

Machine Learning



Amin Golzari Oskouei

a.golzari@azaruniv.ac.ir

a.golzari@tabrizu.ac.ir

<https://github.com/Amin-Golzari-Oskouei>

Azərbaycan Şahid Mədani Universiteti
2023

Unsupervised Learning : Clustering



Table of Contents

2

- Unsupervised Learning
- Clustering
- K-Means Algorithm
- Improved clustering algorithm
- Two-part generator algorithm
- Hierarchical clustering

Reminder: Supervised learning

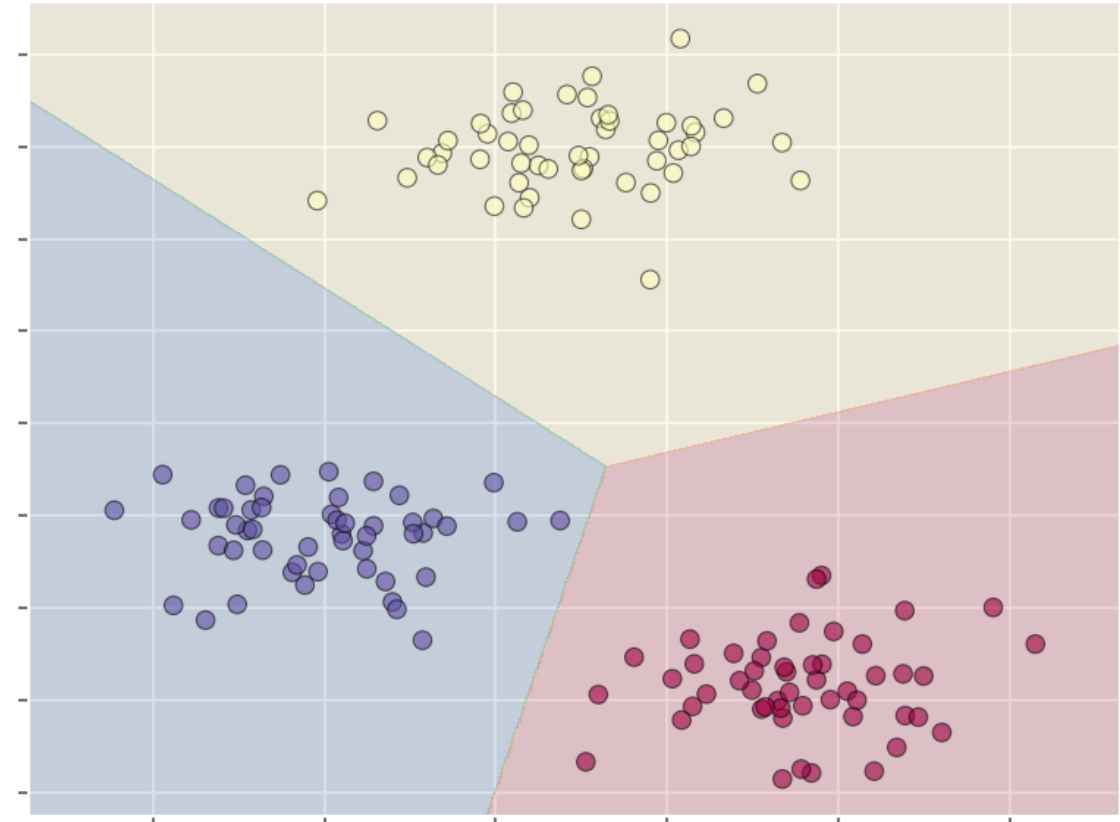
3

- Supervised Learning. For each example, the correct answer is given.

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

↑
Training set

- Types Of Supervised Learning.
 - Regression: Estimation of a continuous quantity.
 - Classification: Estimation of a discrete quantity.



