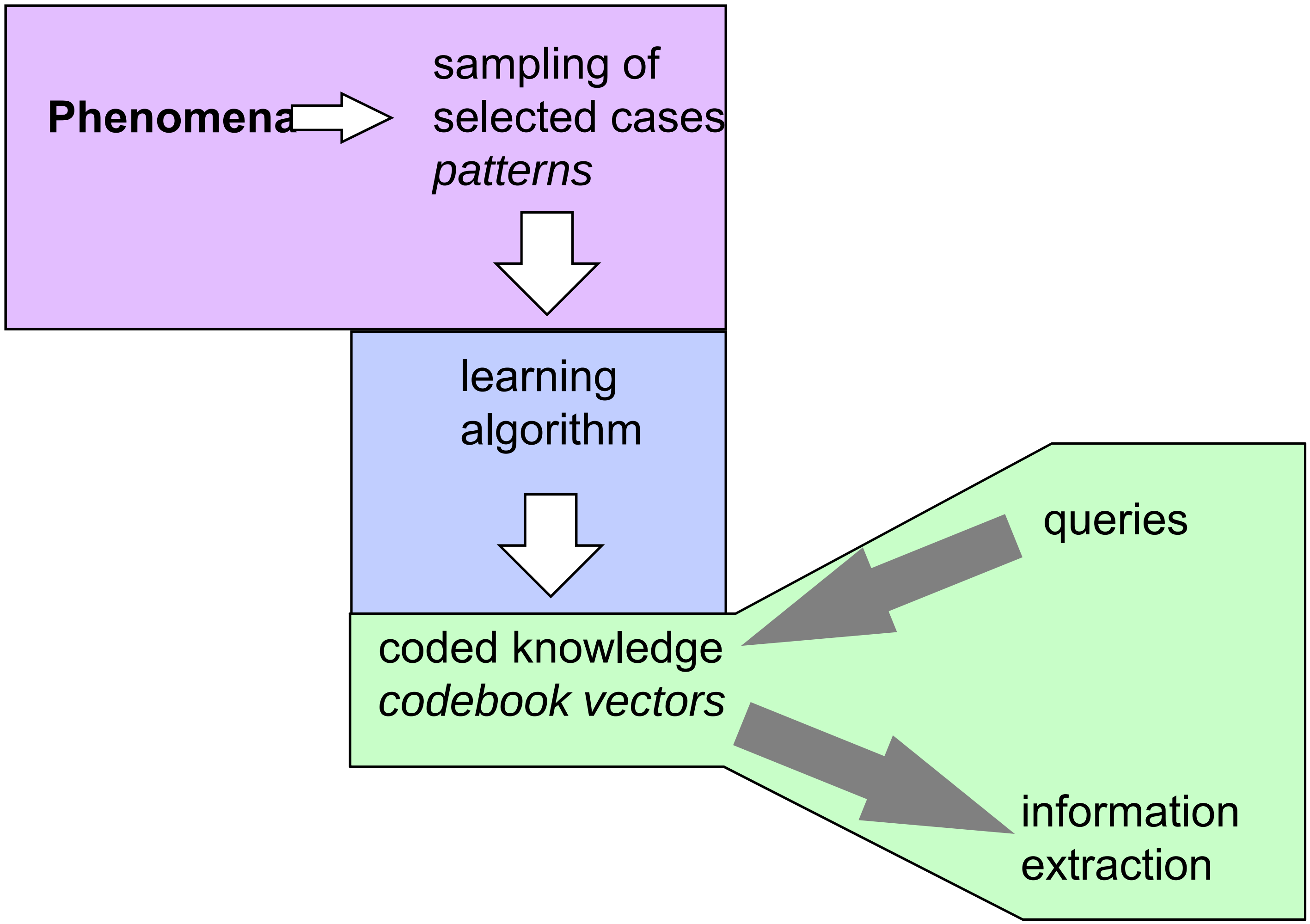
$$w_r^{new} = w_r^{old} + \alpha \cdot h_{rs}\left(z - w_r^{old}\right)$$

h is the neighbour function defining the extension of SOM which participate to the adaptation process

