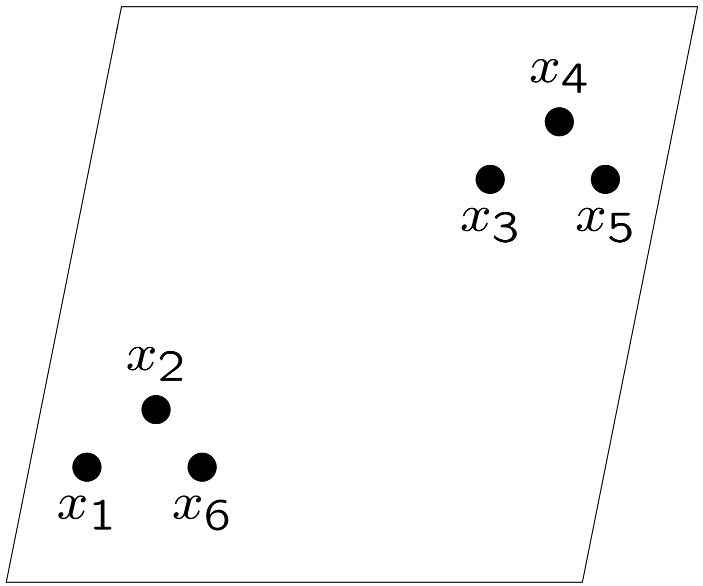
Chapter 9: Clustering

1. Sequential Clustering
2. Prototype–Based Clustering
3. Fuzzy Clustering
4. Relational Clustering
5. Cluster Tendency Assessment
6. Cluster Validity
7. Self–Organizing Map

Cluster Structure

numerical data set

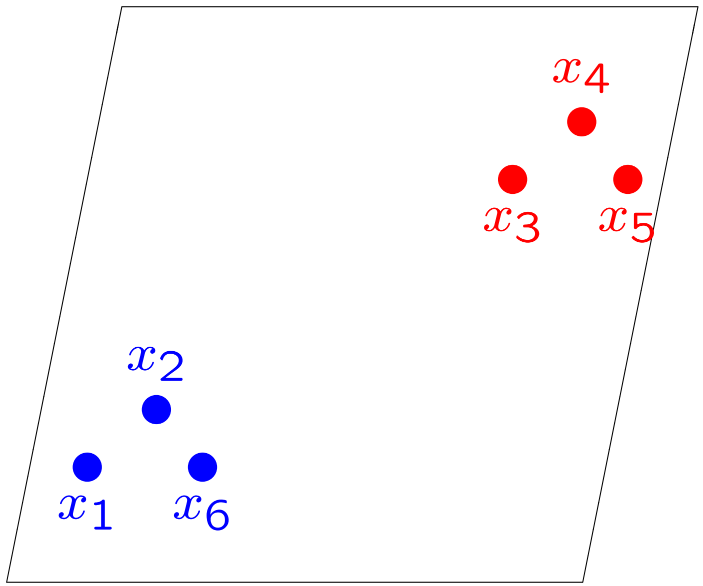
$$X = \{x_1, \ldots, x_n\} \subset \mathrm{IR}^p$$



$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

partition into cluster structure

$$X = C_1 \cup \ldots \cup C_c,\ C_i \neq \{\}\ \forall i$$
$$C_i \cap C_j = \{\}\ \forall i \neq j$$



$$C_1 = \{x_1, x_2, x_6\},\ C_2 = \{x_3, x_4, x_5\}$$

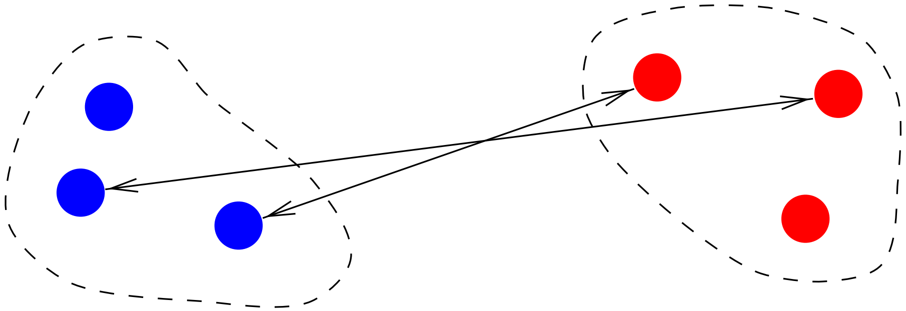
Prof. Dr. Thomas A. Runkler

SAHN Models (Sneath & Sokal '73)

