K-Means algorithm and related

***K. Gibert*[(1)]**

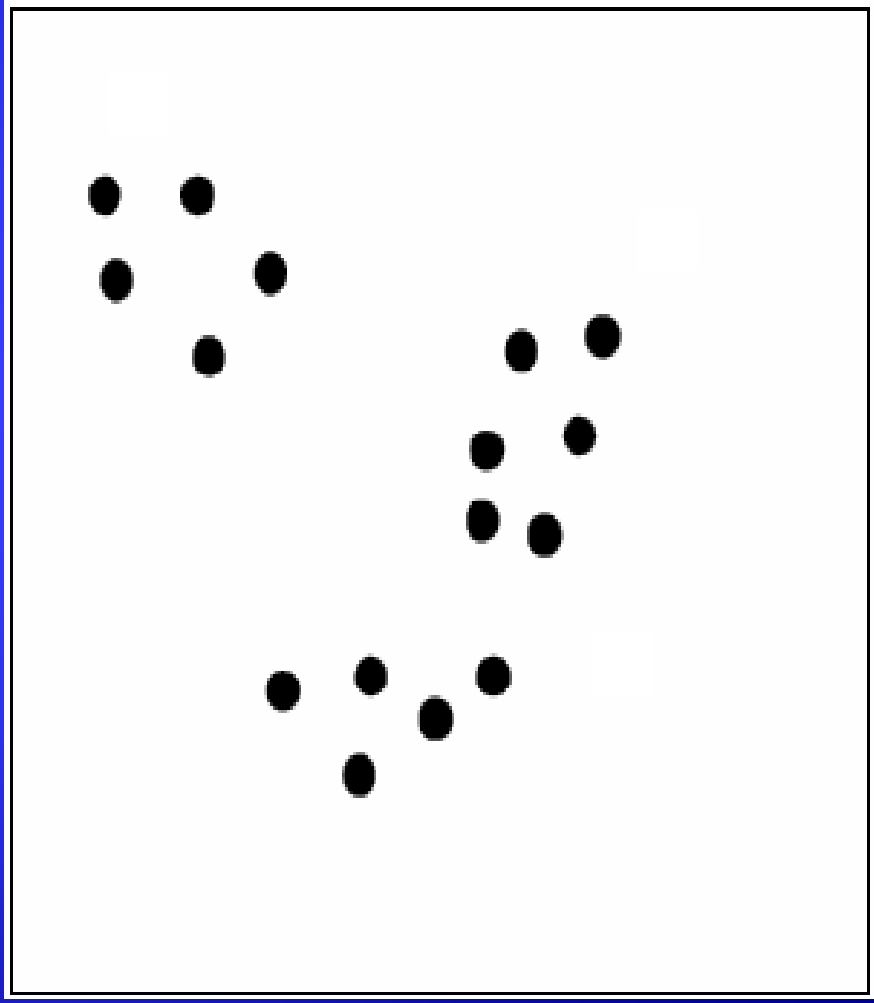
[(1)]Department of Statistics and Operation Research

Knowledge Engineering and Machine Learning group
Universitat Politècnica de Catalunya, Barcelona

©K. Gibert

K-Means algorithm *[McQueen 67]*

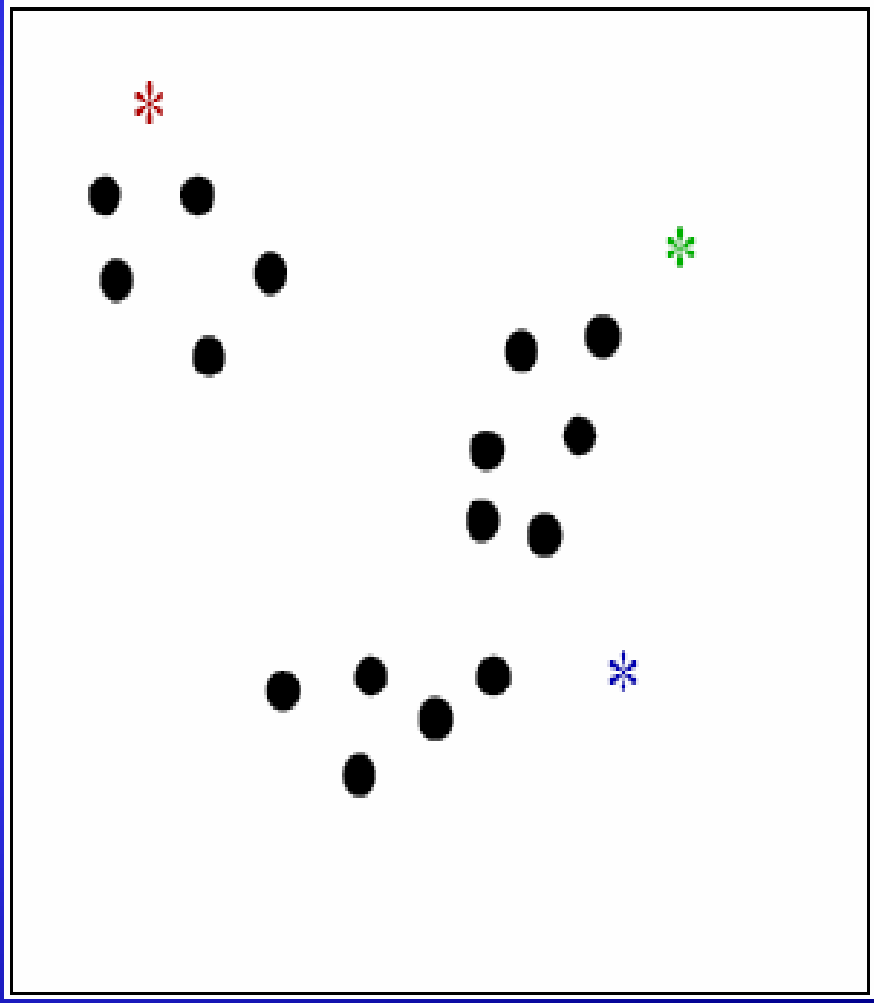
- Select k
- Select k seeds (random?...)
  - *K-Medoids*
    (use Medians as seeds)



