

Pattern Recognition V

Clustering

Michal Haindl

Faculty of Information Technology, KTI
Czech Technical University in Prague
Institute of Information Theory and Automation
Czech Academy of Sciences
Prague, Czechia



<http://ro.utia.cz/>



MI/MIE-ROZ 2023-2024

© Haindl

MI-ROZ - 05

1/61

Outline

- 1 Segmentation
 - Preprocessing
- 2 Segmentation Approaches
 - Thresholding
 - Clustering
 - Hierarchical Clustering
 - Partitional Clustering
- 3 Cluster Validity
- 4 Image Segmentation
 - Region Growing
 - Edge Detection
 - Discontinuity Detection I
 - Split and Merge
 - Motion Segmentation



© Haindl

MI-ROZ - 05

2/61

S SA CV IS P

Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$	data space
$X_i = [x_{i,1}, \dots, x_{i,d}]$	feature vector
$\Omega = \{\omega_1, \dots, \omega_K\}$	class label space
$df : \Theta \rightarrow \Omega$	segmentation / classification / clustering / unsupervised learning / partition

- $\Theta = \bigcup_{i=1}^K \Theta_i$
- $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- $f(\Theta_i) = \text{True} \quad \forall i$
- $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta|\omega_i) = \sum_{i=1}^K f(\Theta|\omega_i)$$



© Haindl

MI-ROZ - 05

3/61

S SA CV IS P

Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$	data space
$X_i = [x_{i,1}, \dots, x_{i,d}]$	feature vector
$\Omega = \{\omega_1, \dots, \omega_K\}$	class label space
$df : \Theta \rightarrow \Omega$	segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta|\omega_i) = \sum_{i=1}^K f(\Theta|\omega_i)$$



© Haindl

MI-ROZ - 05

3/61

Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Segmentation

$\Theta = \{X_i, i = 1, \dots, n\}$ data space
 $X_i = [x_{i,1}, \dots, x_{i,d}]$ feature vector
 $\Omega = \{\omega_1, \dots, \omega_K\}$ class label space
 $df : \Theta \rightarrow \Omega$ segmentation / classification / clustering / unsupervised learning / partition

- 1 $\Theta = \bigcup_{i=1}^K \Theta_i$
- 2 $\Theta_i \cap \Theta_j = \emptyset \quad \forall i \neq j$
- 3 $f(\Theta_i) = \text{True} \quad \forall i$
- 4 $f(\Theta_i \cup \Theta_j) = \text{False} \quad i \neq j, \Theta_i \text{ adjacent } \Theta_j$
- 5 $\forall X_r, X_s \in \Theta_i \Rightarrow X_r, X_s \in \text{connected path}$

mixture probability density decomposition

$$p(\Theta) = \sum_{i=1}^K P(\omega_i) p(\Theta | \omega_i) = \sum_{i=1}^K f(\Theta | \omega_i)$$



Number of Possible Clusterings

Stirling number of the second kind:

$$S(n, K) = \frac{1}{K!} \sum_{i=0}^K (-1)^{K-i} \binom{K}{i} i^n$$

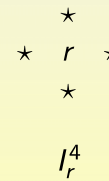
$S(15, 3) = 2\,375\,101$, $S(100, 5) \simeq 10^{68}$, $S(10^6, 10) \simeq 10^{10^6} \Rightarrow$
 evaluate all possible clustering is impossible



Connectivity

pixels r, s are connected if

- r, s are adjacent (4-connectivity, 8-connectivity)
- X_r, X_s are similar in some sense



Connectivity 2

r, s are 4-connected if $s \in I_r^4$

r, s are 8-connected if $s \in I_r^8$

r, s are m-connected if $s \in I_r^4 \vee s \in I_r^D \wedge I_s^4 \cap I_r^4 = \emptyset$

connected path from r_1 to $r_n \Leftrightarrow \forall i = 1, \dots, n-1 \ r_i, r_{i+1}$ are connected

0	0	1
0	1	0
1	1	0

 I_r^8

0	0	1
0	1	0
1	1	0

 I_r^m 

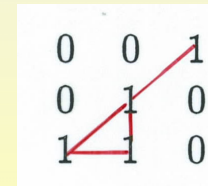
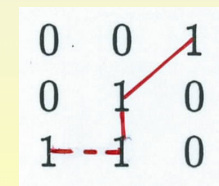
Connectivity 2

r, s are 4-connected if $s \in I_r^4$

r, s are 8-connected if $s \in I_r^8$

r, s are m-connected if $s \in I_r^4 \vee s \in I_r^D \wedge I_s^4 \cap I_r^4 = \emptyset$

connected path from r_1 to $r_n \Leftrightarrow \forall i = 1, \dots, n-1 \ r_i, r_{i+1}$ are connected

 I_r^8  I_r^m 

Preprocessing

- 1 Outlier Removal
- 2 Missing Data
- 3 Data Normalization



Preprocessing

- 1 Outlier Removal
- 2 Missing Data
- 3 Data Normalization

Outlier Removal

$$X \in \mathcal{N} \Rightarrow |X_{\text{outlier}}| > \mu + 3\sigma$$



Preprocessing

- 1 Outlier Removal
- 2 Missing Data
- 3 Data Normalization

Missing Data

- discard
- replace missing values with the corresponding mean / several kNN training samples
- predict from the underlying distribution
- proximity measure robust to missing data

$$b_i = \begin{cases} 0, & \text{both } x_i, y_i \text{ available;} \\ 1, & \text{otherwise.} \end{cases}$$

$$\delta(X, Y) = \frac{d}{d - \sum_{i=1}^d b_i} \sum_{\forall i: b_i=0} |x_i - y_i|$$



Preprocessing

- 1 Outlier Removal
- 2 Missing Data
- 3 Data Normalization

Data Normalization

- have a same scale
- have $\tilde{\mu} = 0$
- have a common standard deviation $\tilde{\sigma} = 1$
- resolve skewness

$$\text{skewness} = \frac{\sum (x_i - \mu)^3}{(n-1)v^{3/2}}$$

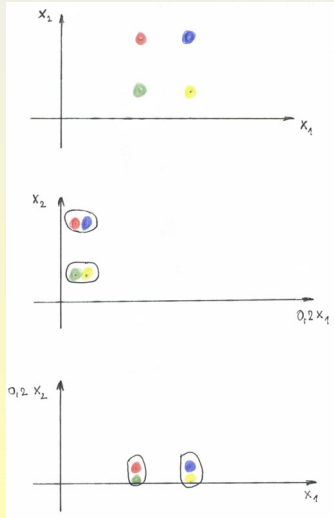
$$v = \frac{\sum (x_i - \mu)^2}{(n-1)}$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$



scaling (4×2 clusters)



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$

centering

$$\tilde{x}_{i,k} = x_{i,k} - \bar{x}_{i,.} \quad \bar{x}_{i,.} = \frac{1}{l} \sum_{j=1}^l x_{i,j}$$

$$\tilde{\mu}_{i,k} = 0$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$

object normalization

$$\begin{aligned}\tilde{x}_{i,k} &= \frac{x_{i,k} - \bar{x}_{i,\cdot}}{\sqrt{\sum_{j=1}^l (x_{i,j} - \bar{x}_{i,\cdot})^2}} & \bar{x}_{i,\cdot} &= \frac{1}{l} \sum_{j=1}^l x_{i,j} \\ \tilde{\mu}_{i,k} &= 0 \\ \|\tilde{x}_{i,\cdot}\|_2 &= \sqrt{\sum_{j=1}^l \tilde{x}_{i,j}^2} = 1\end{aligned}$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$

variable normalization I.

