Fuzzy Clustering with Applications in Pattern Recognition and Data-Driven Modeling

Robert Babuška

Delft Center for Systems and Control
Delft University of Technology
The Netherlands

r.babuska@dcsc.tudelft.nl
www.dcsc.tudelft.nl/˜babuska

Delft Center for Systems and Control

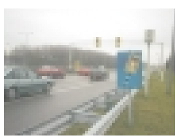
Established recently by merging control groups of:

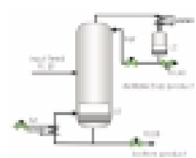
- Information technology and systems
- Applied physics
- Mechanical engineering

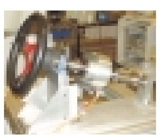
Staff

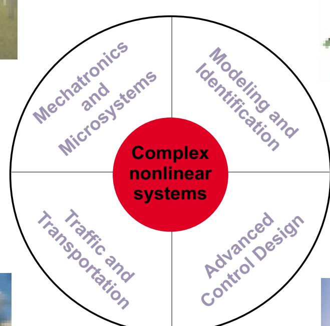
- 6 full professors, 10 associate, 4 assistant professors
- ≈ 25 Ph.D. students
- 10 technical and administrative staff
- ≈25 M.Sc. students per year

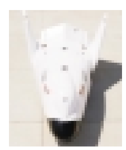
Research Subjects and Applications

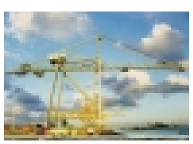


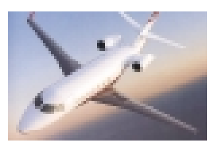












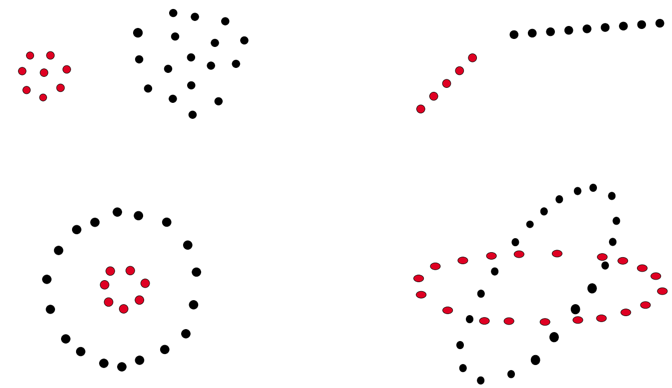
Introduction

- The ability to cluster data (concepts, perceptions, etc.) – essential feature of human intelligence.
- A cluster is a set of objects that are more similar to each other than to objects from other clusters.
- Applications of clustering techniques in pattern recognition and image processing.
- Some machine-learning techniques are based on the notion of similarity (decision trees, case-based reasoning).
- Nonlinear regression and black-box modeling can be based on the partitioning data into clusters.

Outline

- Basic concepts in clustering
  - data set
  - partition matrix
  - distance measures
- Clustering algorithms
  - fuzzy c-means
  - Gustafson–Kessel
- Application examples
  - system identification and modeling
  - diagnosis
- Concluding remarks

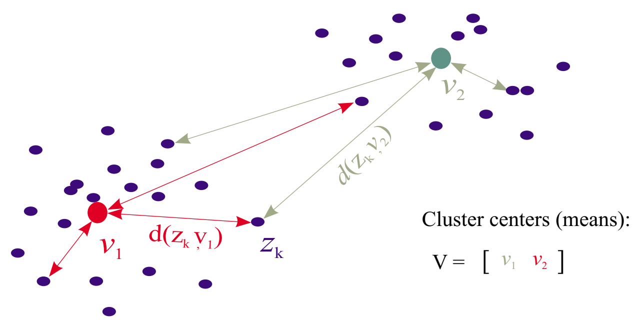
Examples of Clusters



Problem Formulation

- Given is a set of data in $\mathbb{R}^n$ and the (estimated) number of clusters to look for (a difficult problem, more on this later).
- Find the partitioning of the data into subsets (clusters), such that samples within a subset are more similar to each other than to samples from other subsets.
- Similarity is mathematically formulated by using a distance measure (i.e., a dissimilarity function).
- Usually, each cluster will have a prototype and the distance is measured from this prototype.

