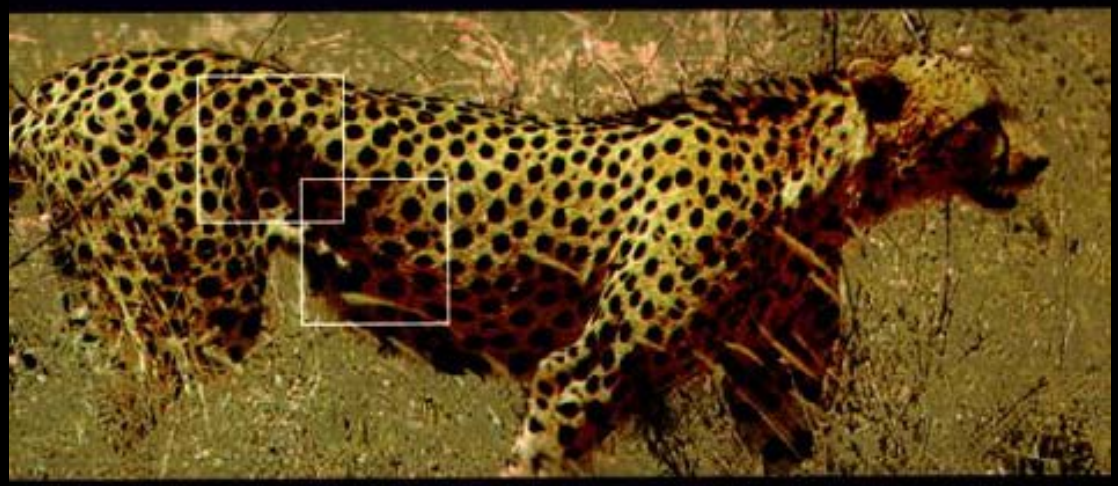


Statistical Pattern Recognition

Second Edition

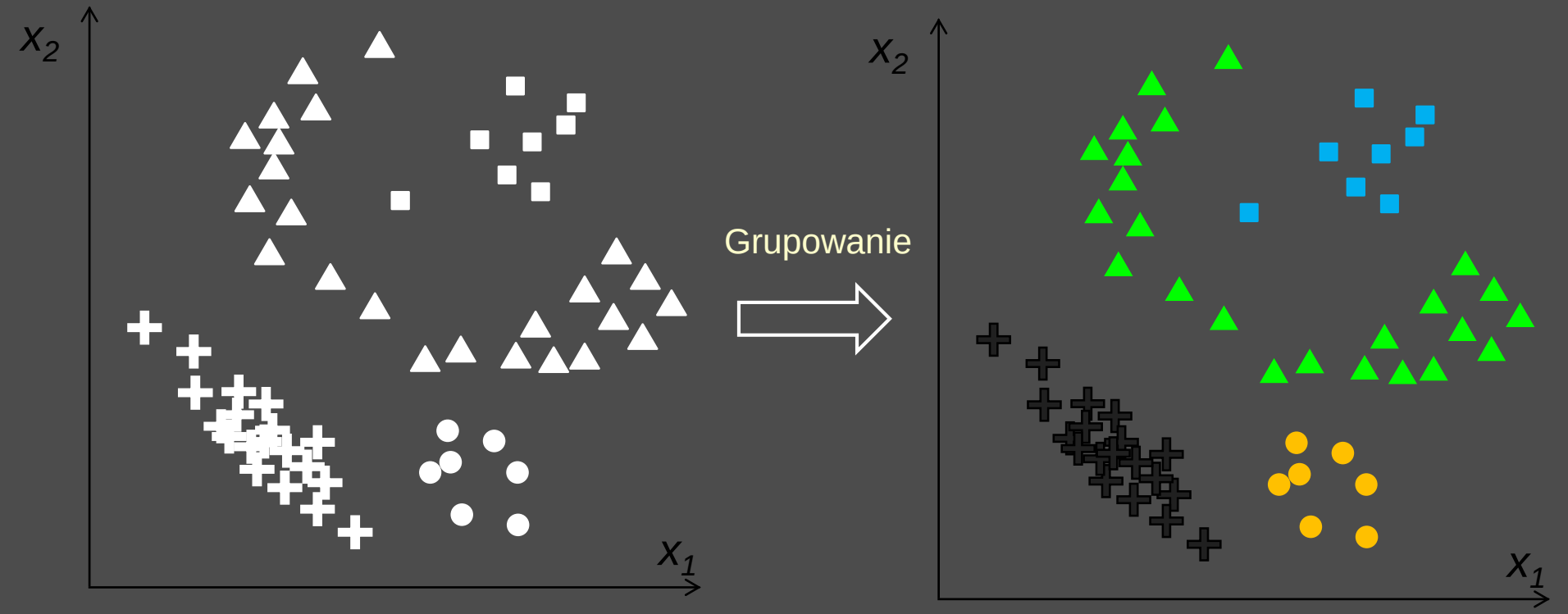
Andrew Webb

Grupowanie

- Wprowadzenie
- Metody hierarchiczne
- Modele mieszane (*mixture models*)
  - Metoda *Expectation-maximization* (EM)
- Metody najmniejszych kwadratów
  - Kryteria jakości grupowania
  - Algorytm *k*-średnich
- Zastosowania

Wprowadzenie

- Grupowanie: podział zbioru danych pomiarowych na rozłączne i zwarte grupy



- Grupa powinna zawierać obiekty podobne do siebie
- Obiekty należące do różnych grup powinny być od siebie znacząco różnić
- Przestrzeń rozwiązań (liczba możliwych podziałów *n* obiektów na *g* grup) jest zbyt duża, nawet dla procedur typu *branch and bound.* Liczba podziałów *n* obiektów na *g* grup jest równa

$$\frac{1}{g!}\sum_{i=1}^{g}(-1)^{g-i}\binom{g}{i}i^n$$

  - Przykładowo, dla n=100, g=5 rozwiązań jest 10^{67}

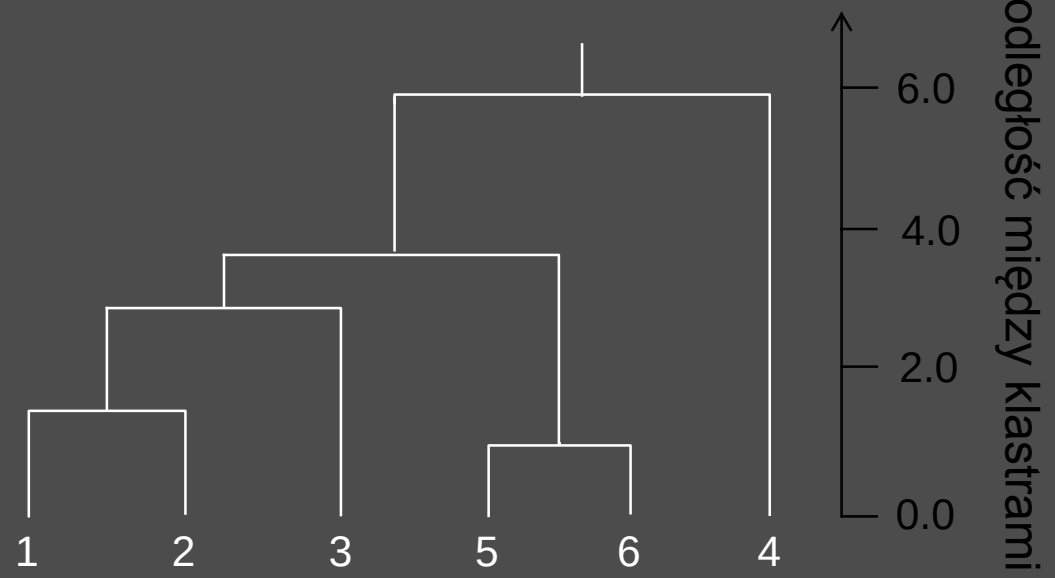
dla n=60, g=2, rozwiązań jest 6×10^{17}

Korzyści

- redukcja ciągu uczącego (np. przez zastąpienie grupy jej reprezentantem) upraszcza procedury rozpoznawania
- uzyskanie „naturalnej” struktury danych ułatwia m.in. dobór postaci modeli

