Expectation Maximization
Chapter 5
k-means

T.L. Grobler

Autor: 01.04.22

Supervised versus Unsupervised

- Supervised Learning:
  - Chapter 4
  - Class label comes with observations
- Unsupervised Learning:
  - Chapter 5
  - Class label inferred from observations

General Expectation Maximization

- Expectation:
  - Chapter 5
  - Current estimate of parameters
  - Estimate class labels
- Maximization (MLE):
  - Chapter 4
  - Current estimate of class labels
  - Estimate parameters
- Alternate E and M step till convergence

Preamble

- Assume we have N-observations
- They belong to one of k-clusters
- The value of k is known

Intuitive Description

- Step 1: Choose initial cluster means
  - Choose *k* random observations
- Step 2: Assign class labels
  - Assign observations to cluster mean closest to it
- Step 3: Update cluster means
  - Compute cluster means using assigned observations
- Repeat steps until convergence

Objective Function

Label

$$J = \sum_{n=1}^{N} \sum_{j=1}^{k} t_{nj} \|\mathbf{x}_n - \mathbf{u}_j\|^2$$

1-of-*k* coding

Obs

Cluster Mean (initialized)

$$\mathbf{t}_n = [t_{n1} \ \cdots \ t_{nk}]^T$$

$$t_{nj} = \begin{cases} 1 & \text{if } \mathbf{x}_n \in C_j \\ 0 & \text{otherwise} \end{cases}$$

Min *J*:
alternate between
fixing and solving for
$\mathbf{t}_n$ and $\mathbf{u}_j$

Alternating Estimation Steps

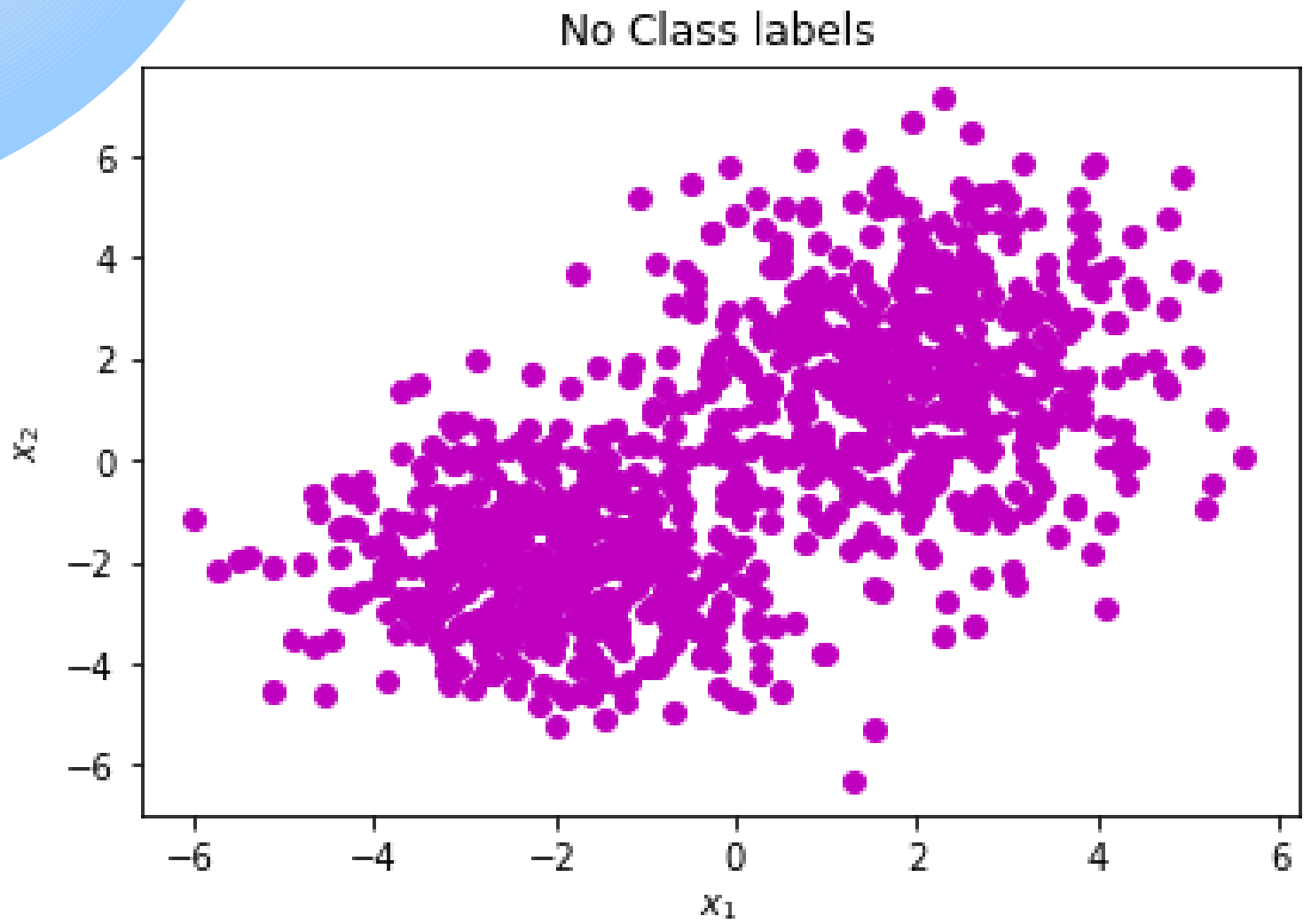
- Determine t_{nj}:
  - $\mathbf{x}_n$ assigned to nearest cluster
  - i.e. $\|\mathbf{x}_n - \mathbf{u}_j\|^2$ smallest
- Determine $\mathbf{u}_j$:
  - set to mean of newly formed clusters