$$\begin{aligned}\hat{x}_{i,k} &= \frac{x_{i,k} - \bar{x}_{\cdot,k}}{\sqrt{\sum_{j=1}^n (x_{j,k} - \bar{x}_{\cdot,k})^2}} & \bar{x}_{\cdot,k} &= \frac{1}{n} \sum_{j=1}^n x_{j,k} \\ \|\hat{x}_{\cdot,k}\|_2 &= 1\end{aligned}$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$

z-score normalization

$$\hat{x}_{i,k} = \frac{x_{i,k} - \bar{x}_{\cdot,k}}{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{j,k} - \bar{x}_{\cdot,k})^2}}$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).

$$X_i = [x_{i,1}, \dots, x_{i,l}]^T$$

z-score normalization

$$\hat{x}_{i,k} = \frac{x_{i,k} - \bar{x}_{\cdot,k}}{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{j,k} - \bar{x}_{\cdot,k})^2}}$$

variable normalization II.

$$\hat{x}_{i,k} = f_k x_{i,k} \quad f_k > 0$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).
 $X_i = [x_{i,1}, \dots, x_{i,l}]^T$

range normalization ($< 0, 1 >$)

$$\hat{x}_{i,k} = \frac{x_{i,k} - \min_i \{x_{i,k}\}}{\max_i \{x_{i,k}\} - \min_i \{x_{i,k}\}}$$



Data Normalization

Euclidean metric requires that all variables are measured on the same scale ($\mu = 0, \sigma = 1$).
 $X_i = [x_{i,1}, \dots, x_{i,l}]^T$

- ⊖ appropriate scaling method not known in advance
- ⊖ object normalization is seldom effective



Segmentation of Multispectral / Multisensor Images

X_r $r = [r_1, r_2, r_3] \rightarrow 3D \text{ histogram(cube)}$
 $r_3 \in \{R, G, B\}$

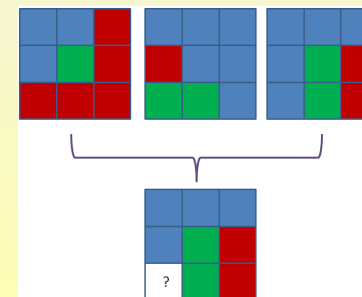
- multi-feature (multispectral) methods (thresholding $\rightarrow$ clustering)
- single-feature (monospectral / single sensor) methods
 - ① $X_{\dots, r_3}$ segmentation $\forall r_3$
 - ② thematic maps fusion (majority, composition)



Segmentation of Multispectral / Multisensor Images

X_r $r = [r_1, r_2, r_3] \rightarrow 3D \text{ histogram(cube)}$
 $r_3 \in \{R, G, B\}$

- multi-feature (multispectral) methods (thresholding $\rightarrow$ clustering)
- single-feature (monospectral / single sensor) methods
 - ① $X_{\dots, r_3}$ segmentation $\forall r_3$
 - ② thematic maps fusion (majority, composition)

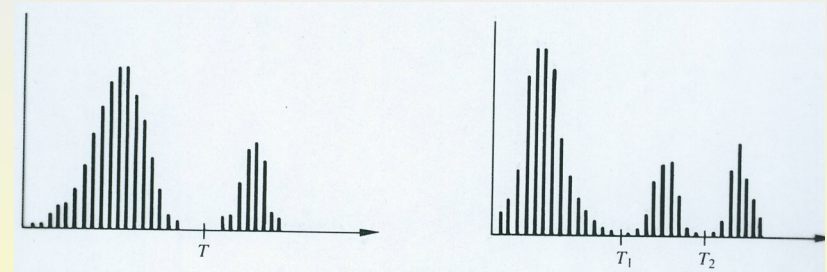


Segmentation Approaches

- spatial clustering
 - thresholding
 - multidimensional measurement SC
 - hierarchical SC
 - agglomerative
 - divisive
 - partitional SC
 - SC with thematic map modelling
- region growing
 - single linkage RG
 - split and merge
- discontinuity (edge) detection
 - derivative operators
 - combined detection
 - probabilistic relaxation
 - linkage combination techniques
- spatial model based methods
- motion segmentation



Thresholding



$$X_r > T \rightarrow \omega_1$$

$$X_r \leq T \rightarrow \omega_2$$

$$T_1 < X_r \leq T_2 \rightarrow \omega_1$$

$$X_r > T_2 \rightarrow \omega_2$$

$$X_r \leq T_1 \rightarrow \omega_3 \text{ (background)}$$



Thresholding

problems:

- ⊖ no contextual properties considered
- ⊖ multimodal histograms

threshold

- global
- local

band thresholding

$$X_r \in \langle \alpha; \beta \rangle \rightarrow \omega_1$$

$$X_r \notin \langle \alpha; \beta \rangle \rightarrow \omega_2$$



Histogram Based Methods

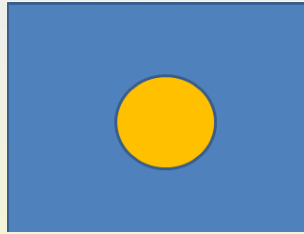
- advantages
 - ⊕ ignore spatial distribution
 - stable against orientation changes
 - stable against template misalignment
 - stable against occlusion
- disadvantages
 - ⊖ inaccurate localization
 - ⊖ dependent on relative objects sizes



Optimal Thresholding

assumptions:

- 2 brightness regions
- 2 values + additive Gaussian noise
- $\mu_1 < \mu_2$



mixture probability density

$$p(X) = P_1 p_1(X) + P_2 p_2(X) = \sum_{i=1}^2 \frac{P_i}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(X - \mu_i)^2}{2\sigma_i^2}\right]$$

$$P_1 + P_2 = 1$$

$$P(\text{error}) = P_2 \int_{-\infty}^T p_2(X) dX + P_1 \int_T^{\infty} p_1(X) dX$$



Optimal Thresholding 2

$$\frac{\partial P(\text{error})}{\partial T} = 0 \Rightarrow P_1 p_1(T) = P_2 p_2(T)$$

$$T_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$a = \sigma_1^2 - \sigma_2^2$$

$$b = 2(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2)$$

$$c = \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + 2\sigma_1^2 \sigma_2^2 \ln\left(\frac{\sigma_2 P_1}{\sigma_1 P_2}\right)$$

$$\text{if } \sigma_1^2 = \sigma_2^2 \Rightarrow T_1 = T_2 = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right)$$

LS parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, P_1)$ estimation

$$e = \frac{1}{n} \sum_{i=1}^n [p(X_i) - h(X_i)]^2$$



⊖ difficult, numerical solution only

Clustering

cluster - set of patterns whose inter-pattern distances are smaller than inter-pattern distances for patterns not in the same cluster

- a homogeneity and uniformity criterion
- no connectivity
- feature normalization - all features in equal range