Metody hierarchiczne

- Drzewo hierarchiczne – dendrogram zawierający zagnieżdżone grupy
- Na najwyższym poziomie wszystkie obiekty należą do jednej grupy
- Na najniższym poziomie każdy obiekt stanowi osobną grupę



- Na osi pionowej odkładana jest odległość pomiędzy połączonymi grupami
- Obcięcie drzewa w określonym miejscu daje podział na *g* rozłącznych grup
- Krawędzie drzewa są uporządkowane tak, aby się nie przecinały

- Algorytmy aglomeracyjne (*agglomerative algorithms*)
  - rozpoczynają od tylu grup, ile jest obiektów
  - w każdym kroku łączone są dwie najbardziej podobne grupy
  - warunek stopu: wszystkie obiekty należą do jednej grupy
- Algorytmy deglomeracyjne (*divisive algorithms*)
  - rozpoczynają od jednej grupy obejmującej wszystkie obiekty
  - w każdym kroku dzielą grupę na dwie, najbardziej od siebie odległe
  - warunek stopu: każdy obiekt stanowi osobną grupę
  - rzadko stosowane ze względu na złożoność obliczeniową

Ultrametryka

- Na podstawie dendrogramu można określić nową macierz niepodobieństwa między obiektami, w której odległość między obiektami i, j jest odległością pomiędzy ich grupami
- Odległość liczona na tej wysokości dendrogramu, na której grupy są połączone pojedynczą ścieżką
- Procedurę poszukiwania dendrogramu można rozpatrywać jako transformację macierzy niepodobieństwa, która zawiera odległości d_{ij}, w macierz zawierającą odległości $\hat{d}_{ij}$, które spełniają nierówność ultrametryki:

$$\forall_{i,j,k}\ \hat{d}_{ij} \leq \max\left(\hat{d}_{ik}, \hat{d}_{jk}\right)$$

Algorytm Single-link

- Przyporządkowuje dwa obiekty o indeksach i_0 i i_m do jednej grupy na poziomie *d*, jeżeli istnieje łańcuch obiektów pośrednich $i_0, i_1, i_2, \ldots, i_{m-1}, i_m$ spełniający $\forall_{k=1,\ldots,m-1} d_{i_k, i_{k+1}} \leq d$
- Stosując podejście aglomeracyjne oraz przyjmując miarę odległości między grupami A i B jako odległość między ich najbliższymi sąsiadami: $d_{AB} = \min_{i \in A, j \in B} d_{ij}$

  otrzymujemy następującą sekwencję macierzy niepodobieństwa:

- Macierz początkowa

	1	2	3	4	5	6
1	0	4	13	24	12	8
2		0	10	22	11	10
3			0	7	3	9
4				0	6	18
5					0	8.5
6						0

- Najbliższe grupy to (3) i (5), więc łączone są w pojedynczą grupę (3,5)
- Należy przeliczyć odległości:

$$d_{1,(3,5)} = \min\{d_{13}, d_{15}\} = 12$$

$$d_{2,(3,5)} = \min\{d_{23}, d_{25}\} = 10$$

- Macierz w kolejnym kroku

$$d_{4,(3,5)} = 6 \qquad d_{6,(3,5)} = 8.5$$

	1	2	$(3,5)$	4	6
1	0	4	12	24	8
2		0	10	22	10
$(3,5)$			0	6	8.5
4				0	18
6					0

	1	2	(3,5)	4	6
1	0	4	12	24	8
2		0	10	22	10
(3,5)			0	6	8.5
4				0	18
6					0

- Najbliższe grupy to (1) i (2), więc łączone są w pojedynczą grupę (1,2)
- Należy przeliczyć odległości:

$$d_{(1,2),(3,5)} = \min\{d_{13}, d_{23}, d_{15}, d_{25}\} = 10$$

$$d_{(1,2),4} = \min\{d_{14}, d_{24}\} = 22$$

$$d_{(1,2),6} = \min\{d_{16}, d_{26}\} = 8$$

- Macierz niepodobieństwa w kolejnym kroku:

	(1,2)	(3,5)	4	6
(1,2)	0	10	22	8
(3,5)		0	6	8.5
4			0	18
6				0

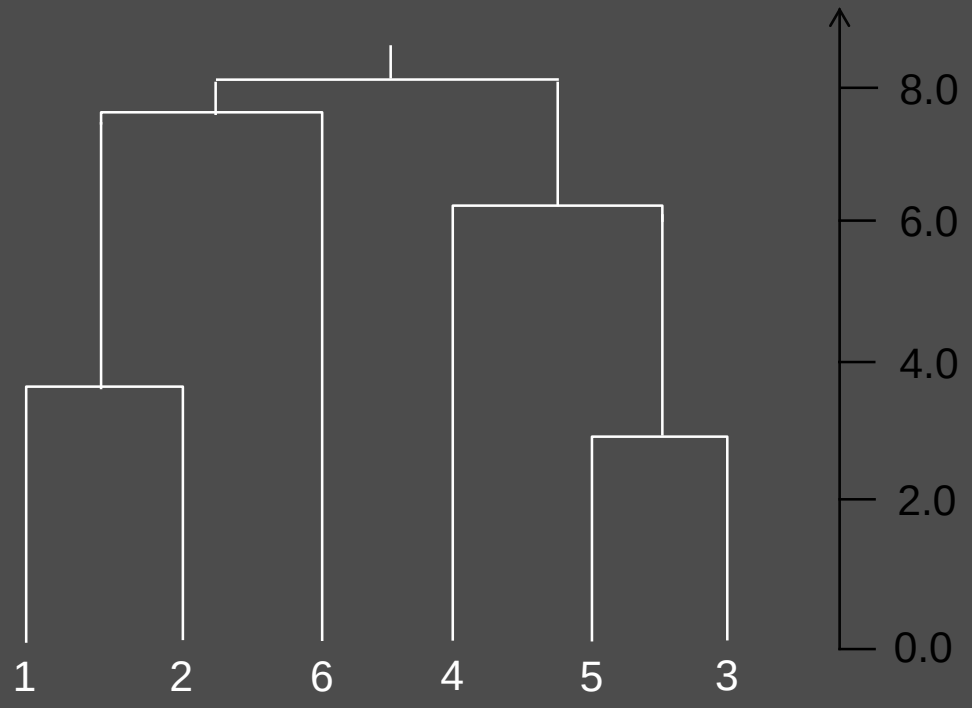
- Najbliższe grupy to (4) i (3,5). Po ich połączeniu otrzymujemy macierz:

	(1,2)	(3,4,5)	6
(1,2)	0	10	8
(3,4,5)		0	8.5
6			0

- Końcowy wynik

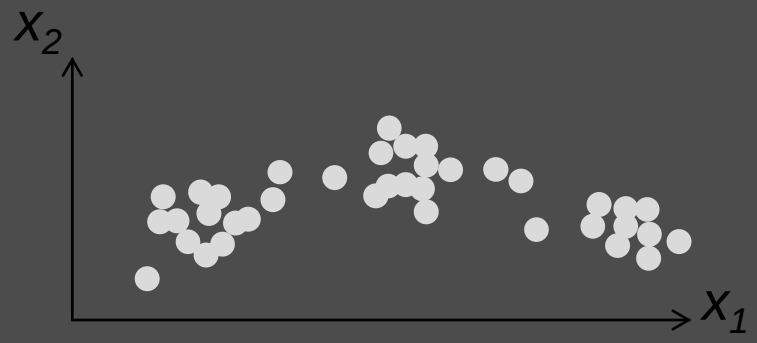
	(1,2,6)	(3,4,5)
(1,2,6)	0	8.5
(3,4,5)		0

- W rezultacie otrzymujemy dendrogram

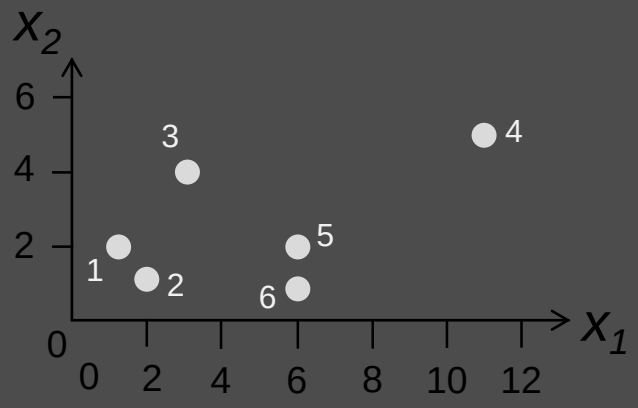


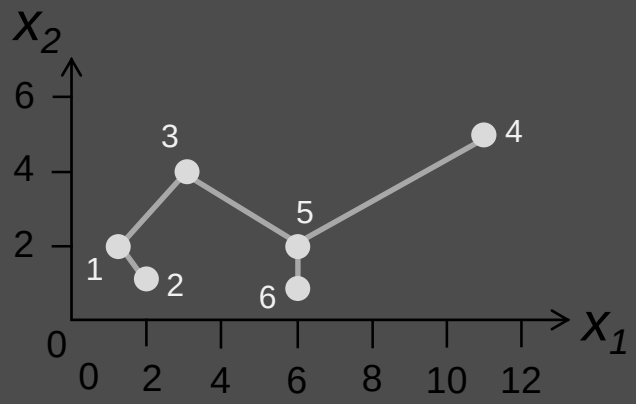
- Algorytm Single-link łączy kolejno grupy, które leżą blisko siebie

  Może to spowodować łączenie odrębnych grup, jeżeli pomiędzy nimi znajdą się przypadkiem jakieś nieliczne obiekty:



- Można zrealizować algorytm Single-link korzystając z minimalnego drzewa rozpinającego (*minimum spanning tree*)
- Drzewo rozpinające: istnieje połączenie (za pomocą jednej lub wielu krawędzi) między każdą parą wierzchołków oraz nie występują cykle
- Minimalne drzewo rozpinające: dodatkowo suma długości wszystkich krawędzi jest minimalna





- Usunięcie z minimalnego drzewa rozpinającego wszystkich krawędzi dłuższych niż *h* daje grupowanie na poziomie *h*

Algorytm Complete-link

- Różnica między algorytmem Complete-ling a Single-link polega na stosowaniu innej miary odległości.
- W algorytmie Complete-link odległości między grupami A i B wyznaczana jest jako odległość między ich najbardziej odległymi obiektami:

$$d_{AB} = \max_{i \in A,\, j \in B} d_{ij}$$

- Dla danych z poprzedniego zadania pierwszy krok procedury daje następujący wynik:

	1	2	3	4	5	6
1	0	4	13	24	12	8
2		0	10	22	11	10
3			0	7	3	9
4				0	6	18
5					0	8.5
6						0

Single-link

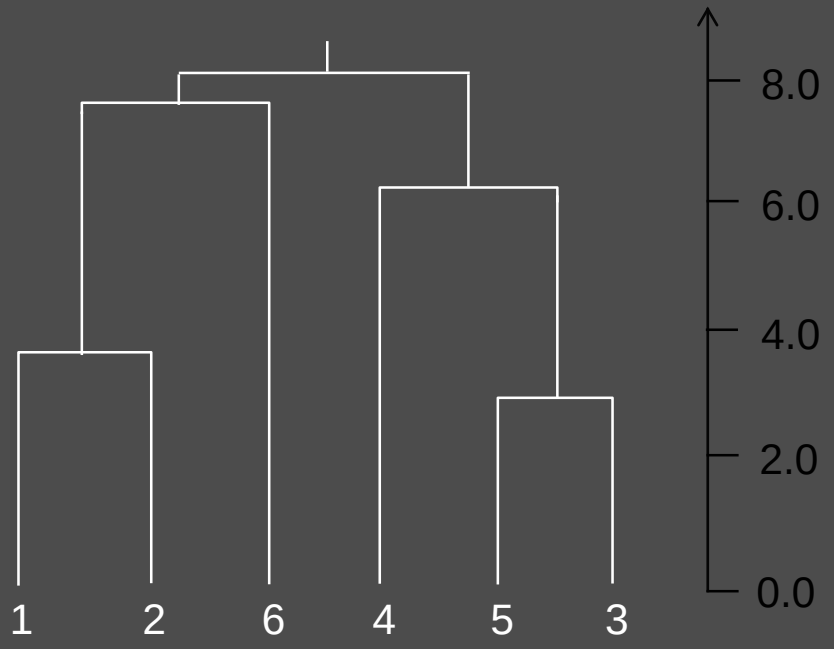
	1	2	$(3,5)$	4	6
1	0	4	12	24	8
2		0	10	22	10
$(3,5)$			0	6	8.5
4				0	18
6					0