Unsupervised Learning

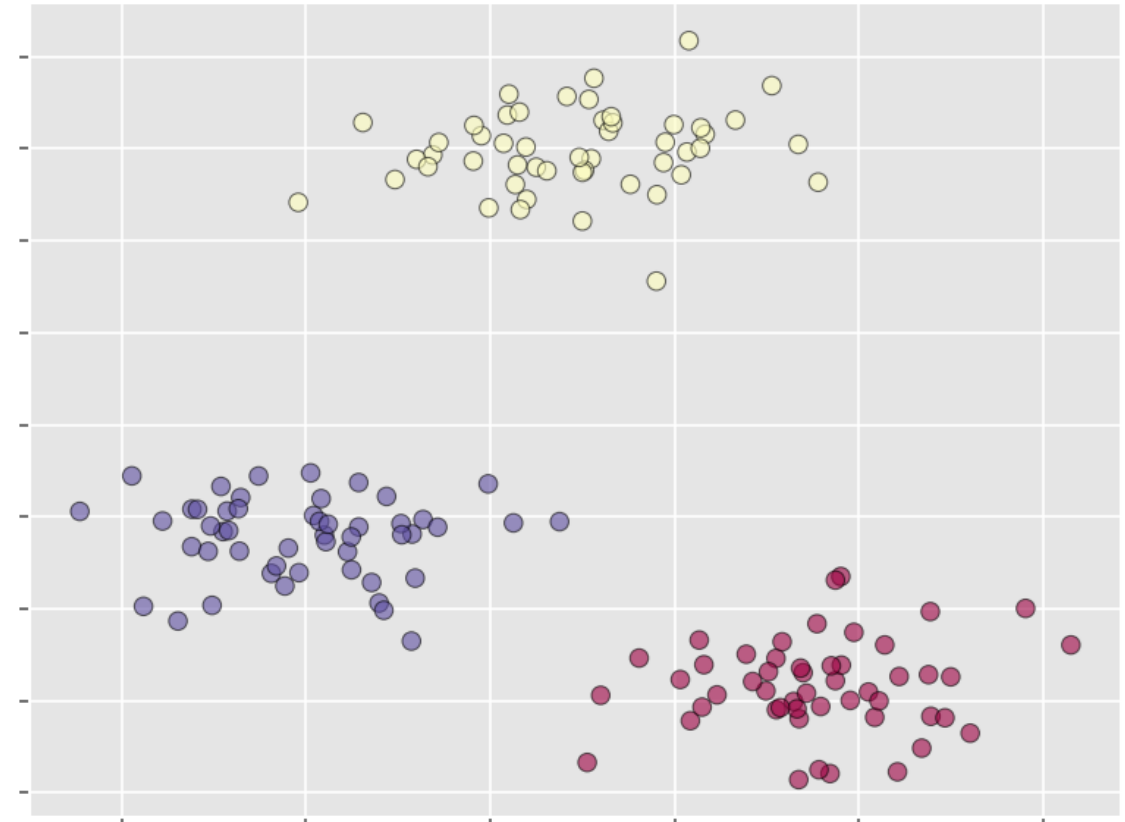
4

- Unsupervised Learning. Not knowing the correct answers.