$$\frac{\partial J}{\partial \mathbf{u}_j} = 0 \implies \mathbf{u}_j = \frac{\sum_{n=1}^{N} t_{nj} \mathbf{x}_n}{\sum_{n=1}^{N} t_{nj}}$$

Expectation Maximization

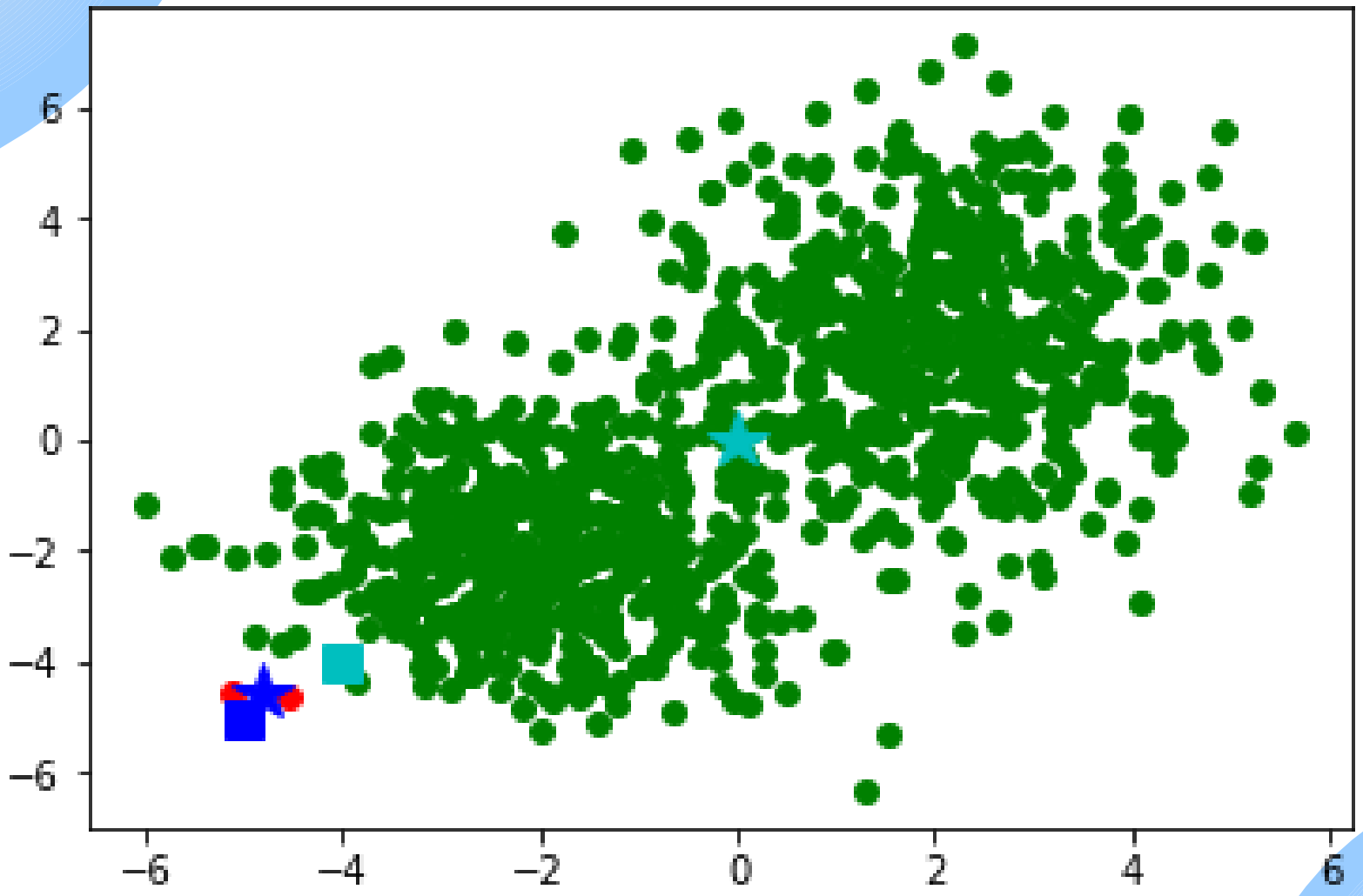
- Expectation Step:
  - class labels hidden
  - $\mathbf{t}_n$ are latent variables
  - given class means
  - calc expected values of $\mathbf{t}_n$
- Maximization Step:
  - use $\mathbf{t}_n$
  - compute ~ likelihood
  - maximize likelihoood
  - determine $\mathbf{u}_j$