Complete-link

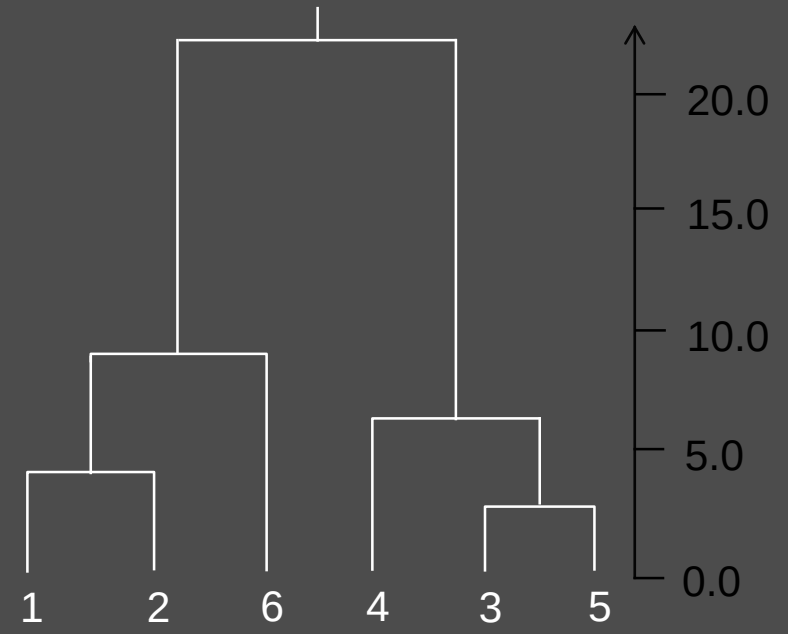
	1	2	$(3,5)$	4	6
1	0	4	13	24	8
2		0	11	22	10
$(3,5)$			0	7	9
4				0	18
6					0

- Dendrogramy:

dendrogram algorytmu
Single-link



dendrogram algorytmu
Complete-link



- Algorytm Single-link wykrywa odizolowane grupy
- Dendrogram wygenerowany przez algorytm Single-link, po przecięciu na poziomie *h*, daje grupy odizolowane od siebie przynajmniej o odległość *h*
- Algorytm Complete-link wykrywa zwarte grupy

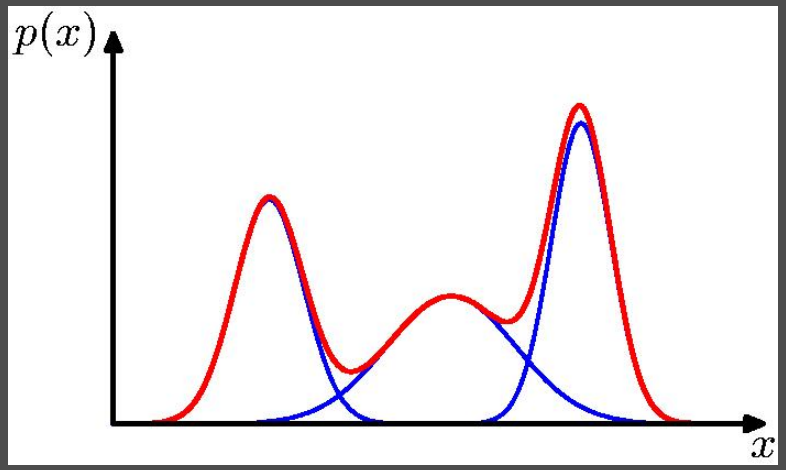
Inne metody

- *Centroid distance* : odległość między grupami jest zdefiniowana jako odległość pomiędzy średnimi w grupach
  - Uwaga: przy łączeniu grupy bardzo licznej z grupą mającą niewiele obiektów, średnia z całości leży blisko średniej z bardziej licznego klastra
- *Median distance* : odległość między grupami jest zdefiniowana jako odległość pomiędzy medianami w grupach
- *Group average link* : odległość między dwiema grupami jest zdefiniowana jako średnia wartość niepodobieństw między wszystkimi parami obiektów z różnych grup:

$$d_{AB} = \frac{1}{n_i n_j} \sum_{i \in A,\, j \in B} d_{ij}$$

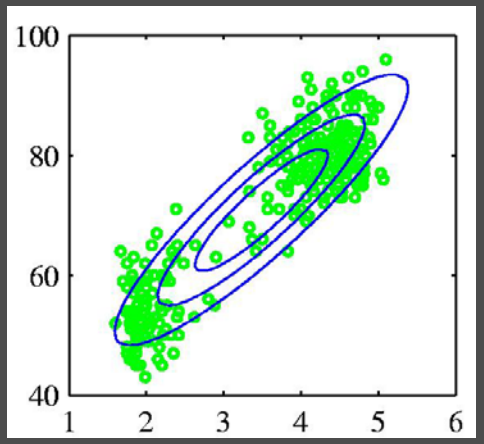
Modele mieszane (*mixture models*)

- Każda grupa opisana jest innym rozkładem prawdopodobieństwa
- Model mieszany jest sumą rozkładów poszczególnych grup

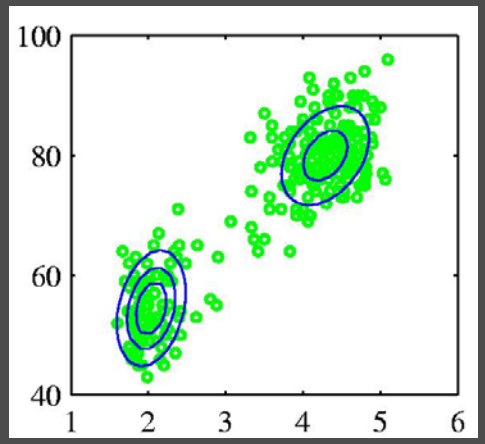


liczba grup = 3

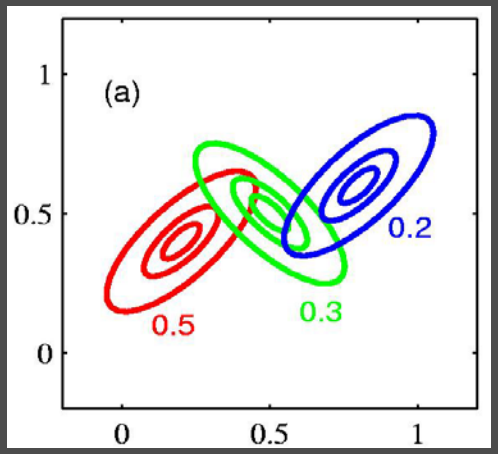
- Mieszanina wielowymiarowych rozkładów Gaussa – najczęściej stosowana w grupowaniu

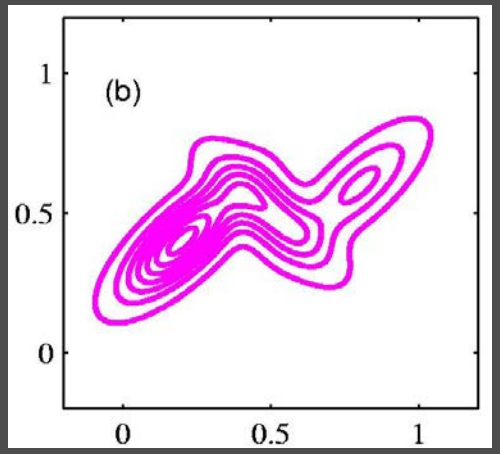


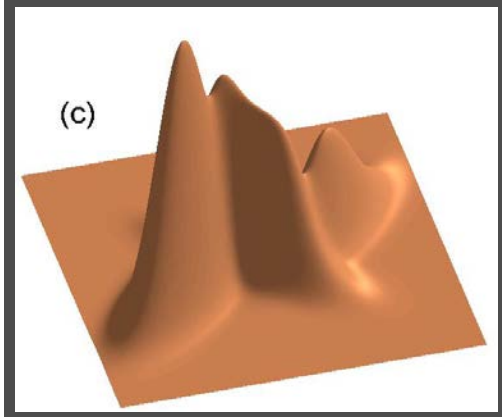
Pojedynczy rozkład Gaussa



Mieszanina dwóch rozkładów Gaussa







Model mieszany

$$p(\mathbf{x}) = \sum_{i=1}^{g} \pi_i \, p(\mathbf{x}; \boldsymbol{\theta}_i)$$

π_i – współczynniki określające udział rozkładu *i*-tej grupy w modelu mieszanym, przy czym $\pi_i \geq 0, \sum_{i=1}^{g} \pi_i = 1$