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$$

↑
Training set

- Goal. Structure detection in input data



Applications of clustering: grouping related news

5

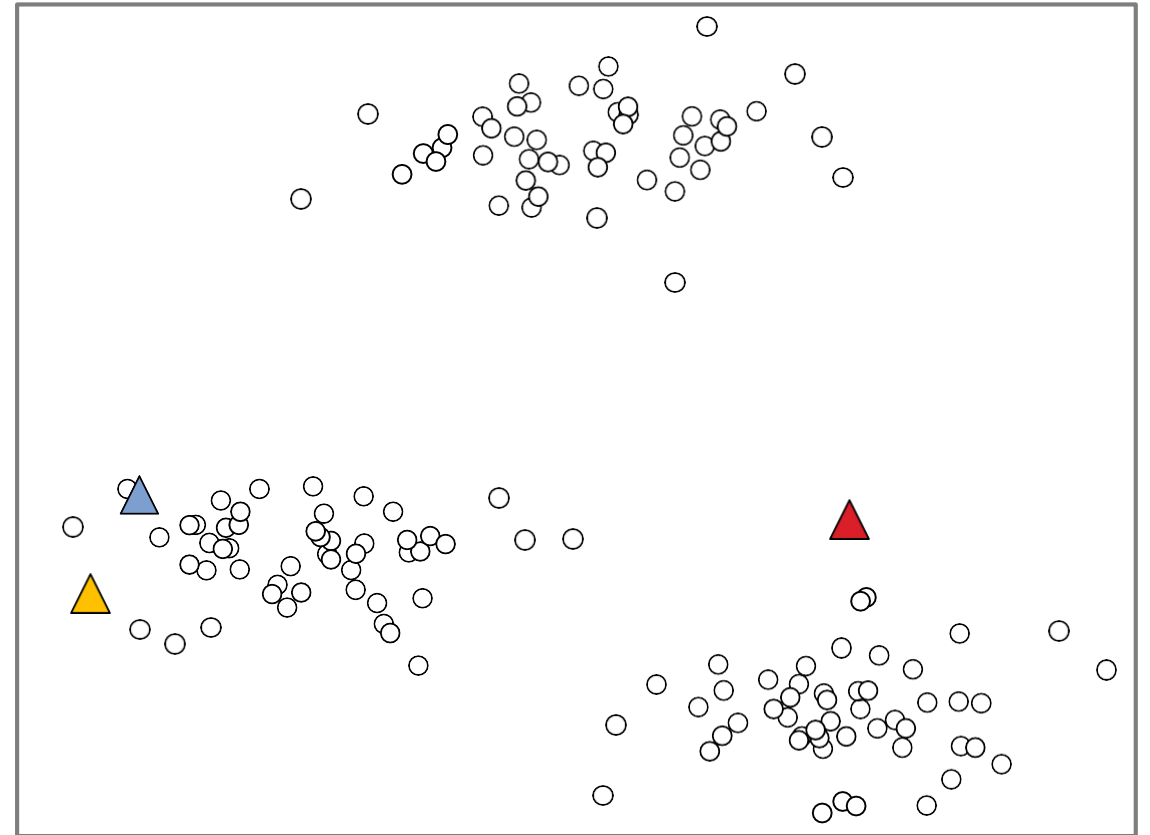
The image displays four screenshots from news websites, illustrating the application of clustering to group related news. Red arrows indicate the flow of information:

- Top Left:** A screenshot of Google News showing search results for "BP Oil Well, Site of National Catastrophe, Dies at One". A red box highlights this result, with an arrow pointing to the CNN article.
- Top Right:** A screenshot of The Wall Street Journal's "THE SOURCE" section, showing an article titled "BP Kills Macondo, But Its Legacy Lives On". A red arrow points from this article to the Guardian article.
- Bottom Left:** A screenshot of the CNN website showing the full article "Allen: Well is dead, but much Gulf Coast work remains". A red arrow points from the Google News result to this article.
- Bottom Right:** A screenshot of the Guardian website showing the article "BP oil spill cost hits nearly \$10bn". A red arrow points from the WSJ article to this one.

K-means Clustering Algorithm

6

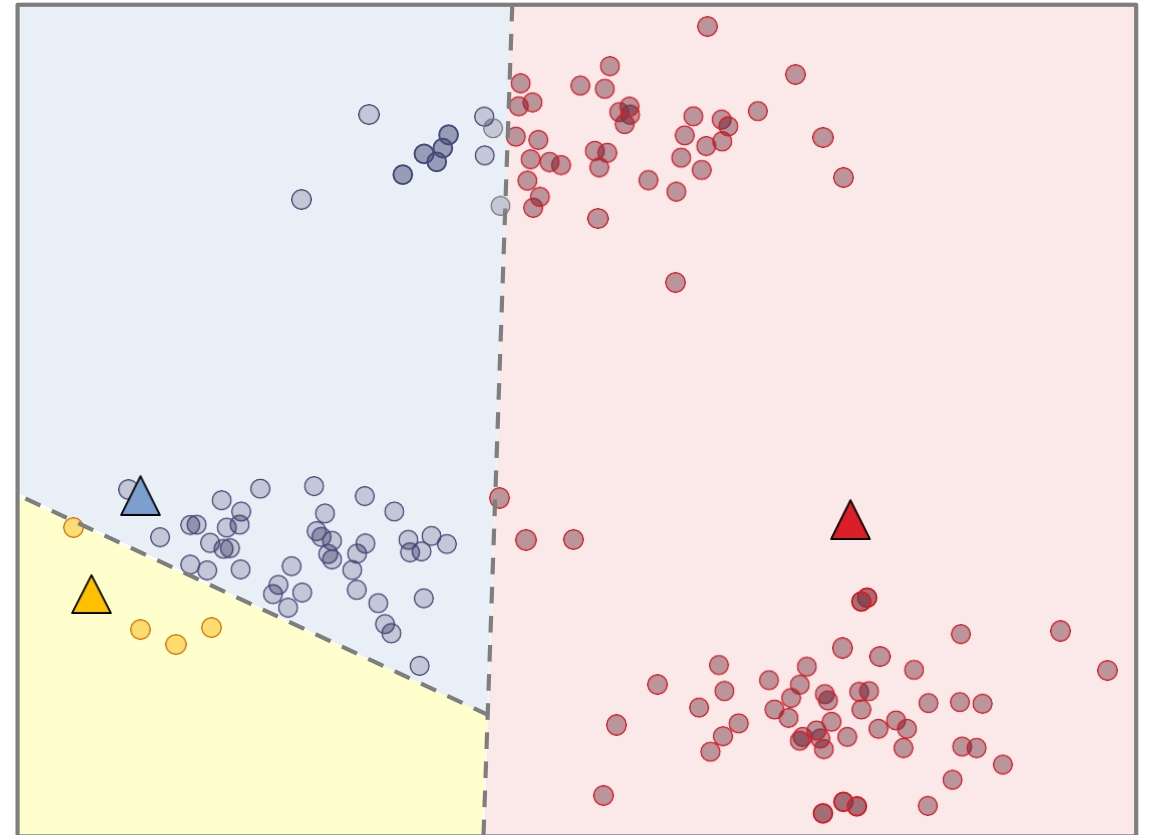
- Its an iterative clustering algorithm
 - It choosing **K** points randomly as the center of the clusters.
 - Repeat the following steps:
 - Assign each data to a cluster with the closest center.
 - Update the centroid of each cluster by averaging the data assigned to that cluster
 - Stop time : When no data changes its cluster in an iteration.