©K. Gibert

K-Means algorithm

- Select k
- Select k seeds (random?...)



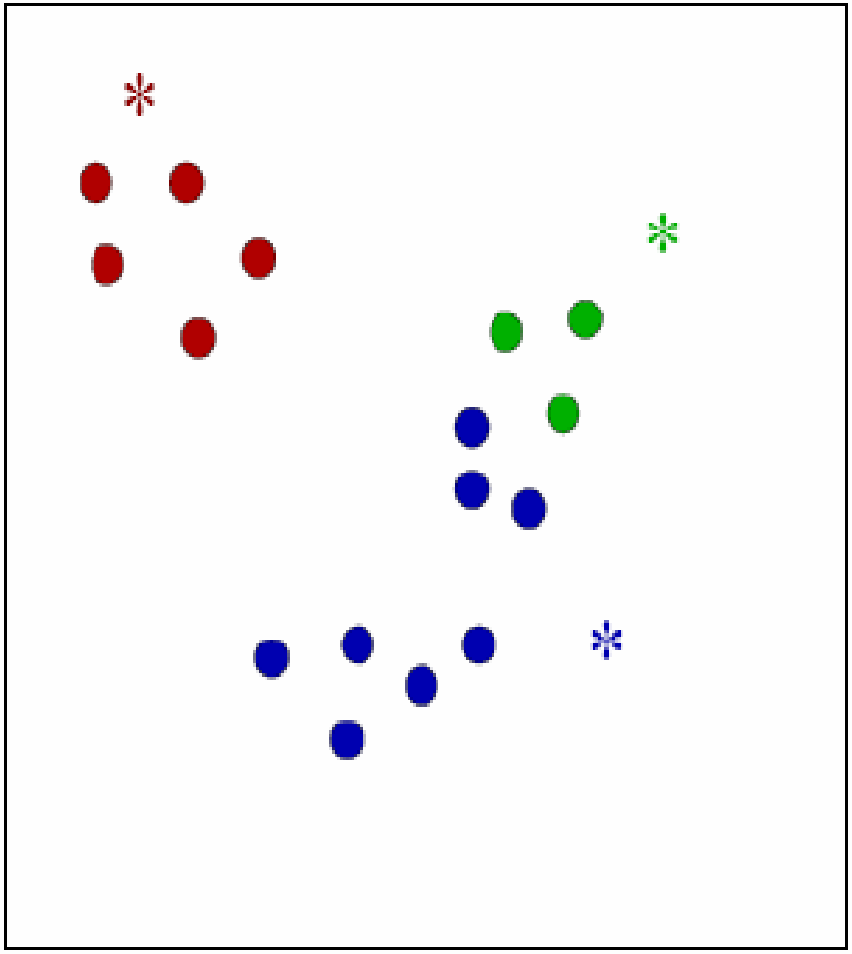
©K. Gibert

K-Means algorithm

- Select k
- Select k seeds (random?...)

- Assign objects to seeds (class)