⊕ little assumptions about data - exploratory data analysis

⊕ intrinsic data dimensionality

⊖ optimal number of classes

⊖ no optimal clustering method

⊖ usually no contextual properties considered

⊖ clustering algorithms find clusters even in entirely random data

⊖ results validation



Clustering

cluster - set of patterns whose inter-pattern distances are smaller than inter-pattern distances for patterns not in the same cluster

- a homogeneity and uniformity criterion
- no connectivity
- feature normalization - all features in equal range

⊕ little assumptions about data - exploratory data analysis

⊕ intrinsic data dimensionality

⊖ optimal number of classes

⊖ no optimal clustering method

⊖ usually no contextual properties considered

⊖ clustering algorithms find clusters even in entirely random data

⊖ results validation



Clustering

cluster - set of patterns whose inter-pattern distances are smaller than inter-pattern distances for patterns not in the same cluster

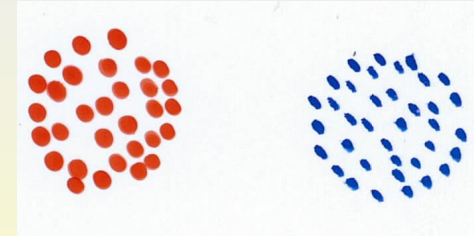
- a homogeneity and uniformity criterion
- no connectivity
- feature normalization - all features in equal range
- ⊕ little assumptions about data - exploratory data analysis
- ⊕ intrinsic data dimensionality
- ⊖ optimal number of classes
- ⊖ no optimal clustering method
- ⊖ usually no contextual properties considered
- ⊖ clustering algorithms find clusters even in entirely random data
- ⊖ results validation



Clustering Structures

$$K = 2 \quad I = 2$$

- compact & well-separated



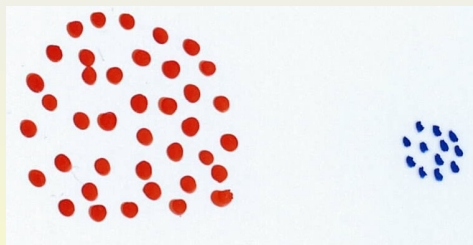
- unequal size
- elongated
- concentric
- touching
- clusters within clusters
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact & well-separated



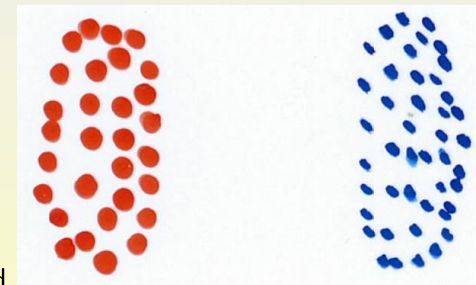
- unequal size
- elongated
- concentric
- touching
- clusters within clusters
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact & well-separated
- unequal size



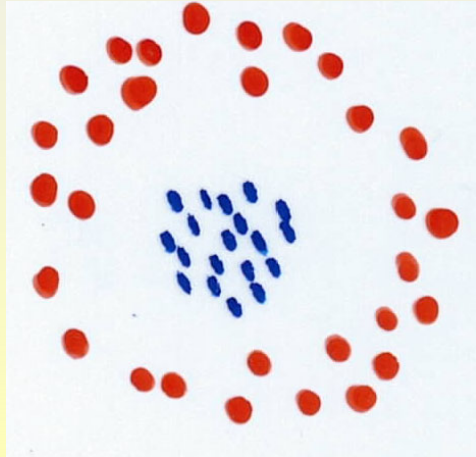
- elongated
- concentric
- touching
- clusters within clusters
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact& well-separated
- unequal size
- elongated



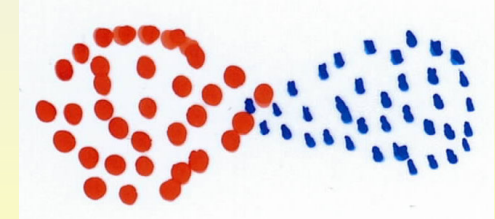
- concentric



Clustering Structures

$$K = 2 \quad I = 2$$

- compact& well-separated
- unequal size
- elongated
- concentric



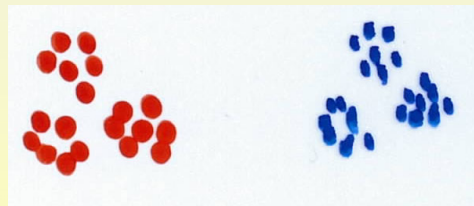
- touching
- clusters within clusters
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact& well-separated
- unequal size
- elongated
- concentric
- touching



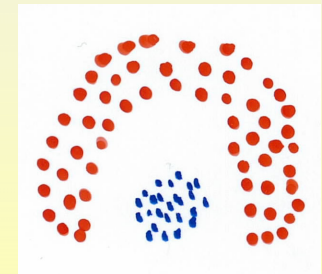
- clusters within clusters
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact& well-separated
- unequal size
- elongated
- concentric
- touching
- clusters within clusters



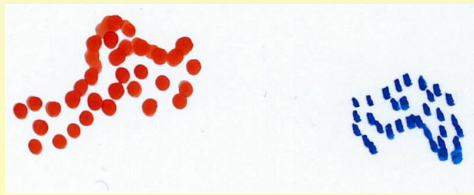
- linearly unseparable
- non compact



Clustering Structures

$$K = 2 \quad I = 2$$

- compact & well-separated
- unequal size
- elongated
- concentric
- touching
- clusters within clusters
- linearly unseparable



- non compact



Variables

- qualitative
 - **nominal** - every of its possible different state can be given a numerical code, e.g. marital status (1 - single, 2 - married, 3 - divorced, 4 - widowed)
 - binary, sex (man, woman)
 - **ordinal** - the states of nominal variable represented by numerical values can be arranged in a meaningful order, (1 - very much, 2 - much, 3 - useful, 4 - not very useful, 5 - unsuitable; A - excellent, B - very good, C - good, D - satisfactory, E - sufficient, F - failed)
- quantitative
 - **interval-scaled** - ordinal & real values permitted & differences between successive values are of equal size, e.g. temperature in degrees centigrade
 - **metric** - interval-scaled with natural zero
- mixed



Variables - Mappings

- nominal vector $\rightarrow$ binary vector

$${}^N X \quad {}^N x_i \in \{1, j_i\} \quad {}^B I = \sum_{i=1}^I j_i$$

$${}^N X \rightarrow {}^B X = \begin{matrix} & {}^N x_1 & & \dots & & {}^N x_I \\ (0, \dots, 1, \dots, 0; & \dots & & 0, \dots, 1, \dots, 0) \end{matrix}$$

$${}^N X = (2, 3, 5)^T \quad {}^B X = (0, 1, 0; 0, 0, 1, 0; 0, 0, 0, 0, 1)^T$$

- ordinal vector $\rightarrow$ binary vector preserves ordering

$${}^O X \rightarrow {}^B X = \begin{matrix} & {}^O x_1 & & \dots & & {}^O x_I \\ (1, 1, \dots, 1, 0, \dots, 0; & \dots & & 1, 1, \dots, 1, 0, \dots, 0) \end{matrix}$$

$${}^O X = (2, 3, 5)^T \quad \forall j_i = 5 \quad {}^B X = (1, 1, 0, 0, 0; 1, 1, 1, 0, 0; 1, 1, 1, 1, 1)^T$$