K-means Clustering Algorithm

7

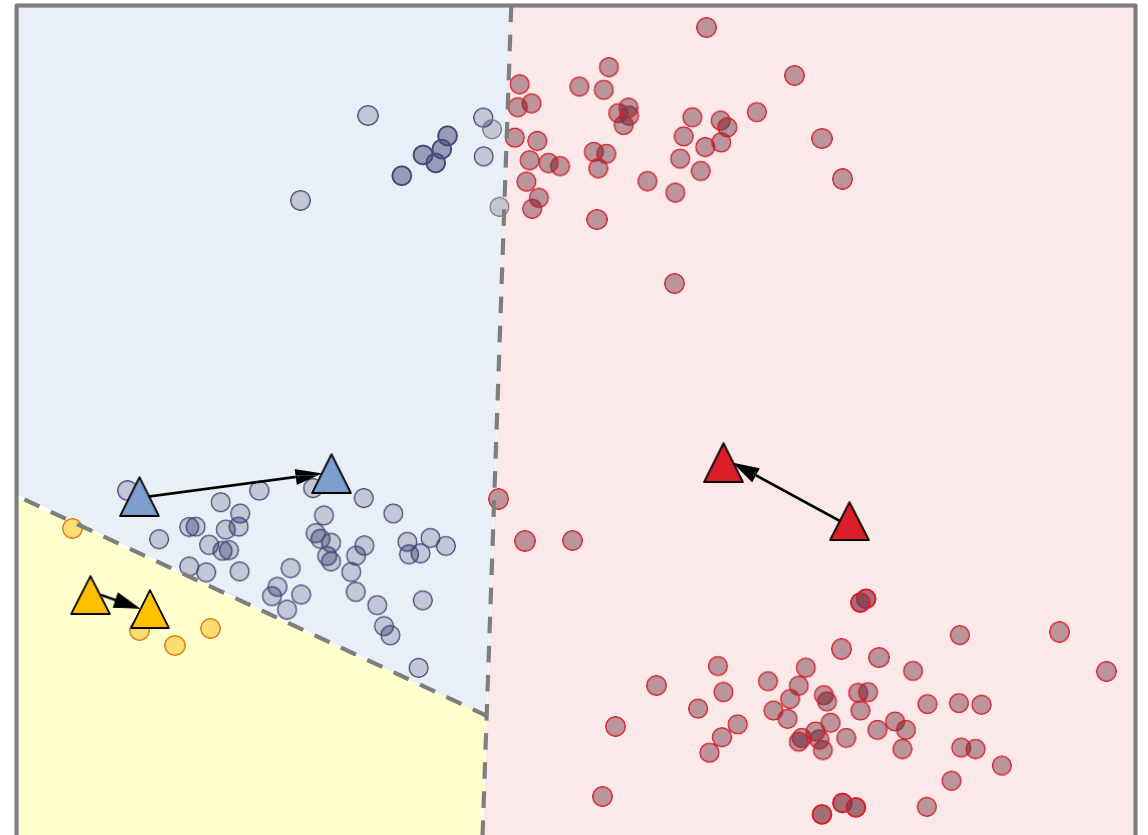
- Its an iterative clustering algorithm
 - It choosing **K** points randomly as the center of the clusters.
 - Repeat the following steps:
 - Assign each data to a cluster with the closest center.
 - Update the centroid of each cluster by averaging the data assigned to that cluster
 - Stop time : When no data changes its cluster in an iteration.



