13. Fuzzy Cluster Analysis

- Classification of a given dataset $X=\{x_1, ..., x_n\}$ into c clusters.
- The membership degree of datum x_j to cluster c_i is u_{ij}.
- A cluster is defined by its prototype β_i.

- Minimization of the following objective function:

$$J(X,U,\beta) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d^2(\beta_i, x_j)$$

with respect to

$$\sum_{i=1}^{c} u_{ij} = 1 \qquad \forall\ j \in \{1,\ldots,n\},$$

$$\sum_{j=1}^{n} u_{ij} > 0 \qquad \forall\ i \in \{1,\ldots,c\}$$

Computation of a Classification

A classification is obtained by alternating optimization:

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a dataset and c the number of clusters.

Choose c, ε.

Initialize the membership degrees uij of data to clusters.

REPEAT

> Compute the clusters $\beta=\{\beta_1,\ldots, \beta_c\}$ to minimize the given objective function $J(X,U, \beta)$.
>
> Compute the membership degrees $U=\{u_{11},\ldots,u_{cn}\}$ based on the new clusters.

UNTIL the change of membership degrees U is less than ε.

Fuzzy C-Means Algorithm

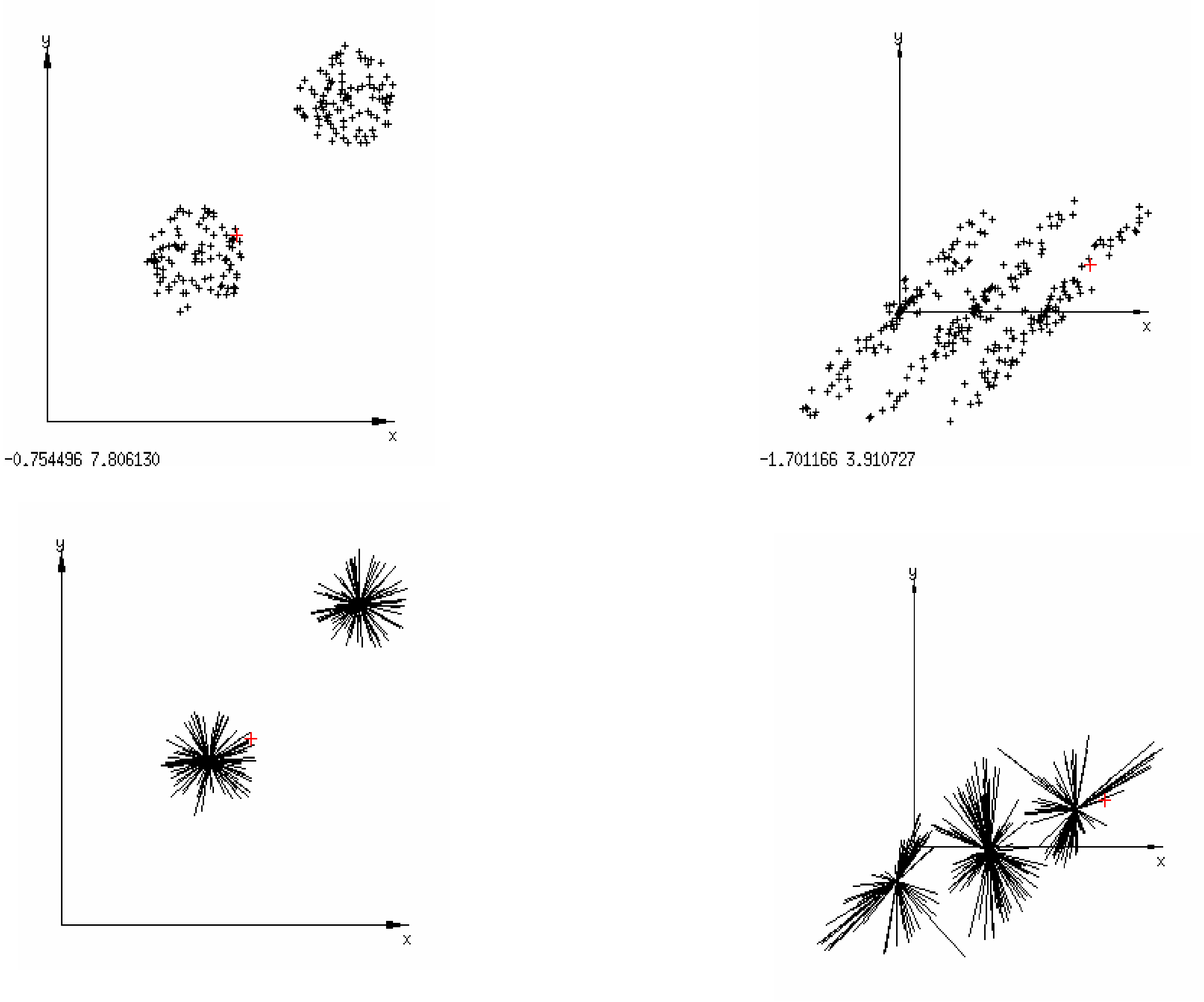
- Computation of clusters and membership degrees:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}} \qquad u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d^2(c_i, x_k)}{d^2(c_i, x_j)} \right)^{\frac{1}{m-1}}}$$