Variables - Mappings

- nominal vector $\rightarrow$ binary vector

$${}^N X \quad {}^N x_i \in \{1, j_i\} \quad {}^B I = \sum_{i=1}^I j_i$$

$${}^N X \rightarrow {}^B X = \begin{matrix} & {}^N x_1 & & \dots & & {}^N x_I \\ (0, \dots, 1, \dots, 0; & \dots & & 0, \dots, 1, \dots, 0) \end{matrix}$$

$${}^N X = (2, 3, 5)^T \quad {}^B X = (0, 1, 0; 0, 0, 1, 0; 0, 0, 0, 0, 1)^T$$

- ordinal vector $\rightarrow$ binary vector preserves ordering

$${}^O X \rightarrow {}^B X = \begin{matrix} & {}^O x_1 & & \dots & & {}^O x_I \\ (1, 1, \dots, 1, 0, \dots, 0; & \dots & & 1, 1, \dots, 1, 0, \dots, 0) \end{matrix}$$

$${}^O X = (2, 3, 5)^T \quad \forall j_i = 5 \quad {}^B X = (1, 1, 0, 0, 0; 1, 1, 1, 0, 0; 1, 1, 1, 1, 1)^T$$



Metrics

distance function - real-valued function satisfying:

- 1 $\delta(r, s) = \delta(s, r)$ symmetry
- 2 $\delta(r, s) \geq 0$ non-negativity
- 3 $\delta(r, r) = 0$ (for similarity index $\delta(r, r) \geq \max_s \{\delta(r, s)\}$)

metric - 1., 2., 3. &:

- 5 If $\delta(r, s) = 0 \Leftrightarrow X_r = X_s$ identity
- 6 $\delta(r, t) \leq \delta(r, s) + \delta(s, t)$ triangle inequality

If δ is a metric so is $\ln \delta$, $-\ln(\max \delta - \delta)$



Metrics 2

- Minkowski metric

$$\delta(r, s) = \left(\sum_{j=1}^I |x_{r,j} - x_{s,j}|^m \right)^{\frac{1}{m}}$$

invariant to translations $Y = X + A$

- $m = 1$ city-block (Manhattan, chessboard) distance
- $m = 2$ Euclidean distance
 - $\oplus$ invariant to orthogonal mappings (rotations) $Y = QX$
 $Q^T Q = I$
- $m \rightarrow \infty$ sup distance, Chebychev distance
 $\delta_\infty(r, s) = \max_i |x_{r,i} - x_{s,i}|$

- $\delta(r, s) = (X_r - X_s)^T Q (X_r - X_s)$ quadratic distance Q a PD scaling matrix



Metrics 3

- $\delta(r, s) = \sqrt{(X_r - X_s)^T Q^{-1} (X_r - X_s)}$ Mahalanobis distance (metric), Q covariance matrix,
 - $\oplus$ invariant to non-singular mappings, if $Q = \text{diag}[\]$
scale invariant too
 - $\ominus$ numerically complex
 - $\ominus$ Q based on all samples not on separate clusters

- $\delta(r, s) = \begin{cases} L & \text{if } \delta(r, s) > T \\ 0 & \text{otherwise} \end{cases}$ nonlinear distance



Non-Metric Distances

- sample correlation coefficient distance (2 sets of data)
 $\delta(r, s) = 1 - \rho_{r,s}$

$$\rho_{r,s} = \frac{\sum_{i=1}^n X_{r,i} X_{s,i} - n \mu_r \mu_s}{n \sigma_r \sigma_s} = \frac{\Sigma_{r,s}}{\sqrt{\Sigma_{r,r} \Sigma_{s,s}}}$$

$$\delta(r, s) = \sqrt{\frac{1}{I} \sum_{j=1}^I \left(\frac{x_{r,j} - x_{s,j}}{x_{r,j} + x_{s,j}} \right)^2}$$

$$\delta(r, s) = \frac{\sum_{j=1}^I |x_{r,j} - x_{s,j}|}{\sum_{j=1}^I |x_{r,j} + x_{s,j}|}$$

$$\delta(r, s) = \sum_{j=1}^I \frac{|x_{r,j} - x_{s,j}|}{|x_{r,j}| + |x_{s,j}|}$$



Nominal, Ordinal and Mixed Data Metrics

assumption - all data are binary vectors

notation

$$\alpha_1 = \sum_{i=1}^I \min\{x_{r,i}, x_{s,i}\}$$

$$\beta = \sum_{i=1}^I x_{r,i} - \alpha_1$$

$$\gamma = \sum_{i=1}^I x_{s,i} - \alpha_1$$

$$\alpha_0 = I - (\alpha_1 + \beta + \gamma)$$

similarity measures $\xi = \begin{cases} 0 & \text{matching 1s} \\ \alpha_0 & \text{0s and 1s equally weighted} \end{cases}$



Nominal, Ordinal and Mixed Data Metrics 2

$$\tau(r, r) = 1 \Rightarrow \delta(r, s) = 1 - \tau(r, s)$$

-

$$\tau_1(r, s) = \frac{\alpha_1 + \xi}{I}$$

- δ is metric

$$\tau_2(r, s) = \frac{\alpha_1 + \xi}{\alpha_1 + \beta + \gamma + \xi}$$

- δ is metric

$$\tau_3(r, s) = \frac{\alpha_1 + \xi}{\alpha_1 + \xi + 2(\beta + \gamma)}$$

-

$$\tau_4(r, s) = \frac{2(\alpha_1 + \xi)}{2(\alpha_1 + \xi) + \beta + \gamma}$$

-

$$\tau_5(r, s) = \frac{1}{4} \left(\frac{\alpha_1}{\alpha_1 + \beta} + \frac{\alpha_1}{\alpha_1 + \gamma} + \frac{\xi}{\gamma + \xi} + \frac{\xi}{\beta + \xi} \right)$$



Nominal, Ordinal and Mixed Data Metrics 3

- corresponds to correlation measures

$$\tau_6(r, s) = \frac{\alpha_1 + \xi}{\sqrt{(\alpha_1 + \beta)(\alpha_1 + \gamma)(\xi + \beta)(\xi + \gamma)}}$$

mixed data $X = (X_{nominal}^T, X_{ordinal}^T, X_{metric}^T)$

$$\delta(r, s) = c_1 \delta_n(r, s; n) + c_2 \delta_o(r, s; o) + c_3 \delta_m(r, s; m)$$

$c_i > 0$ and δ_i are corresponding (metric) distance functions (δ is also metric)



Nominal, Ordinal and Mixed Data Metrics - Example

$$I = 10 \quad X_r = [1, 0, 1, 1, 1, 0, 1, 1, 0, 0] \quad X_s = [0, 1, 1, 0, 1, 0, 1, 0, 1, 0]$$

$$\alpha_1 = 3, \quad \beta = 3, \quad \gamma = 2, \quad \alpha_0 = 2, \quad \xi = 2 = 0$$

