Tema 11

Classificação

Não Supervisionada

Professora:
Ariane Machado Lima

SISTEMAS DE INFORMAÇÃO

EACH

Classificação ou aprendizado não supervisionado

- Objetivo: classificar elementos onde:
  - Não se sabe a classe dos elementos (não há uma amostra de treinamento)
  - Não se conhece sobre o processo de geração dos padrões (das classes)
  - Às vezes não se sabe nem quantas classes estão envolvidas

- Normalmente a única informação são os vetores de características dos elementos

Agrupamentos (*Clustering*)

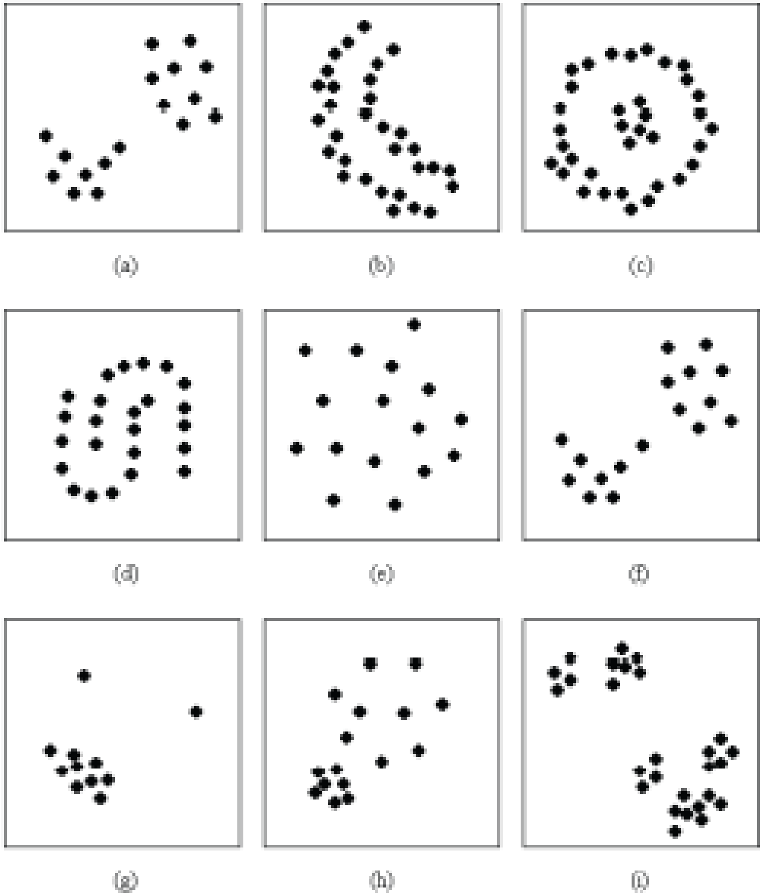
- Realização de partições do espaço com base em um critério

O que é dado?

Os vetores de características dos elementos a serem classificados

Podem apresentar diferentes distribuições no espaço

Exemplos - como você separaria?