K-means Clustering Algorithm

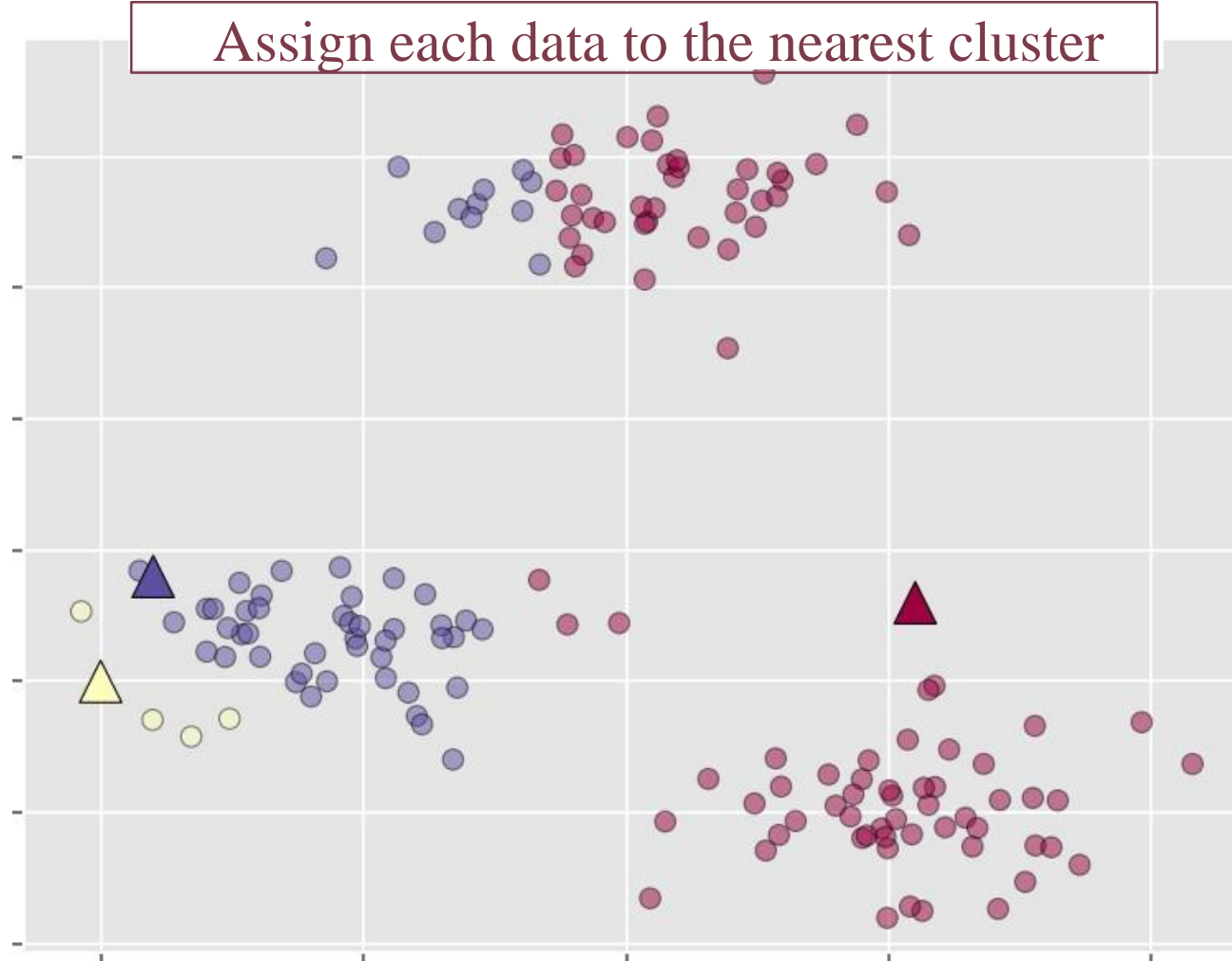
8

- Its an iterative clustering algorithm
 - It choosing **K** points randomly as the center of the clusters.
 - Repeat the following steps:
 - Assign each data to a cluster with the closest center.
 - Update the centroid of each cluster by averaging the data assigned to that cluster
 - Stop time : When no data changes its cluster in an iteration.