$p(\mathbf{x}; \boldsymbol{\theta}_i)$ – wielowymiarowy rozkład prawdopodobieństwa, zależny od wektora parametrów $\boldsymbol{\theta}_i$

- Estymować należy trzy zestawy parametrów:
  - π_i , $\boldsymbol{\theta}_i$, g - liczba grup

- Mieszanina rozkładów Gaussa

$$p(\mathbf{x}) = \sum_{i=1}^{g} \pi_i \, p(\mathbf{x};\, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

gdzie $$p(\mathbf{x};\, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]$$

- Estymacja parametrów
  - metoda maksymalnej wiarygodności
  - metoda Expectation-Maximization (EM)
- Grupowanie
  - przypisanie obiektów do grup na podstawie estymowanych prawdopodobieństw a'posteriori przynależności do grup

  Obiekt **x** należy do grupy *i*, jeżeli

$$\forall_{j \neq i,\, j=1,\ldots,g} \; \pi_i \, p(\mathbf{x}; \boldsymbol{\theta}_i) \geq \pi_j \, p(\mathbf{x}; \boldsymbol{\theta}_j)$$

- Metoda *Expectation-Maximization* (EM)
  - W modelu mieszanym postaci

$$p(\mathbf{x}) = \sum_{i=1}^{g} \pi_i \, p(\mathbf{x}; \boldsymbol{\theta}_i)$$

przyjmujemy rozkład Gaussa $p(\mathbf{x}; \boldsymbol{\theta}_i)$, zatem $\boldsymbol{\theta}_i = \{ \boldsymbol{\mu}_j ; \boldsymbol{\Sigma}_j \}$

  - Dla *n*-elementowej próby $\mathbf{X}=\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ konstruujemy funkcję wiarygodności o postaci:

$$L(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \sum_{j=1}^{g} \pi_i \, p(\mathbf{x}_i \mid \boldsymbol{\theta}_j)$$

gdzie $\boldsymbol{\Psi} = \{\pi_1, \ldots, \pi_g ; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g\}$ jest zbiorem parametrów

- Dla ogólnego przypadku nie jest możliwe rozwiązanie układu

$$\nabla_{\Psi} L = \mathbf{0}$$

- Funkcję wiarygodności *L* maksymalizuje się z wykorzystaniem ogólnej klasy procedur iteracyjnych, znanych jako EM (*expectation – maximization*).
- Zostały one wprowadzone w kontekście estymacji brakujących danych
- W kontekście zadania grupowania brakującymi danymi są numery klas (grup)

- Dany jest zestaw *n* niekompletnych danych pomiarowych:

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix}, \quad i = 1, \ldots, n,$$

gdzie $\mathbf{z}_i$ jest wektorem zawierającym brakujące wartości oraz

$\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\} = \mathbf{Y}$

- Dla danego wektora $\mathbf{x}_i$ można zaproponować wiele różnych wektorów $\mathbf{y}_i$.
- Jedne mogą być bardziej, drugie mniej prawdopodobne.
- W zadaniu grupowania z użyciem modeli mieszanych wektor $\mathbf{z}_i = [z_{1i} \quad \cdots \quad z_{gi}]^T$ wskazuje numer grupy, tj. $\mathbf{z}_{ji} = 1$ jeżeli $\mathbf{x}_i$ należy do *j*-tej grupy, w przeciwnym przypadku $\mathbf{z}_{ji} = 0$

Ogólna postać procedury EM

- Maksymalizowana będzie funkcja wiarygodności

$$L(\mathbf{\Psi}) = p(\mathbf{X} \mid \mathbf{\Psi})$$

- Funkcję $p(\mathbf{X} \mid \mathbf{\Psi})$ można wyznaczyć na podstawie funkcji $p(\mathbf{Y} \mid \mathbf{\Psi})$, której postać jest znana
- Funkcję $p(\mathbf{Y} \mid \mathbf{\Psi})$ uzyskujemy przez scałkowanie po wszystkich zbiorach **Y**, które mogą być uzupełnieniem zbioru **X**:

$$L(\mathbf{\Psi}) = p(\mathbf{X}, \mathbf{\Psi}) = \int \prod_{i=1}^{n} g(\mathbf{x}_i, \mathbf{z} \mid \mathbf{\Psi}) dz$$

Ogólna postać procedury EM

- Począwszy od pewnego początkowego rozwiązania $\boldsymbol{\Psi}^{(0)}$ generowana jest sekwencja $\left\{\boldsymbol{\Psi}^{(m)}\right\}$ oszacowań parametrów $\boldsymbol{\Psi}$
- Procedura polega na naprzemiennych wykonywaniu dwóch kroków:

  1) E-step: wyznaczenie wartości kryterium jakości

$$Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(m)}\right) \equiv \mathrm{E}\left[\log\left(g\left(\mathbf{Y} \mid \boldsymbol{\Psi}\right)\right) \mid \mathbf{X}, \boldsymbol{\Psi}^{(m)}\right]$$

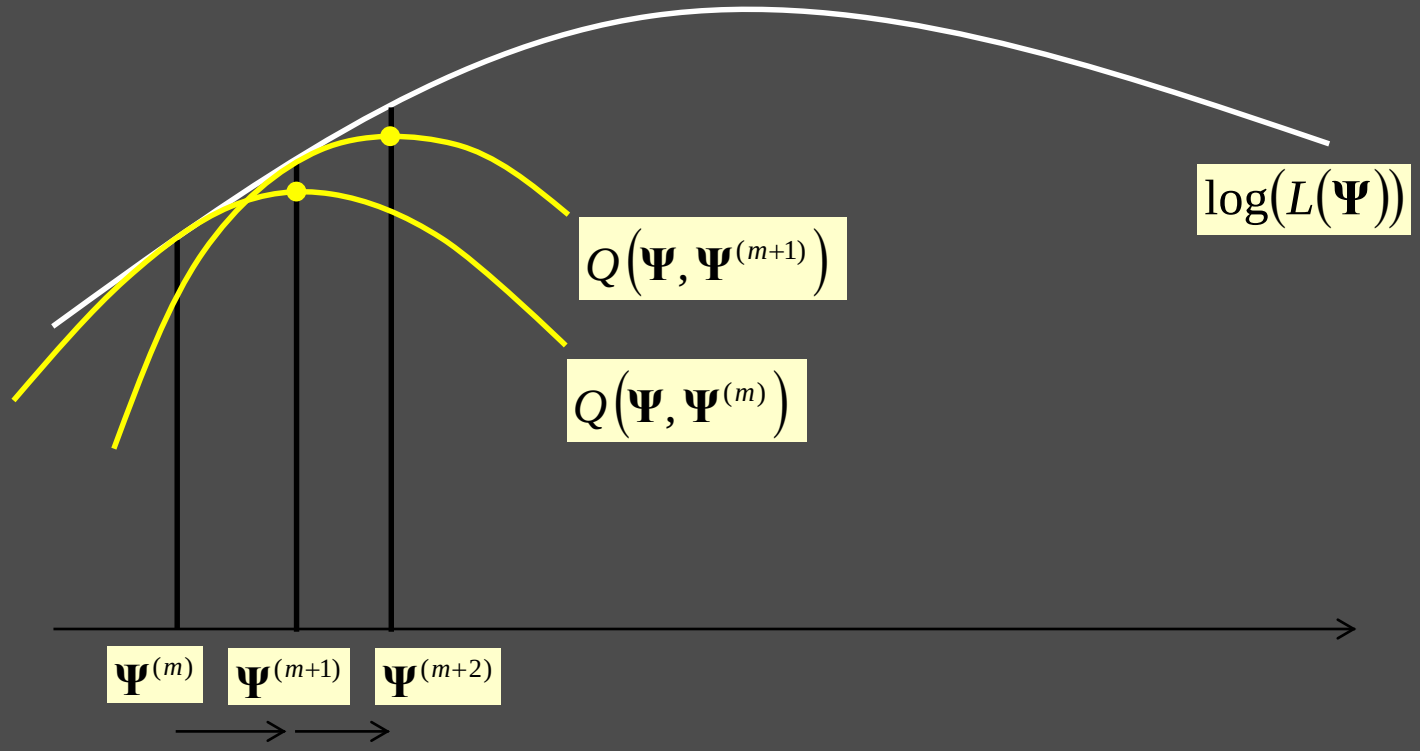
  tzn.