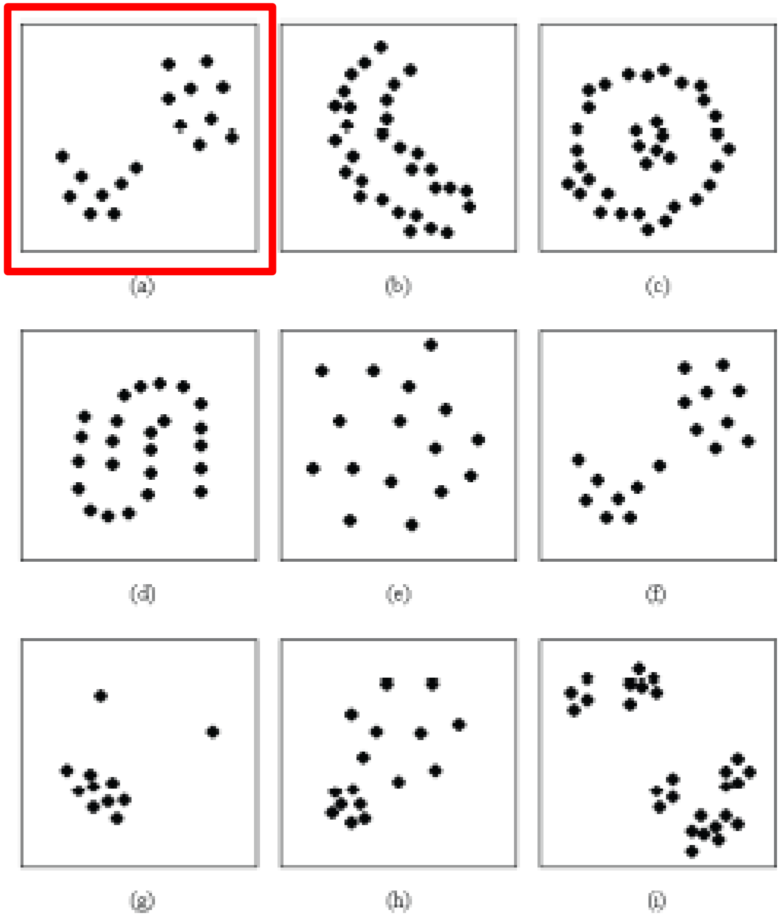


[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

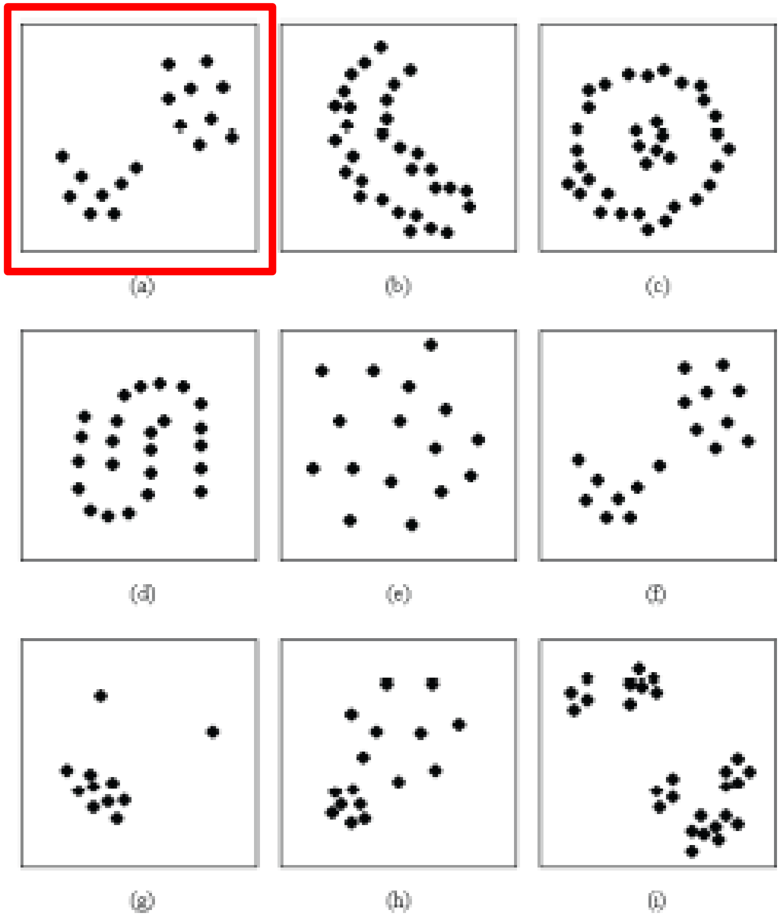
EACH

Exemplos

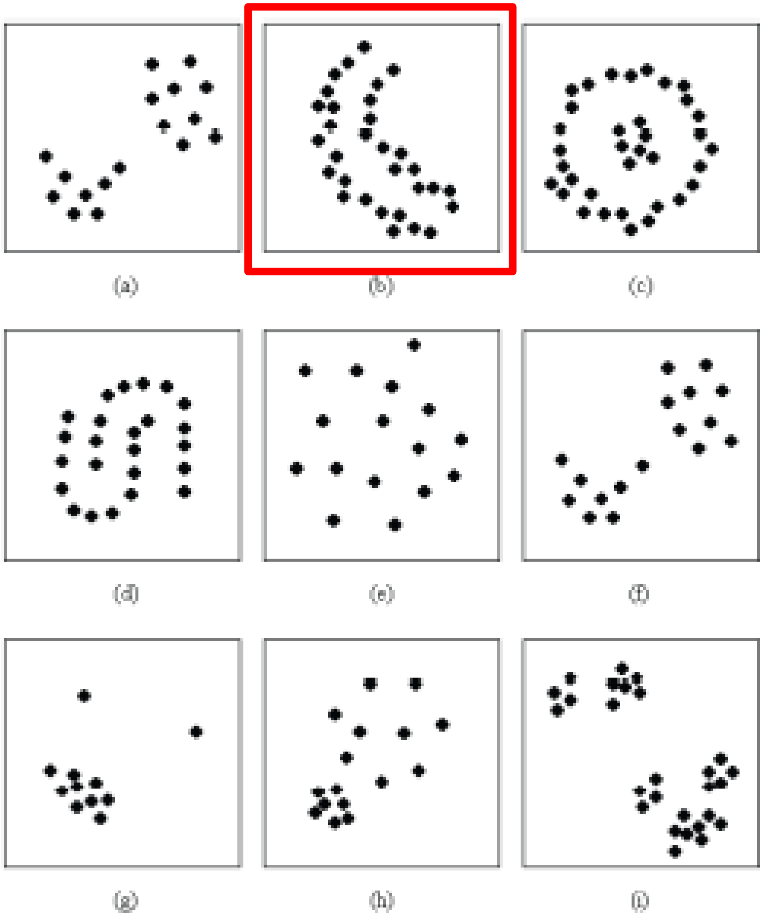


Exemplos

2 grupos linearmente separáveis

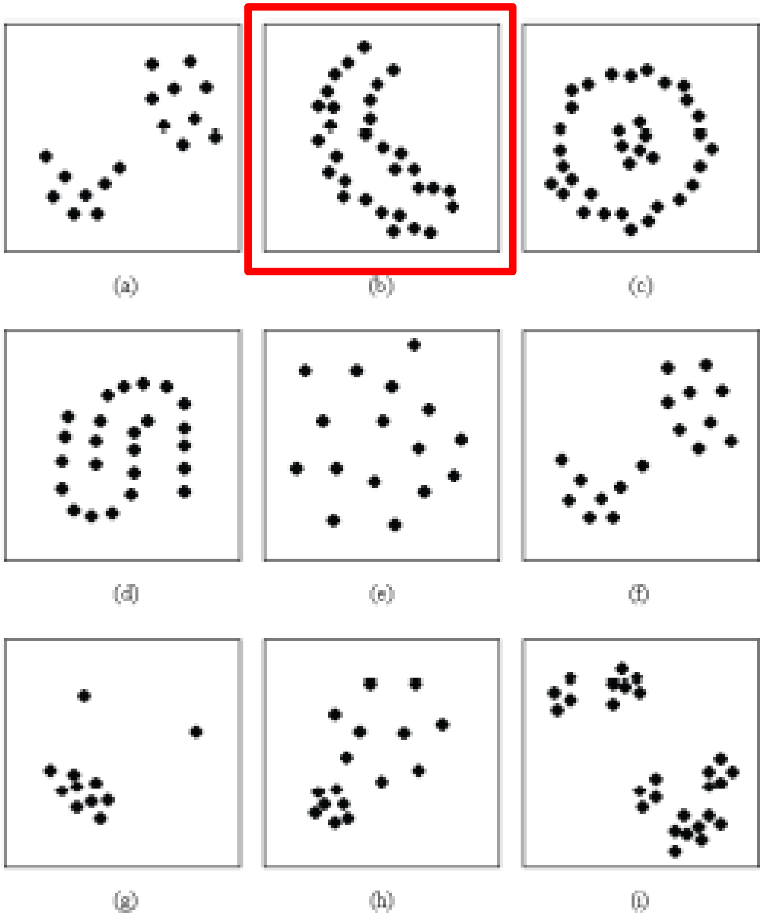


Exemplos

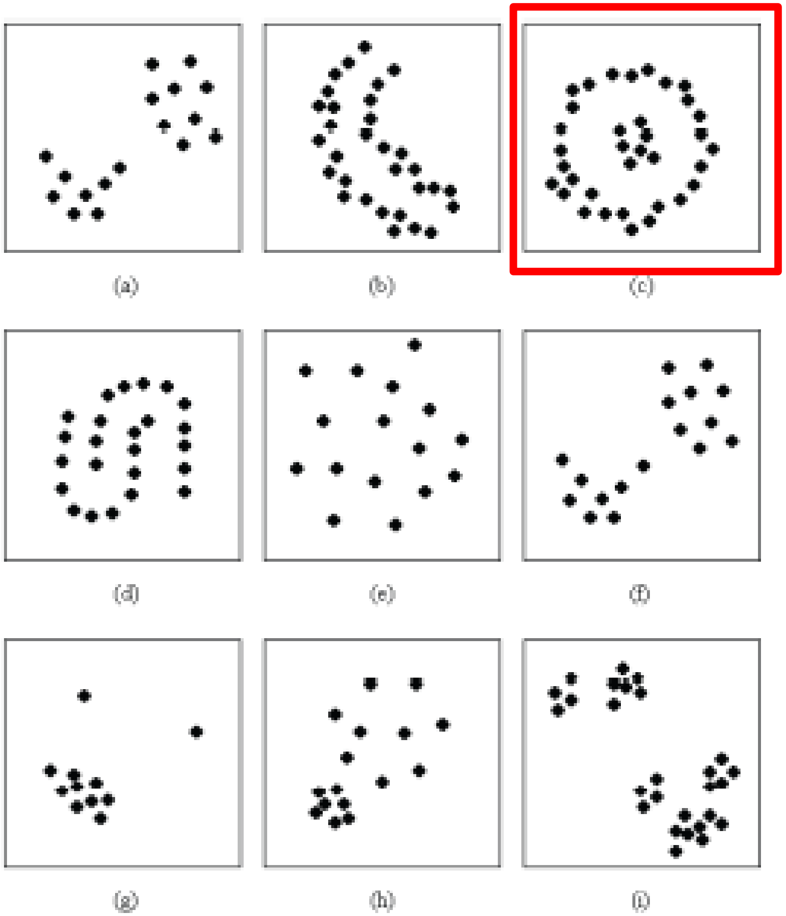


Exemplos

2 grupos
separação
não linear

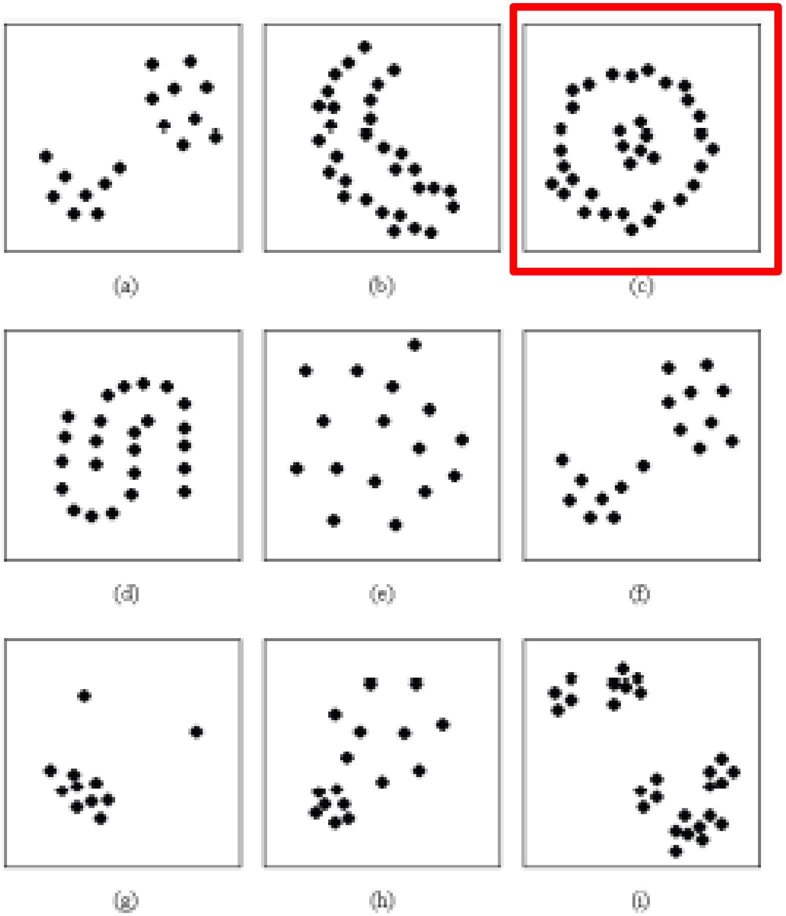


Exemplos

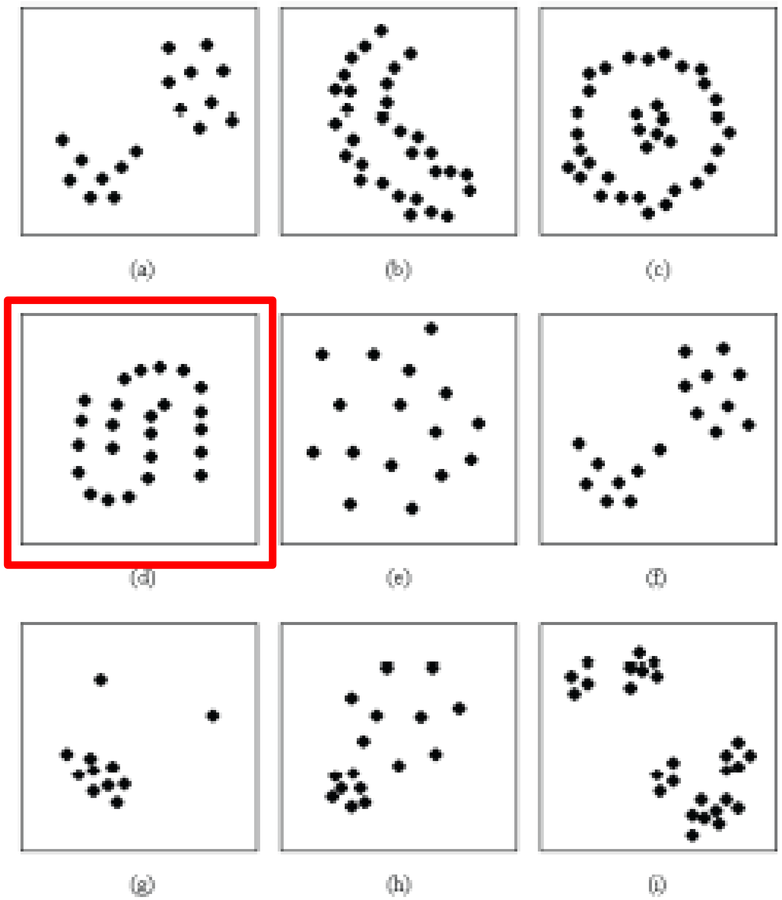


Exemplos

2 grupos
separação
não linear

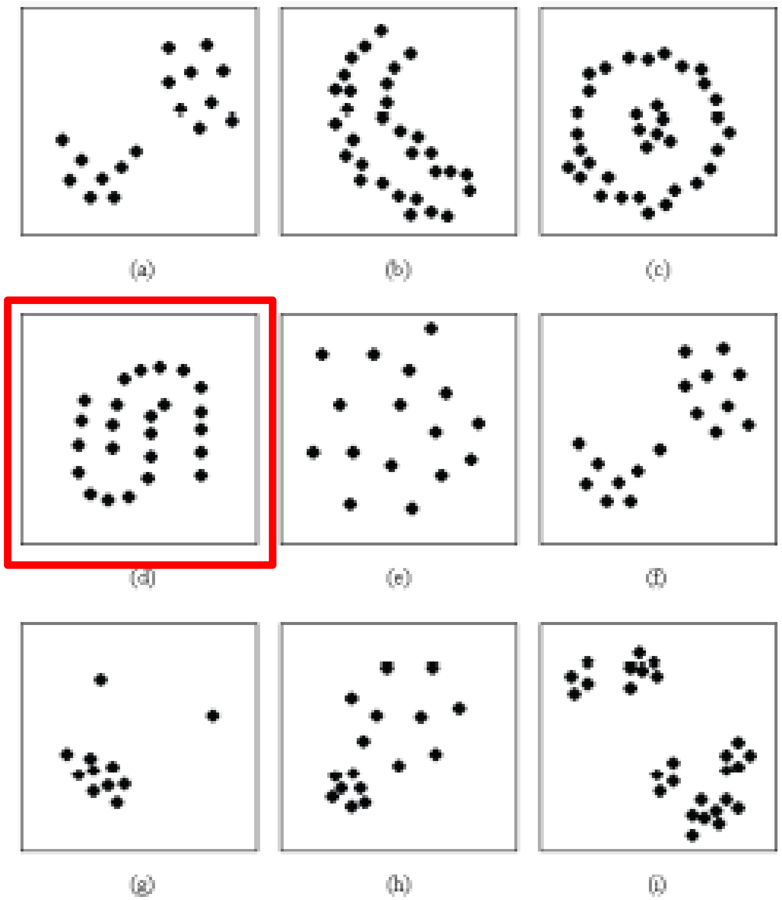


Exemplos