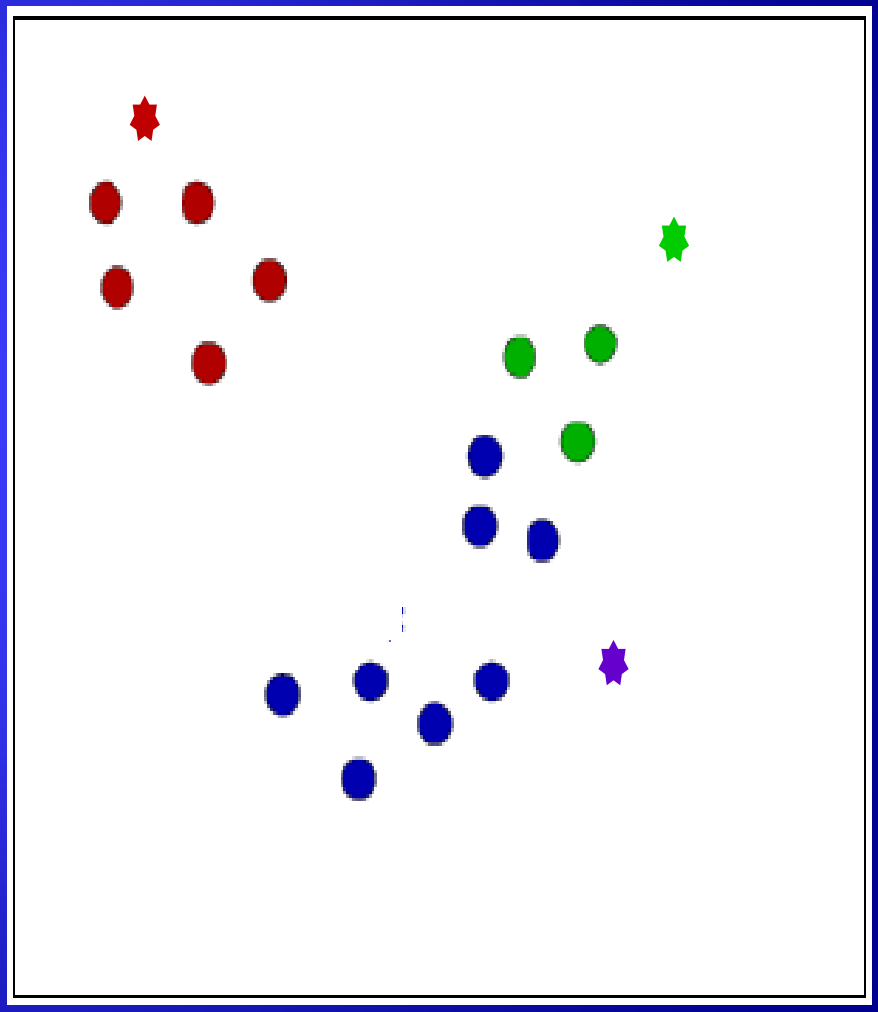
$$\min_{\{\mu_1,\cdots,\mu_k\}} \sum_{h=1}^{k} \sum_{x \in X_h} \|x - \mu_h\|^2$$



©K. Gibert

K-Means algorithm

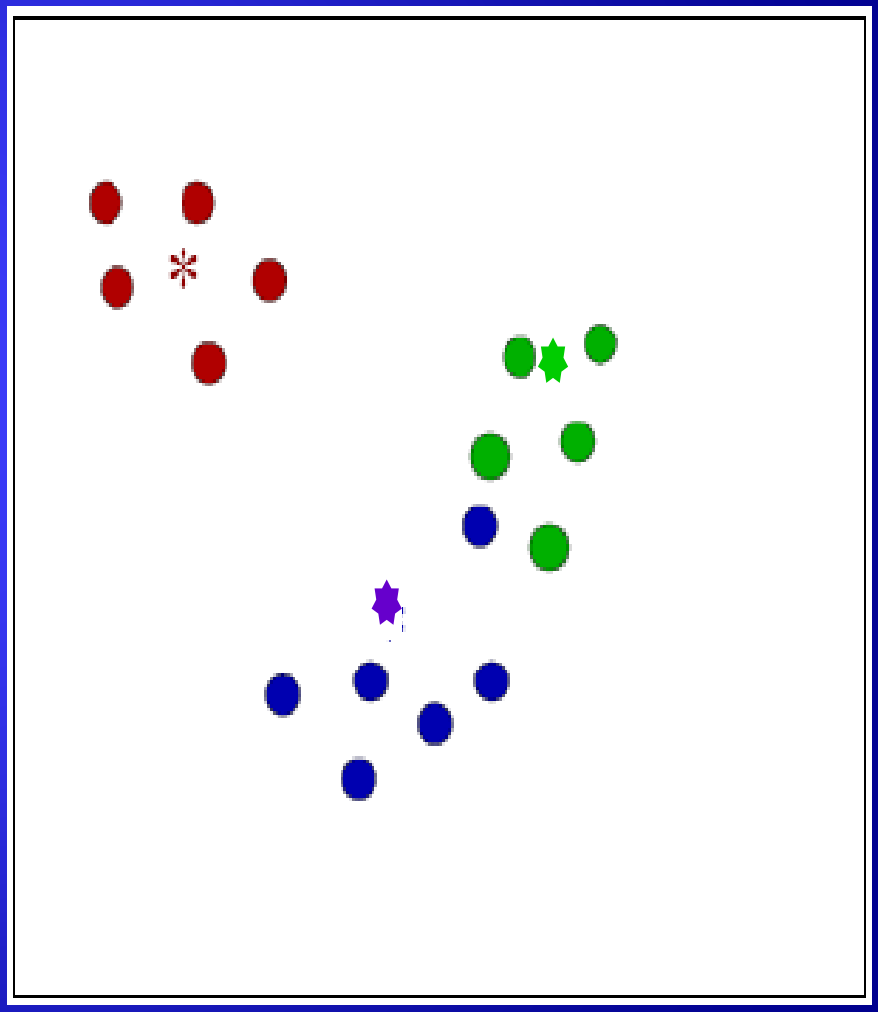
- Select k
- Select k seeds (random?...)
- Assign objects to seeds (class)
- Update seeds accordingly