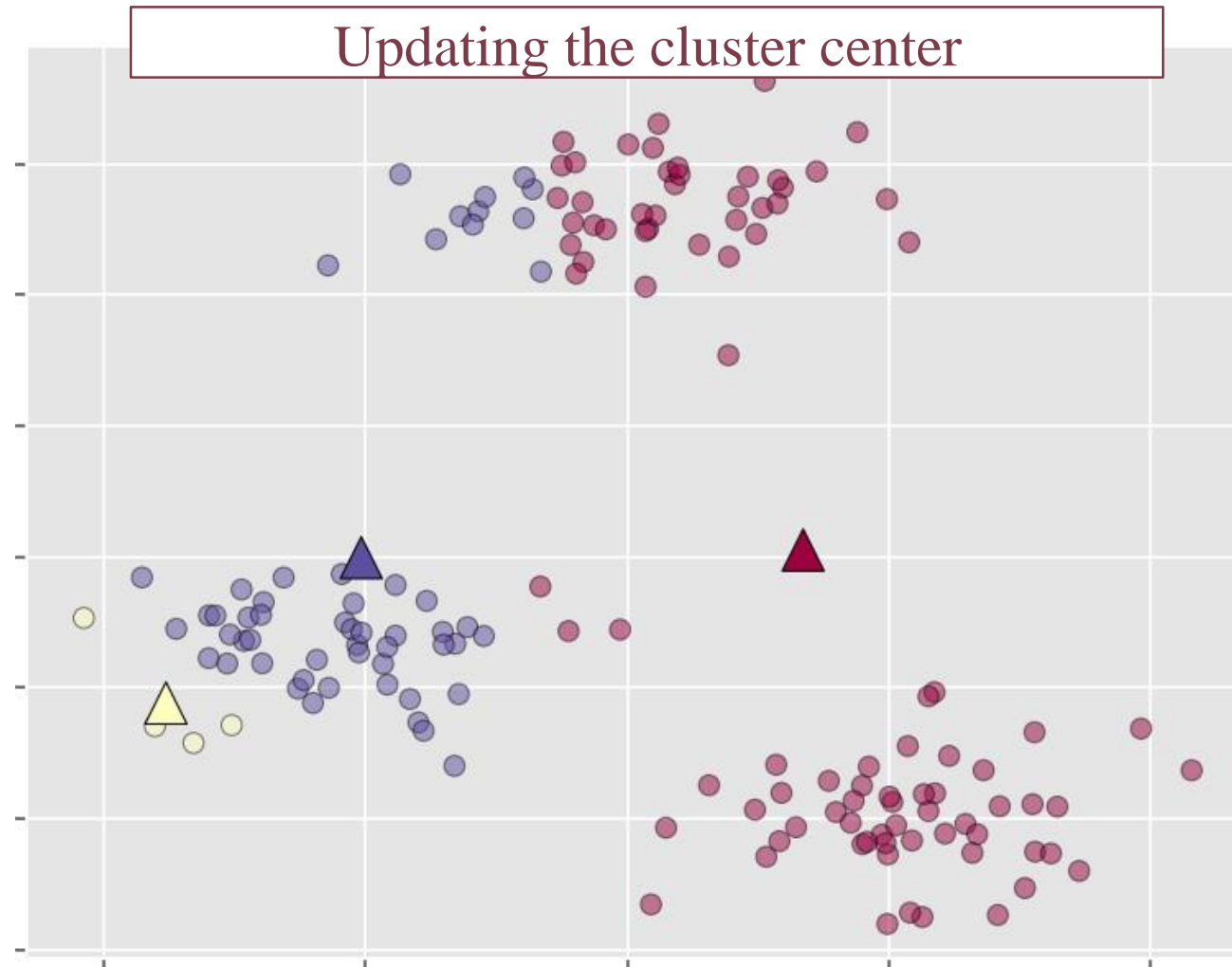
Illustrations of algorithm process

9