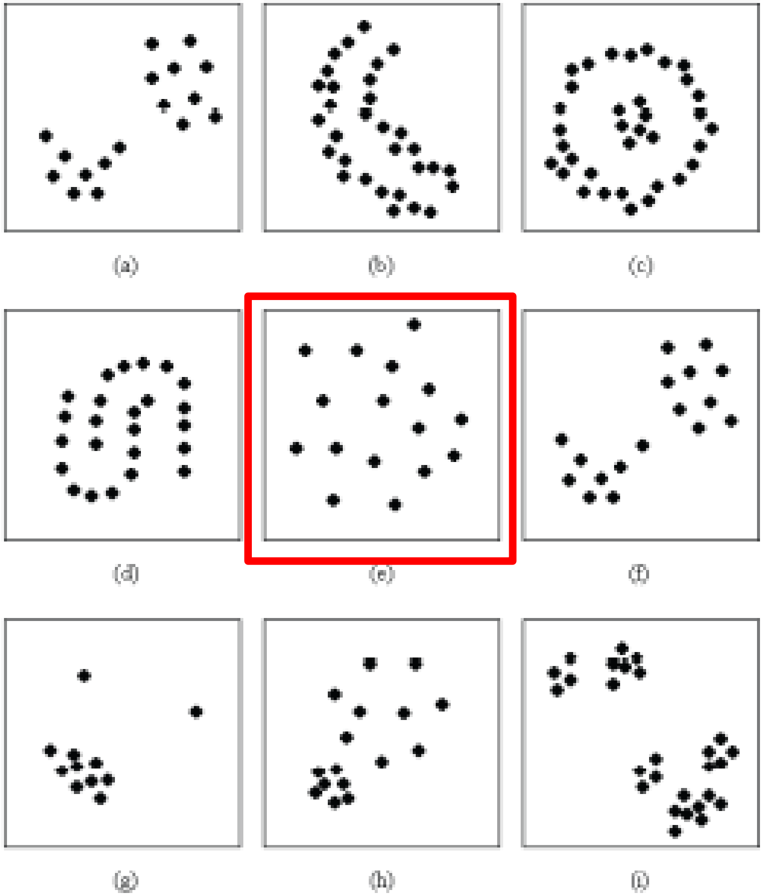


Exemplos

2 grupos
separação
não linear

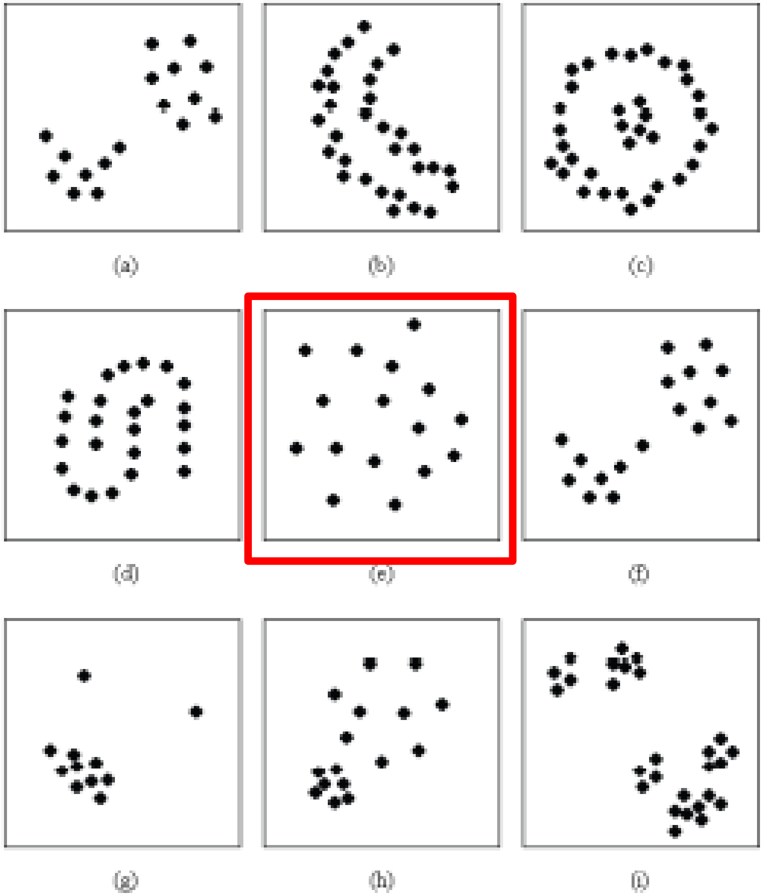


Exemplos



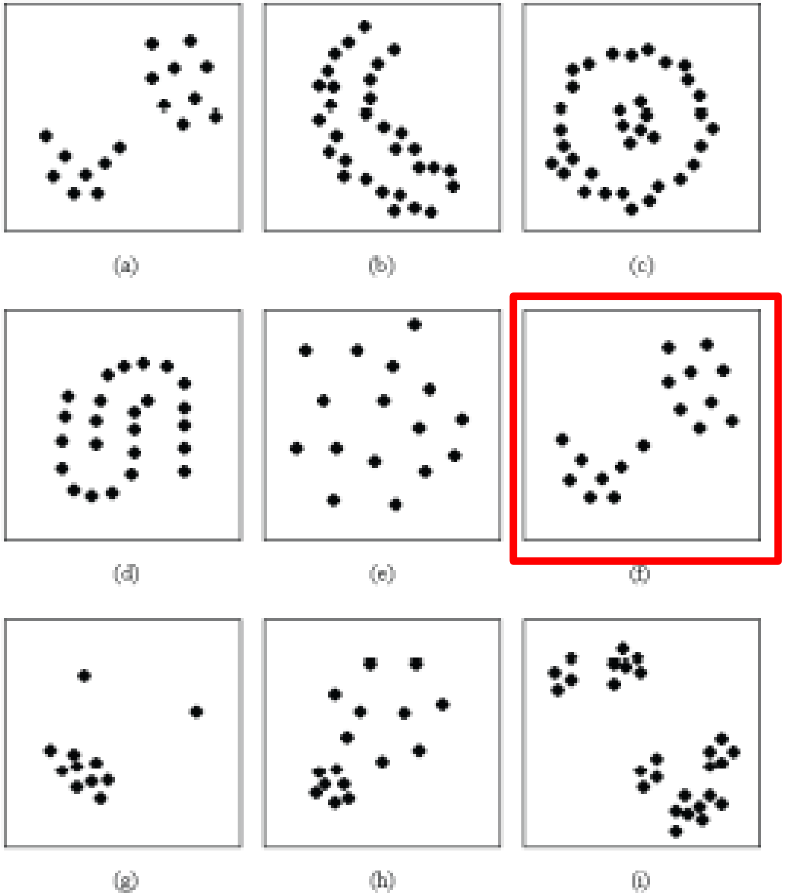
Exemplos

Quantos grupos?
Distribuição
uniforme



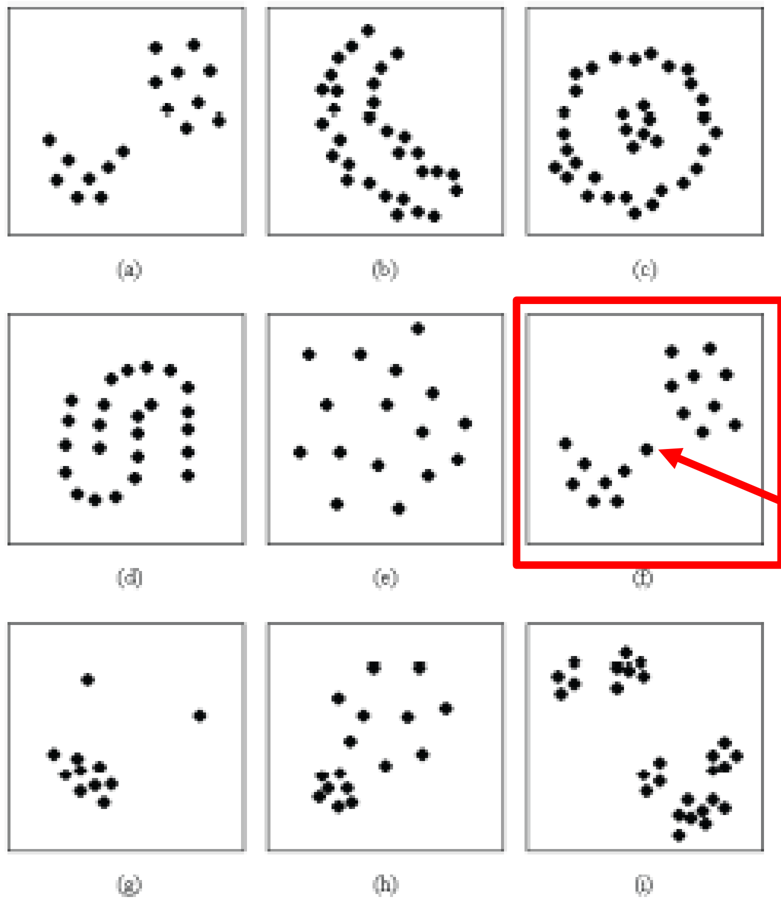
Exemplos

(variação da figura a)



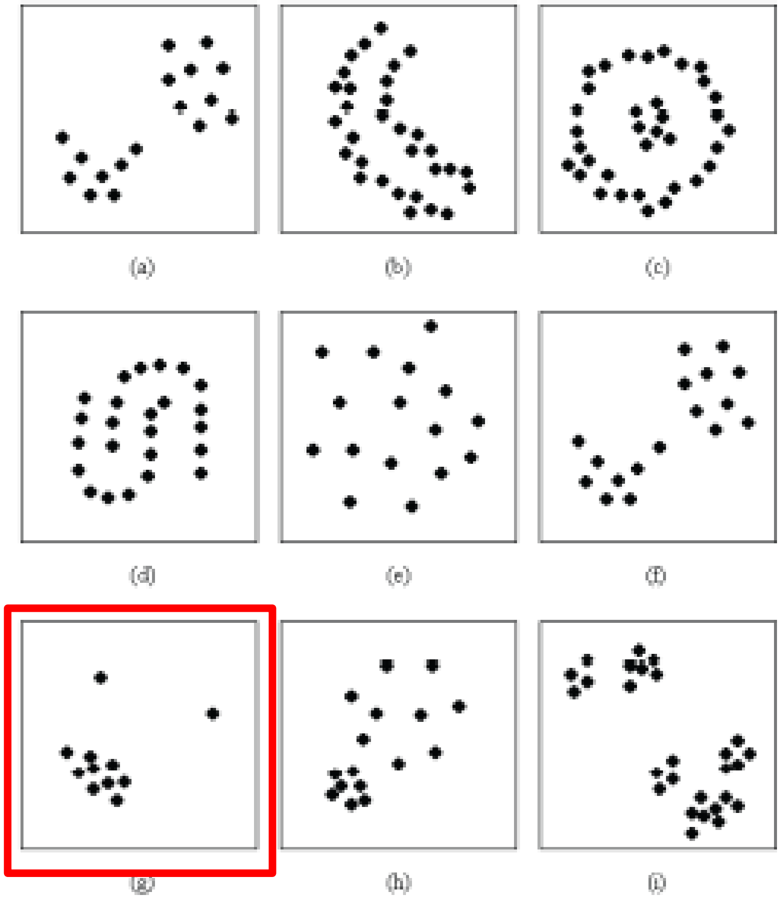
Exemplos

(variação da figura a) Dependendo do deslocamento de um ponto, podemos imaginar 1 ou 2 grupos

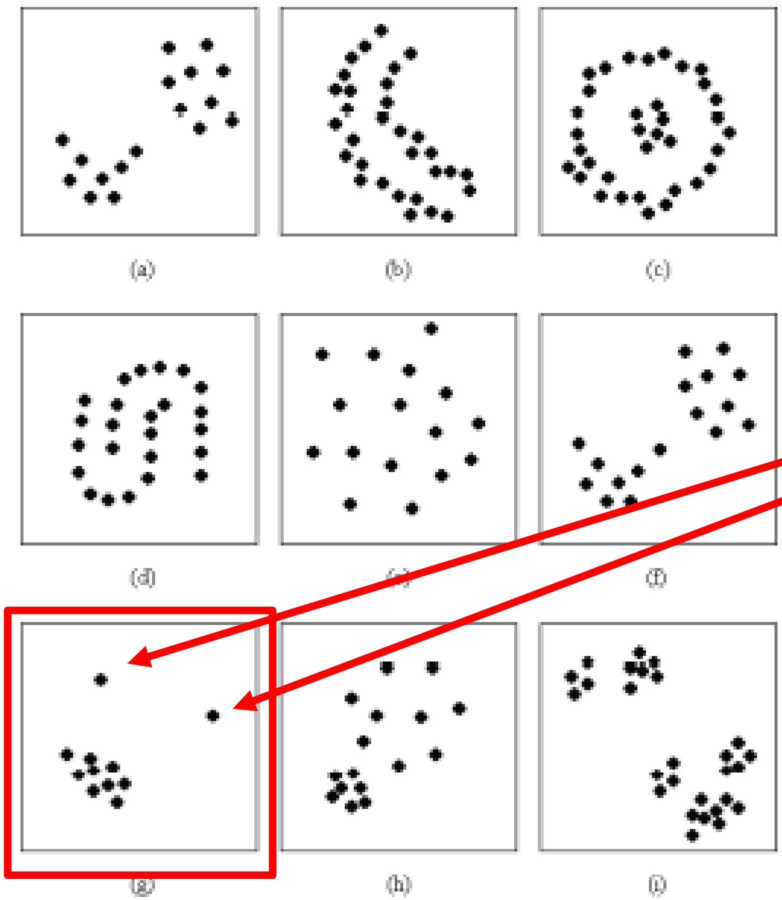


Ponto de ruído

Exemplos



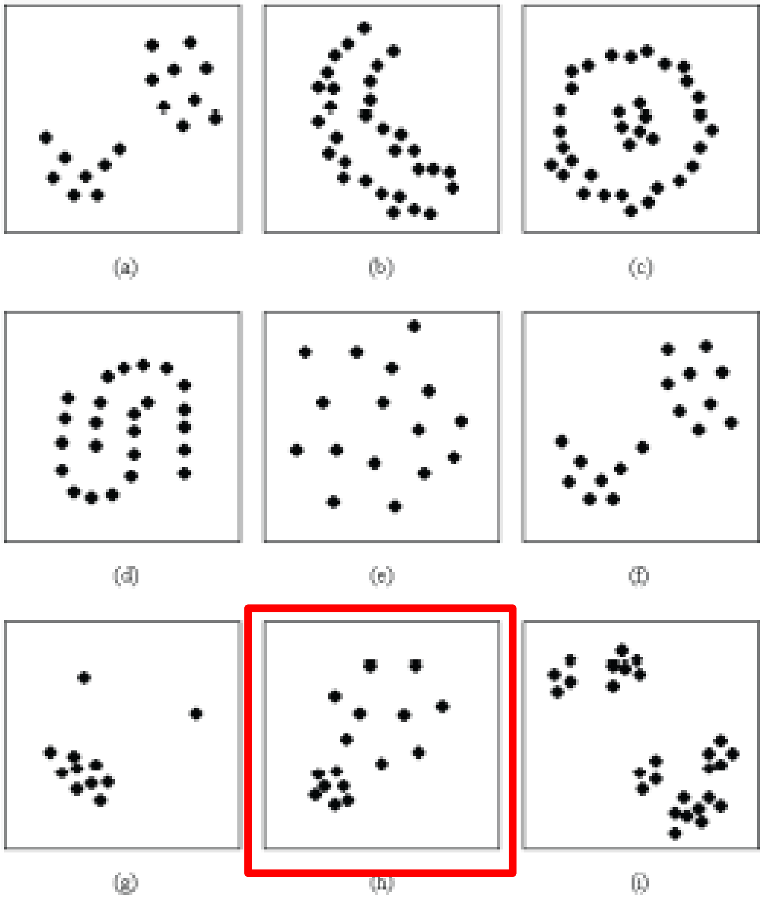
Exemplos



Outliers?

1, 2 ou 3 grupos?