©K. Gibert

K-Means algorithm

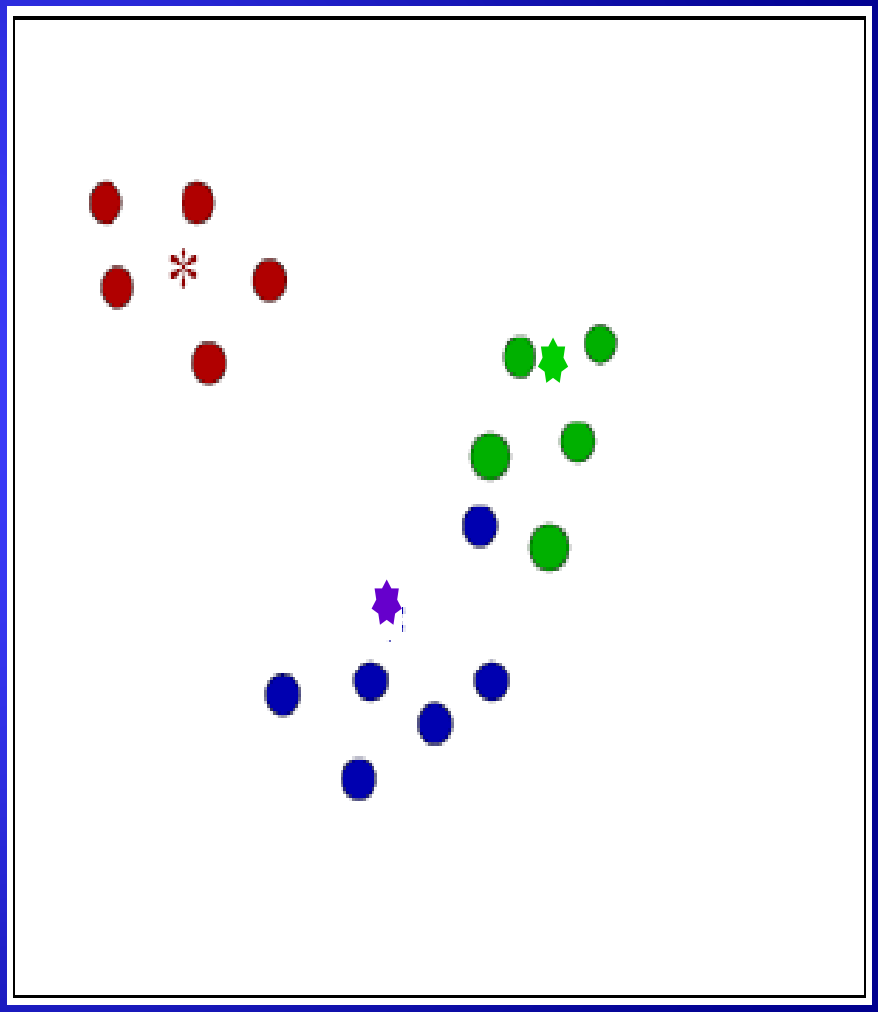
- Select k
- Select k seeds (random?...)
- Assign objects to seeds (class)
- Update seeds accordingly
- Iterate till no changes found



©K. Gibert

K-Means algorithm

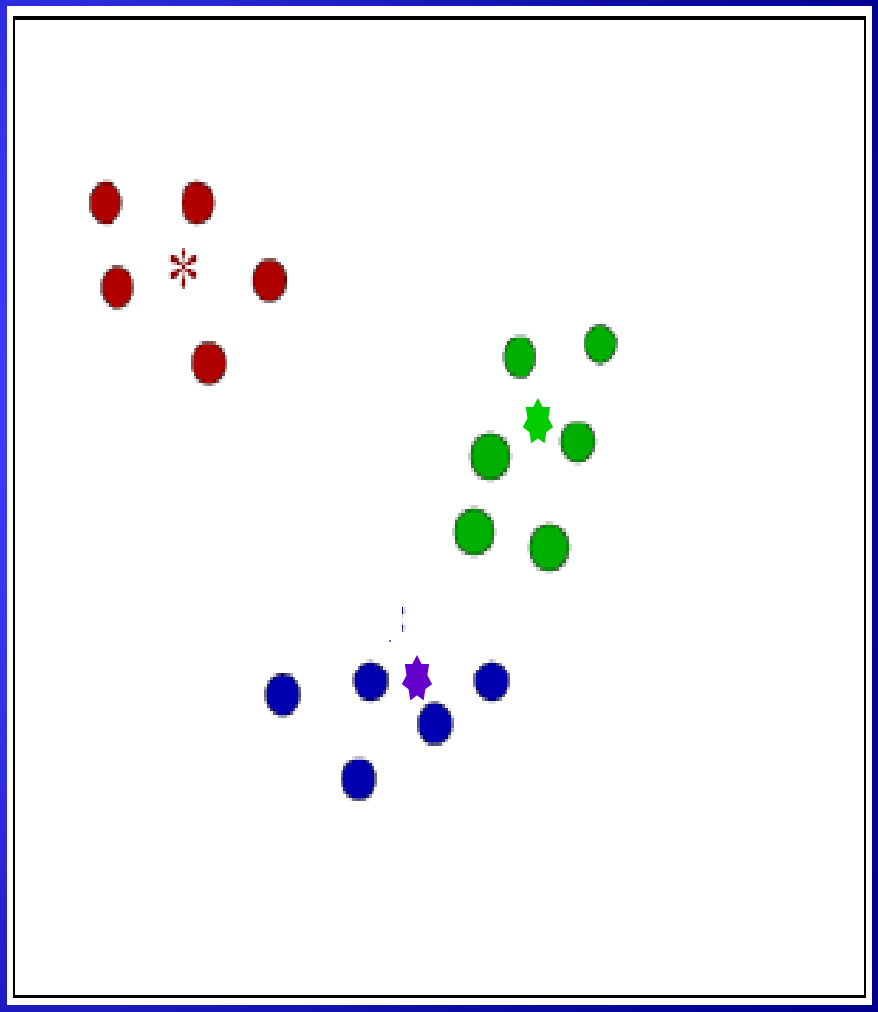
- Select k
- Select k seeds (random?...)
- Assign objects to seeds (class)
- Update seeds accordingly
- Iterate till no changes found