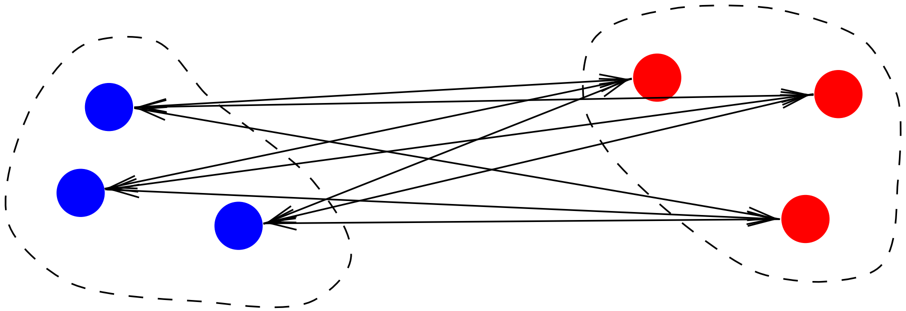
- sequential agglomerative hierarchical non–overlapping (SAHN)
- SAHN algorithm (bottom–up):
  1. initialize $\Gamma_n = \{\{o_1\}, \ldots, \{o_n\}\}$
  2. for $c = n-1, n-2, \ldots, 1$
     - $(i,j) = \underset{(C_r,C_s)\in\Gamma_c}{\text{argmin}}\ d(C_r, C_s)$
     - $\Gamma_c = (\Gamma_{c+1} - C_i - C_j) \cup (C_i \cup C_j)$
  3. output partitions $\Gamma_1, \ldots, \Gamma_n$
- sequential divisive hierarchical non–overlapping (SDHN)
- SDHN algorithm (top–down): correspondingly with initialization $\Gamma_0 = \{\{x_1, \ldots, x_n\}\}$, $c_0 = 1$

Distance Measures between Clusters

- single linkage: $d(C_r, C_s) = \min_{x \in C_r,\ y \in C_s} d(x, y)$
- complete linkage: $d(C_r, C_s) = \max_{x \in C_r,\ y \in C_s} d(x, y)$



- average linkage: $d(C_r, C_s) = \frac{1}{|C_r| \cdot |C_s|} \sum_{x \in C_r,\ y \in C_s} d(x, y)$