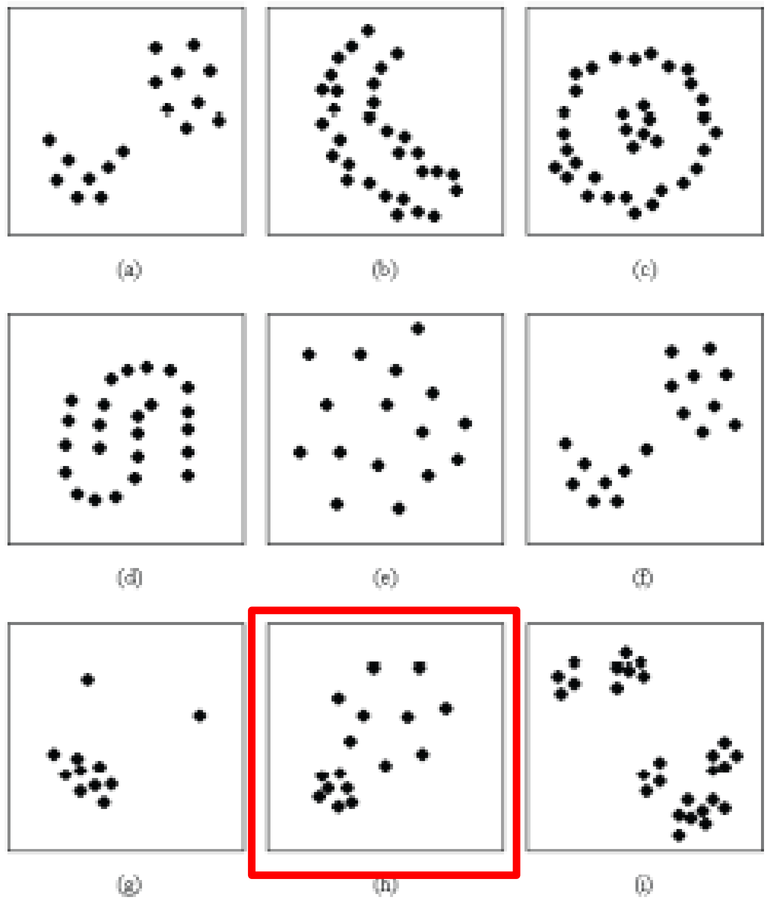
Exemplos



SISTEMAS DE INFORMAÇÃO

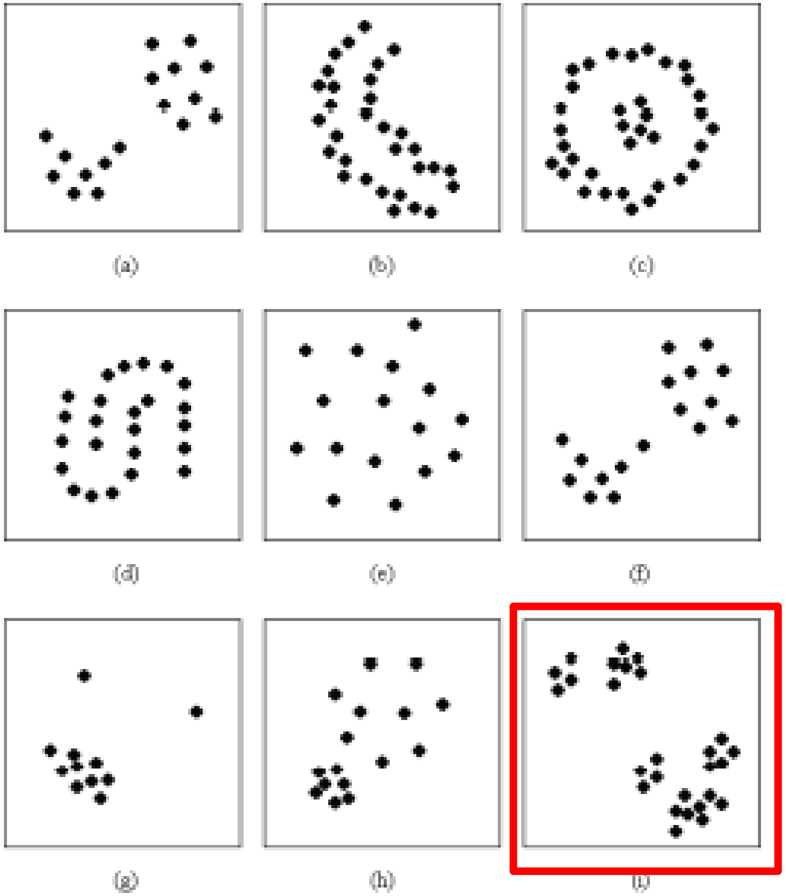
EACH

Exemplos

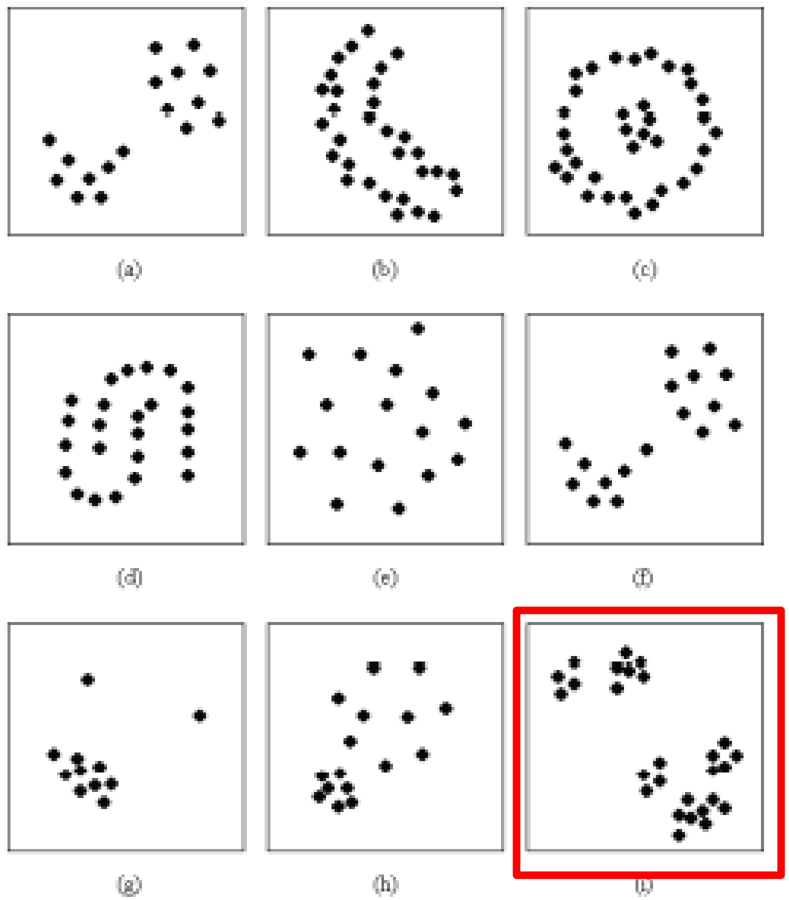


2 grupos...
… ou 1, se usar uma escala diferente (logaritmica)

Exemplos



Exemplos



2 ou 5 grupos?
clusters de
clusters

Esteja consciente de que...

- O agrupamento é feito usando características e critérios que podem não ser adequados para a classificação verdadeira
- Qualquer critério vai impor uma estrutura sobre os dados, que pode não ser a real
- POR ISSO: qualquer informação extra que você tenha é VALIOSA (por ex: nr de classes)

Qual critério utilizar?

Considerar vários critérios, e analisar os resultados.

Em geral, os critérios seguem a ideia:

Critério de Similaridade: agrupar elementos de tal forma que

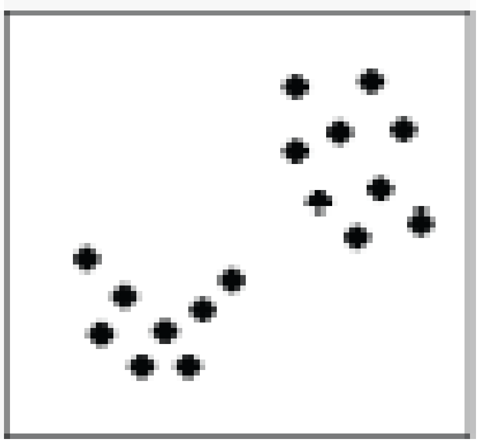
- elementos da mesma classe sejam o mais similares possível entre si
- elementos de classes distintas sejam o mais diferentes possível entre si

Cabe então definir o que é a similaridade utilizada

Critério de similaridade

Uma das formas de formas de formalizar o critério de similaridade é através da dispersão dos dados

- Cada grupo deve conter elementos com uma dispersão mais baixa (baixa dispersão intraclasse)
- Elementos de grupos distintos devem apresentar maior dispersão (alta dispersão interclasse)



[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

Matrizes de Dispersão (*scatter matrices*)

- Dada uma matriz dataset NxM (elementos nas linhas, características nas colunas)
- K classes
- Cada classe C_i com N_i elementos
- Cada elemento f_j é um vetor Mx1
- $\vec{M}$ é o vetor médio Mx1 (valor médio de cada caraterística)
- $\vec{\mu}_i$ é o vetor médio Mx1 da classe C_i

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo

K = 3 classes:

Object #	Class	Feature 1	Feature 2
1	C_3	9.2	33.2
2	C_2	5.3	21.4
3	C_3	8.8	31.9
4	C_1	2.9	12.7
5	C_3	9.0	32.4
6	C_1	1.5	12.0
7	C_1	1.2	11.5

Matriz de dados:

$$F = \begin{bmatrix} 9.2 & 33.2 \\ 5.3 & 21.4 \\ 8.8 & 31.9 \\ 2.9 & 12.7 \\ 9.0 & 32.4 \\ 1.5 & 12.0 \\ 1.2 & 11.5 \end{bmatrix}$$

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo

$$F = \begin{bmatrix} 9.2 & 33.2 \\ 5.3 & 21.4 \\ 8.8 & 31.9 \\ 2.9 & 12.7 \\ 9.0 & 32.4 \\ 1.5 & 12.0 \\ 1.2 & 11.5 \end{bmatrix}$$

$$\vec{M} = \begin{bmatrix} 5.4143 \\ 22.1571 \end{bmatrix}$$

$$\vec{f_1} = \begin{bmatrix} 9.2 \\ 33.2 \end{bmatrix}; \quad \vec{f_2} = \begin{bmatrix} 5.3 \\ 21.4 \end{bmatrix}; \quad \vec{f_3} = \begin{bmatrix} 8.8 \\ 31.9 \end{bmatrix}; \quad \vec{f_4} = \begin{bmatrix} 2.9 \\ 12.7 \end{bmatrix};$$

$$\vec{f_5} = \begin{bmatrix} 9.0 \\ 32.4 \end{bmatrix}; \quad \vec{f_6} = \begin{bmatrix} 1.5 \\ 12.0 \end{bmatrix}; \quad \vec{f_7} = \begin{bmatrix} 1.2 \\ 11.5 \end{bmatrix}$$

Exemplo

Matrizes de dados para as classes 1, 2 e 3:

$$F_1 = \begin{bmatrix} 2.9 & 12.7 \\ 1.5 & 12.0 \\ 1.2 & 11.5 \end{bmatrix}; \qquad F_2 = \begin{bmatrix} 5.3 & 21.4 \end{bmatrix}; \qquad F_3 = \begin{bmatrix} 9.2 & 33.2 \\ 8.8 & 31.9 \\ 9.0 & 32.4 \end{bmatrix}$$

$$\vec{\mu}_1 = \begin{bmatrix} 1.8667 \\ 12.0667 \end{bmatrix}; \qquad \vec{\mu}_2 = \begin{bmatrix} 5.3 \\ 21.4 \end{bmatrix}; \qquad \vec{\mu}_3 = \begin{bmatrix} 9.0 \\ 32.5 \end{bmatrix}$$

Matrizes de dispersão

- Matriz de dispersão total:

$$S = \sum_{i=1}^{N} \left(\vec{f_i} - \vec{\mathrm{M}}\right)\left(\vec{f_i} - \vec{\mathrm{M}}\right)^T$$

Matrizes de dispersão

- Matriz de dispersão total:

$$S = \sum_{i=1}^{N} \left(\vec{f_i} - \vec{\mathrm{M}}\right)\left(\vec{f_i} - \vec{\mathrm{M}}\right)^T$$