$$Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(m)}\right) \equiv \int \sum_i \log\left(g\left(\mathbf{x}_i, \mathbf{z}_i \mid \boldsymbol{\Psi}\right)\right) p\left(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\Psi}^{(m)}\right) d\mathbf{z}_1 \ldots d\mathbf{z}_n$$

  2) M-step: znalezienie $\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(m+1)}$ maksymalizującego $Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(m)}\right)$

- W kolejnych iteracjach zachodzi

$$L\left(\boldsymbol{\Psi}^{(m+1)}\right) \geq L\left(\boldsymbol{\Psi}^{(m)}\right)$$


$\log(L(\Psi))$
$Q(\Psi, \Psi^{(m+1)})$
$Q(\Psi, \Psi^{(m)})$
$\Psi^{(m)}$
$\Psi^{(m+1)}$
$\Psi^{(m+2)}$

- Algorytm EM dla mieszaniny rozkładów Gaussa
  - Oznaczmy $w_{ji} \equiv \mathrm{E}\left[z_{ji} \mid \mathbf{x}_i, \boldsymbol{\Psi}^{(m)}\right]$

1) E-step:

$$w_{ji} = \frac{\pi_j^{(m)} p\left(\mathbf{x}_i \mid \boldsymbol{\theta}_j^{(m)}\right)}{\sum_k \pi_k^{(m)} p\left(\mathbf{x}_i \mid \boldsymbol{\theta}_k^{(m)}\right)}$$

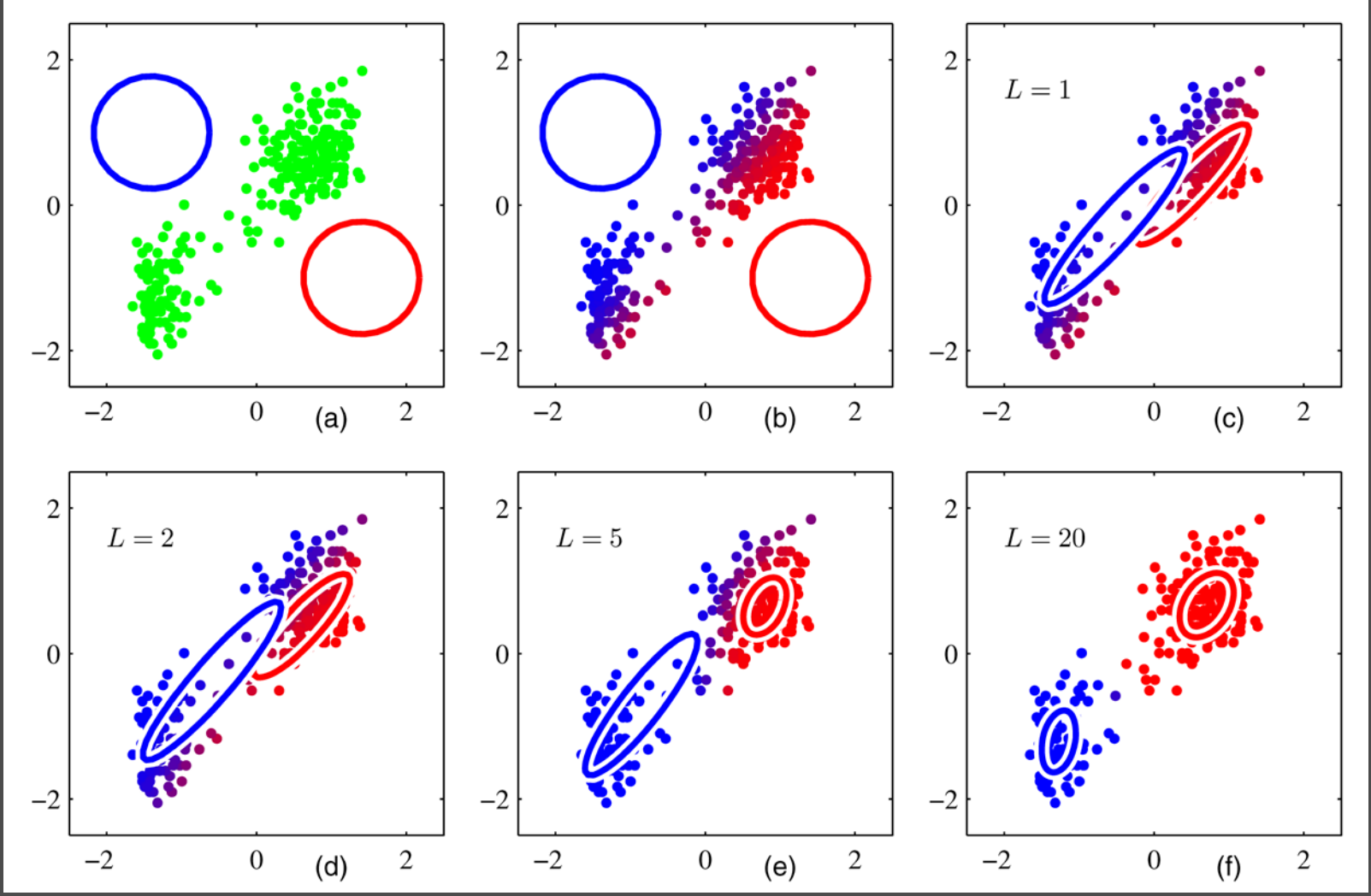
2) M-step:

$$\hat{\pi}_j = \frac{1}{n}\sum_{i=1}^{n} w_{ji}$$

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{n\hat{\pi}_j}\sum_{i=1}^{n} w_{ji}\mathbf{x}_i$$

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n\hat{\pi}_j}\sum_{i=1}^{n} w_{ji}\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_j\right)\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_j\right)^T$$