Prof. Dr. Thomas A. Runkler

Distance Measures between Clusters

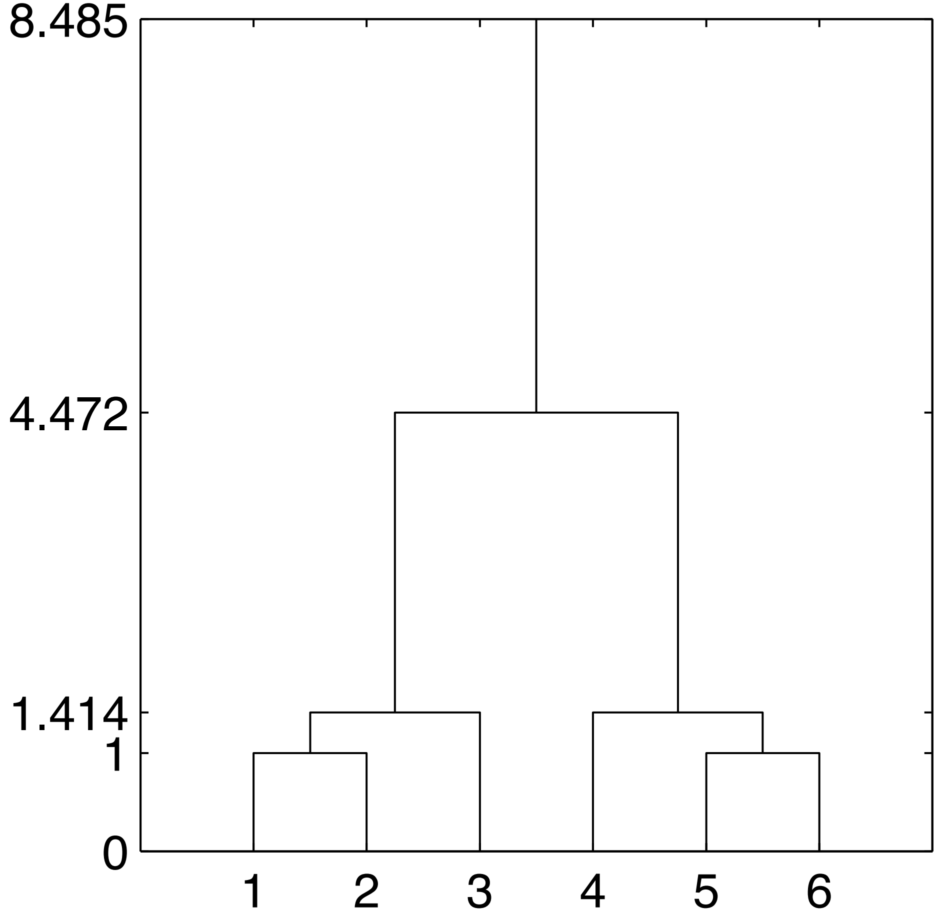
- distance of the centers

$$d(C_r, C_s) = \left\| \frac{1}{|C_r|} \sum_{x \in C_r} x - \frac{1}{|C_s|} \sum_{x \in C_s} x \right\|$$

- Ward's measure

$$d(C_r, C_s) = \frac{|C_r| \cdot |C_s|}{|C_r| + |C_s|} \left\| \frac{1}{|C_r|} \sum_{x \in C_r} x - \frac{1}{|C_s|} \sum_{x \in C_s} x \right\|$$

Dendrogram



Partition Matrix

partition matrix U represents cluster structure

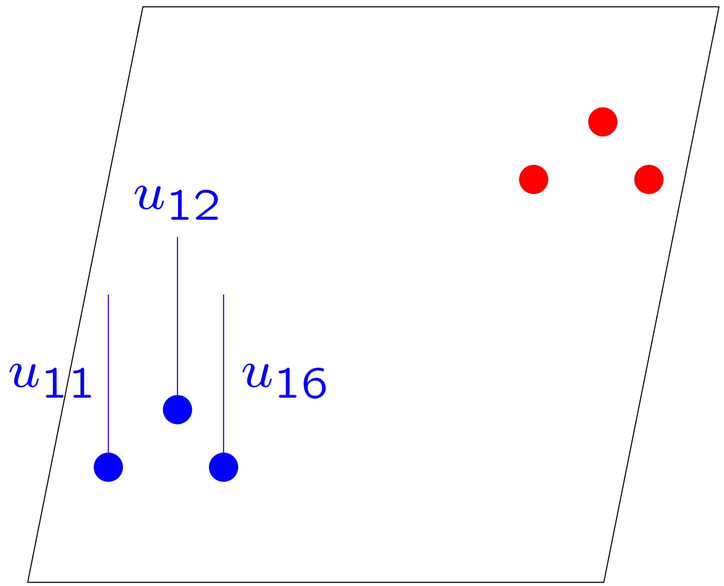
$$u_{ik} = \begin{cases} 1 & \text{if } x_k \in C_i \\ 0 & \text{if } x_k \notin C_i \end{cases}$$

each cluster contains at least one object

$$\sum_{k=1}^{n} u_{ik} > 0 \ \forall i$$

each object is assigned to exactly one cluster

$$\sum_{i=1}^{c} u_{ik} = 1 \ \forall k$$

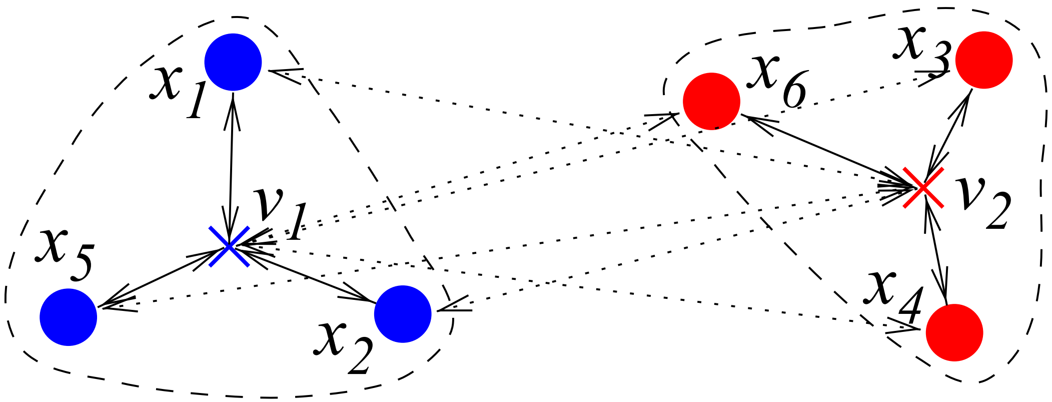


$$U = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Cluster Centers

- clusters are defined by cluster centers (prototypes) $V = \{v_1, \ldots, v_c\}$, hence c–means
- nearest neighbor rule: each x_k belongs to cluster i, iff