No nosso exemplo:

$$= \begin{bmatrix} 9.2 - 5.4143 \\ 33.20 - 22.1571 \end{bmatrix} \begin{bmatrix} 9.2 - 5.4143 & 33.20 - 22.1571 \end{bmatrix} + \cdots +$$

$$+ \begin{bmatrix} 1.2 - 5.4143 \\ 11.5 - 22.1571 \end{bmatrix} \begin{bmatrix} 1.2 - 5.4143 & 11.5 - 22.1571 \end{bmatrix}$$

$$= \begin{bmatrix} 78.0686 & 220.0543 \\ 220.0543 & 628.5371 \end{bmatrix}.$$

SISTEMAS DE INFORMAÇÃO

Matrizes de dispersão

- Matriz de dispersão total:

$$S = \sum_{i=1}^{N} \left(\vec{f_i} - \vec{\mathrm{M}}\right)\left(\vec{f_i} - \vec{\mathrm{M}}\right)^T$$

- Matriz de dispersão da classe C_i :

$$S_i = \sum_{i \in C_i} \left(\vec{f_i} - \vec{\mu}_i\right)\left(\vec{f_i} - \mu_i\right)^T$$

Matrizes de dispersão

- Matriz de dispersão total:

$$S = \sum_{i=1}^{N} \left(\vec{f_i} - \vec{\mathrm{M}}\right)\left(\vec{f_i} - \vec{\mathrm{M}}\right)^T$$

- Matriz de dispersão da classe C_i :

$$S_i = \sum_{i \in C_i} \left(\vec{f_i} - \vec{\mu}_i\right)\left(\vec{f_i} - \mu_i\right)^T$$

- Matriz de dispersão intraclasse:

$$S_{\mathrm{intra}} = \sum_{i=1}^{K} S_i$$

Matrizes de dispersão

- Matriz de dispersão total:

$$S = \sum_{i=1}^{N} \left(\vec{f_i} - \vec{\mathrm{M}}\right)\left(\vec{f_i} - \vec{\mathrm{M}}\right)^T$$

- Matriz de dispersão da classe C_i :

$$S_i = \sum_{i \in C_i} \left(\vec{f_i} - \vec{\mu}_i\right)\left(\vec{f_i} - \mu_i\right)^T$$

- Matriz de dispersão intraclasse:

$$S_{\mathrm{intra}} = \sum_{i=1}^{K} S_i$$

- Matriz de dispersão interclasse:

$$S_{\mathrm{inter}} = \sum_{i=1}^{K} N_i \left(\vec{\mu}_i - \vec{\mathrm{M}}\right)\left(\vec{\mu}_i - \vec{\mathrm{M}}\right)^T$$

Matrizes de dispersão

- $S = S_{intra} + S_{inter}$

Matrizes de dispersão

- $S = S_{intra} + S_{inter}$

As matrizes são simétricas e a posição ii mostra o quadrado da dispersão de f_i

- Trace(S): soma da diagonal principal de S

trace (S) = $trace(S_{intra}) + trace(S_{inter})$

Consequência:

SISTEMAS DE INFORMAÇÃO

EACH

Matrizes de dispersão

- $S = S_{intra} + S_{inter}$

As matrizes são simétricas e a posição ii mostra o quadrado da dispersão de f_i

- Trace(S): soma da diagonal principal de S

trace (S) = trace(S_{intra}) + trace(S_{inter})

Consequência: pode-se atentar a apenas diminuir S_{intra} ou aumentar S_{inter}

Técnicas de agrupamentos

- Particional (ou não hierárquico)
  - Grupos obtidos a partir de uma partição do espaço de características com respeito a um número fixo de grupos
- Hierárquico
  - Agrupamento progressivo de elementos (formação de subclasses)

Agrupamento Particional

- Um algoritmo possível baseado nas dispersões (para um dado número fixo de classes):

Ideia: um bom agrupamento deveria exibir

- Baixa dispersão intraclasse
- Alta dispersão interclasse

Agrupamento Particional

- Um algoritmo possível baseado nas dispersões (para um dado número fixo de classes):

Associe aleatoriamente uma classe a cada objeto

Enquanto não satisfizer o critério de parada

    Selecione aleatoriamente um objeto

    Mude a classe desse objeto (aleatoriamente, mas sem deixar classes vazias)

    Se $trace(S_{intra}^{nova}) > trace(S_{intra}^{anterior})$

        Volte o objeto à sua classe anterior

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional Simples

- Critério de parada:
  - quando os grupos estabilizarem, por exemplo, quando o número de interações sem alteração de classificação for acima de um limiar

Agrupamento Particional Simples

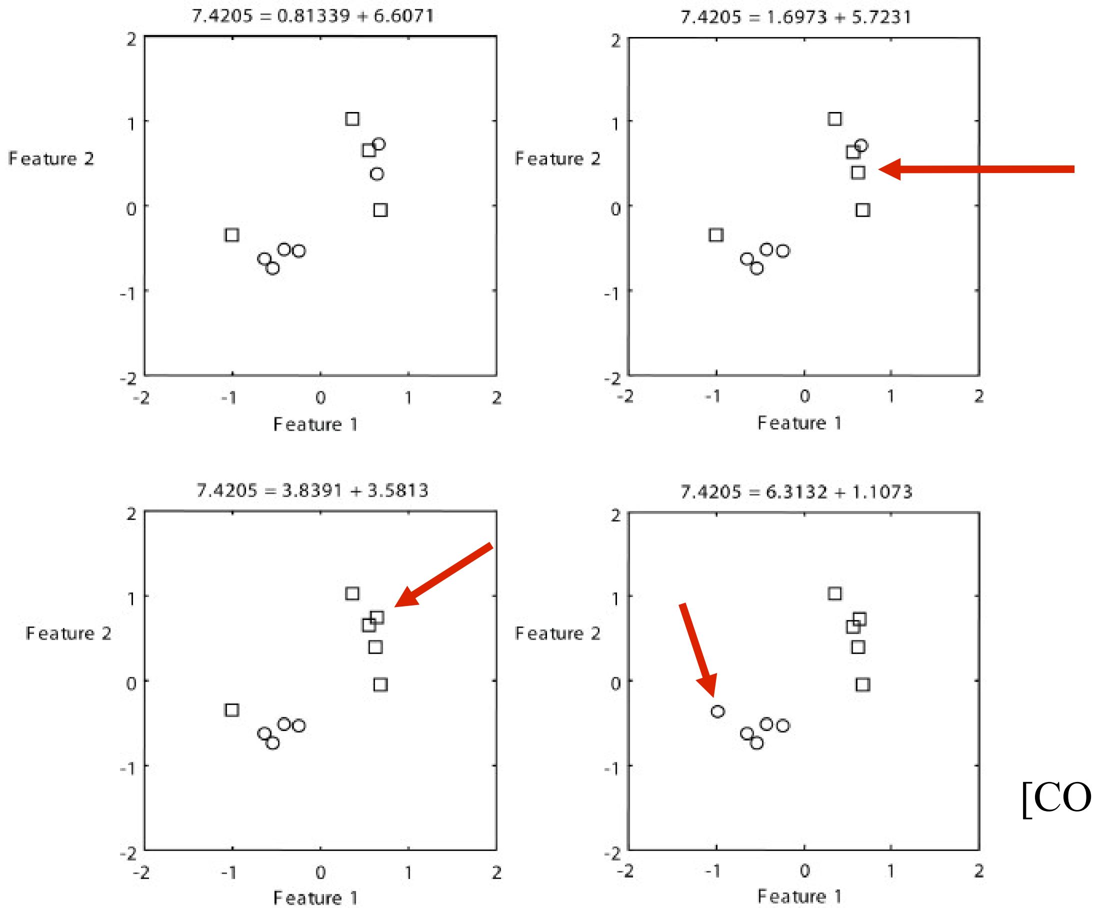
- Critério de parada:
  - quando os grupos estabilizarem, por exemplo, quando o número de interações sem alteração de classificação for acima de um limiar

- Note a importância do número de grupos ser fixo! Caso contrário haveria uma tendência a aumentar o número de grupos (grupos menores tendem a ter menores dispersões intraclasse

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional Simples – Ex: (apenas sitações intermediárias de decréscimo de S_{intra})



[COSTA& CESAR, 2009]

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional Simples

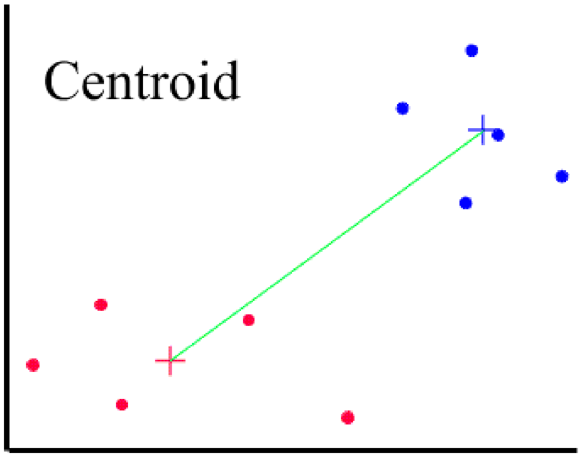
- Rápida convergência mas...
- Sofre do problema de mínimos locais

Agrupamento Particional K-Médias

- Número de classes pré-definido (K)
- Distância entre pontos (objetos)
  - Matriz de Distâncias ao invés de Matriz de Dispersão
- K pontos iniciais para representar cada classe (sementes)
  - Não necessariamente pontos de objetos
  - Oportunidade para conhecer algum conhecimento a priori (senão, seleção aleatória)

Agrupamento Particional K-Médias

- Centróide: centro de massa



- Cada centróide define uma área de influência
  - Pontos mais próximos dele do que de qualquer outro centróide

Agrupamento Particional
K-Médias

Escolha os K pontos protótipos iniciais (sementes) e guarde-os na lista W

Enquanto não satisfizer o critério de parada

- Calcule todas as distâncias entre cada objeto e os pontos protótipos P_i (matriz D KxN)
- Use a matriz D para identificar os objetos mais próximos de cada protótipo P_i Guarde-los na lista L_i
- Obtenha como novos protótipos os centróides dos objetos de L_i

Agrupamento Particional K-Médias

- Critério de parada:
  - Ex: Deslocamento de cada centróide é menor que um limiar

Agrupamento Particional
K-Médias

- Critério de parada:
  - Ex: Deslocamento de cada centróide é menor que um limiar

- No final, os objetos mais próximos de um centróide pertencem à sua classe

Agrupamento Particional
K-Médias

- Critério de parada:
  - Ex: Deslocamento de cada centróide é menor que um limiar
- No final, os objetos mais próximos de um centróide pertencem à sua classe
- Note que um grupo pode ficar vazio!
  - Possíveis soluções:
    - Reconhecer que há uma classe a menos
    - Escolher outra semente para aquela classe e re-rodar o algoritmo (a partir do loop enquanto)

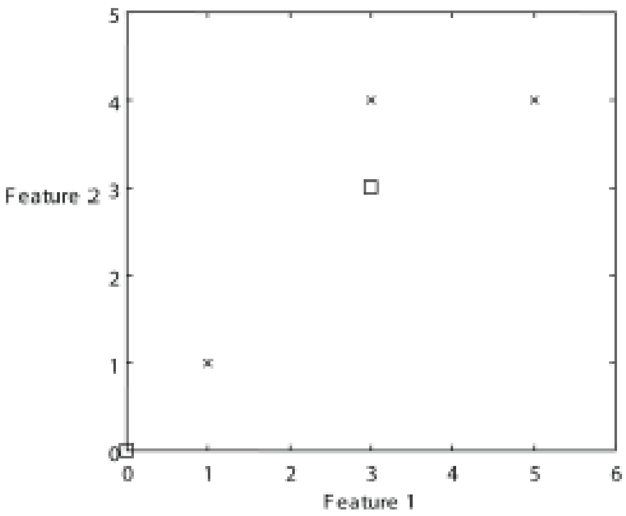
Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos iniciais:

$P_1 = (0,0)$ e $P_2 = (3,3)$

Critério de Parada: maior deslocamento de centróide m < 0,25



SISTEMAS DE INFORMAÇÃO

EACH

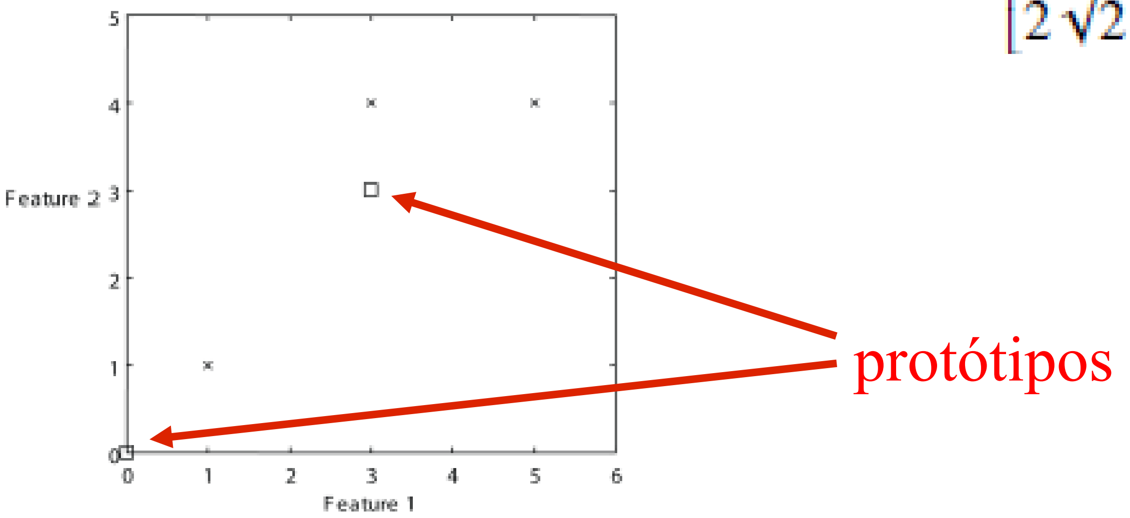
Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos iniciais:

$P_1 = (0,0)$ e $P_2 = (3,3)$

$$D = \begin{bmatrix} \sqrt{2} & 5 & \sqrt{41} \\ 2\sqrt{2} & 1 & \sqrt{5} \end{bmatrix}$$



SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional
K-Médias - Exemplo

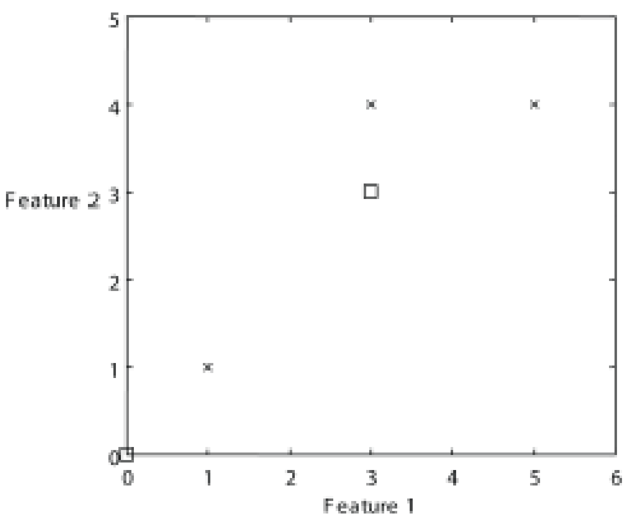
Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos iniciais:

$P_1 = (0,0)$ e $P_2 = (3,3)$

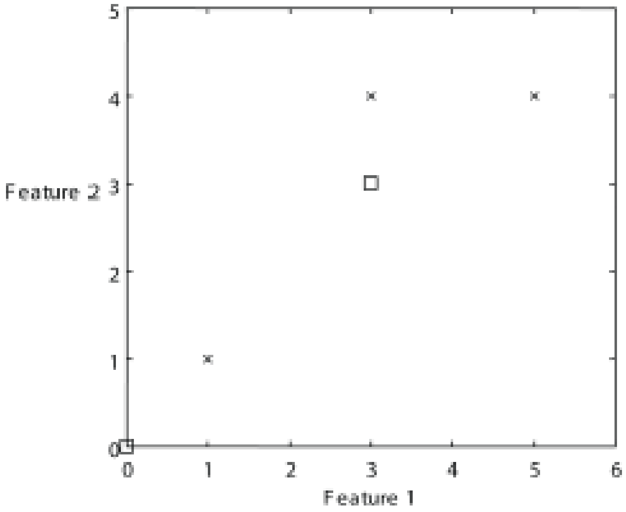
$$D = \begin{bmatrix} \sqrt{2} & 5 & \sqrt{41} \\ 2\sqrt{2} & 1 & \sqrt{5} \end{bmatrix}$$

$$L_1 = (X_1) \quad \text{and} \quad L_2 = (X_2, X_3)$$



Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4



Protótipos iniciais:

$P_1 = (0,0)$ e $P_2 = (3,3)$

$$D = \begin{bmatrix} \sqrt{2} & 5 & \sqrt{41} \\ 2\sqrt{2} & 1 & \sqrt{5} \end{bmatrix}$$

$$L_1 = (X_1) \quad \text{and} \quad L_2 = (X_2, X_3)$$

Protótipos novos:

P'_1 = média $\{X_1\}$ = (1,1) e

P'_2 = média $\{X_2,X_3\}$ = (4,4)

Agrupamento Particional
K-Médias - Exemplo

Protótipos iniciais:

$P_1 = (0,0)$ e $P_2 = (3,3)$

Protótipos novos:

$P'_1 = \text{média } \{X_1\} = (1,1)$ e

$P'_2 = \text{média } \{X_2, X_3\} = (4,4)$

$m = \max\{\|P'_1 - P_1\|, \|P'_2 - P_2\|\}$

$= \max\{\text{raiz}(2), \text{raiz}(2)\} > 0.25$

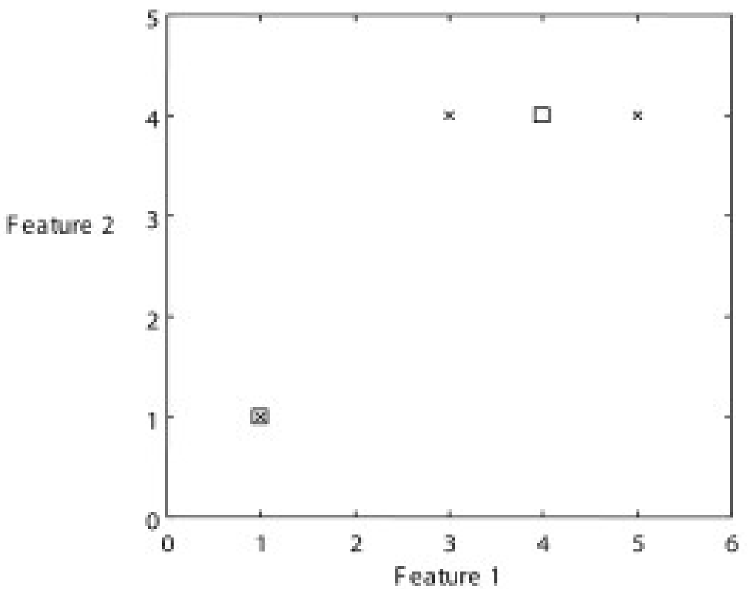
CONTINUA!

Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos:

$P_1 = (1,1)$ e $P_2 = (4,4)$



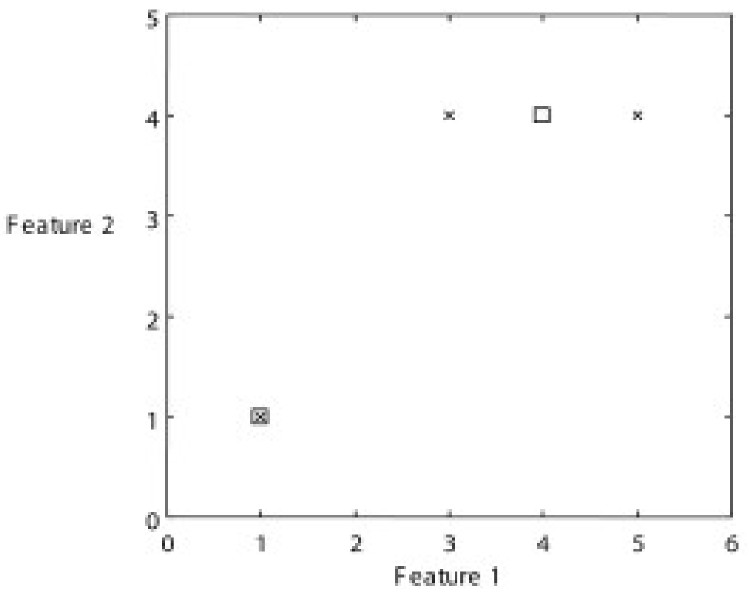
Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos:

$P_1 = (1,1)$ e $P_2 = (4,4)$

$$D = \begin{bmatrix} 0 & \sqrt{13} & 5 \\ \sqrt{18} & 1 & 1 \end{bmatrix}$$



Agrupamento Particional
K-Médias - Exemplo

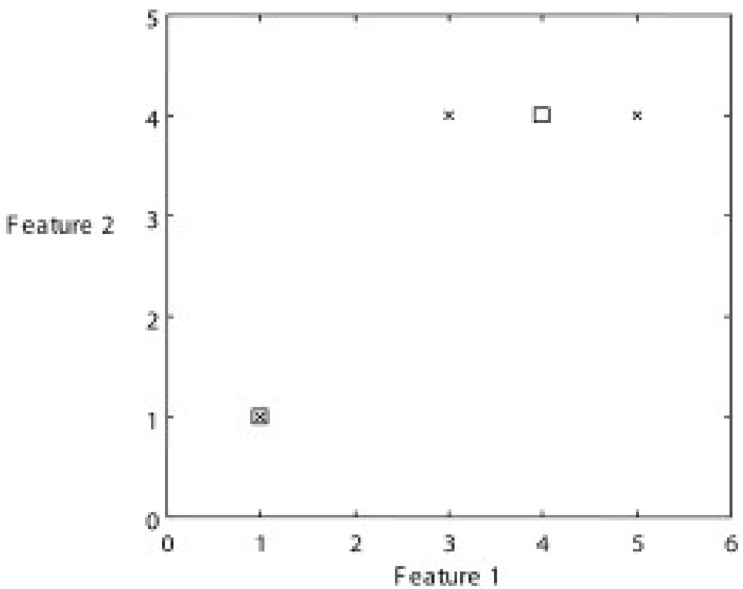
Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4

Protótipos:

$P_1 = (1,1)$ e $P_2 = (4,4)$

$$D = \begin{bmatrix} 0 & \sqrt{13} & 5 \\ \sqrt{18} & 1 & 1 \end{bmatrix}$$

$$L_1 = (X_1) \quad \text{and} \quad L_2 = (X_2, X_3)$$

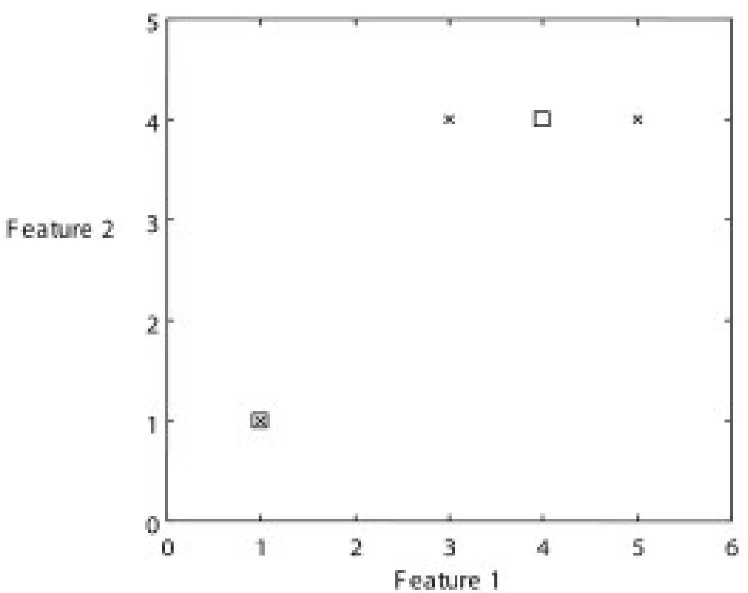


SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional
K-Médias - Exemplo

Object	Feature 1	Feature 2
X_1	1	1
X_2	3	4
X_3	5	4



Protótipos:

$P_1 = (1,1)$ e $P_2 = (4,4)$

$$D = \begin{bmatrix} 0 & \sqrt{13} & 5 \\ \sqrt{18} & 1 & 1 \end{bmatrix}$$

$$L_1 = (X_1) \quad \text{and} \quad L_2 = (X_2, X_3)$$

Protótipos novos:

$P'_1 = \text{média}\ \{X_1\} = (1,1)$ e

$P'_2 = \text{média}\ \{X_2,X_3\} = (4,4)$

Agrupamento Particional
K-Médias - Exemplo

Protótipos anteriores:

$P_1 = (1,1)$ e $P_2 = (4,4)$

Protótipos novos:

$P'_1 = \text{média } \{X_1\} = (1,1)$ e

$P'_2 = \text{média } \{X_2, X_3\} = (4,4)$

$$m = \max \{\|P'_1 - P_1\|, \| P'_2 - P_2 \|\}$$

$$= \max\{0, 0\} < 0.25$$

TERMINA!