- $\tau_1(r, s) = \frac{3+2}{10} = 0.5 = 0.3$

- $\tau_2(r, s) = \frac{3+2}{3+3+2+2} = 0.5 = 0.38$

- $\tau_3(r, s) = \frac{3+2}{3+2+2(3+2)} = 0.33 = 0.23$

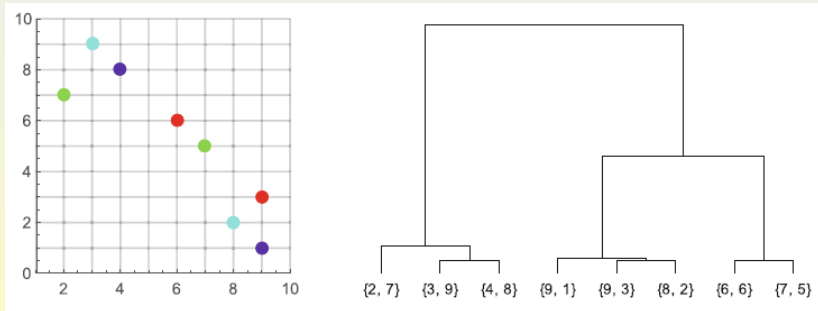
- $\tau_4(r, s) = \frac{2(3+2)}{2(3+2)+3+2} = 0.66 = 0.55$

- $\tau_5(r, s) = \frac{1}{4} \left(\frac{3}{6} + \frac{3}{5} + \frac{2}{4} + \frac{2}{5} \right) = 0.5 = 0.28$

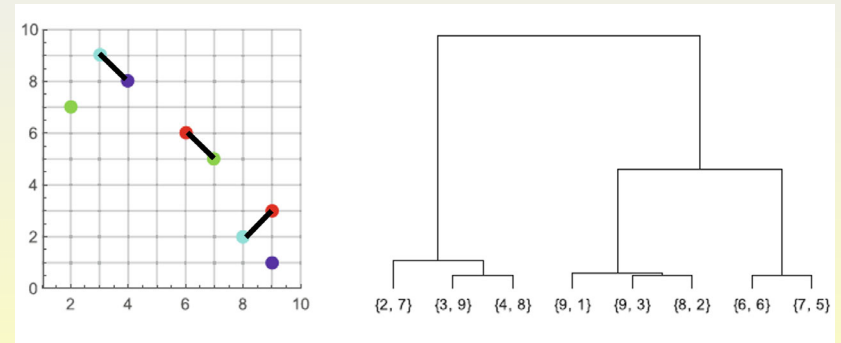
- $\tau_6(r, s) = \frac{3+2}{\sqrt{6 \cdot 5 \cdot 5 \cdot 4}} = 0.204 = 0.22$



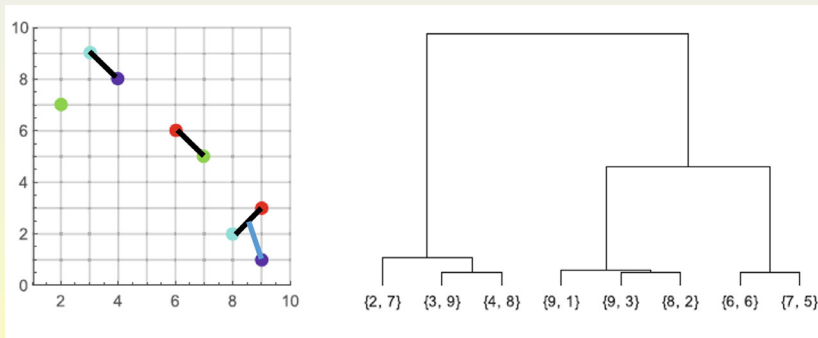
Hierarchical Clustering

 I_1 

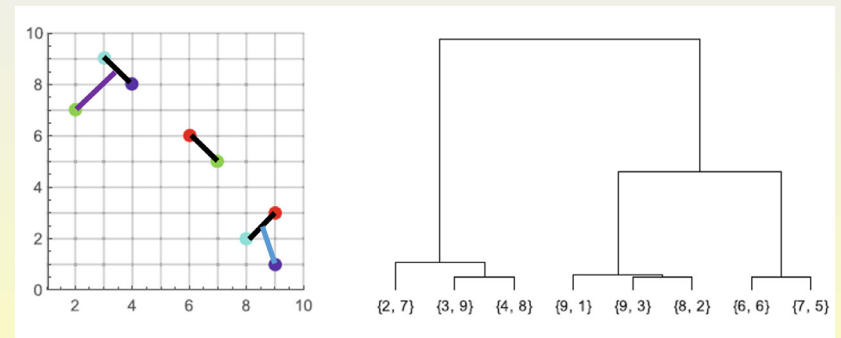
Hierarchical Clustering

 I_1 

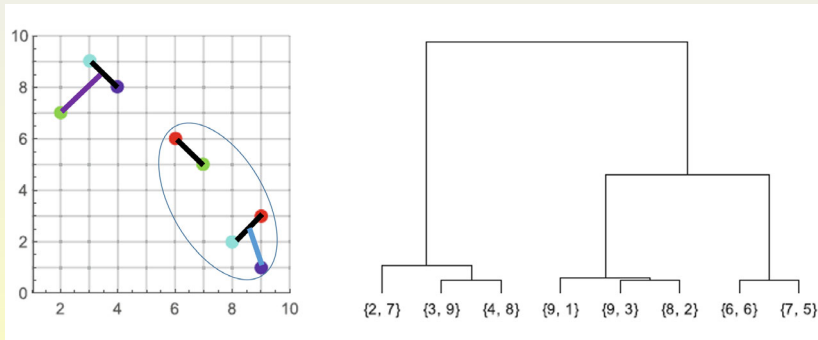
Hierarchical Clustering

 I_1 

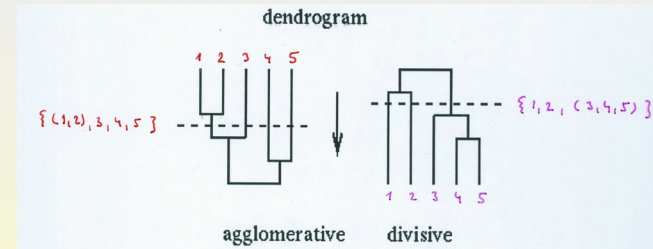
Hierarchical Clustering

 I_1 

Hierarchical Clustering

I₁

Hierarchical Clustering

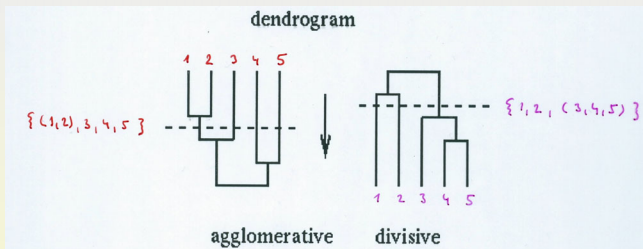


SAHN - sequential, agglomerative, hierarchical, nonoverlapping

- 1 set the iteration number $m = 0$
- 2 find the cluster pair such $\delta[(s), (t)] = \min_{\forall \text{ pairs}} \{\delta[(i), (j)]\}$
- 3 $m = m + 1$ merge s, t clusters, define clustering level
- 4 update dissimilarity matrix (delete entry corresponding to s, t clusters, add new cluster val.)