©K. Gibert

K-Means algorithm

- Select k
- Select k seeds (random?...)
- Assign objects to seeds (class)
- Update seeds accordingly
- Iterate till no changes found



©K. Gibert

K-Means algorithm

- Drawbacks

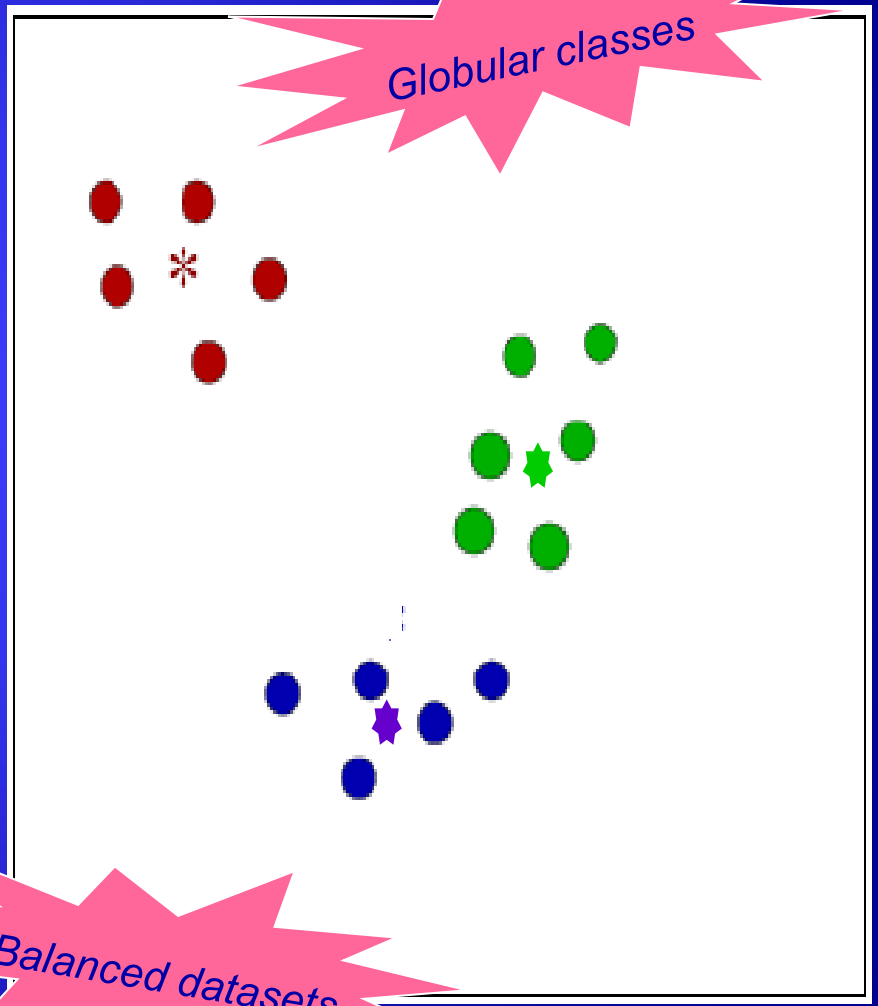
- Order depending
  - Local optima