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Particional
K-Médias

- A convergência para a menor dispersão não é garantida

- Alternativas:
  - várias rodadas (com diferentes sementes) e escolher a configuração com menor matriz de dispersão intraclasse
  - Juntar grupos de centróides próximos e partir em 2 grupos com alta dispersão

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Hierárquico

- Agrupamentos progressivos de N objetos em classes, de acordo com algum critério (similaridade ou distância)
- Objetos mais próximos são agrupados em subgrupos antes de objetos mais distantes
- No final
  - todos os objetos pertencem a um único e grande grupo
  - Você define qual partição usar (número variável de subgrupos - classes)

Agrupamento Hierárquico

- Abordagem aglomerativa: parte de elementos individuais e vai agrupando subgrupos

- Abordagem divisiva: parte do grande grupo (contendo todos os elementos) e vai dividindo em subgrupos

Agrupamento Hierárquico

- Abordagem aglomerativa: parte de elementos individuais e vai agrupando subgrupos
  - Vamos ver aqui

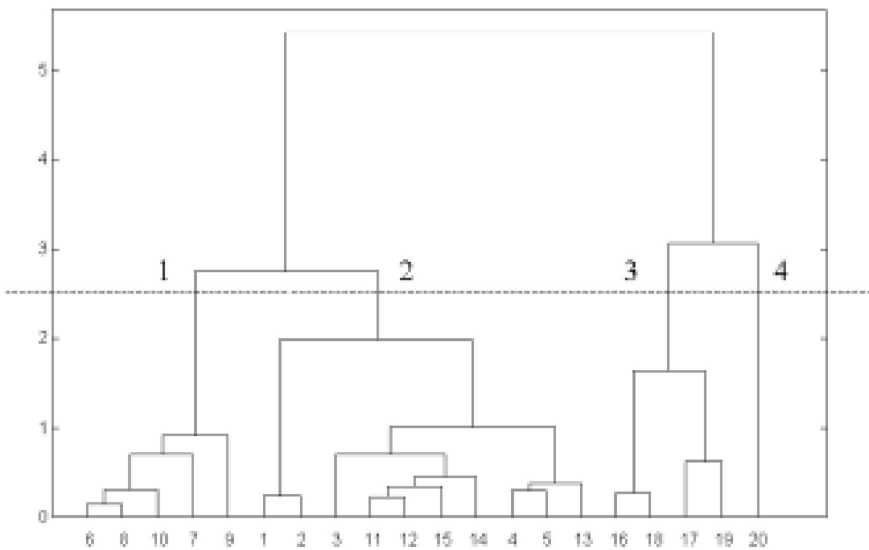
- Abordagem divisiva: parte do grande grupo (contendo todos os elementos) e vai dividindo em subgrupos

Agrupamento Hierárquico

- Inicialmente cada objeto é um grupo
- Em cada iteração:
  - Matriz de distâncias entre os grupos correntes
  - Une os dois grupos mais próximos

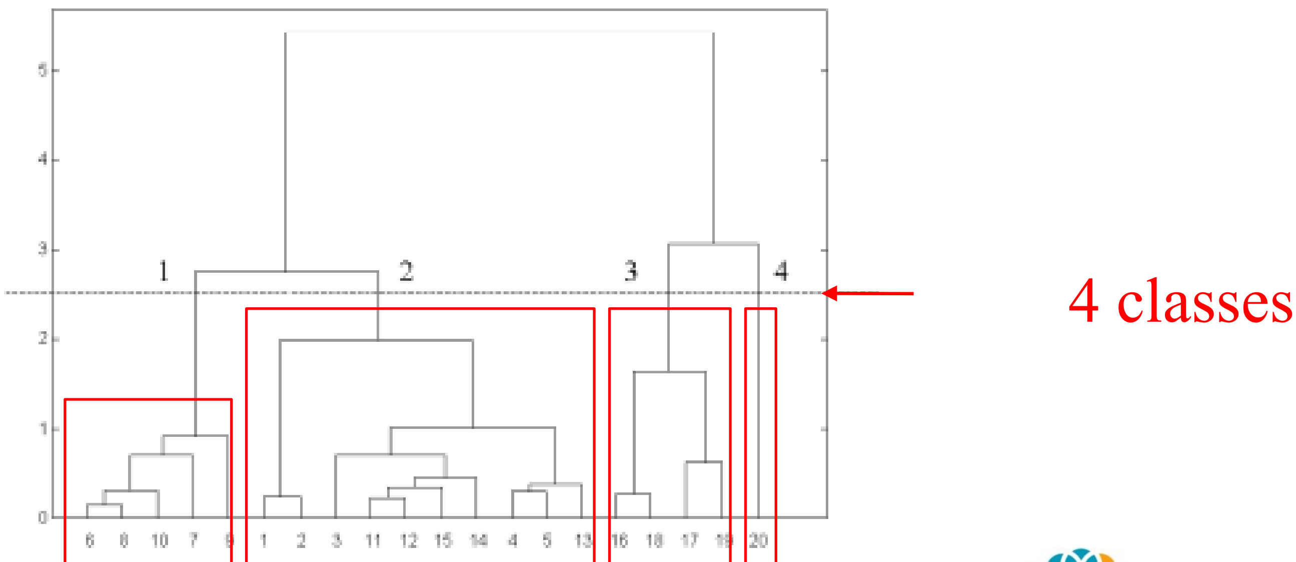
Agrupamento Hierárquico

- Inicialmente cada objeto é um grupo
- Em cada iteração:
  - Matriz de distâncias entre os grupos correntes
  - Une os dois grupos mais próximos
- Dendograma: vários possíveis números de classes



Agrupamento Hierárquico

- Inicialmente cada objeto é um grupo
- Em cada iteração:
  - Matriz de distâncias entre os grupos correntes
  - Une os dois grupos mais próximos
- Dendograma: vários possíveis números de classes

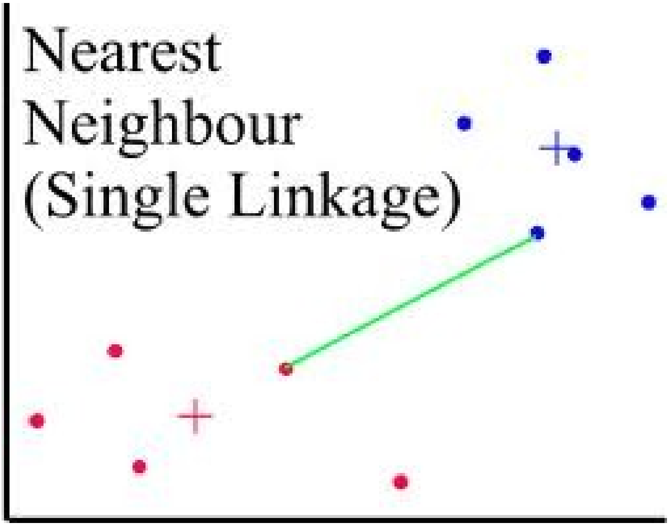


Agrupamento Hierárquico

- Como definir distâncias entre grupos?
- Diferentes formas de se medir essa distância implicam em diferentes tipos de agrupamentos hierárquicos
  - Single linkage
  - Complete linkage
  - Group average
  - Centroid

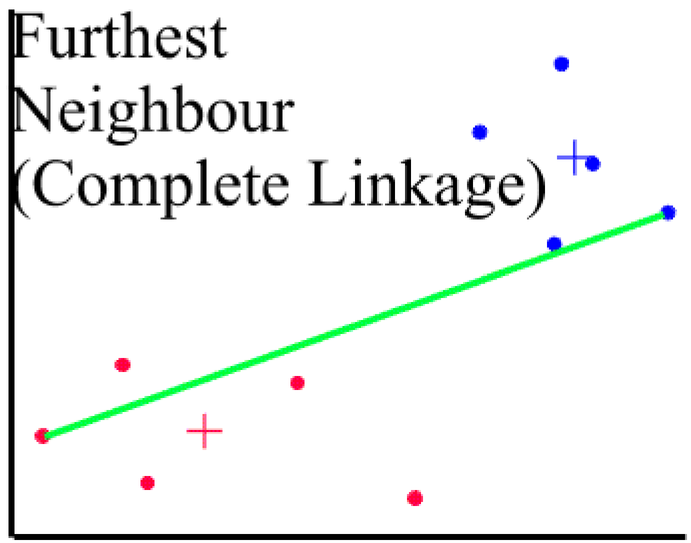
Agrupamento Hierárquico
Single Linkage

- $dist(A,B) = \min dist(a,b), a \in A, b \in B$
- Distância mínima entre qualquer ponto da classe A e qualquer ponto da classe B



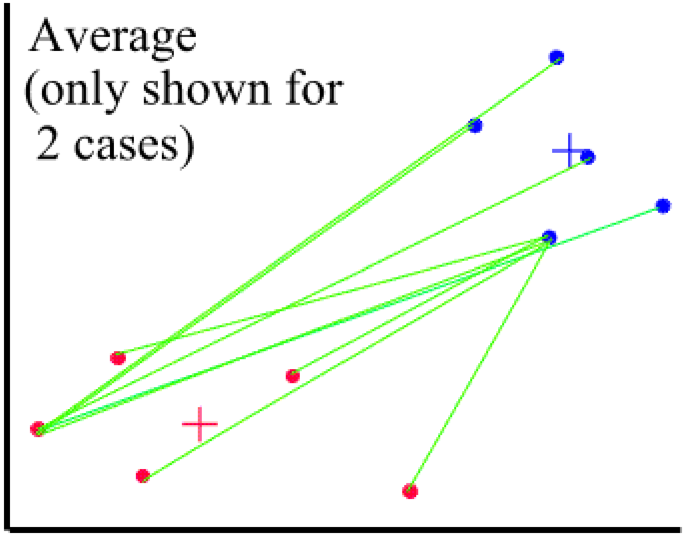
Agrupamento Hierárquico Complete Linkage

- dist(A,B) = max dist(a,b), a ∈ A, b ∈ B
- Distância máxima entre qualquer ponto da classe A e qualquer ponto da classe B



Agrupamento Hierárquico Group Average

- $\text{dist}(A,B) = 1/(N_A N_B) \sum \text{dist}(a,b),\ a \in A,\ b \in B$
- Distância média de todos os pares de pontos das classes A e B

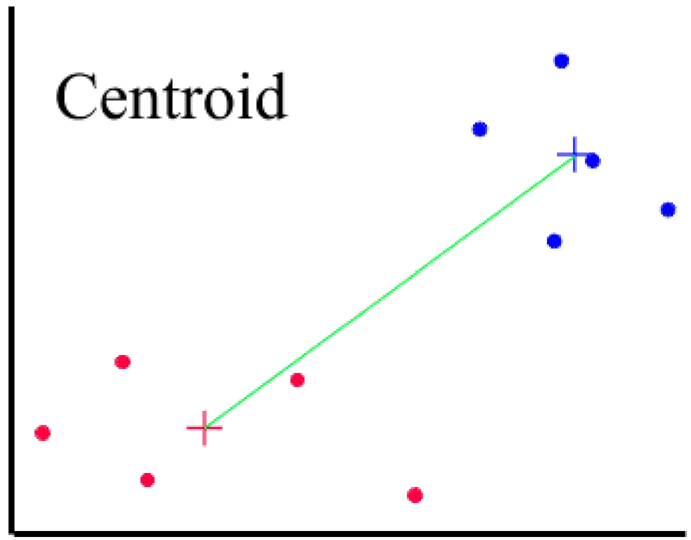


Agrupamento Hierárquico Centróide

dist(A,B) = dist (C^A, C^B)

- Onde C^A_i é a média dos componentes i dos pontos do grupo A

Distância entre os centros de massa (centróides) da classe A (C^A) e da classe B (C^B)



Agrupamento Hierárquico

Definido pelo tipo de distância utilizada e a métrica de distância (ex: Euclidiana, chessboard, etc.)

- Distância euclidiana (**a**,**b**) = raiz($\Sigma_{i=1..n}(a_i-b_i)^2$)
- Distância chessboard (**a**,**b**) = $\max_i (|a_i-b_i|)$

...