- Algorytm EM dla mieszaniny rozkładów Gaussa



Metody najmniejszych kwadratów

- Podział na grupy uzyskiwany jest w drodze maksymalizacji zadanego kryterium jakości grupowania
- Kryterium jakości uwzględnia macierze rozproszeń wewnątrz grup i między grupami

- Średnie i kowariancje
  - Dana jest *n*-elementowa próba $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
  - Średnia z próby

$$\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

  - Macierz kowariancji z próby

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

  - Liczba obiektów w *j*-tej grupie

$$n_j = \sum_{i=1}^{n} z_{ij}$$, gdzie $$z_{ij} = \begin{cases} 1 & \text{dla} \quad \mathbf{x}_i \in \text{grupa } j \\ 0 & \text{w przec. przyp} \end{cases}$$

  - Średnia w *j*-tej grupie

$$\mathbf{m}_j = \frac{1}{n_j}\sum_{i=1}^{n} z_{ij}\mathbf{x}_i$$

- Macierze rozproszenia (*scatter matrices*)
  - Macierz rozproszenia wewnątrz grup

    (*pooled within-group scatter matrix*)

$$\mathbf{S}_W = \frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n} z_{ij}\left(\mathbf{x}_i - \mathbf{m}_j\right)\left(\mathbf{x}_i - \mathbf{m}_j\right)^T$$

  - Macierz rozproszenia między grupami

    (*between-group scatter matrix*)

$$\mathbf{S}_B = \hat{\boldsymbol{\Sigma}} - \mathbf{S}_W = \sum_{j=1}^{g}\frac{n_j}{n}\left(\mathbf{m}_j - \mathbf{m}\right)\left(\mathbf{m}_j - \mathbf{m}\right)^T$$

- Kryteria jakości grupowania
  - Ślad macierzy $\boldsymbol{S}_W$

$$\mathrm{Tr}(\mathbf{S}_W) = \frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n} z_{ij}\left|\mathbf{x}_i - \mathbf{m}_j\right|^2 = \frac{1}{n}\sum_{j=1}^{g} S_j$$

gdzie S_j to suma kwadratów odchyleń od średniej w *j*-tej grupie. Minimalizacja kryterium $\mathrm{Tr}(\boldsymbol{S}_W)$ oznacza minimalizację całkowitej sumy kwadratów odchyleń od średniej wewnątrz grup.

  - Iloraz $\left|\mathbf{S}_W\right| \Big/ \left|\hat{\boldsymbol{\Sigma}}\right|$

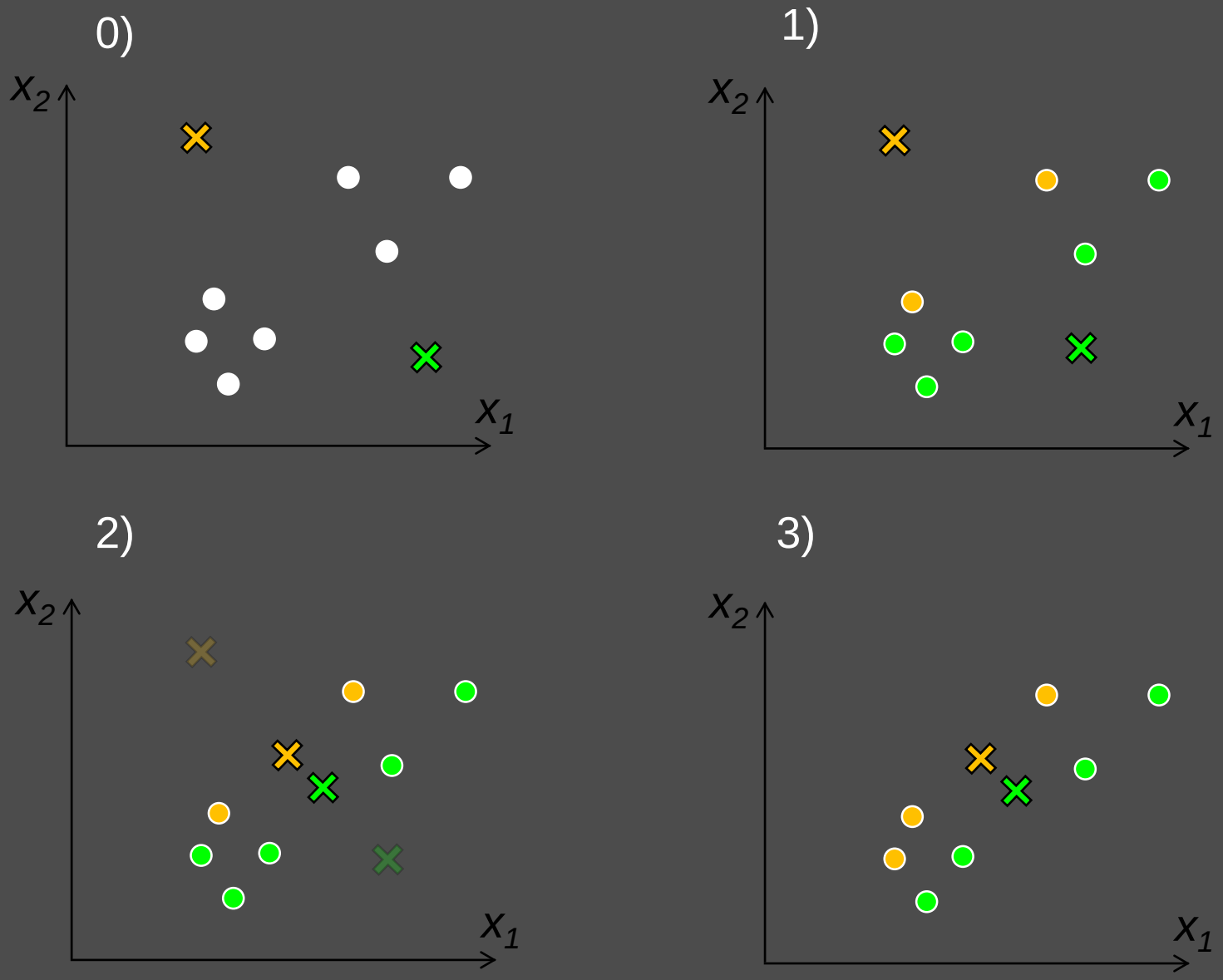
Jest to kryterium niezmiennicze ze względu na nieosobliwe (*nonsingular*) transformacje liniowe (czyli takie, które mają odwrotność). Dla danej próby minimalizacja kryterium $\left|\mathbf{S}_W\right| \Big/ \left|\hat{\boldsymbol{\Sigma}}\right|$ jest równoważna podziałowi minimalizującemu $\mathbf{S}_W$, ponieważ macierz $\hat{\boldsymbol{\Sigma}}$ nie zależy od podziału na grupy.