Distance Measure



Distance Measures

- Euclidean norm:

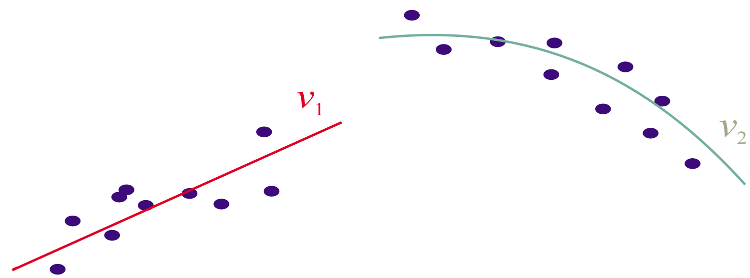
$$d^2(\mathbf{z}_j, \mathbf{v}_i) = (\mathbf{z}_j - \mathbf{v}_i)^T(\mathbf{z}_j - \mathbf{v}_i)$$

- Inner-product norm:

$$d^2_{\mathbf{A}_i}(\mathbf{z}_j, \mathbf{v}_i) = (\mathbf{z}_j - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{z}_j - \mathbf{v}_i)$$

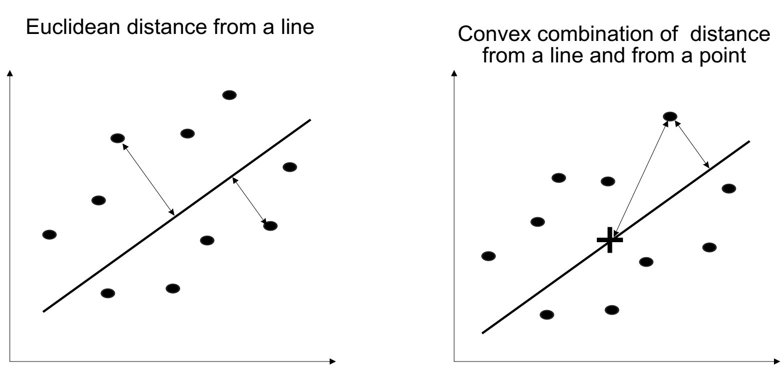
- Many other possibilities . . .

Generalized Prototypes (Varieties)



lines, circles, ellipses, regression functions, etc.

Corresponding Distance Measures



Mathematical Formulation of Clustering

Given the data:

$$\mathbf{z}_k = [z_{1k}, z_{2k}, \ldots, z_{nk}]^T \in \mathbb{R}^n, \quad k = 1, \ldots, N$$

Find:

the partition matrix:

$$\mathbf{U} = \begin{bmatrix} \mu_{11} & \ldots & \mu_{1k} & \ldots & \mu_{1N} \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ \mu_{c1} & \ldots & \mu_{ck} & \ldots & \mu_{cN} \end{bmatrix}$$

and the cluster prototype (centers):

$$\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}, \quad \mathbf{v}_i \in \mathbb{R}^n$$

Fuzzy Clustering: an Optimization Approach

Objective function (least-squares criterion):

$$J(\mathbf{Z}; \mathbf{V}, \mathbf{U}, \mathbf{A}) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{i,j}^{m} d_{\mathbf{A}_i}^{2}(\mathbf{z}_j, \mathbf{v}_i)$$

subject to constraints:

$0 \leq \mu_{i,j} \leq 1,$	$i = 1, \ldots, c, \; j = 1, \ldots, N$	membership degree
$0 < \sum_{j=1}^{N} \mu_{i,j} < 1,$	$i = 1, \ldots, c$	no cluster empty
$\sum_{i=1}^{c} \mu_{i,j} = 1,$	$j = 1, \ldots, N$	total membership

Fuzzy c-Means Algorithm

Repeat:

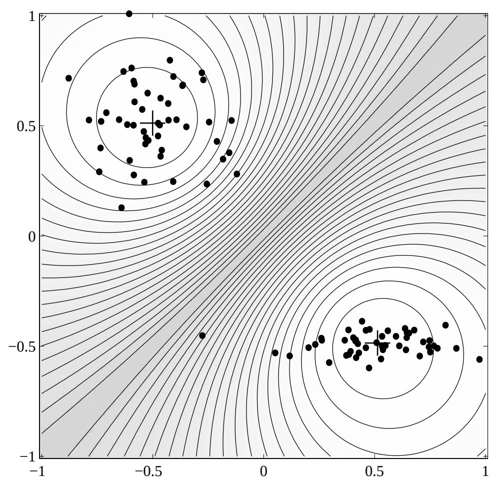
1. Compute cluster prototypes (means): $v_i = \frac{\sum_{k=1}^{N} \mu_{i,k}^{m} \mathbf{z}_k}{\sum_{k=1}^{N} \mu_{i,k}^{m}}$