Hierarchical Clustering



SAHN - sequential, agglomerative, hierarchical, nonoverlapping

- 1 set the iteration number $m = 0$
- 2 find the cluster pair such $\delta[(s), (t)] = \min_{\forall \text{ pairs}} \{\delta[(i), (j)]\}$
- 3 $m = m + 1$ merge s, t clusters, define clustering level
- 4 update dissimilarity matrix (delete entry corresponding to s, t clusters, add new cluster val.)



Hierarchical Clustering

- ⊖ difficult for large number of patterns
- ⊖ patterns cannot move between clusters
- ⊖ problem where to cut the dendrogram $\Rightarrow$ number of classes

$$\delta[(k), (s, t)]_{SAHN} = \alpha_s \delta[(k), (s)] + \alpha_t \delta[(k), (t)] + \beta \delta[(s), (t)] + \gamma |\delta[(k), (s)] - \delta[(k), (t)]|$$



method	α_s	α_t	β	γ
Single-link	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Complete-link	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
UPGMA	$\frac{n_s}{n_s+n_t}$	$\frac{n_s}{n_s+n_t}$	0	0
WPGMA	$\frac{1}{2}$	$\frac{1}{2}$	0	0
UPGMC	$\frac{n_s}{n_s+n_t}$	$\frac{n_s}{n_s+n_t}$	$-\frac{n_s n_t}{(n_s+n_t)^2}$	0
WPGMC	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{-1}{4}$	0
Ward's	$\frac{n_s+n_k}{n_s+n_t+n_k}$	$\frac{n_t+n_k}{n_s+n_t+n_k}$	$-\frac{n_k}{n_s+n_t+n_k}$	0

- PGM - pair group method (clusters merged in pairs)
- U - unweighted method (all clusters members treated equally)
- W - weighted method (all clusters the same importance, e.g. smaller clusters members bigger influence)
- A - arithmetic average

$$\delta[(s), (t)] = \frac{1}{n_t n_s} \sum_{i=1}^{n_t} \sum_{j=1}^{n_s} \delta(x_{t,i} x_{s,j})$$

- C - centroids

$$\delta[(s), (t)] = ||\bar{x}_{(s)} - \bar{x}_{(t)}||_2$$

$$\delta[(k), (s, t)]_{S, \textcolor{blue}{C}} = \frac{1}{2} \delta[(k), (s)] + \frac{1}{2} \delta[(k), (t)] \mp \frac{1}{2} |\delta[(k), (s)] - \delta[(k), (t)]|$$

dissimilarity matrix (-S, + C)

$I.$	1	2	3	4
1	0			
2	3	0		
3	12	6	0	
4	8	9	10	0
5	7	13	5	11

1	2	3	5	4
↓	↙			
$(1, 2)_{S, C}$		3	5	4

$$\delta[(k), (s, t)] = \frac{1}{2}\delta[(k), (s)] + \frac{1}{2}\delta[(k), (t)] \mp \frac{1}{2}|\delta[(k), (s)] - \delta[(k), (t)]|$$

$$\delta[(3), (1, 2)] = \frac{1}{2}12 + \frac{1}{2}6 \mp \frac{1}{2}|12 - 6| = \begin{cases} 6, \\ 12, \end{cases}$$

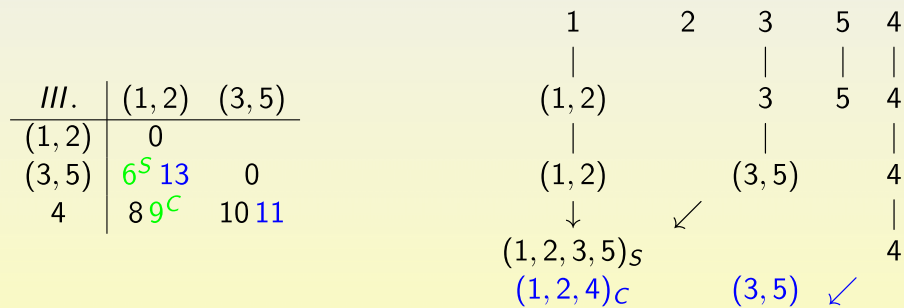
$$\delta[(4), (1, 2)] = \frac{1}{2}8 + \frac{1}{2}9 \mp \frac{1}{2}|8 - 9| = \begin{cases} 8, \\ 9, \end{cases}$$

$$\delta[(5), (1, 2)] = \frac{1}{2}7 + \frac{1}{2}13 \mp \frac{1}{2}|7 - 13| = \begin{cases} 7, \\ 13, \end{cases}$$

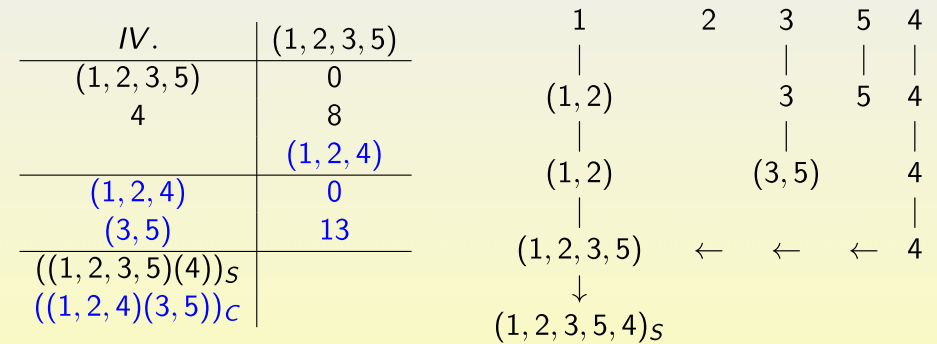
$II.$	$(1, 2)$	3	4
$(1, 2)$	0		
3	6 12	0	
4	8 9	10	0
5	7 13	5	11

$$\begin{array}{ccccc}
 1 & 2 & 3 & 5 & 4 \\
 | & & | & | & | \\
 (1, 2) & & 3 & 5 & 4 \\
 | & & \downarrow & \swarrow & | \\
 (1, 2) & & (3, 5)_{S.C} & & 4
 \end{array}$$

Single (-) / Complete (+)-Link Example



Single (-) / Complete (+)-Link Example

S tree $\neq$ C tree

Partitional Clustering

objective to find the optimal clustering ($\forall$ partitions - computationally unfeasible)

K has to be known in advance



Partitional Clustering

objective to find the optimal clustering ($\forall$ partitions - computationally unfeasible)

K has to be known in advance

K-means (Isodata)

- 1 Select K Θ_i $i = 1, \dots, K$.
- 2 Compute μ_i^{j-1} $i = 1, \dots, K$.
- 3 $\forall \{X_l, i\}$ $\delta(X_l, \mu_i) = [(X_l - \mu_i)^T (X_l - \mu_i)]^{\frac{1}{2}}$
- 4 $\forall l$ $\delta(X_l, \mu_i) = \min_j \{\delta(X_l, \mu_j)\} \Rightarrow X_l \rightarrow \omega_i$
- 5 stopping criterion (no of iteration, changes ...)
- 6 $j = j + 1$ go to 2.

(7.) Apply heuristic splitting / merging criteria.