- shape of all clusters is equal (common: spherical clusters)
- size of all clusters is equal
- widely used

Fuzzy C-Means algorithm searches for equally large clusters in form of (hyper)balls

Examples



Fuzzy Cluster Analysis

- Fuzzy C-Means: simple, looks for spherical clusters of same size, uses Euclidean distance
- Gustafson & Kessel: looks for hyper-ellipsoidal clusters of same size, distance via matrices
- Gath & Geva: looks for hyper-ellipsoidal clusters of arbitrary size, distance via matrices
- Axis-parallel variations exist that use diagonal matrices (computationally less expensive and less loss of information when rules are created)

Improvement by Gustafson and Kessel

Transformation of the distance function *d* for each cluster with a symmetric, positive definite matrix A_i.

$$d^2\left(\beta_i, \mathbf{x}_j\right) = (\mathbf{c}_i - \mathbf{x}_j)^{\mathrm{T}} \mathbf{A}_i (\mathbf{c}_i - \mathbf{x}_j)$$

Computation of A_i

$$\mathbf{A}_i = (\rho_i \det(\mathbf{S}_i))^{\frac{1}{p}} \mathbf{S}_i^{-1}$$

$$\mathbf{S}_i = \sum_{j=1}^{n} u_{ij}{}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T$$

$\det(A_i)=\rho$ avoids the trivial solution $A_i=0$.

The Gustafson-Kessel algorithm searches for hyper ellipsoidal clusters of fixed size.

Improvement by Gath and Geva

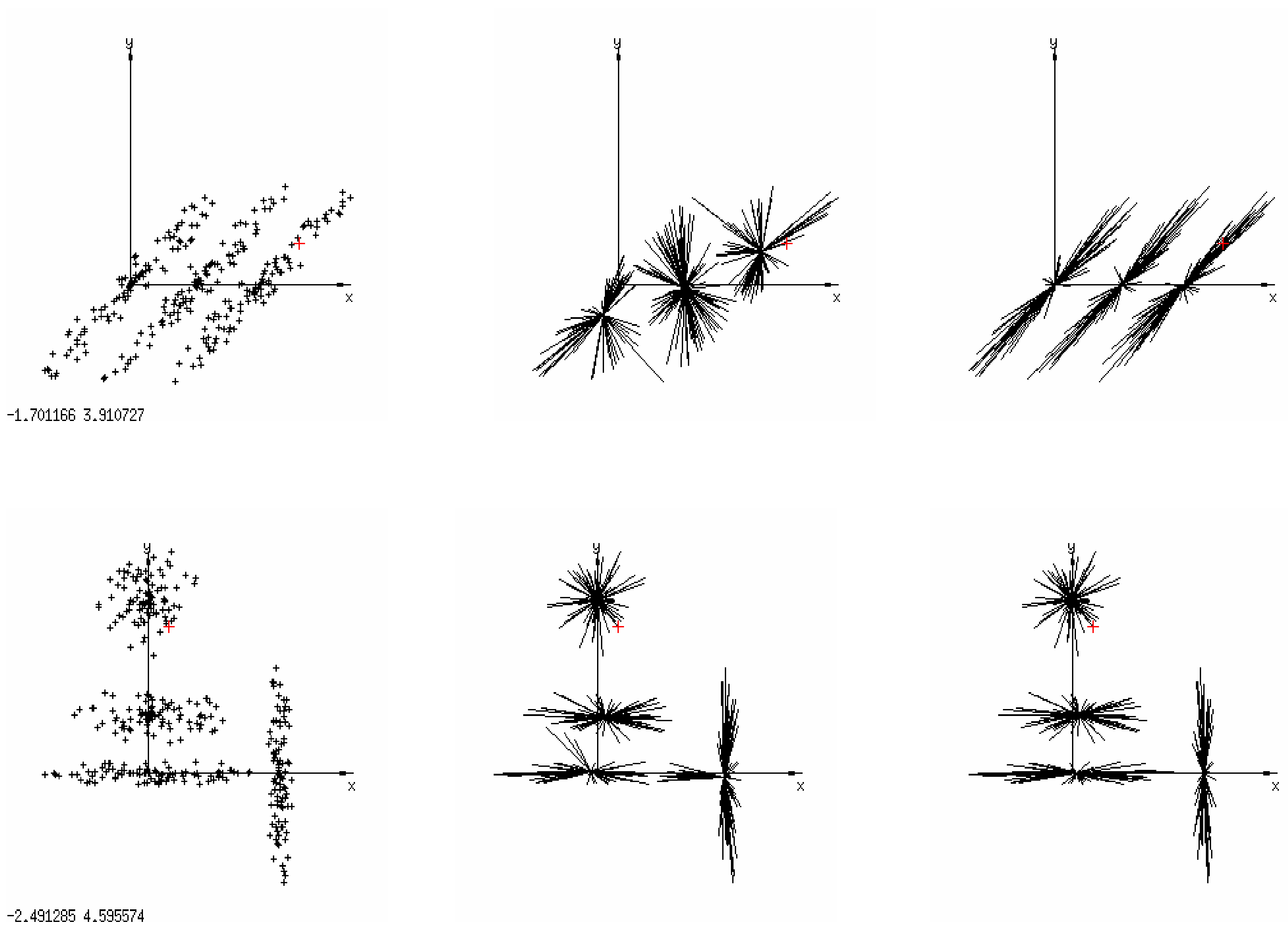
Idea: The dataset is interpreted as a realization of a collection of p-dimensional normal distributions.