$$\|x_k - v_i\| = \min_{j=1,\ldots,c} \|x_k - v_j\|$$



Prof. Dr. Thomas A. Runkler

c–Means Model (Ball & Hall '65)

- determine cluster centers by minimizing

$$J_{CM}(U,V;X) = \sum_{i=1}^{c} \sum_{x_k \in C_i} \|x_k - v_i\|^2 = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \|x_k - v_i\|^2$$

- necessary condition for extrema: $\dfrac{\partial J_{CM}(U,V;X)}{\partial v_i} = 0$
- (local) extremum at

$$v_i = \frac{1}{|C_i|} \sum_{x_k \in C_i} x_k = \frac{\sum\limits_{k=1}^{n} u_{ik}\, x_k}{\sum\limits_{k=1}^{n} u_{ik}}$$

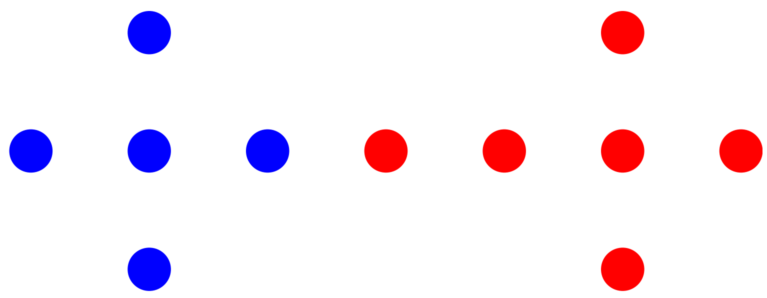
- v_i is first moment (center of gravity) of cluster i

Alternating Optimization

1. input data $X = \{x_1, \ldots, x_n\} \subset R^p$,
   cluster number $c \in \{2, \ldots, n-1\}$,
   maximum number of steps t_{max},
   distance measure $\|.\|$,
   distance measure for termination $\|.\|_\varepsilon$,
   termination threshold ε
2. initialize prototypes $V^{(0)} \subset R^p$
3. <u>for</u> $t = 1, \ldots, t_{\text{max}}$
   - compute $U^{(t)}(V^{(t-1)}, X)$
   - compute $V^{(t)}(U^{(t)}, X)$
   - <u>if</u> $\|V^{(t)} - V^{(t-1)}\|_\varepsilon \leq \varepsilon$, <u>stop</u>
4. output partition matrix $U \in [0, 1]^{c \times n}$,
   prototypes $V = \{v_1, \ldots, v_c\} \in R^p$

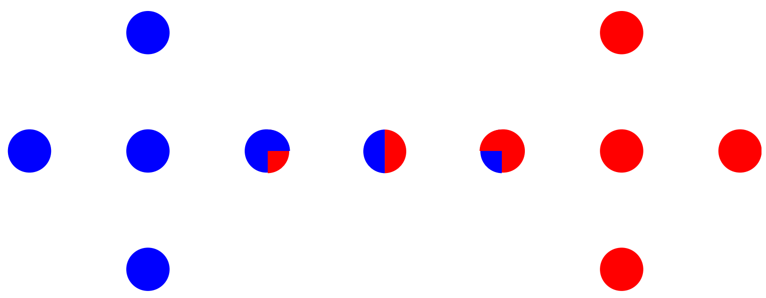
Butterfly Example

crisp partition:



center point is assigned to one cluster

fuzzy partition:



fuzzy assignment to different clusters

Fuzzy Partition Matrix

semantics of membership values:

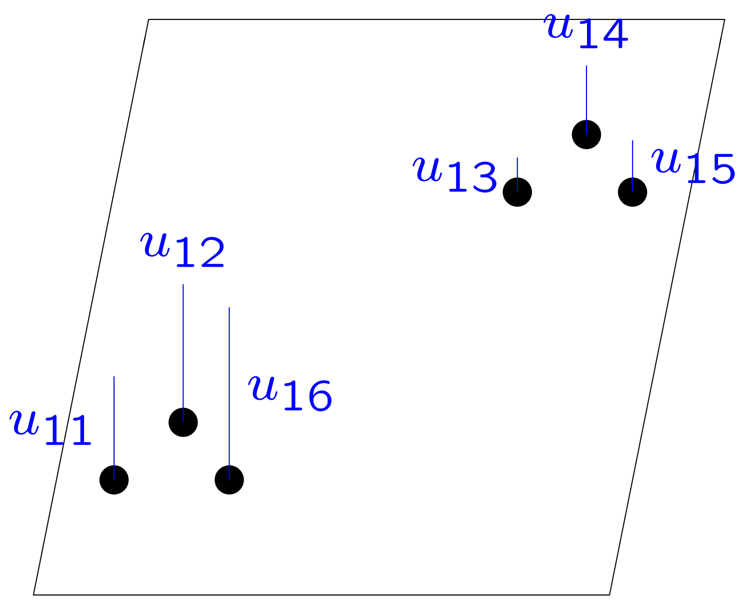
$$u_{ik} \begin{cases} = 0 & \text{if } x_k \notin C_i \\ = 1 & \text{if } x_k \in C_i \\ \in (0,1) & \text{otherwise} \end{cases}$$

fuzzy clusters are not empty

$$\sum_{k=1}^{n} u_{ik} > 0 \ \forall i$$

normalization condition

$$\sum_{i=1}^{c} u_{ik} = 1 \ \forall k$$



$$U = \begin{pmatrix} 0.6 & 0.8 & 0.2 & 0.4 & 0.3 & 1 \\ 0.4 & 0.2 & 0.8 & 0.6 & 0.7 & 0 \end{pmatrix}$$

Prof. Dr. Thomas A. Runkler

Fuzzy c–Means (Bezdek ’74)

Minimize

$$J_{FCM}(U,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} \|x_k - v_i\|^2$$

under the constraints

$$\sum_{k=1}^{n} u_{ik} > 0 \ \forall i$$

and

$$\sum_{i=1}^{c} u_{ik} = 1 \ \forall k.$$

fuzziness parameter $m \in (1,\infty)$, typically $m = 2$

Optimization of the FCM Model

Lagrange function

$$F_{FCM}(U,V,\lambda;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \|x_k - v_i\|^2 - \sum_{k=1}^{n}\lambda_k\left(\sum_{i=1}^{c} u_{ik} - 1\right)$$

necessary conditions for local extrema

$$\left.\begin{array}{l} \frac{\partial F_{FCM}}{\partial \lambda_k} = 0 \\[2em] \frac{\partial F_{FCM}}{\partial u_{ik}} = 0 \end{array}\right\} \Rightarrow \quad u_{ik} = 1 \Bigg/ \sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{\frac{2}{m-1}}$$

$$\frac{\partial F_{FCM}}{\partial v_i} = 0 \quad \Rightarrow \quad v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}$$

algorithm: alternating optimization

Possibilistic c–Means (Krishnapuram '93)

- relaxation of the constraint

$$\sum_{i=1}^{c} u_{ik} = 1 \ \forall k.$$

- trivial solution $U = 0$ is avoided by penalty term

$$J_{PCM}(U,V;X) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} \|x_k - v_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m$$

- solution: Cauchy function

$$u_{ik} = 1 \Bigg/ \left(1 + \left(\frac{\|x_k - v_i\|^2}{\eta_i}\right)^{\frac{1}{m-1}}\right)$$

Prof. Dr. Thomas A. Runkler

Noise Clustering (Davé '91)

- all data are partially assigned to a noise cluster
- distance to noise cluster constant δ
- cost function

$$J_{NC}(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \|x_k - v_i\|^2 + \sum_{k=1}^{n}\left(1 - \sum_{j=1}^{c} u_{jk}\right)^m \delta^2$$

- solution

$$u_{ik} = 1 \Bigg/ \left(\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{\frac{2}{m-1}} + \left(\frac{\|x_k - v_i\|}{\delta}\right)^{\frac{2}{m-1}}\right)$$