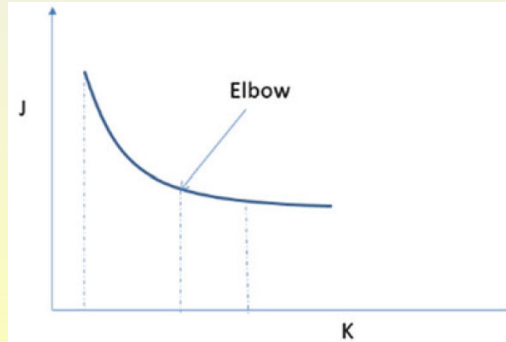


Partitional Clustering

objective to find the optimal clustering ($\forall$ partitions - computationally unfeasible)

K has to be known in advance

K estimation (Elbow method)



$$J(K) = \frac{1}{n} \sum_{i=1}^n \|x_i - \mu_j^*\|$$

μ_j^* assigned center for vector x_i



Partitional Clustering

K-means (Isodata)

- 1 Select $K \quad \Theta_i \quad i = 1, \dots, K.$
- 2 Compute $\mu_i^{j-1} \quad i = 1, \dots, K.$
- 3 $\forall \{X_l, i\} \quad \delta(X_l, \mu_i) = [(X_l - \mu_i)^T (X_l - \mu_i)]^{\frac{1}{2}}$
- 4 $\forall l \quad \delta(X_l, \mu_i) = \min_j \{\delta(X_l, \mu_j)\} \Rightarrow X_l \rightarrow \omega_i$
- 5 stopping criterion (no of iteration, changes ...)
- 6 $j = j + 1$ go to 2.

(7.) Apply heuristic splitting / merging criteria.

- ⊕ patterns can move between clusters
- ⊖ suboptimal partitions resulting from used heuristics
- ⊖ only metric data
- ⊖ sensitive to the initial μ_i^0 selection (rerun many times)
- ⊖ sensitive to outliers



Heuristic Partitioning

- 1 $K = 1, i = 1$
- 2 Assign X_i into ω_K .
- 3 $i = i + 1$ stop if $i > \text{card}(S)$
- 4 Assign X_i to the first of previously generated clusters (j) for which the first (leading) element $\delta(X_i, \mu_j) < \text{threshold}$.
- 5 If there is no such a cluster and $K < K_{\max}$ set $K = K + 1$ go to 2., otherwise leave X_i unallocated (ω_0) and go to 3.

- ⊕ can classify also non-metric data
- ⊕ fast
- ⊖ partitioning depends on data ordering



Heuristic Partitioning 2

- 1 $K = 0$
- 2 $K = K + 1$ stop if $K > K_{\max} \vee \forall X$ assigned
- 3 Assign X_i into ω_K if is not a member of a cluster and its Euclidean distance from all other yet unassigned X_j is maximum. $X_i \rightarrow \mu_K$
- 4 Assign X_i into ω_K if not assigned in 3. and $\delta(X_i, \mu_K) \leq \text{threshold}$ and X_i closest to μ_K ; otherwise go to 2.
- 5 Recalculate μ_K , go to 4.



Clustering by the Affinity Propagation

- ⊕ unsupervised,
- ⊕ number of clusters K unknown,
- ⊕ simultaneously considers all data points as potential centers,
- ⊖ recursive,
- ⊙ magnitude of each message reflects the current affinity that one data point has for choosing another data point as its center

Frey, B.J. and Dueck, D., Clustering by passing messages between data points, Science, vol. 315, no. 5814, pp. 972–976, 2007



Clustering by the Affinity Propagation

Message passing:

- 1 similarities (x_k center for x_i ?)

$$s(i, k) = -||x_i - x_k||^2$$

- 2 responsibility ($x_i \rightarrow x_k$ to be a center, alternative centers ?)

$$r(i, k) \leftarrow s(i, k) - \max_{k's.t. k \neq k} \{a(i, k) + s(i, k)\}$$

- 3 availability ($x_k \rightarrow x_i$ to join a center, other points support ?)

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i's.t. i \notin \{i, k\}} \max\{0, r(i, k)\} \right\}$$

$$a(k, k) \leftarrow \sum_{i's.t. i \notin \{i, k\}} \max\{0, r(i, k)\}$$



Clustering by the Affinity Propagation

Message passing termination:

- fixed number of iterations
- changes in the messages fall below a threshold
- local decisions stay constant for some number of iterations



Cluster Validity

problems

- Are the resulting clusters "real" or merely an artifact of the clustering algorithm ?
- How many clusters are present in data?
- How well does a hierarchy fit to data?

problems with the definition of cluster and the meaning of validity

Validation

- external - compares recovered structure to an a priori structure
- internal - no a priori information, tries to determine if the structure is intrinsically appropriate for the data
- relative - compares two structures and measures their relative merit



Cluster Validity

problems

- Are the resulting clusters "real" or merely an artifact of the clustering algorithm ?
- How many clusters are present in data?
- How well does a hierarchy fit to data?

problems with the definition of cluster and the meaning of validity

Validation

- external - compares recovered structure to an a priori structure
- internal - no a priori information, tries to determine if the structure is intrinsically appropriate for the data
- relative - compares two structures and measures their relative merit



Cluster Validity 2

Validity Paradigm

- 1 Identify the clustering structure (hierarchy, clustering, cluster) and the type of validation (external, internal, relative).
- 2 Select an index.
- 3 Select a hypothesis of "no structure".
- 4 Obtain (by theory or simulation) the baseline distribution of the index under the "no structure" hypothesis.
- 5 Compute the index for the structure being tested.
- 6 Formally test the hypothesis of "no structure" by determining whether the observed index is "unusual".



Cluster Validity 3

- clustering tendency - a hypothesis that the data have a continuous uniform distribution over some compact convex sampling window
 - scan test - unusually large count of number of patterns in the most populous region of the sampling window indicates a clustering
- Calinski Harabash index - relative test, optimal number of clusters (more compact and separated) $\max_K CH(K)$

$$CH(K) = \frac{n - K}{K - 1} \left[\frac{E_1^2}{E_K^2} - 1 \right]$$

$E_K^2 = \sum_{k=1}^K e_k^2$, e_k^2 is the sum of squared Euclidean distances between patterns in k and μ_k



Cluster Validity 3

- clustering tendency - a hypothesis that the data have a continuous uniform distribution over some compact convex sampling window
 - scan test - unusually large count of number of patterns in the most populous region of the sampling window indicates a clustering
- Calinski Harabash index - relative test, optimal number of clusters (more compact and separated) $\max_K CH(K)$

$$CH(K) = \frac{n - K}{K - 1} \left[\frac{E_1^2}{E_K^2} - 1 \right]$$

$E_K^2 = \sum_{k=1}^K e_k^2$, e_k^2 is the sum of squared Euclidean distances between patterns in k and μ_k