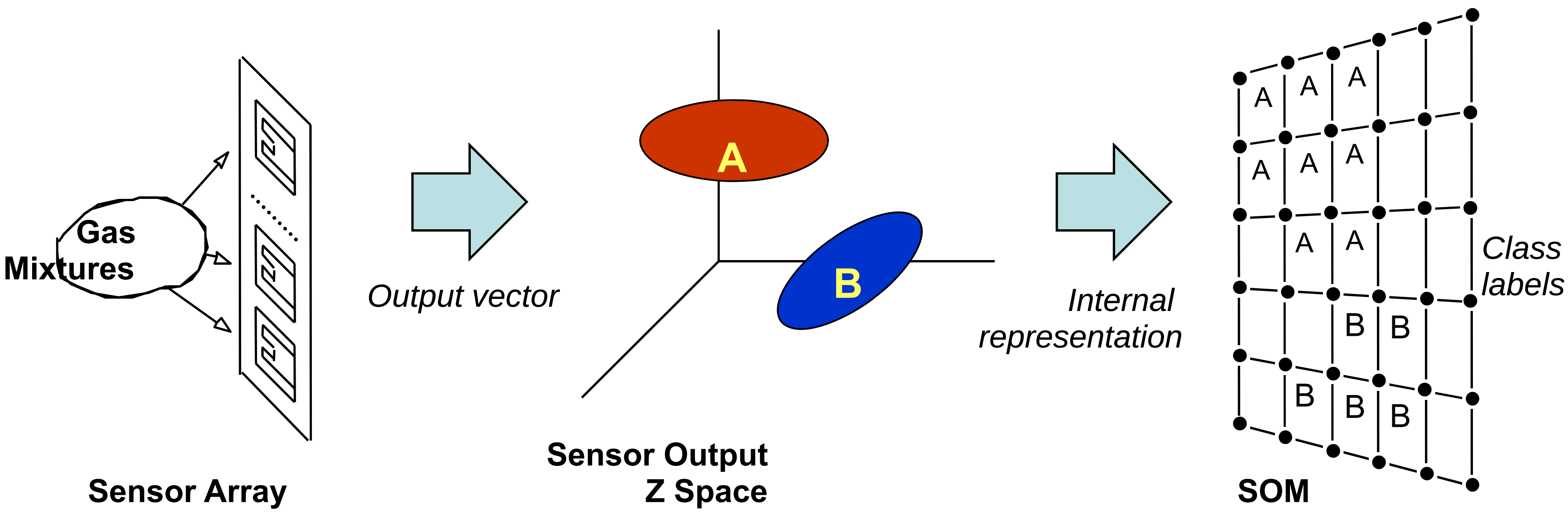
$$h_{rs} = exp\left(\frac{-\|r-s\|^2}{2\,\sigma^2}\right)$$