- Sensitive to outliers

- SSE decreases with k
  no goodness indicator

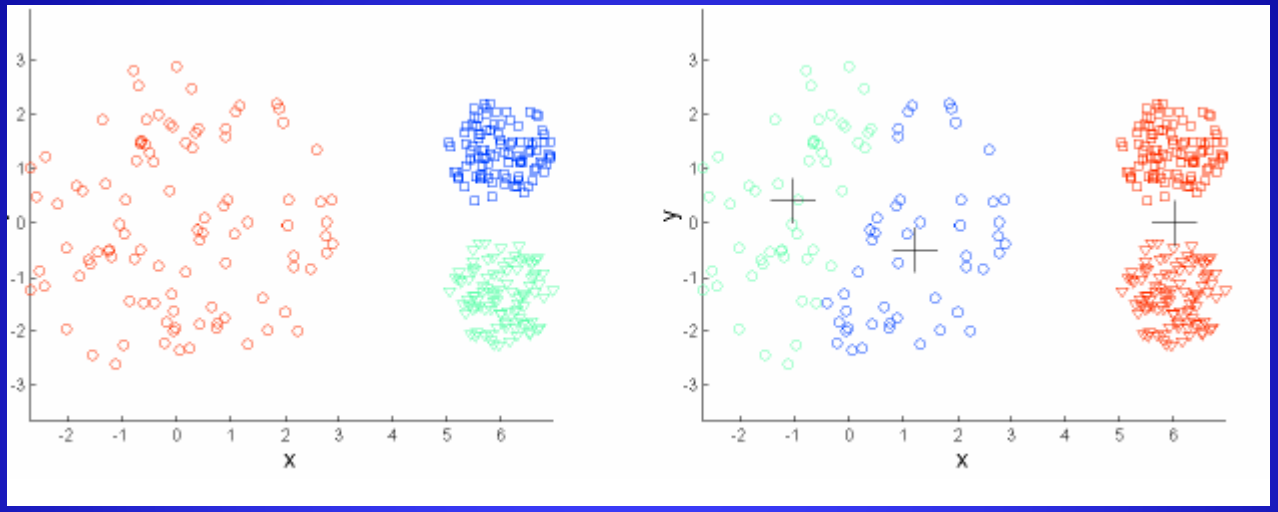
- Fast, but not scalable
  Fast K-Means (Triangular inequality)

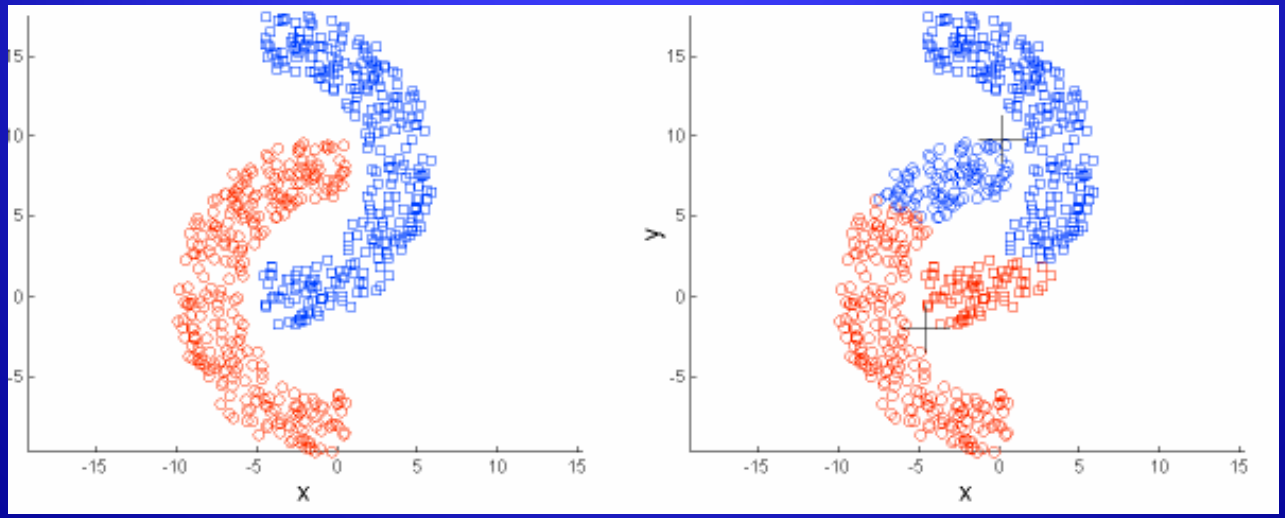
- K is an input
  (X-Means: kd-trees and
  Bayesian information criterion)