2. Calculate distances: $d_{ik} = (\mathbf{z}_k - \mathbf{v}_i)^T (\mathbf{z}_k - \mathbf{v}_i)$

3. Update partition matrix: $\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (d_{ik}/d_{jk})^{1/(m-1)}}$

until $\|\Delta \mathbf{U}\| < \epsilon$

Failure to Discover Non-Spherical Clusters



Adaptive Distance Measure

Inner-product norm:

$$d_{\mathbf{A}_i}^{2}(\mathbf{z}_j, \mathbf{v}_i) = (\mathbf{z}_j - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{z}_j - \mathbf{v}_i)$$

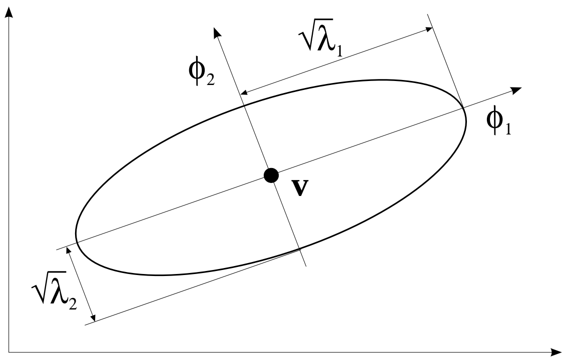
norm-inducing matrix

$$\mathbf{A}_i = \rho_i \det(\mathbf{F}_i)^{1/n} \mathbf{F}_i^{-1}$$

covariance matrix

$$\mathbf{F}_i = \frac{\sum_{k=1}^{N} \mu_{ik}^{m} (\mathbf{z}_k - \mathbf{v}_i)(\mathbf{z}_k - \mathbf{v}_i)^T}{\sum_{k=1}^{N} \mu_{ik}^{m}}$$

Inner-Product Norm



ellipsoid: $(\mathbf{z}-\mathbf{v})^T\mathbf{F}^{-1}(\mathbf{z}-\mathbf{v}) = \text{const}$

Gustafson–Kessel Algorithm

Repeat:

1. Compute cluster prototypes (means): $v_i = \dfrac{\sum_{k=1}^N \mu_{i,k}^m \mathbf{z}_k}{\sum_{k=1}^N \mu_{i,k}^m}$

2. Compute covariance matrices: $\mathbf{F}_i = \dfrac{\sum_{k=1}^N \mu_{ik}^m (\mathbf{z}_k-\mathbf{v}_i)(\mathbf{z}_k-\mathbf{v}_i)^T}{\sum_{k=1}^N \mu_{ik}^m}$

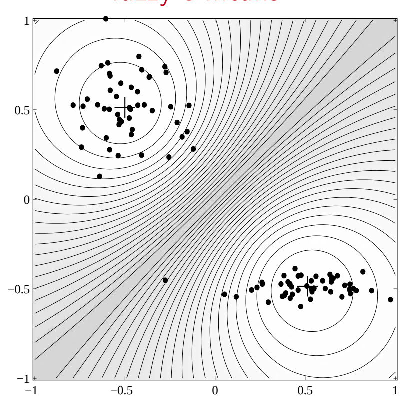
3. Compute distances: $d_{ik} = (\mathbf{z}_k-\mathbf{v}_i)^T \rho_i \det(\mathbf{F}_i)^{1/n}\mathbf{F}_i^{-1}(\mathbf{z}_k-\mathbf{v}_i)$

4. Compute partition matrix: $\mu_{ik} = \dfrac{1}{\sum_{j=1}^c (d_{ik}/d_{jk})^{1/(m-1)}}$

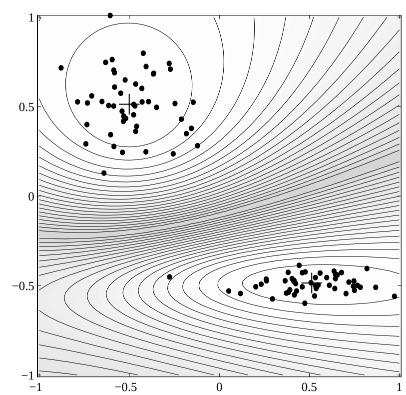
until $\|\Delta \mathbf{U}\| < \epsilon$