α is a learning rate

SOM: Flow of Data

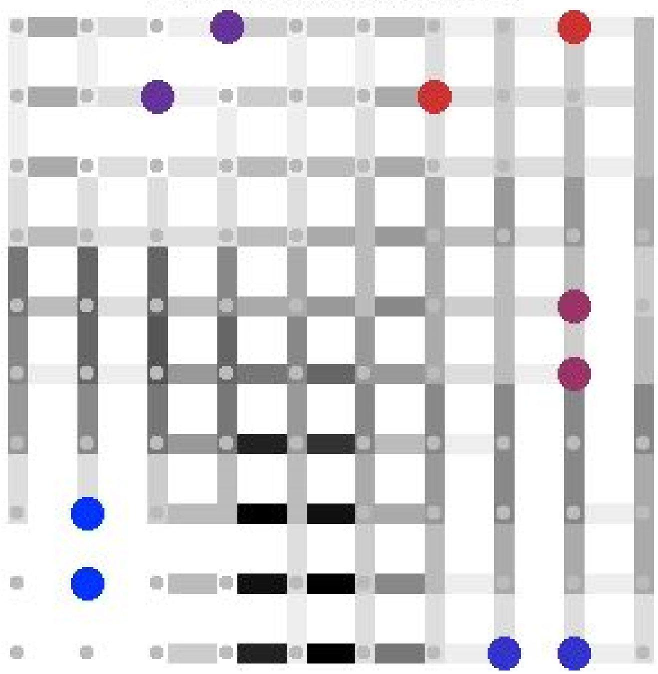


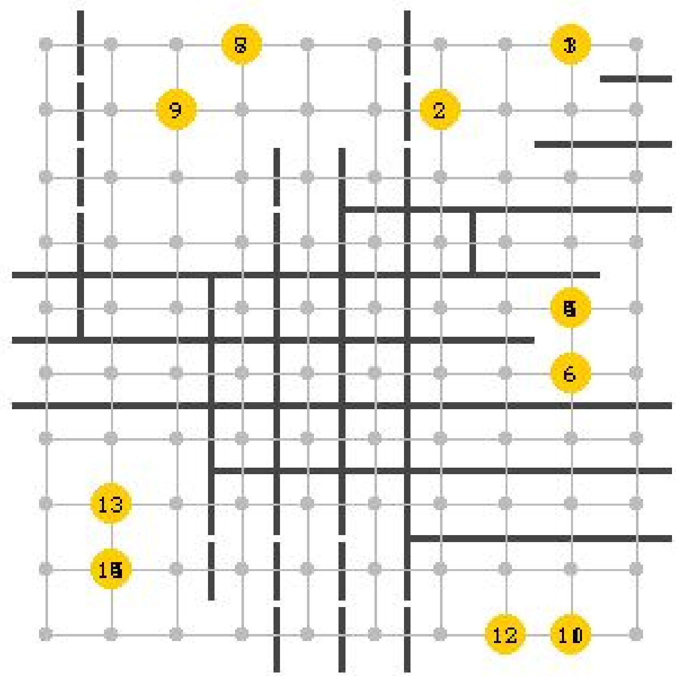
SOM and Sensor Array



SOM: Representation of Clustering

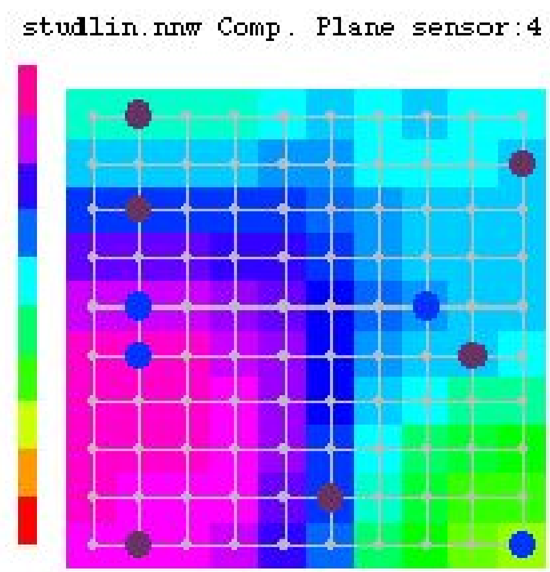
- Distance between neurons can be represented drawing lines, connecting adjacent neurons, in a color-scale proportional to the distance between the codebook vectors. Clusters can be formed fixing a threshold value to the distance.