©K. Gibert

K-Means algorithm





©K. Gibert

Case Study: Fisher's Iris Flower

Iris setosa

iris.csv ***

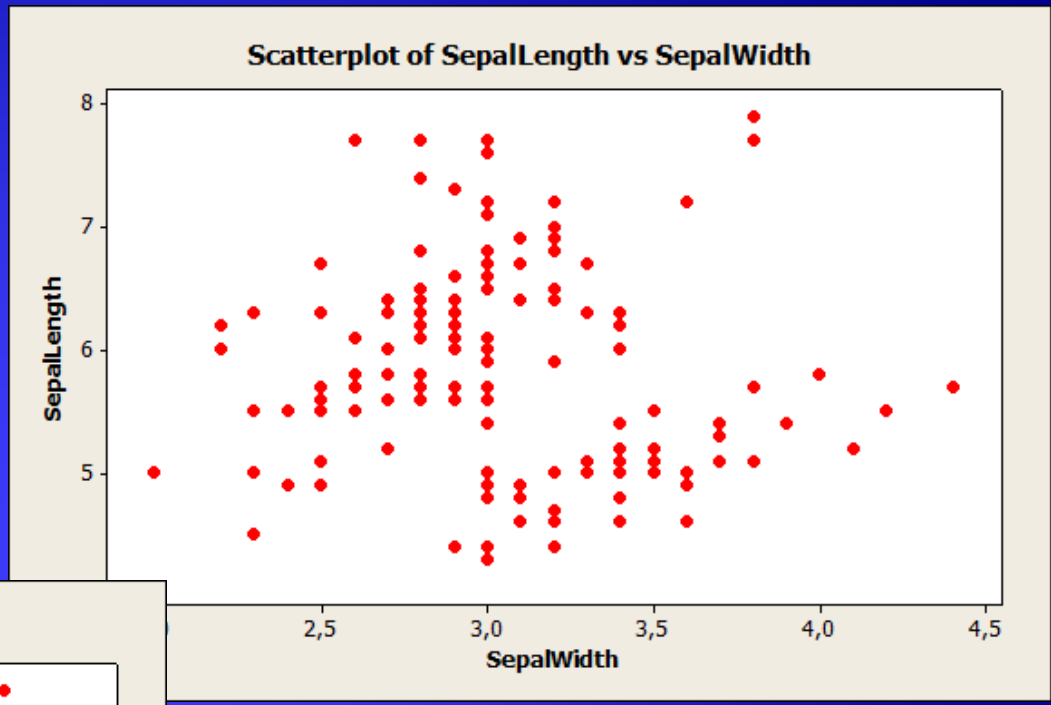
↓	C1	C2	C3	C4	C5-T
	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
1	5,1	3,5	1,4	0,2	setosa
2	4,9	3,0	1,4	0,2	setosa
3	4,7	3,2	1,3	0,2	setosa
4	4,6	3,1	1,5	0,2	setosa
5	5,0	3,6	1,4	0,2	setosa
6	5,4	3,9	1,7	0,4	setosa
7	4,6	3,4	1,4	0,3	setosa
8	5,0	3,4	1,5	0,2	setosa
9	4,4	2,9	1,4	0,2	setosa
10	4,9	3,1	1,5	0,1	setosa
11	5,4	3,7	1,5	0,2	setosa
12	4,8	3,4	1,6	0,2	setosa
13	4,8	3,0	1,4	0,1	setosa
14	4,3	3,0	1,1	0,1	setosa
15	5,8	4,0	1,2	0,2	setosa
16	5,7	4,4	1,5	0,4	setosa
17	5,4			0,4	setosa