SISTEMAS DE INFORMAÇÃO

EACH

Agrupamento Hierárquico
Algoritmo

Construa uma matriz de distâncias D NxN

para n de 1 até N-1

(a) Determine a distância mínima em D, e os grupos C_j e C_k (j<k) que determinam essa distância

(b) $C_{N+n} \leftarrow C_j \cup C_k$

(c) Atualize a matriz, agora sem a linha e coluna k, e considerando que a linha e coluna j correspondem agora ao grupo C_{N+n}

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo – Single Linkage

N = 5 objetos

$$F = \begin{bmatrix} 1.2 & 2.0 \\ 3.0 & 3.7 \\ 1.5 & 2.7 \\ 2.3 & 2.0 \\ 3.1 & 3.3 \end{bmatrix}$$

$$D^{(1)} = \begin{array}{c} \begin{bmatrix} 0 & & & & \\ 2.4759 & 0 & & & \\ 0.7616 & 1.8028 & 0 & & \\ 1.1000 & 1.8385 & 1.0630 & 0 & \\ 2.3022 & 0.4123 & 1.7088 & 1.5264 & 0 \end{bmatrix} \end{array} \begin{array}{c} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{array}$$

n = 1, e dmin = 0.4123 (j =2 e k = 5)

$C_6 \leftarrow C_2 \cup C_5$

Exemplo – Single Linkage (cont.)

$$D^{(2)} = \begin{array}{c} \left[\begin{array}{cccc} 0 & & & \\ 2.3022 & 0 & & \\ 0.7616 & 1.7088 & 0 & \\ 1.1000 & 1.5264 & 1.0630 & 0 \end{array}\right] \end{array} \begin{array}{c} C_1 \\ C_2C_5 = C_6 \\ C_3 \\ C_4 \end{array}$$

n = 2, e dmin = 0.7616 (j =1 e k = 3)

$C_7 \leftarrow C_1 \cup C_3$

Exemplo – Single Linkage (cont.)

$$D^{(2)} = \begin{array}{c} \begin{bmatrix} 0 & & \\ 1.7088 & 0 & \\ 1.0630 & 1.5264 & 0 \end{bmatrix} \end{array} \begin{array}{c} C_1C_3 = C_7 \\ C_2C_5 = C_6 \\ C_4 \end{array}$$

n = 3, e dmin = 1.0630 (j =1 e k = 3)

$C_8 \leftarrow C_7 \cup C_4$

SISTEMAS DE INFORMAÇÃO

EACH

Exemplo – Single Linkage (cont.)

$$D^{(2)} = \begin{bmatrix} 0 & \\ 1.5264 & 0 \end{bmatrix} \begin{array}{l} C_1C_3C_4 = C_8 \\ C_2C_5 = C_6 \end{array}$$

n = 4, e dmin = 1.5264 (j =1 e k = 2)

$C_9 \leftarrow C_8 \cup C_6$

Exemplo – Single Linkage (cont.)

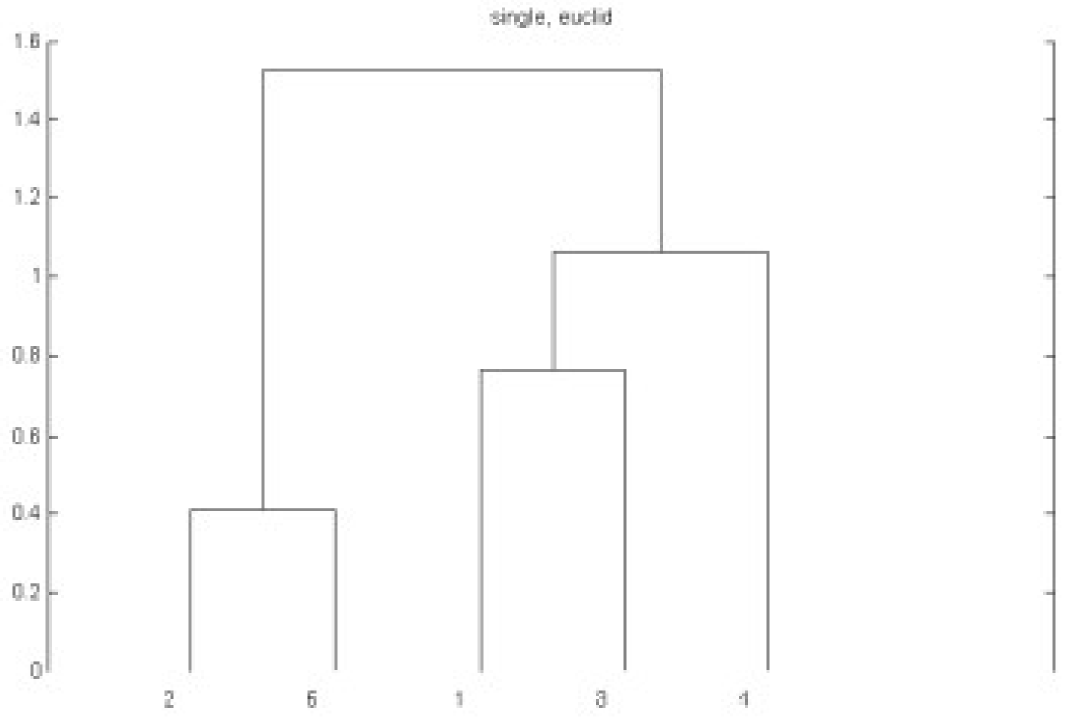


Figure 8.34: *The obtained dendrogram.*

SISTEMAS DE INFORMAÇÃO

EACH

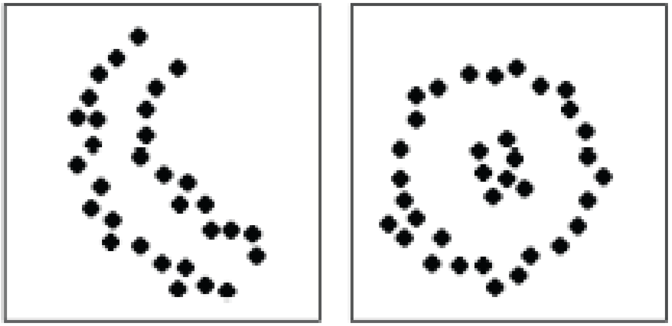
Agrupamento Hierárquico Baseado em Dispersão

- Ao invés de unir os dois grupos mais próximos une os dois grupos que menos aumentam a dispersão intraclasse

- Ward´s linkage

Comparação entre métodos hierárquicos

- Single Linkage:
  - Tendência de *chaining*, o que implica na união de grupo bem separados mas conectados por alguns poucos pontos
  - Não adequado para dados gaussianos (BAYNE, 1980)
  - Menos afetado por *outliers*
  - Um dos poucos que funcionam bem para dados não elipsoides
  - Desempenho pobre



Comparação entre métodos hierárquicos

- Complete linkage:
  - Procura por grupos compactos, elipsoides
  - Pobre desempenho quando há alta densidade de grupos

Comparação entre métodos hierárquicos

- Complete linkage:
  - Procura por grupos compactos, elipsoides
  - Pobre desempenho quando há alta densidade de grupos

- Group average linkage:
  - Tende a produzir resultados parecidos com o do complete linkage
  - Resultados piores na presença de *outliers*

Comparação entre métodos hierárquicos

- Centroid linkage:
  - Sugerido apenas para distância Euclidiana
  - Adequado para tratar grupos de diferentes tamanhos

Comparação entre métodos hierárquicos

- Ward´s linkage (baseado em dispersão):
  - Procura grupos elipsoides e compactos
  - Melhores resultados quando os grupos são de mesmo tamanho (grupos maiores tendem a absorver grupos menores)
  - Piores resultados na presença de *outliers*
  - Considerado um dos melhores

Comparação entre métodos hierárquicos

- Conclusão:
  - Nenhum é o melhor para todos os casos
  - Solução: testar vários métodos e escolher aquele mais compatível com o esperado

Determinando a relevância e o número de grupos

- Agrupamento hierárquico permite a escolha de K grupos, $1 \leq K \leq N$
- Qual K escolher?

Determinando a relevância e o número de grupos

- Agrupamento hierárquico permite a escolha de K grupos, $1 \leq K \leq N$
- Qual K escolher?
- Nenhuma regra de ouro
- Mas há dicas para tentar ter uma idéia da relevância de certos grupos...

Determinando a relevância e o número de grupos

- Dica 1: olhar para os grupos com maior tempo de vida – intervalo de distância entre a criação de um grupo e sua união com outro grupo
- Ex: {1,3,4}   x   {1,3} U {4}

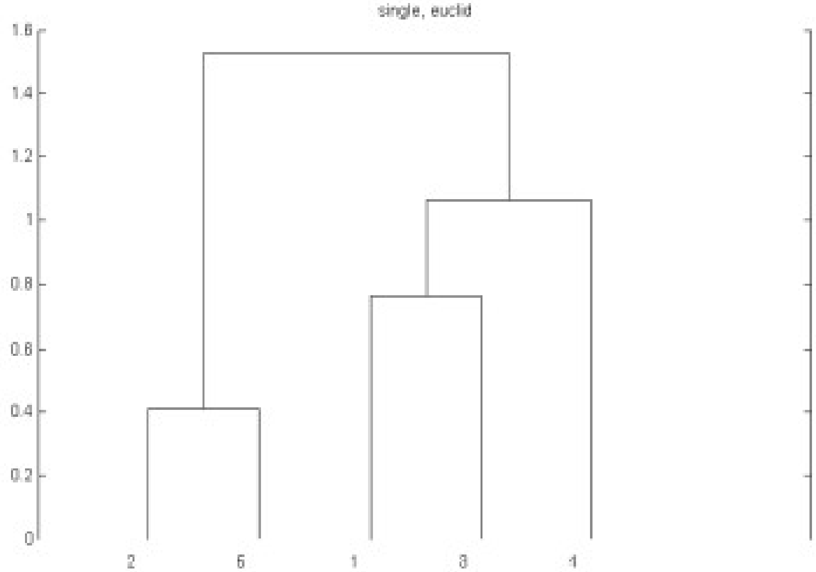


Figure 8.34: *The obtained dendrogram.*

Determinando a relevância e o número de grupos

- Dica 2: olhar para o maior salto entre as distâncias mínimas entre grupos de uma dada escolha de K (se acima de um limiar)
- Ex:

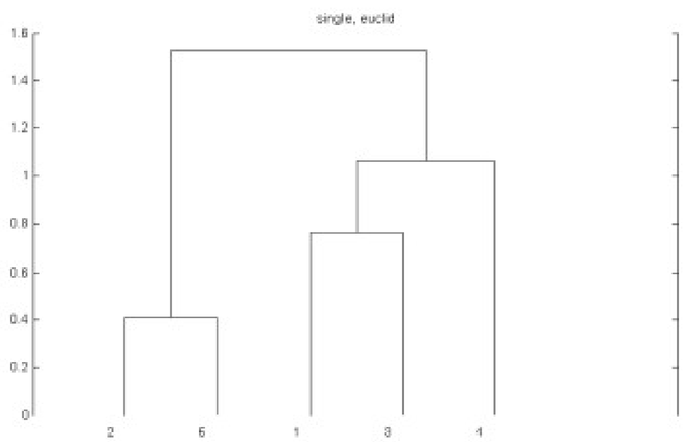


Figure 8.34: *The obtained dendrogram.*

Number of clusters	Distance	Distance jump
4	0.4123	0.3493
3	0.7616	0.3014
2	1.0630	0.4634
1	1.5264	—

Validação

- Como avaliar os resultados?

Validação

- Como avaliar os resultados?

- Uma simples e possível alternativa é replicação
  - Testa com vários subconjuntos
  - Testa com vários algoritmos de agrupamentos
  - Espera-se obter uma certa concordância de grupos

Agrupamento

- É uma técnica de classificação
  - Supervisionada ou não supervisionada?

Agrupamento

- É uma técnica de classificação
  - Supervisionada ou não supervisionada?

Agrupamento

- É uma técnica de classificação
  - Supervisionada ou não supervisionada?
  - Paramétrica ou não paramétrica?

Agrupamento

- É uma técnica de classificação
  - Supervisionada ou não supervisionada?
  - Paramétrica ou não paramétrica?
    - Porém há versões que se utilizam das distribuições condicionais das classes (paramétricas)

R

- Diferentes algoritmos de agrupamentos hierárquicos (pacote stats)
  - hclust

- kmeans

Referências

- BAYNE, C. K. *et al*. Monte Carlo comparisons of selected clustering procedures. **Pattern Recognition**, v. 12, p.51-62, 1980
- COSTA, L. F.; CESAR Jr, R. B**. Shape Classification and Analysis: Theory and Practice**. CRC Press, 2009, 2 ed. (Cap. 8.3)
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical Pattern Recognition: A Review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22 n. 1 p. 4-37, 2000 (seção 8).

SISTEMAS DE INFORMAÇÃO

EACH