Clusters of Different Shape and Orientation

fuzzy c-means



Gustafson–Kessel



Number of Clusters

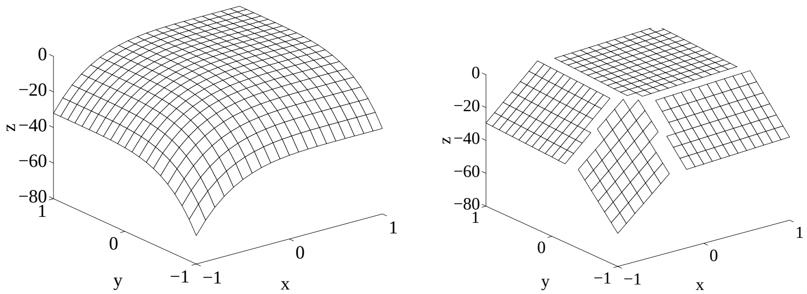
Validity measures

Fuzzy hypervolume: $V_h = \sum_{i=1}^{c} [\det(\mathbf{F}_i)]^{1/2}$

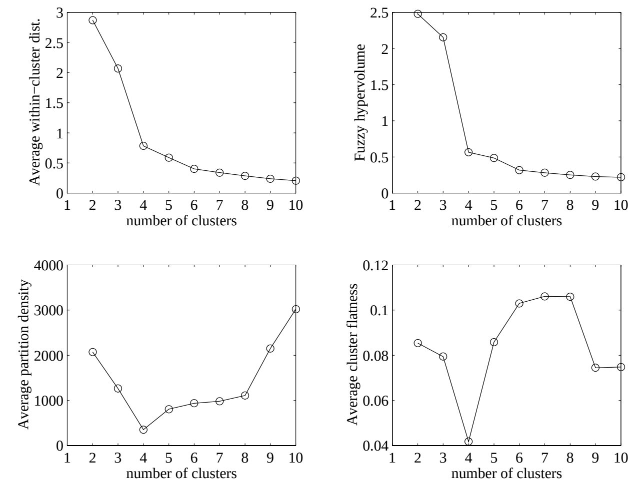
Average within-cluster distance: $D_w = \frac{1}{c}\sum_{i=1}^{c} \dfrac{\sum_{k=1}^N \mu_{ik}^m D_{ik}^2}{\sum_{k=1}^N \mu_{ik}^m}$

Xie-Beni index . . .