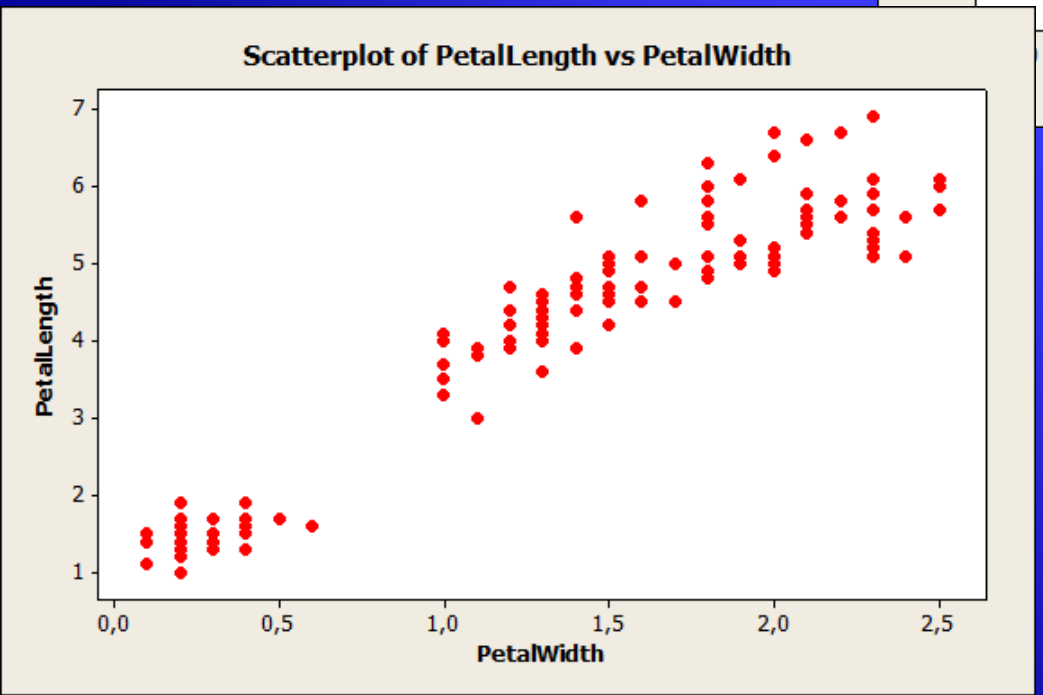
Iris versicolor

Fisher 1936
Anderson measured
150 flowers

Iris virginica

Case Study: Fisher's Iris Flower



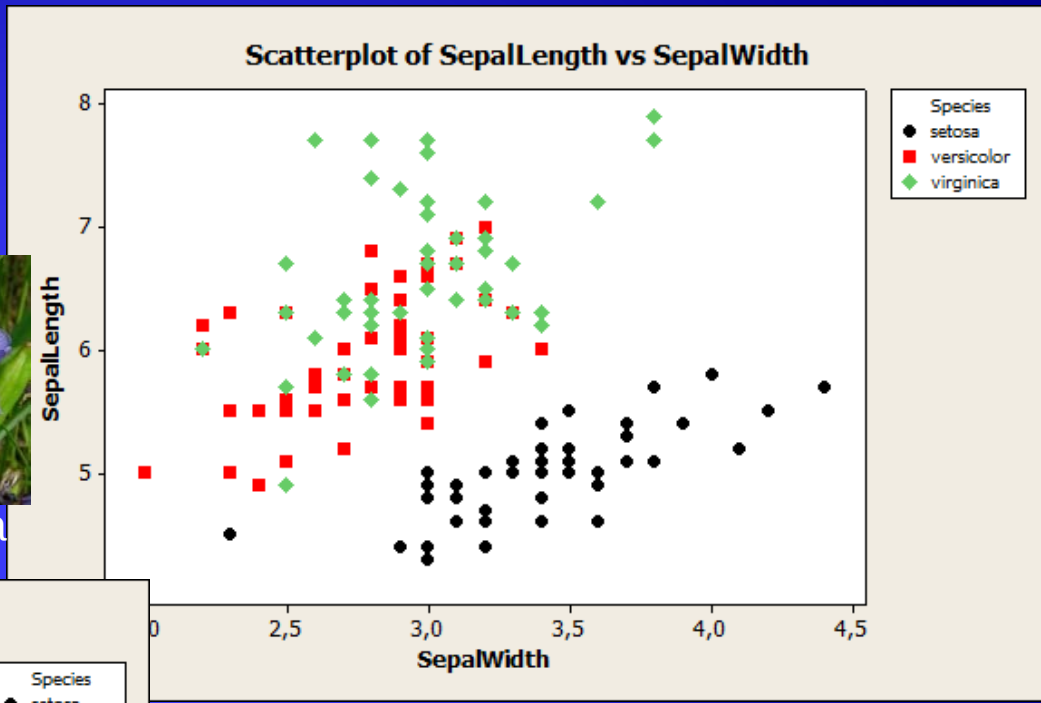


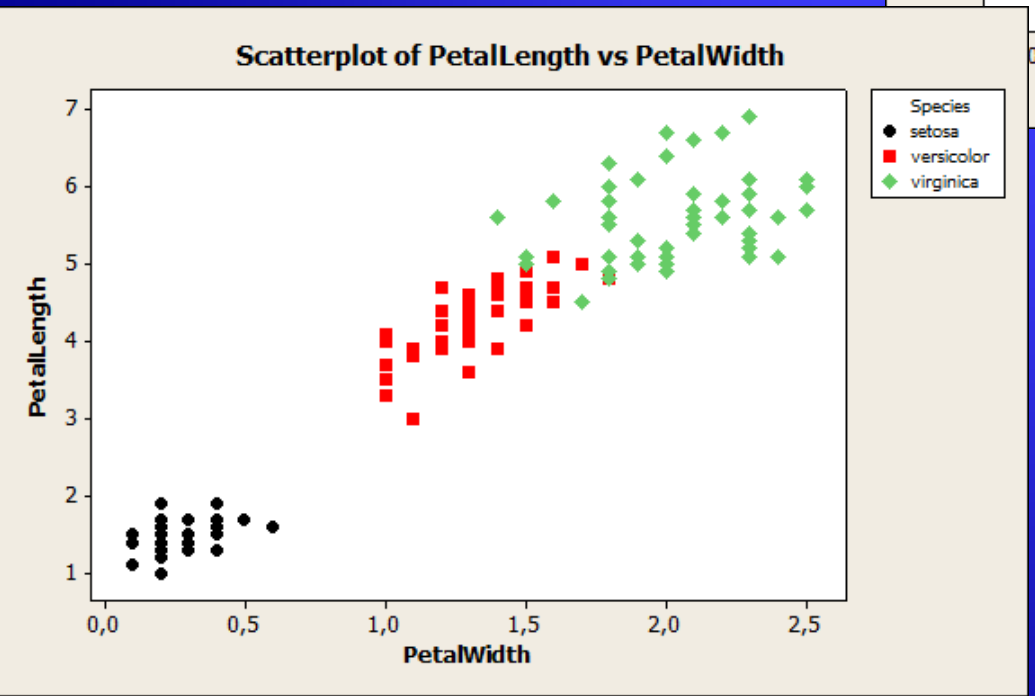
Case Study: Fisher's Iris Flower

Iris setosa

Iris versicolor

Iris virginica





Feature Weighting

PL: 9,08
PW: 9,7
SL: 0.3
SW: 0.05

Case Study: Fisher's Iris Flower


Iris setosa


Iris versicolor


Iris virginica