The distance of a datum to a cluster is inversely proportional to the a-posterior possibility that a datum is the realization of the i^{th} normal distribution.

$$d^2\left(\beta_i, \mathbf{x}_j\right) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m}{\sum_{j=1}^{n} u_{ij}^m} \sqrt{\det \mathbf{C}_i} \exp\left(\frac{(\mathbf{x}_j - \mathbf{c}_i)\mathbf{C}_i^{-1}(\mathbf{x}_j - \mathbf{c}_i)^{\mathrm{T}}}{2} \right)$$

The algorithm searches for hyper ellipsoidal clusters of arbitrary size.

Examples



Noisy data and Outliers

Approaches to deal with noisy data:

- Possibilistic clustering

  Noisy data and outliers can be assigned to none cluster

  Neglection of restriction:

$$\sum_{i=1}^{c} u_{ij} = 1 \quad \forall\, j \in \{1,\ldots,n\}$$

- Noise cluster

  Noisy data and outliers are assigned to an extra cluster.

- Combination of noise clustering and possibilistic clustering

Possibilistic Cluster Analysis

Minimization of the following objective function:

$$J(\mathbf{X}, U, \beta) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d^{2}\left(\beta_{i}, \mathbf{x}_{j}\right) + \sum_{i=1}^{c} \eta_{i} \sum_{j=1}^{n} \left(1 - u_{ij}^{m}\right)^{m}$$

with respect to

$$\sum_{j=1}^{n} u_{ij} > 0 \quad \forall\ i \in \{1, \ldots, c\}$$

Computation of membership degrees:

$$u_{ij} = \frac{1}{1 + \left(\frac{d^{2}\left(\beta_{i}, \mathbf{x}_{j}\right)}{\eta_{i}} \right)^{\frac{1}{m-1}}}$$

Cluster Validity

Judgement of classification by validity measures.

To determine the number of clusters, the algorithm is executed several times with a changing number of clusters. The best solution is chosen.

Validity measures are based on several criteria, e.g.:

- membership degrees should be nearly 0 or 1,

  e.g. partition coefficien t (PC), $\mathrm{PC} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^{2}$