Example



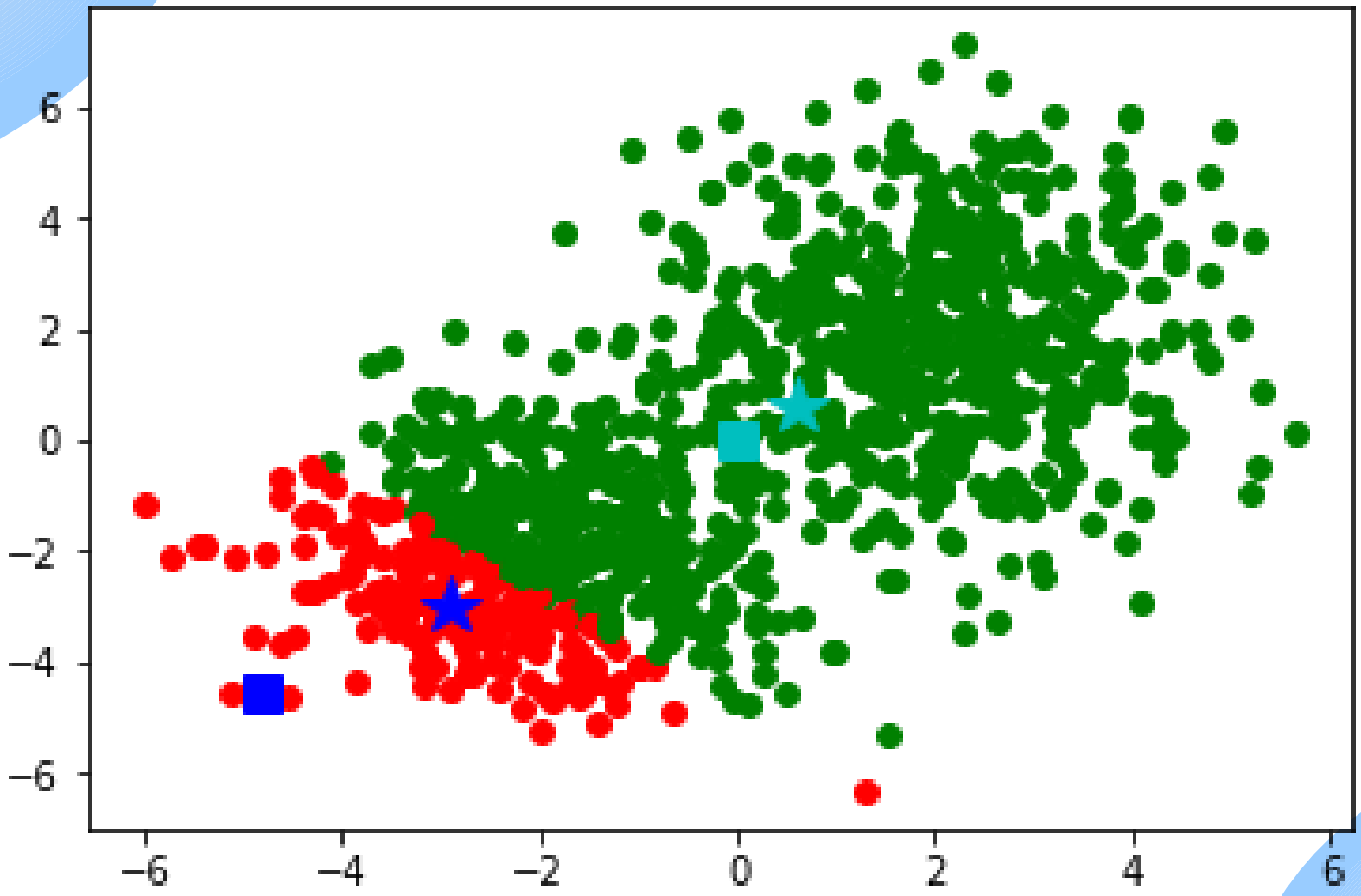
Autor: 01.04.22

1-iteration



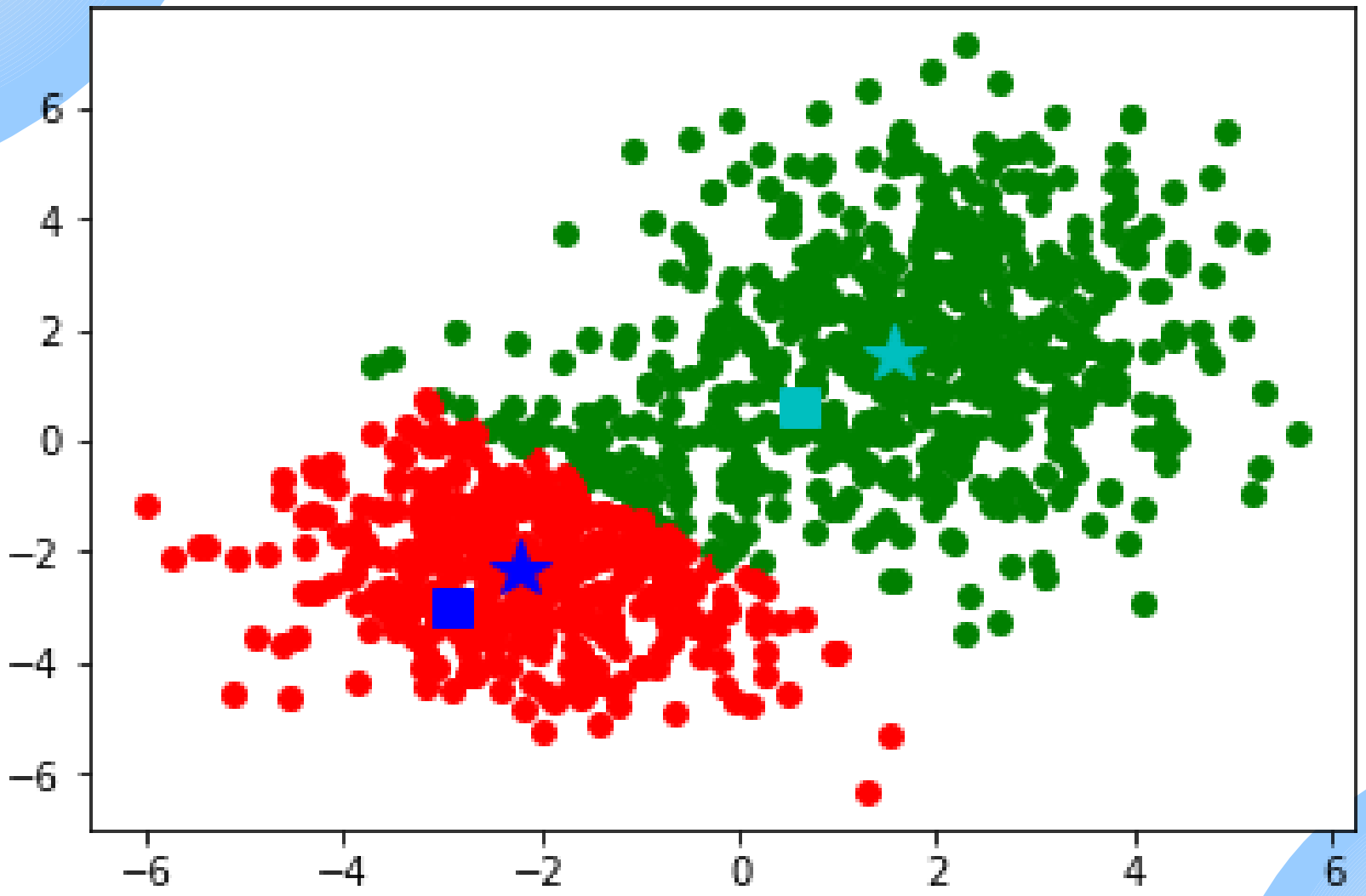
Autor: 01.04.22

2-iteration



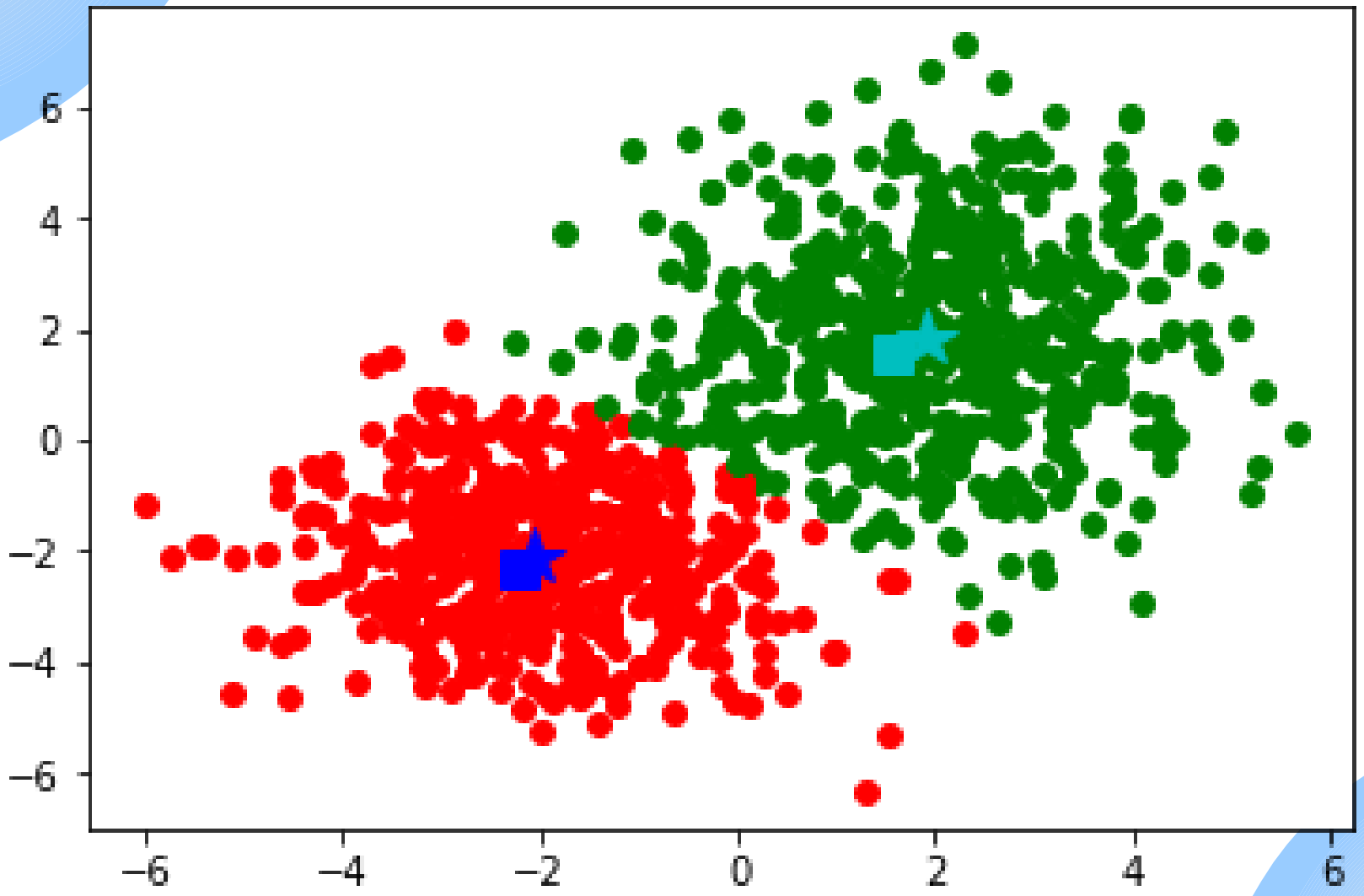