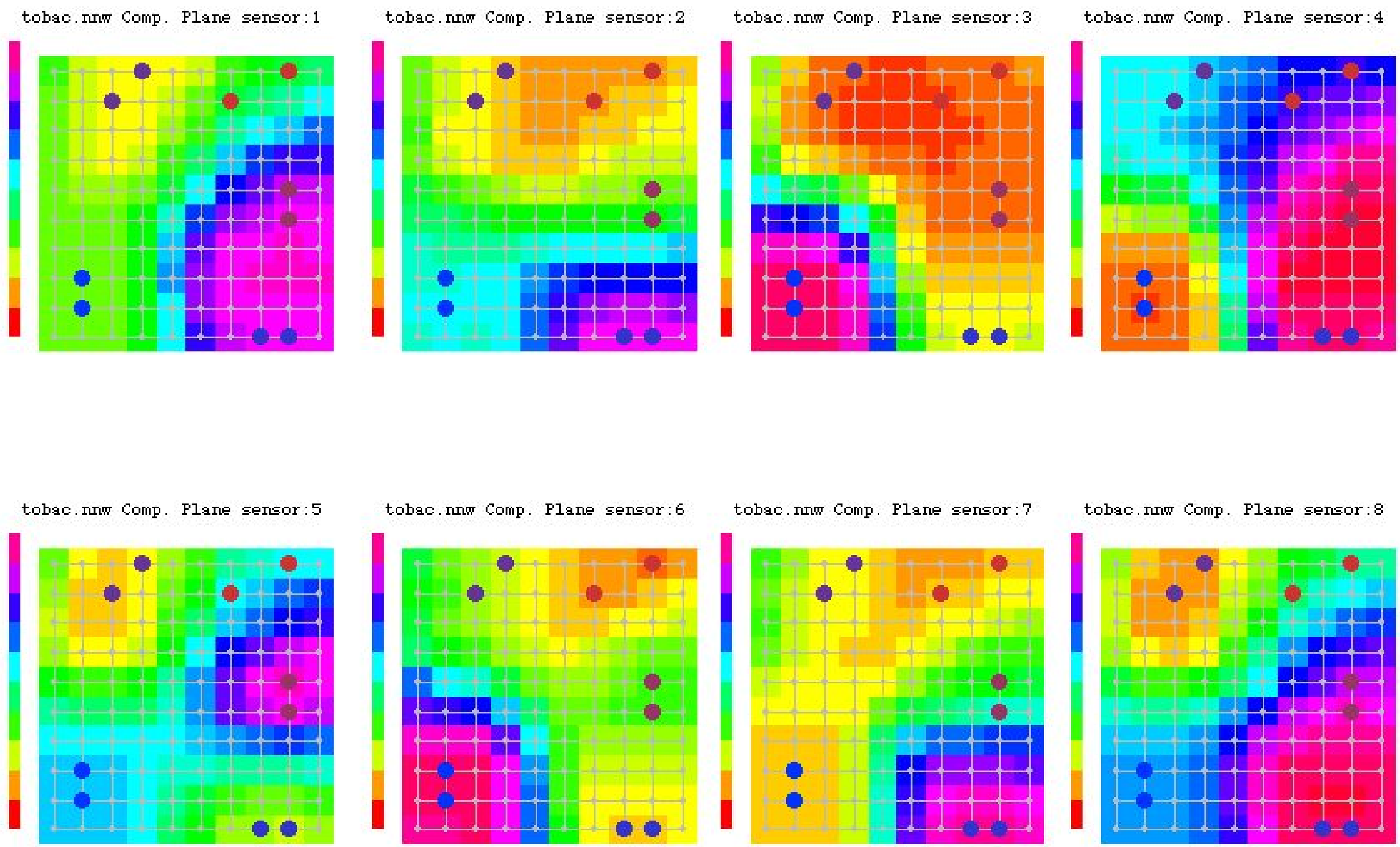


SOM: Component Planes

- The components of the codebook vectors are related to the sensors composing the array. These components can plotted onto the SOM grid giving information about the behaviour of single sensors.



SOM: Component Planes

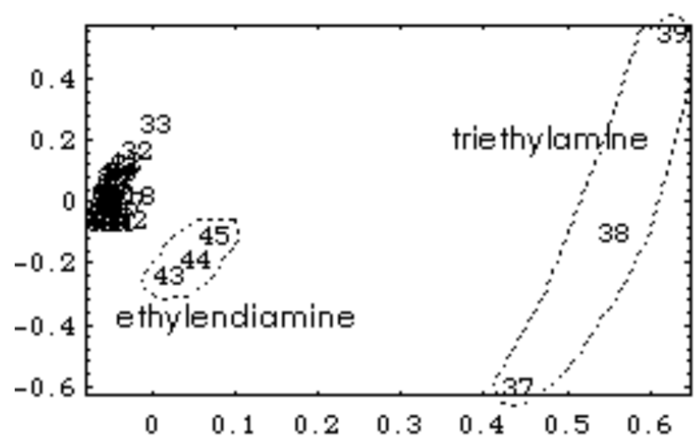


PCA - SOM comparison:
Array of QMB for the detection of VOC

Sensor responses have been measured, at different concentrations, for a number of different volatile compounds chosen as representatives of the following classes: alkanes, aldehydes, alcohols, aromatics and amines.

