

# PROJEKT J

# 23.01.2019

## TERMIN ROZLICZENIA:

## 29.01.2019

## METODA K-ŚREDNICH

### Projekt

Zaimplementuj algorytm K-średnich metodą Lloyd'a przy różnych  $K$  oraz dla różnych ustawień początkowych.

Oceń zestawy współrzędnych centrów uzyskane z różnych przebiegów.

Obliczenia przeprowadź dla następującej tablicy ekspresji.

Time	1 hr	2 hr	3 hr
$g_1$	10.0	8.0	10.0
$g_2$	10.0	0.0	9.0
$g_3$	4.0	8.5	3.0
$g_4$	9.5	0.5	8.5
$g_5$	4.5	8.5	2.5
$g_6$	10.5	9.0	12.0
$g_7$	5.0	8.5	11.0
$g_8$	2.7	8.7	2.0
$g_9$	9.7	2.0	9.0
$g_{10}$	10.2	1.0	9.2

(a) Intensity matrix,  $I$

2

### Problem:

Dla zadanego zbioru  $n$  punktów z  $m$ -wymiarowej przestrzeni oraz danej wartości  $k$  zbudować zbiór  $X$ ,  $k$  centrów klastrowania takich, dla których średni błąd kwadratowy jest minimalny.

---

### $k$ -Means Clustering Problem:

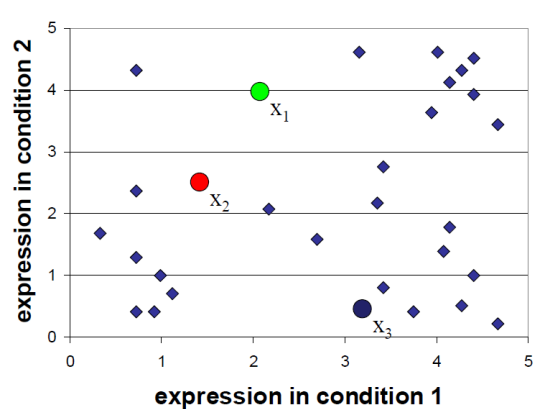
Given  $n$  data points, find  $k$  center points minimizing the squared error distortion.

**Input:** A set,  $\mathcal{V}$ , consisting of  $n$  points and a parameter  $k$ .

**Output:** A set  $\mathcal{X}$  consisting of  $k$  points (called centers) that minimizes  $d(\mathcal{V}, \mathcal{X})$  over all possible choices of  $\mathcal{X}$ .

---

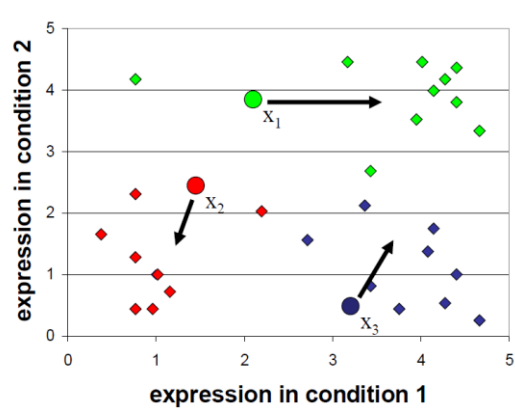
1. Lloyd Algorithm
2. Arbitrarily assign the  $k$  cluster centers
3. while the cluster centers keep changing
4. Assign each data point to the cluster  $C_i$  corresponding to the closest cluster representative (center) ( $1 \leq i \leq k$ )
5. Compute new cluster representatives according to the center of gravity of each cluster, that is, the new cluster representative is  $\sum v / |C|$  for all  $v$  in  $C$  for every cluster  $C$



Zadanie:  
podziel  
punkty o  
dwóch  
cechach  
na trzy klasy

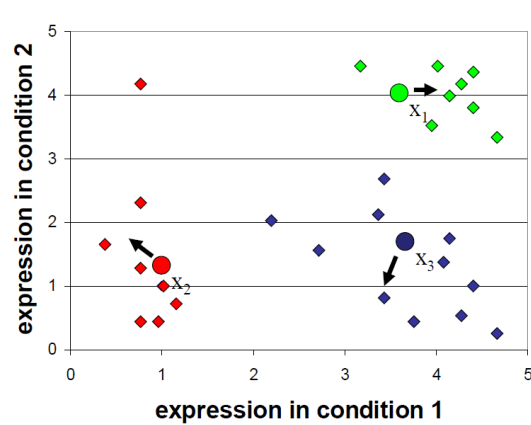
•Krok 1 - wylosuj k położeń dla centrów

5



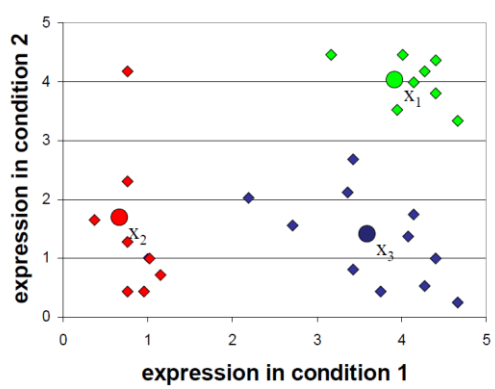
*Zadanie:*  
podziel  
punkty o  
dwóch  
cechach  
na trzy klasy

- Krok 2** - przydziel każdemu badanemu punktowi najbliższy mu punkt centrum przez obliczenie odległości punktu od każdego z centrów.



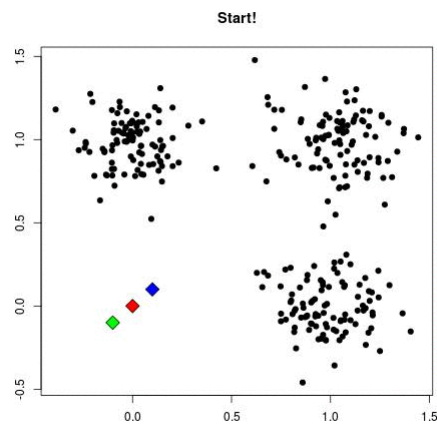
*Zadanie:*  
podziel  
punkty o  
dwóch  
cechach  
na trzy klasy

- Krok 3** - Ustal nowe położenie centrów poprzez uśrednienie współrzędnych przydzielonych punktów



**Zadanie:**  
podziel  
punkty o  
dwóch  
cechach  
na trzy klasy

- Krok 1** - wylosuj k położeń dla centrów
- Krok 2** - przydziel każdemu badanemu punktowi najbliższy mu punkt centrum przez obliczenie odległości punktu od każdego z centrów.
- Krok 3** - ustal nowe położenie centrów poprzez uśrednienie współrzędnych przydzielonych punktów
- Krok 4** - powtarzaj kroki 2 i 3 dopóki nie ustalą się położenia centrów.



- Krok 1** - wylosuj k położeń dla centrów
- Krok 2** - przydziel każdemu badanemu punktowi najbliższy mu punkt centrum przez obliczenie odległości punktu od każdego z centrów.
- Krok 3** - Ustal nowe położenie centrów poprzez uśrednienie współrzędnych przydzielonych punktów
- Krok 4** - powtarzaj kroki 2 i 3 dopóki nie ustalą się położenia centrów.