Validity Measures: Example

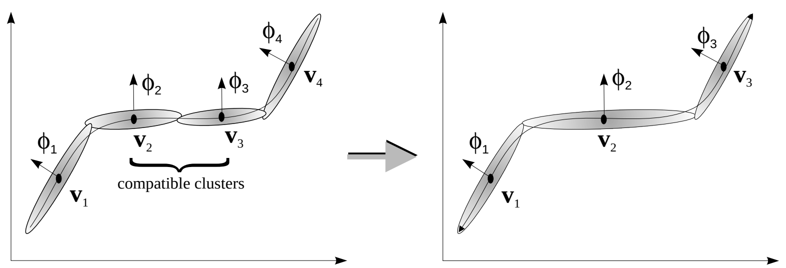


Validity Measures



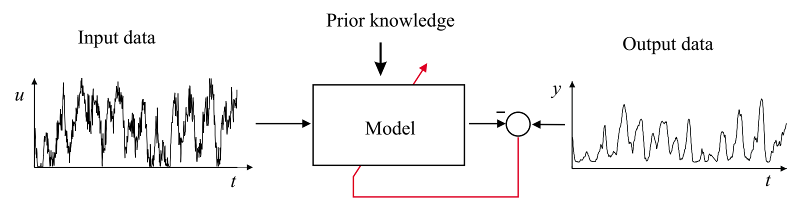
Number of Clusters

Compatible cluster merging



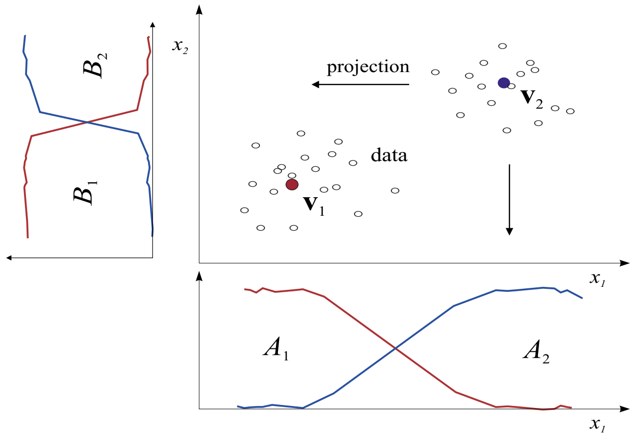
$$|\,\Phi_i \cdot \Phi_j\,| \geq k_1, \quad k_1 \to 1 \quad \text{and} \quad \|v_i - v_j\| \leq k_2, \quad k_2 \to 0$$

Data-Driven (Black-Box) Modeling



- linear model (for linear systems only, limited in use)
- neural network (black box, unreliable extrapolation)
- rule-based model (more transparent, ‘gray-box’)

Extraction of Rules by Fuzzy Clustering



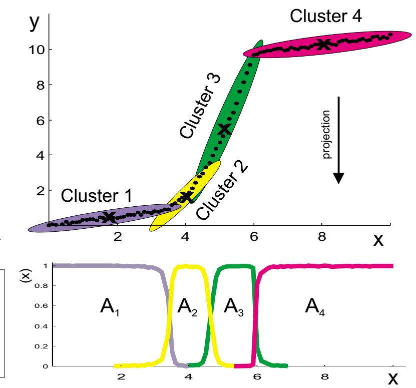
Extraction of Rules by Fuzzy Clustering

Takagi-Sugeno model

Rule-based description:

If x is A_1 then y = a_1x + b_1
If x is A_2 then y = a_2x + b_2

etc...

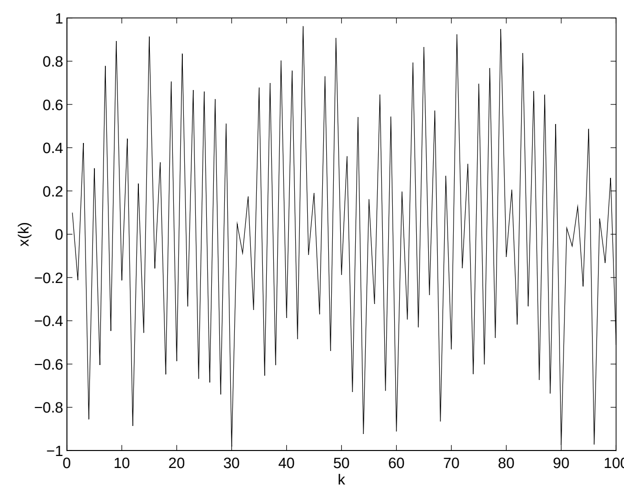


Example: Nonlinear Autoregressive System

$$x(k+1) = f(x(k)) + \epsilon(k)$$

$$f(x) = \begin{cases} 2x - 2, & 0.5 < x \\ -2x, & -0.5 \le x < 0.5 \\ 2x + 2, & x \le -0.5 \end{cases}$$

Identification Data



Structure Selection and Data Preparation

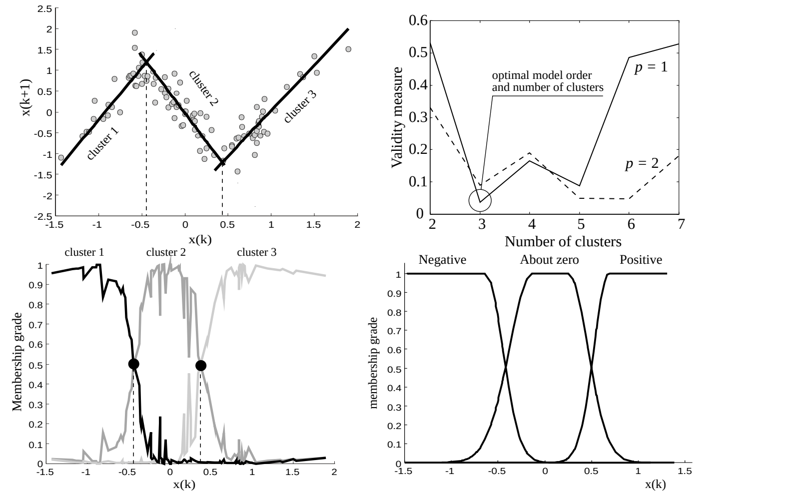
1) Choose model order p

$$x(k+1) = f(\underbrace{x(k), x(k-1), \ldots, x(k-p+1)}_{\mathbf{x}(k)})$$

2) Form pattern matrix $\mathbf{Z}$ to be clustered

$$\mathbf{Z}^T = \begin{bmatrix} x(1) & x(2) & \ldots & x(p) & x(p+1) \\ x(2) & x(3) & \ldots & x(p+1) & x(p+2) \\ \vdots & \vdots & & \vdots & \vdots \\ x(N-p) & x(N-p+1) & \ldots & x(N-1) & x(N) \end{bmatrix}$$