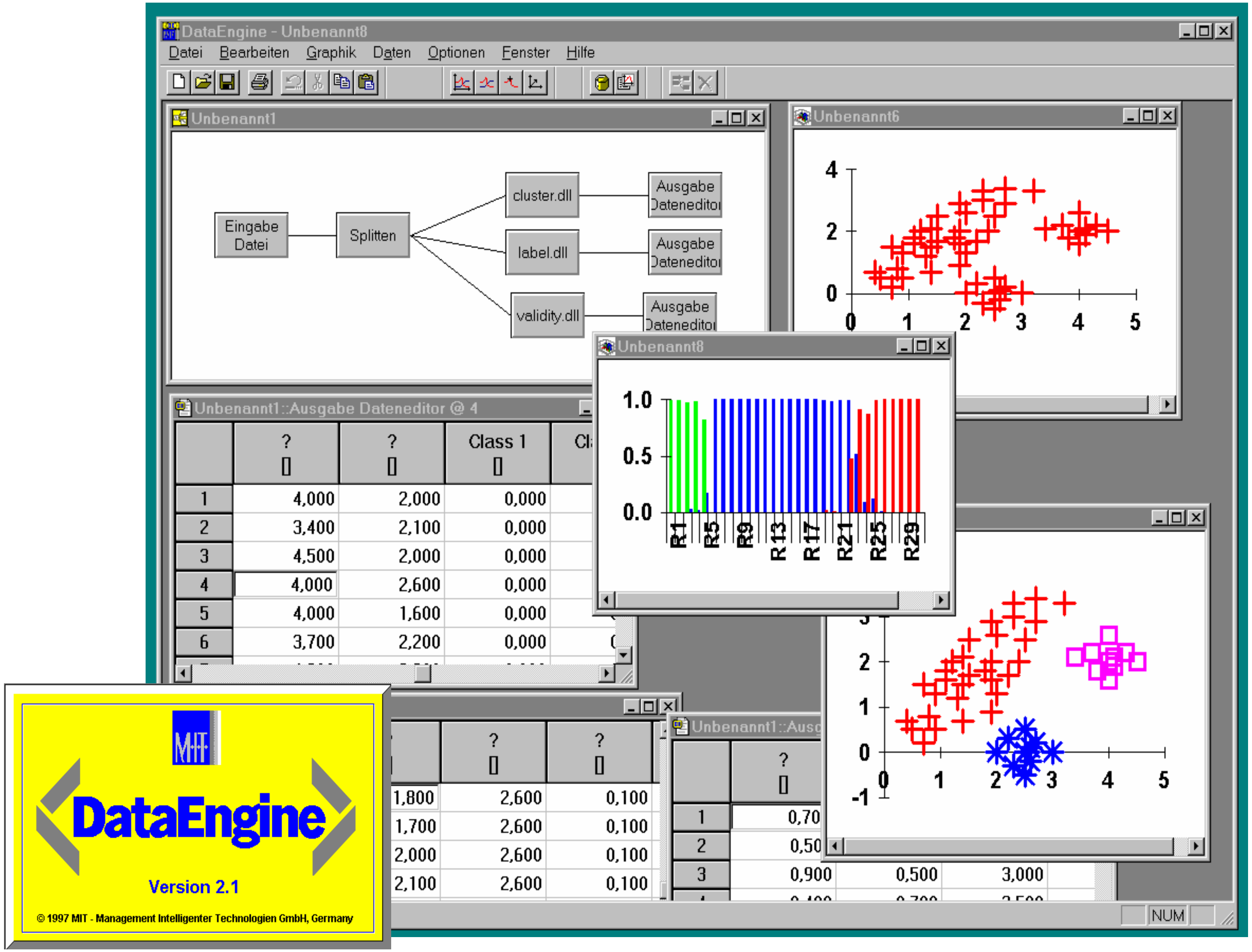
- compactness of clusters,

  e.g. average partition density (APD), $\mathrm{APD} = \frac{1}{c}\sum_{i=1}^{c}\frac{\sum_{j\in Y_i} u_{ij}}{\sqrt{\det\left(A_i\right)}}$