Cluster Validity 4

- validity of partitions
 - external test, $a(d)$ no. of pairs objects in the same (different) cluster for both classifications

$$Rand_index = \frac{a + d}{\binom{n}{2}}$$

$$compactness = \frac{\sum_{i=1}^{n_k} \#_{i,k}}{n_k(n_k - 1)}$$

where n_k is number of points in the k cluster, $\#_{i,k}$ is number of $n_k - 1$ nearest neighbours to a k cluster point belonging to k

$$isolation = 1 - \frac{\sum_{i=1}^{n_k} \bar{\#}_{i,k}}{(n - n_k)(n - n_k - 1)}$$

where n is number of data points, $\bar{\#}_{i,k}$ is number of $n - n_k - 1$ nearest neighbours to a k cluster point not belonging to k

ideally $compactness = isolation = 1$



Cluster Validity 4

- validity of partitions
 - external test, $a(d)$ no. of pairs objects in the same (different) cluster for both classifications

$$Rand_index = \frac{a + d}{\binom{n}{2}}$$

$$compactness = \frac{\sum_{i=1}^{n_k} \#_{i,k}}{n_k(n_k - 1)}$$

where n_k is number of points in the k cluster, $\#_{i,k}$ is number of $n_k - 1$ nearest neighbours to a k cluster point belonging to k

$$isolation = 1 - \frac{\sum_{i=1}^{n_k} \bar{\#}_{i,k}}{(n - n_k)(n - n_k - 1)}$$

where n is number of data points, $\bar{\#}_{i,k}$ is number of $n - n_k - 1$ nearest neighbours to a k cluster point not belonging to k

ideally $compactness = isolation = 1$



Cluster Validity 4

- validity of partitions
 - external test, $a(d)$ no. of pairs objects in the same (different) cluster for both classifications

$$Rand_index = \frac{a + d}{\binom{n}{2}}$$

$$compactness = \frac{\sum_{i=1}^{n_k} \#_{i,k}}{n_k(n_k - 1)}$$

where n_k is number of points in the k cluster, $\#_{i,k}$ is number of $n_k - 1$ nearest neighbours to a k cluster point belonging to k

$$isolation = 1 - \frac{\sum_{i=1}^{n_k} \bar{\#}_{i,k}}{(n - n_k)(n - n_k - 1)}$$

where n is number of data points, $\bar{\#}_{i,k}$ is number of $n - n_k - 1$ nearest neighbours to a k cluster point not belonging to k

ideally $compactness = isolation = 1$



Scatter Matrix Criteria

Scatter matrix for i th cluster

$$S_i = \sum_{X \in S_i} (X - \mu_i)(X - \mu_i)^T$$

Within-cluster scatter matrix - measure of cluster compactness

$$S_W = \sum_{i=1}^K S_i$$

Between-cluster scatter matrix

$$S_B = \sum_{i=1}^K n_i(\mu_i - \mu)(\mu_i - \mu)^T$$

Total scatter matrix - independent on data partitioning

$$S_T = \sum_X (X - \mu)(X - \mu)^T$$

$$S_T = S_W + S_B$$



Scatter Matrix Criteria 2

Criteria invariant to nonsingular linear transformation of data (λ_i eigenvalues of $\mathcal{S}_W^{-1}\mathcal{S}_B$)

- $\min \text{tr}(\mathcal{S}_W)$ proportional to sum of the variances
- $\min \frac{|\mathcal{S}_W|}{|\mathcal{S}_T|} = \prod_{i=1}^I \frac{1}{1+\lambda_i}$
- $\max \text{tr}\{\mathcal{S}_W^{-1}\mathcal{S}_B\} = \sum_{i=1}^I \lambda_i$
equivalent to minimizing the square error under the Mahalanobis metric
- $\max \left\{ \frac{\text{tr}\mathcal{S}_T}{\text{tr}\mathcal{S}_W} \right\}$
- $\max \frac{|\mathcal{S}_B|}{|\mathcal{S}_W|} = \prod_{i=1}^I \lambda_i$
⊖ often 0
- ⊖ invariant criteria are more likely to possess multiple local extrema



Image Segmentation - Region Growing

advantages:

- ⊕ considers contextual properties
- ⊕ more robust in noisy images than edge-based methods

problems:

- ⊖ inherent dependence on neighbourhood definition
- ⊖ dependence on region order pixels examination
- ⊖ dependence on initial seed selection
 - prior information
 - histogram peaks
- ⊖ stop growing criterion



Region Growing 2

Merging

- 1 segment image into regions (e.g., pixels)
- 2 if possible merge iteratively all adjacent regions satisfying a homogeneity criterion

Pixel aggregation

- 1 select number and location of seeds, homogeneity criterion
- 2 assign every pixel to the best seed according the criterion, for more equal seeds assign randomly



Split and Merge

- Θ subdivided into set of disjoint regions
- iteratively split and/or merge regions according to some criterion

- 1 $f(\Theta_i) = \text{False} \Rightarrow$ split Θ_i into 4 disjoint quadrants
- 2 $f(\Theta_j \cup \Theta_k) = \text{True} \Rightarrow$ merge Θ_j, Θ_k
- 3 no 1. $\vee$ 2. possible $\Rightarrow$ stop

$$f(\Theta_i) = \text{True} \Leftrightarrow X_r = \omega \quad \forall r \in \Theta_i$$



Split and Merge

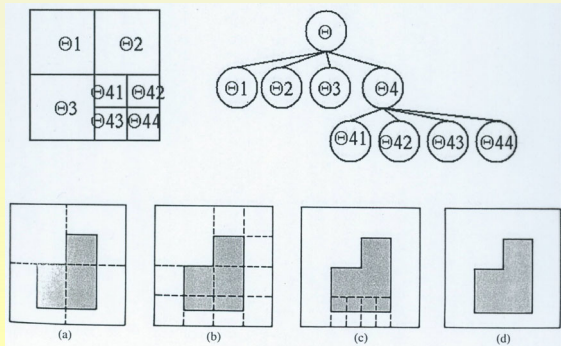
- Θ subdivided into set of disjoint regions
- iteratively split and/or merge regions according to some criterion

1 $f(\Theta_i) = \text{False} \Rightarrow$ split Θ_i into 4 disjoint quadrants

2 $f(\Theta_j \cup \Theta_k) = \text{True} \Rightarrow$ merge Θ_j, Θ_k

3 no 1. $\vee$ 2. possible $\Rightarrow$ stop

$$f(\Theta_i) = \text{True} \Leftrightarrow X_r = \omega \quad \forall r \in \Theta_i$$



UTIA

Split and Merge

- Θ subdivided into set of disjoint regions
- iteratively split and/or merge regions according to some criterion

1 $f(\Theta_i) = \text{False} \Rightarrow$ split Θ_i into 4 disjoint quadrants

2 $f(\Theta_j \cup \Theta_k) = \text{True} \Rightarrow$ merge Θ_j, Θ_k

3 no 1. $\vee$ 2. possible $\Rightarrow$ stop

$$f(\Theta_i) = \text{True} \Leftrightarrow X_r = \omega \quad \forall r \in \Theta_i$$

$$f(\Theta_i) = \text{True} \Leftrightarrow 0.8 \text{card}(\Theta_i) \leq \{r : r \in \Theta_i, |X_r - \mu_i| \leq 2\sigma_i\}$$

UTIA

Motion Segmentation

spatial techniques

- registered images $X_{\dots, r_3}, X_{\dots, r_3+i}$
- constant illumination
- stationary reference image (built if necessary from several frames)

$$\Delta(r_3, i) = \begin{cases} 1 & \text{if } |X_{r_1, r_2, r_3} - X_{r_1, r_2, r_3+i}| > T \\ 0 & \text{otherwise} \end{cases}$$

difference image pixel = 1 denotes motion

Accumulative difference - ignores changes occurring only between 2 subsequent frames

Frequency domain techniques

- static background
- several moving objects

UTIA