Autor: 01.04.22

3-iteration



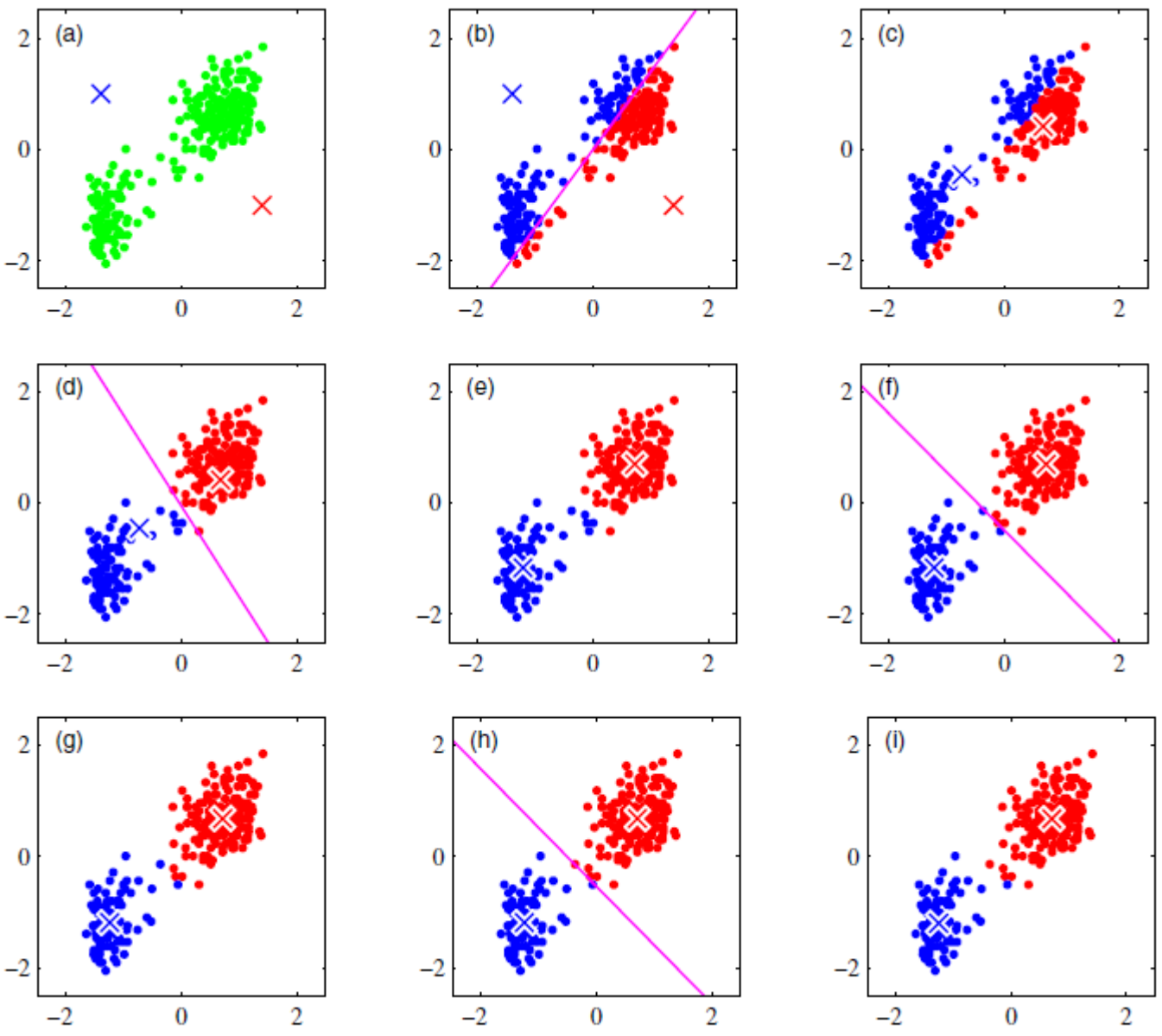
Autor: 01.04.22

4-iteration



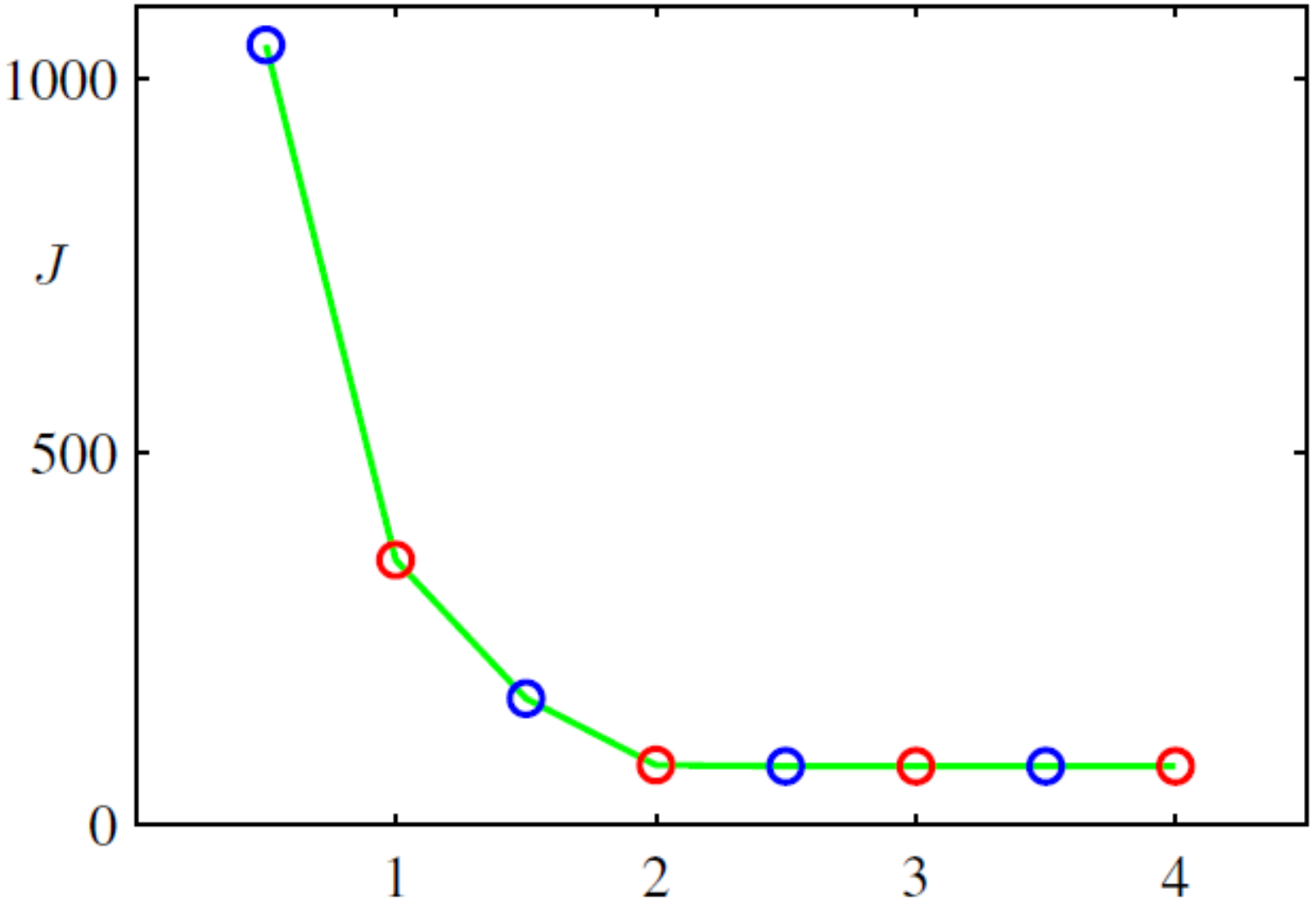
Autor: 01.04.22

Another Example



Autor: 01.04

Objective Function Plot



Autor: 01.04.22

Normalization

- Whitening data
- Normalized data
- Warning
  - Rescales distances
  - Indicative of between cluster seperation

Hard versus Soft

- Hard
  - Ambiguity ignored
  - Current strategy
- Soft
  - Reflects uncertainty

Local Minima

- *J* allows multiple locl minima
  - Initialization critical
  - Find good local minima

Binary Split

- Two random cluster means
- *k*-means (two cluster means)
- Compute scatter of clusters
- Two cluster means from cluster with largest scatter
- *k*-means (three cluster means)
- Continue until clusters match desired number