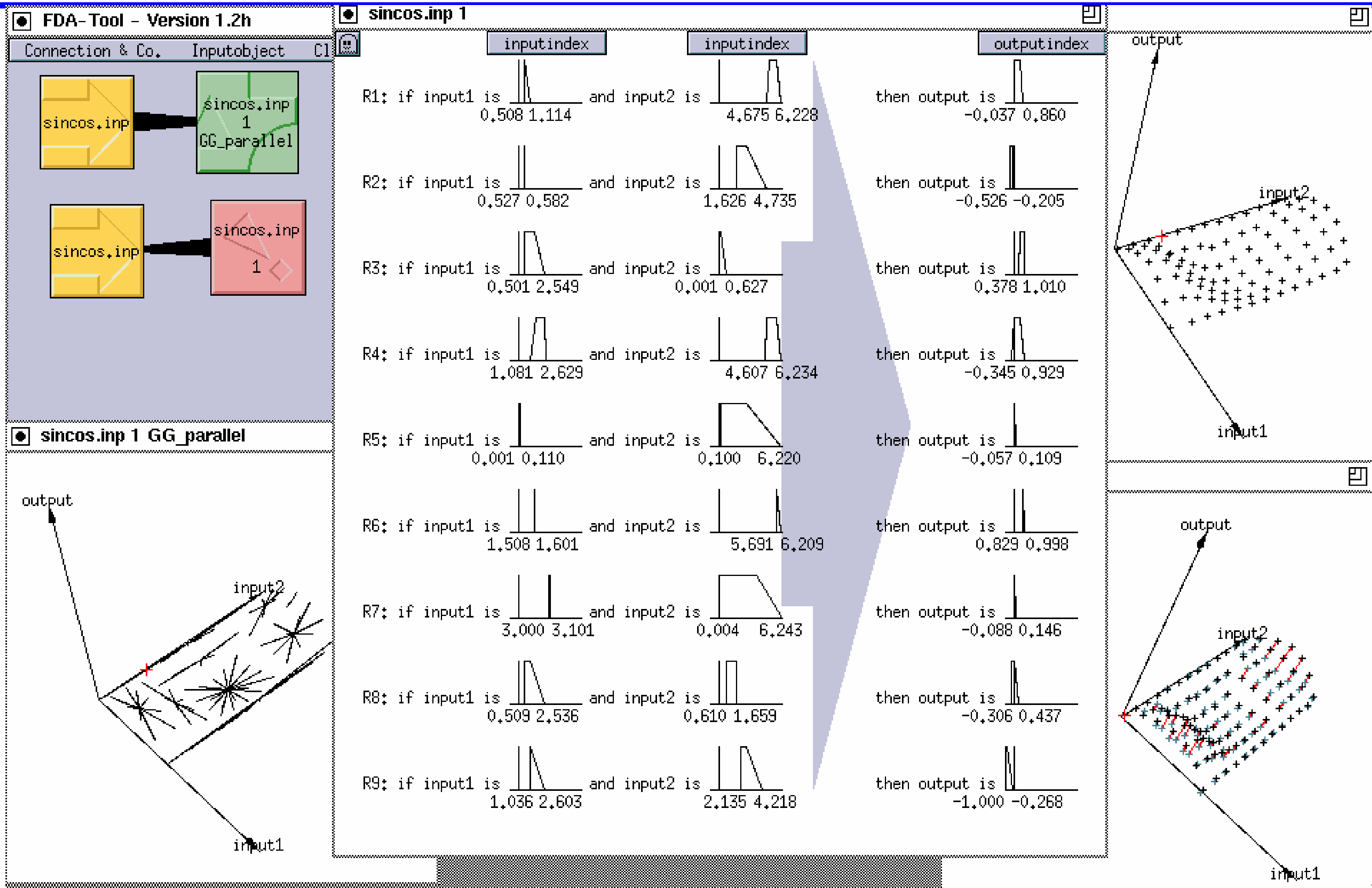
  where $\mathrm{Y_i} = \left\{\mathrm{j}\in\aleph, \mathrm{j}\le\mathrm{n}\,|\,(\mathbf{c}_i-\mathbf{x}_j)^{\mathrm{T}}\mathbf{A}_i(\mathbf{c}_i-\mathbf{x}_j)<1\right\}$

- separation of clusters,
- ...

Fuzzy-Clusteranalyse mit Data Engine



FCLUSTER: Tool for Fuzzy Cluster Analysis

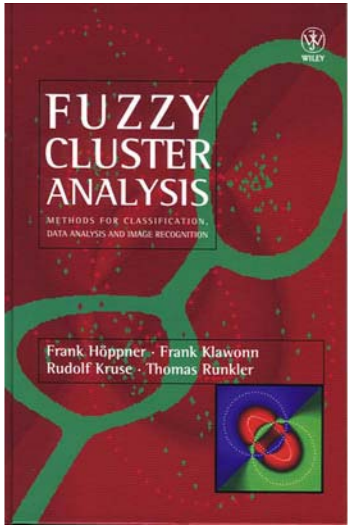


Resources

F. Höppner, F. Klawonn, R.Kruse, T. Runkler:

Fuzzy Cluster Analysis

Wiley, Chichester, 1999, ISBN: 0-471-98864-2



Software Tools:

http://www.fuzzy-clustering.de