Clustering Results



Rules Obtained

1) If $x(k)$ is *Positive* then $x(k+1) = 2.0244x(k) - 2.0289$

2) If $x(k)$ is *About zero* then $x(k+1) = -1.8852x(k) + 0.0005$

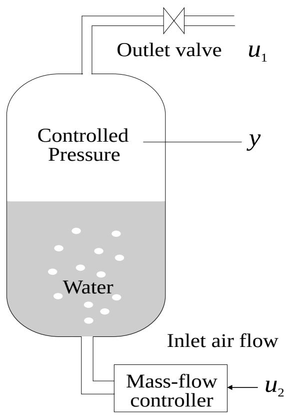
3) If $x(k)$ is *Negative* then $x(k+1) = 1.9050x(k) + 1.9399$

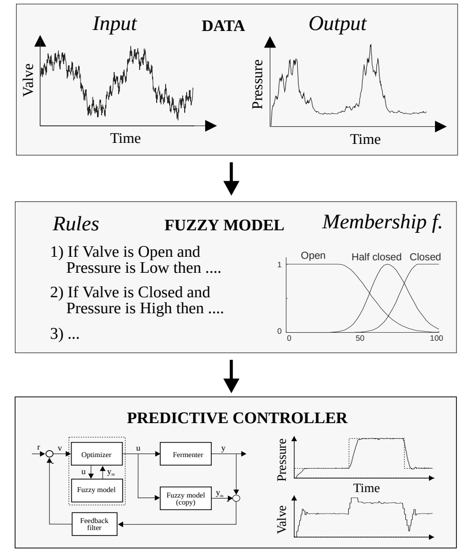
original function: $$f(x) = \begin{cases} 2x - 2, & 0.5 < x \\ -2x, & -0.5 \leq x < 0.5 \\ 2x + 2, & x \leq -0.5 \end{cases}$$

Real-World Applications of Clustering

- Identification of pressure dynamics in a fermenter.
- Engine load prediction for a combine harvester.
- Estimation of respiratory parameters (human lungs).
- Turbidity prediction in a rapid sand filter.
- Beer taste test score analysis.
- Modeling pilot behavior in a lead following task (on-going activity)

Identification of Pressure Dynamics





Obtained If–Then Rules

1. **If** $y(k)$ **is** *Low* **and** $u(k)$ **is** *Open*

   then $y(k+1) = 0.67y(k) + 0.0007u(k) + 0.35$

2. **If** $y(k)$ **is** *Medium* **and** $u(k)$ **is** *Half closed*

   then $y(k+1) = 0.80y(k) + 0.0028u(k) + 0.07$

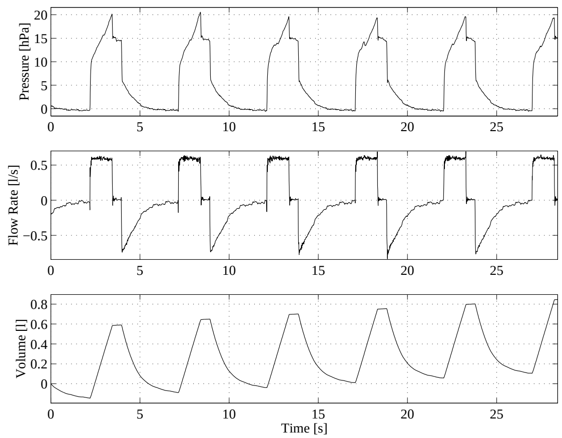
3. **If** $y(k)$ **is** *High* **and** $u(k)$ **is** *Closed*

   then $y(k+1) = 0.90y(k) + 0.0071u(k) - 0.39$

Estimation of respiratory parameters

- pressure–flow records of ventilated patients
- nonlinear model of respiratory mechanics
- local parameter estimation problem via fuzzy clustering
- experimental results for two groups of patients (with and without COPD)