Gustafson and Kessel's Algorithm ('79)

clusters are hyperspheres with similar volumes

"natural" cluster shape from scatter matrix (cf. covariance matrix)

$$S_i = \sum_{k=1}^{n} u_{ik}^m (x_k - v_i)^T (x_k - v_i)$$

matrix norm with (cf. Mahalanobis distance)

$$A_i = \sqrt[p]{\rho_i \det(S_i)}\, S_i^{-1} \Rightarrow \det(A_i) = \rho_i$$

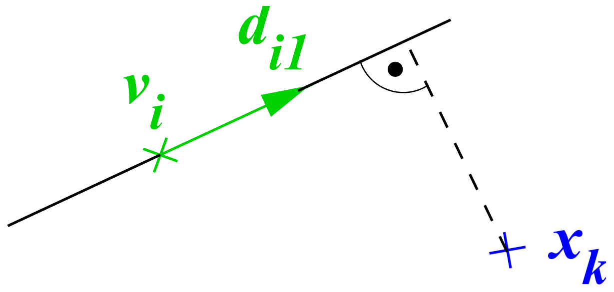
cluster volumines z.B. $\rho_1 = \ldots = \rho_c = 1$

e.g. FCM Gustafson Kessel cost function

$$J_{FCMGK}(U, V; X) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m (x_k - v_i)^T A_i (x_k - v_i)$$

Detection of Lines / Hyperplanes

- prototype: $(v_i, d_{i,1}, \ldots, d_{i,q})$, $i = 1, \ldots, c$, $q \in \{1, \ldots, p-1\}$



- distance measure (c–lines):

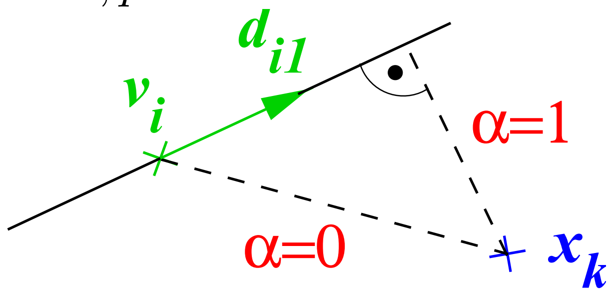
$$d(x_k, v_i, d_{i1}, \ldots, d_{iq}) = \sqrt{\|x_k - v_i\|^2 - \sum_{j=1}^{q} (x_k - v_i) d_{ij}^T}$$

- computation of the direction vectors:

$$d_{ij} = \underset{j}{\text{eig}} \sum_{k=1}^{n} u_{ik}^m (x_k - v_i)^T (x_k - v_i)$$

Detection of Line Segments

- prototype: $(v_i, d_{i,1}, \ldots, d_{i,q})$, $i = 1, \ldots, c$, $q \in \{1, \ldots, p-1\}$



- distance measure (c–elliptotypes):

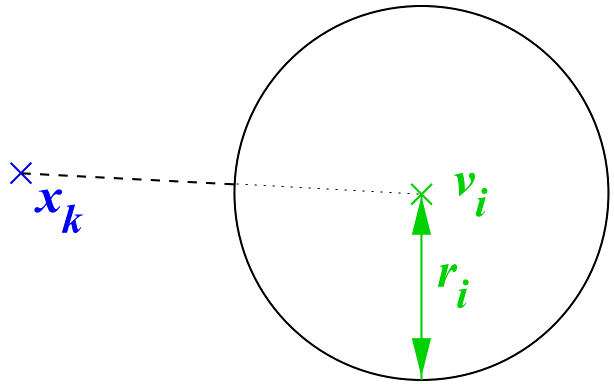
$$d(x_k, v_i, d_{i1}, \ldots, d_{iq}) = \sqrt{\|x_k - v_i\|^2 - \alpha \cdot \sum_{j=1}^{q} (x_k - v_i) d_{ij}^T}$$

- computation of the direction vectors:

$$d_{ij} = \underset{j}{\text{eig}} \sum_{k=1}^{n} u_{ik}^m (x_k - v_i)^T (x_k - v_i)$$

Detection of Circles

- prototype: (v_i, r_i), $i = 1, \ldots, c$



- distance measure (c–shells):

$$d(x_k, v_i, r_i) = \Big| \|x_k - v_i\| - r_i \Big|$$

- computation of the radii:

$$r_i = \frac{\sum_{k=1}^{n} u_{ik}^m \|x_k - v_i\|}{\sum_{k=1}^{n} u_{ik}^m}$$