PCA

not preprocessed data $\Delta f_i \cong S_{ij} \cdot c_j$



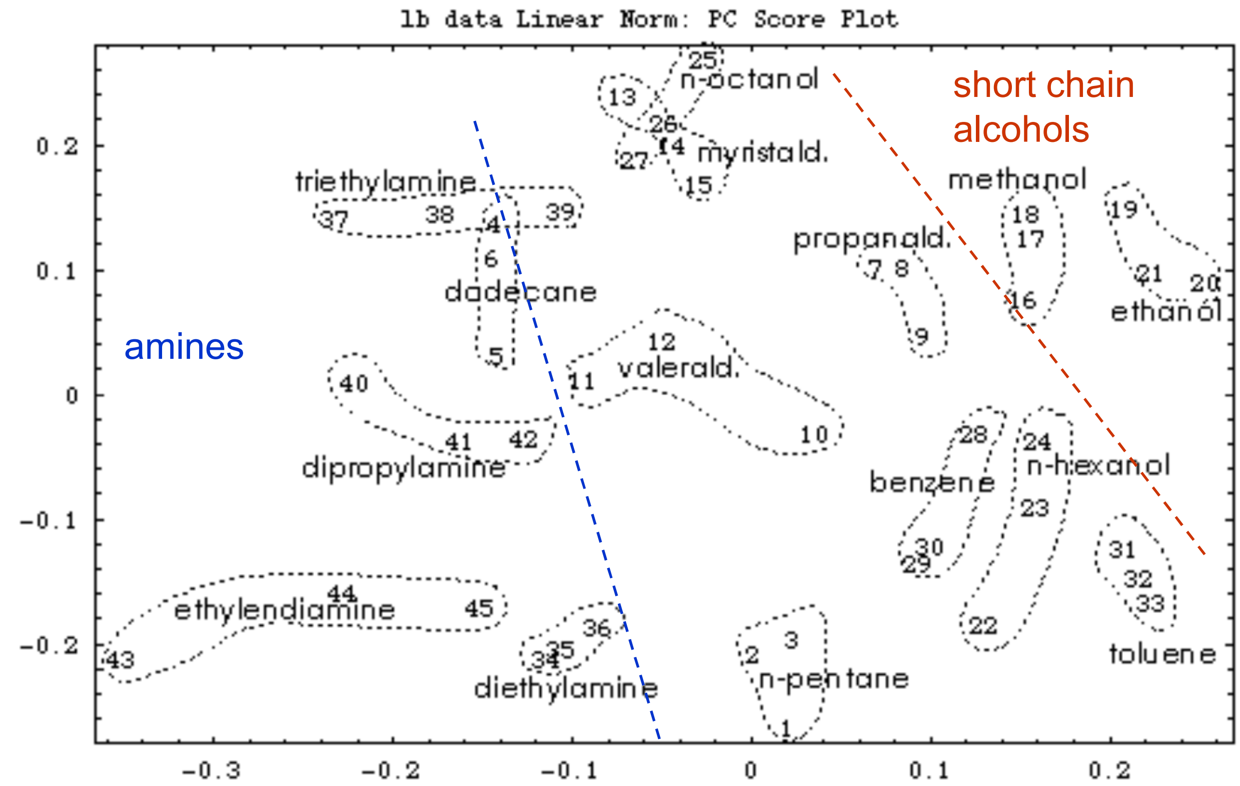
score plot

PCA - SOM comparison:

Array of QMB for the detection of VOC

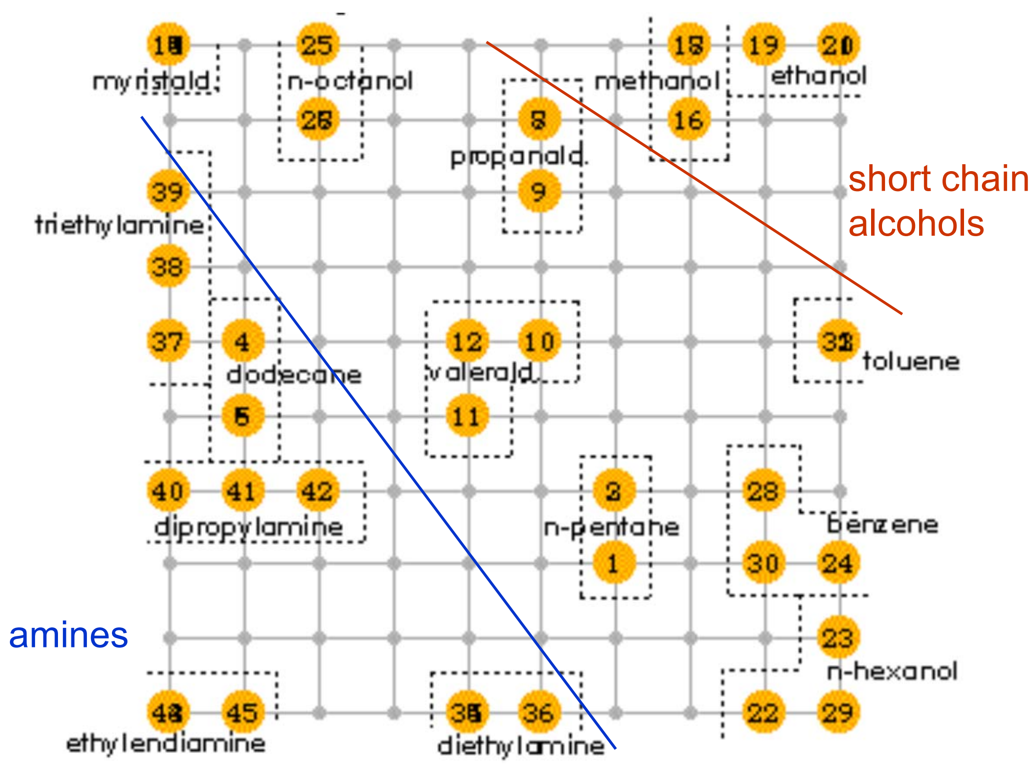
Linear Normalization

$$\Delta f_i \Rightarrow \frac{\Delta f_i}{\sum_k f_k} \cong \frac{S_{ij} \cdot c_j}{\sum_k S_{kj} \cdot c_j} = \frac{S_{ij}}{\sum_k S_{kj}}$$



PCA score plot

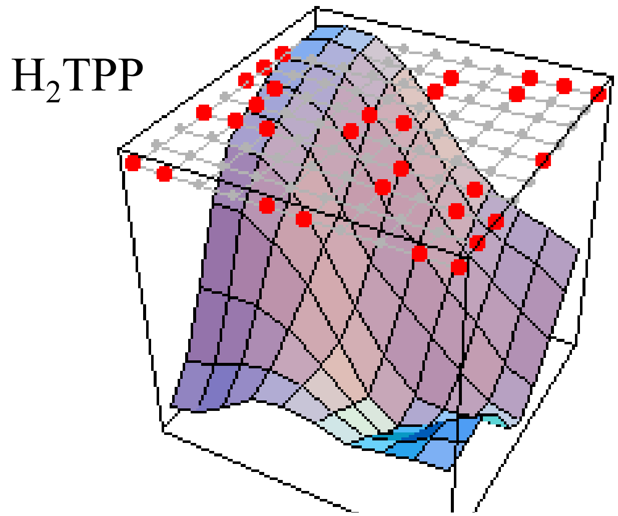
PCA - SOM comparison:
Array of QMB for the detection of VOC

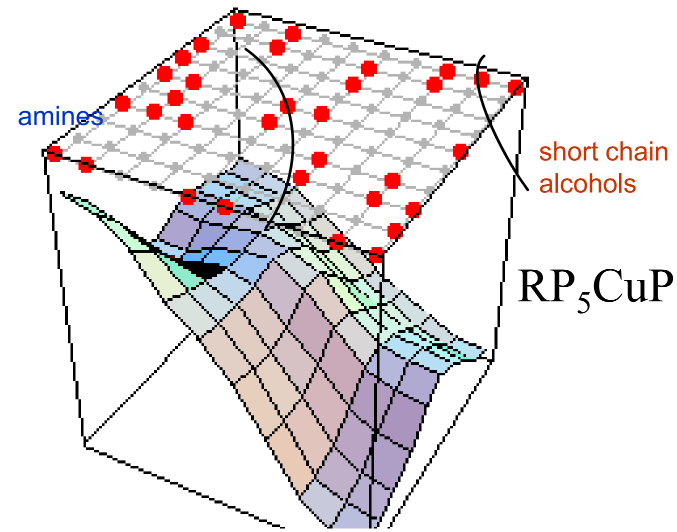


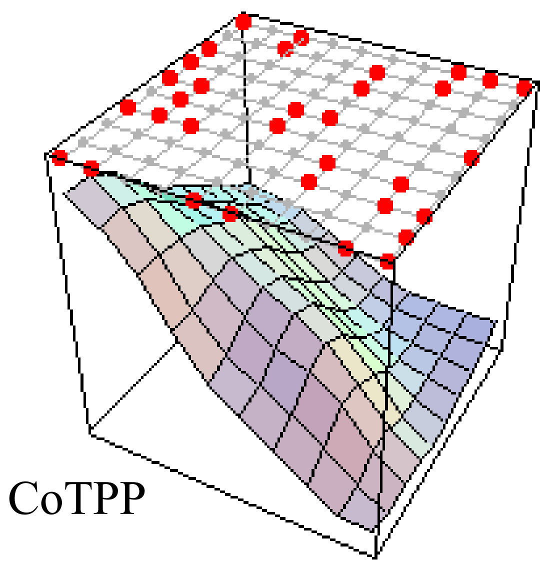
PCA - SOM comparison:

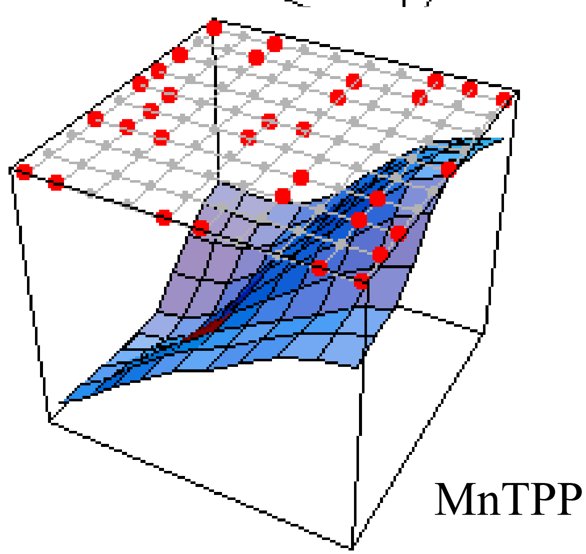
Array of QMB for the detection of VOC

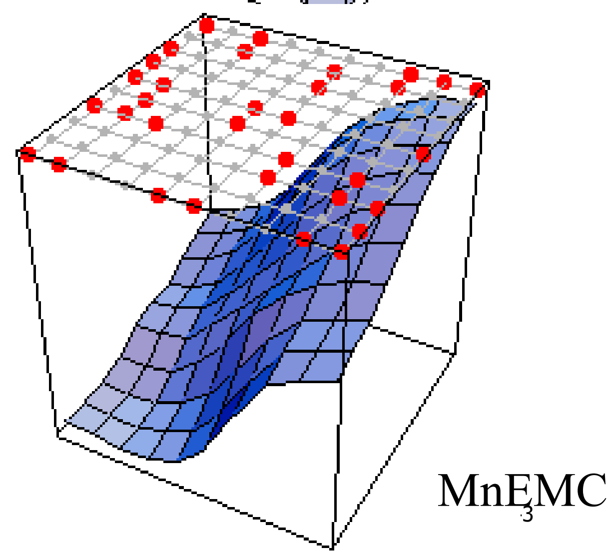
- **Study of the Component Planes**











Nonlinear classification methods: Learning Vector Quantization

- Vector Quantization: an approximation of the probability density functions of vectorial variables by finite sets of codebook vectors.
- Basic LVQ algorithm:

m_i codebook vectors assigned to each class
an input **x** is assigned to the class to which the closest m_i belongs

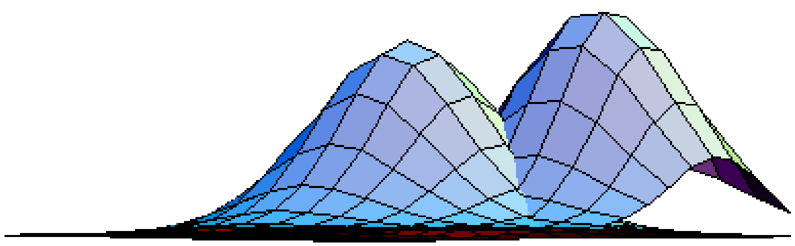
$$m_c(t+1) = m_c(t) + \alpha(t)\left[x(t) - m_c(t)\right] \qquad \textit{If x belongs to the class of } m_c$$

$$m_c(t+1) = m_c(t) - \alpha(t)\left[x(t) - m_c(t)\right] \qquad \textit{If x does not belong to the class of } m_c$$

$$m_i(t+1) = m_i(t) \qquad \textit{If } i \neq c$$

Nonlinear classification methods: Learning Vector Quantization

Example with two classes with 2D vectors



Codebook vectors