- Kryteria jakości grupowania
  - Kryterium $\mathrm{Tr}\left(\mathbf{S}_W^{-1}\mathbf{S}_B\right)$

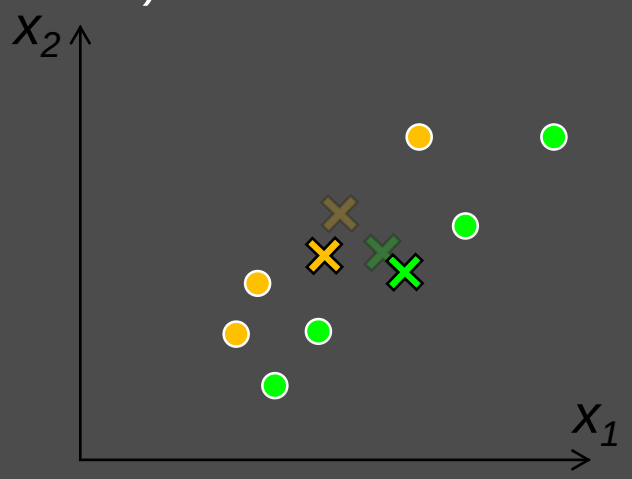
    Maksymalizacja tego kryterium prowadzi go grup o kształcie hiperelipsoidalnym (ogólniejsze od grup hipersferycznych). Odpowiada to przyjęciu metryki Mahalanobisa w miejsce Euklidesowej. Jest to kryterium niezmiennicze ze względu na nieosobliwe transformacje.
  - Kryterium $\mathrm{Tr}\left(\hat{\Sigma}^{-1}\mathbf{S}_W\right)$

    Minimalizacja kryterium kwadratowego dla danych unormowanych w taki sposób, że macierz rozproszenia jest macierzą jednostkową

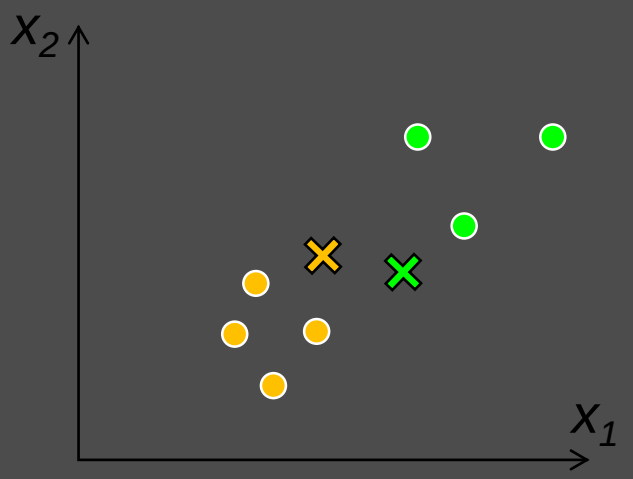
- Algorytm *k*-średnich (*k-means*)
  - Algorytm znajduje rozwiązanie suboptymalne
  - Minimalizacja kryterium Tr($\boldsymbol{S}_W$) przez podział danych na *k* grup
  - Wykonywane są naprzemiennie dwie procedury
    - obiekty przydzielane są do grupy, której środek leży najbliżej (odległość Euklidesowa)
    - dla wykonanego przydziału wyznaczane są nowe średnie grup
  - Warunek stopu: w kolejnych iteracjach nie udaje się zmniejszyć wartości kryterium jakości


0)
x_2
x_1
1)
x_2
x_1
2)
x_2
x_1
3)
x_2
x_1

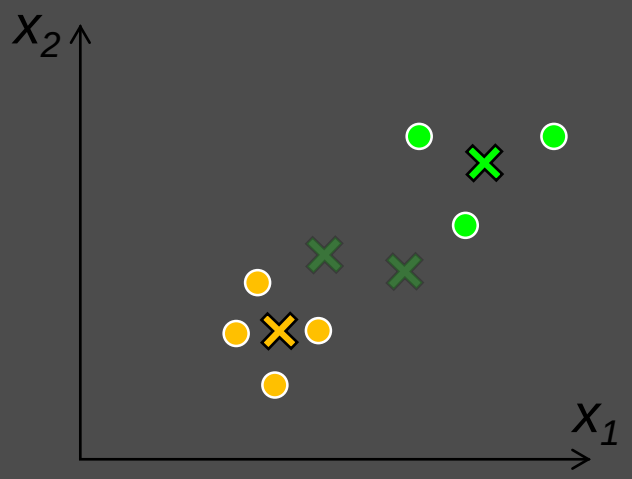
4)



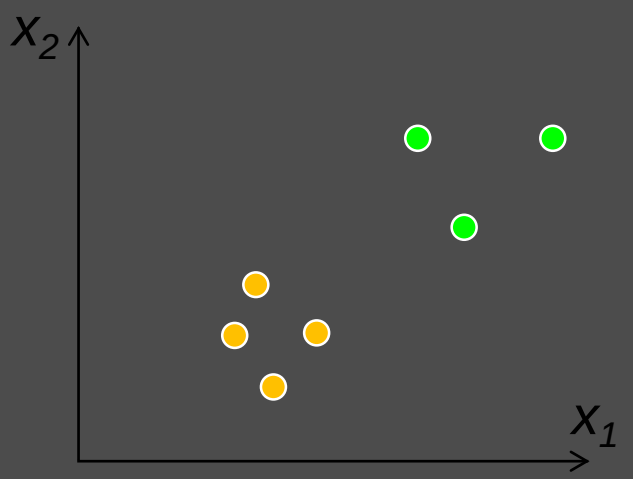
5)



6)



7)



- Algorytm *k*-średnich jest szczególnym przypadkiem algorytmu Expectation-Maximization

- Rozmyty algorytm *k*-średnich (*Fuzzy k-means*)
  - Przynależność obiektów do grup określona jest za pomocą funkcji przynależności, którą reprezentuje zestaw parametrów μ_{ji}

    ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, g$)

    μ_{ji} – stopień przynależności *i*-tego obiektu do *j*-tej grupy
  - Minimalizowane jest kryterium jakości:

$$J_r = \sum_{i=1}^{n}\sum_{j=1}^{g} \mu_{ji}^{r} \left|\mathbf{x}_i - \mathbf{m}_j\right|^2, \qquad \text{gdzie } \mathbf{m}_j = \frac{\sum_{i=1}^{n} \mu_{ji}^{r}\mathbf{x}_i}{\sum_{i=1}^{n} \mu_{ji}^{r}},$$

    $\mathbf{m}_j$ jest centroidą *j*-tej grupy oraz $r \geq 1$ jest stopniem rozmycia grup (dla r = 1 μ_{ji} przyjmuje wyłącznie wartości *0* i *1*, co sprowadza algorytm do zwykłego algorytmu k-średnich)

1) Ustal wartość r, wybierz początkowe wartości μ_{ji}

2) Wyznacz centra $\mathbf{m}_j$ grup

3) Oblicz odległości

$$d_{ij} = \left| \mathbf{x}_i - \mathbf{m}_j \right|$$

4) Wyznacz wartości funkcji przynależności

- Jeżeli $d_{il} = 0$ dla pewnego l to $\mu_{li} = 1$
- Jeżeli $j \neq i$ to $\mu_{li} = 0$
- W pozostałych przypadkach

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{g} \left(d_{ij} / d_{ik} \right)^{\frac{2}{r-1}}}$$

5) Jeżeli nie zachodzi warunek stopu, przejdź do kroku 2)

Zastosowania

- Segmentacja obrazów
- Rozpoznawanie pisma odręcznego (jedna litera może być pisana na różne sposoby – każdemu sposobowi odpowiada w przestrzeni cech pewna grupa)
- Grupowanie książek i dokumentów
- Eksploracja danych w sieci www
- Odnajdywanie struktur w dużych bazach danych
- Podział pacjentów na grupy
- Grupowanie genów na podstawie danych o ich ekspresji