Data set (patient without COPD)



Model of respiratory mechanics

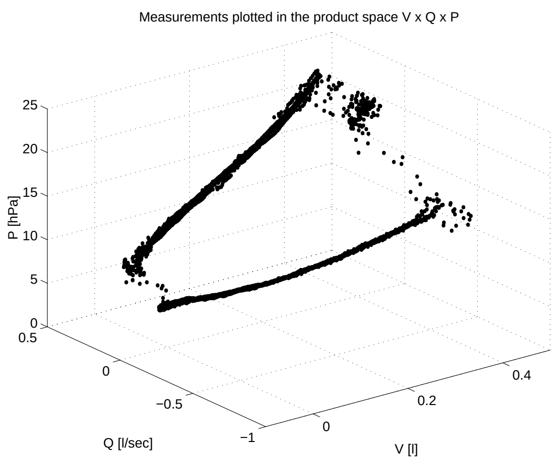
A classical single-compartment model

$$P = P_0 + EV + RV'$$

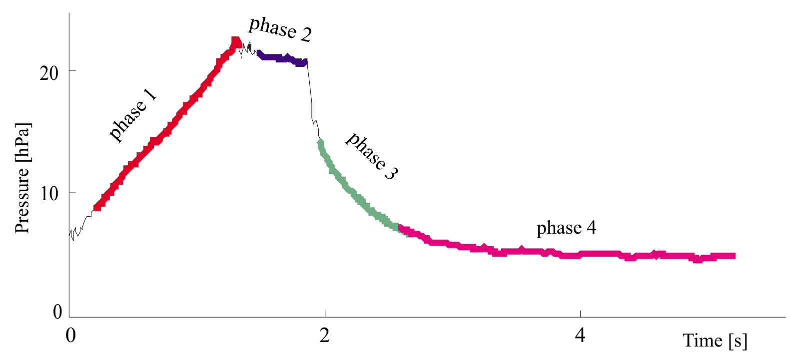
More general nonlinear model

$$P = f(P_0, V, V')$$

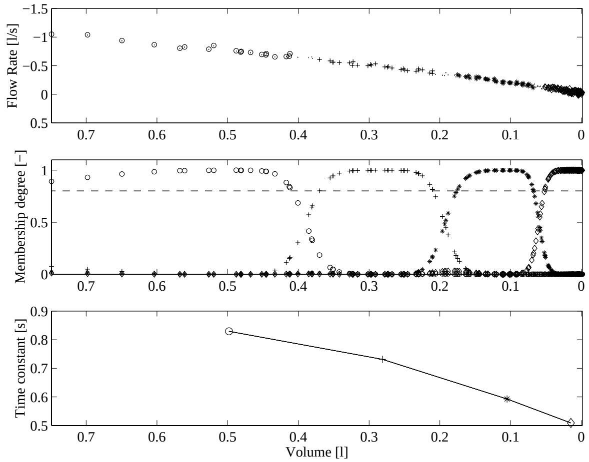
Data in three dimensions



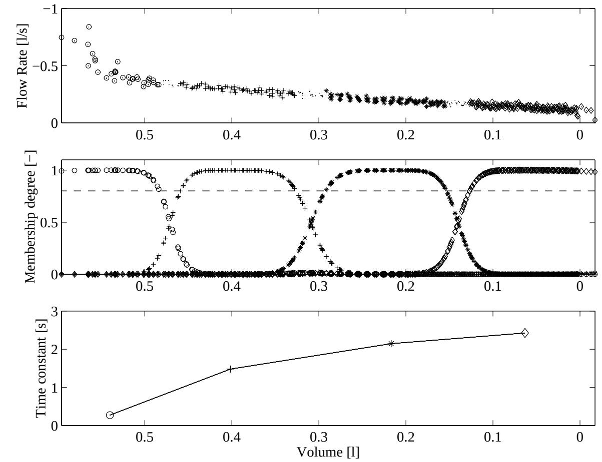
Partitioning of a respiratory cycle



Expiration – patient without COPD



Expiration – patient with COPD



Concluding Remarks

- Optimization approach to clustering
  - effective for metric (e.g., real-valued) data
  - accurate results for small to medium complexity problems
  - for large problems, convergence to local optima, slow
- Many other techniques
  - agglomerative methods
  - hierarchical splitting methods
  - graph-theoretic methods
- Variety of applications