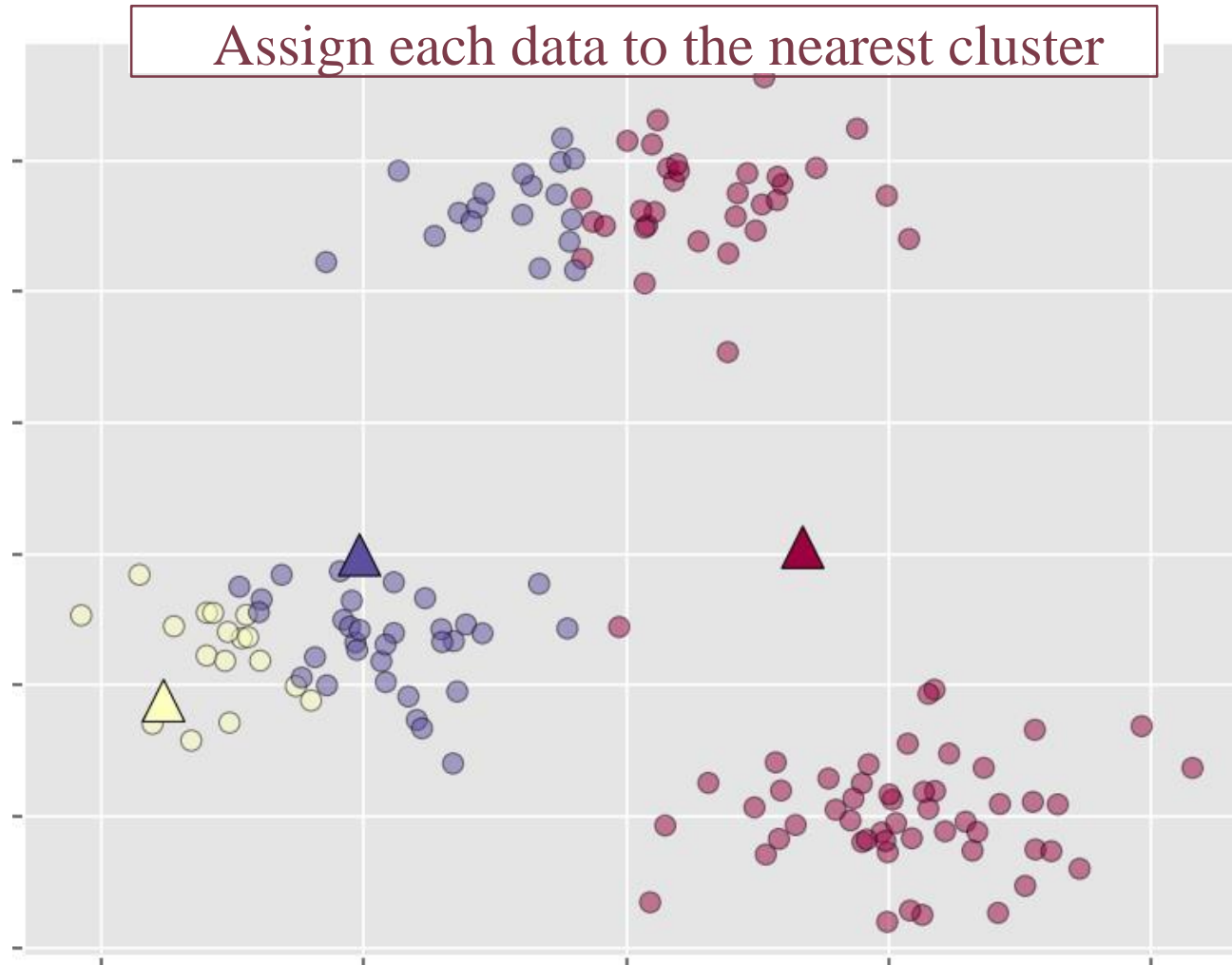
Illustrations of algorithm process

10



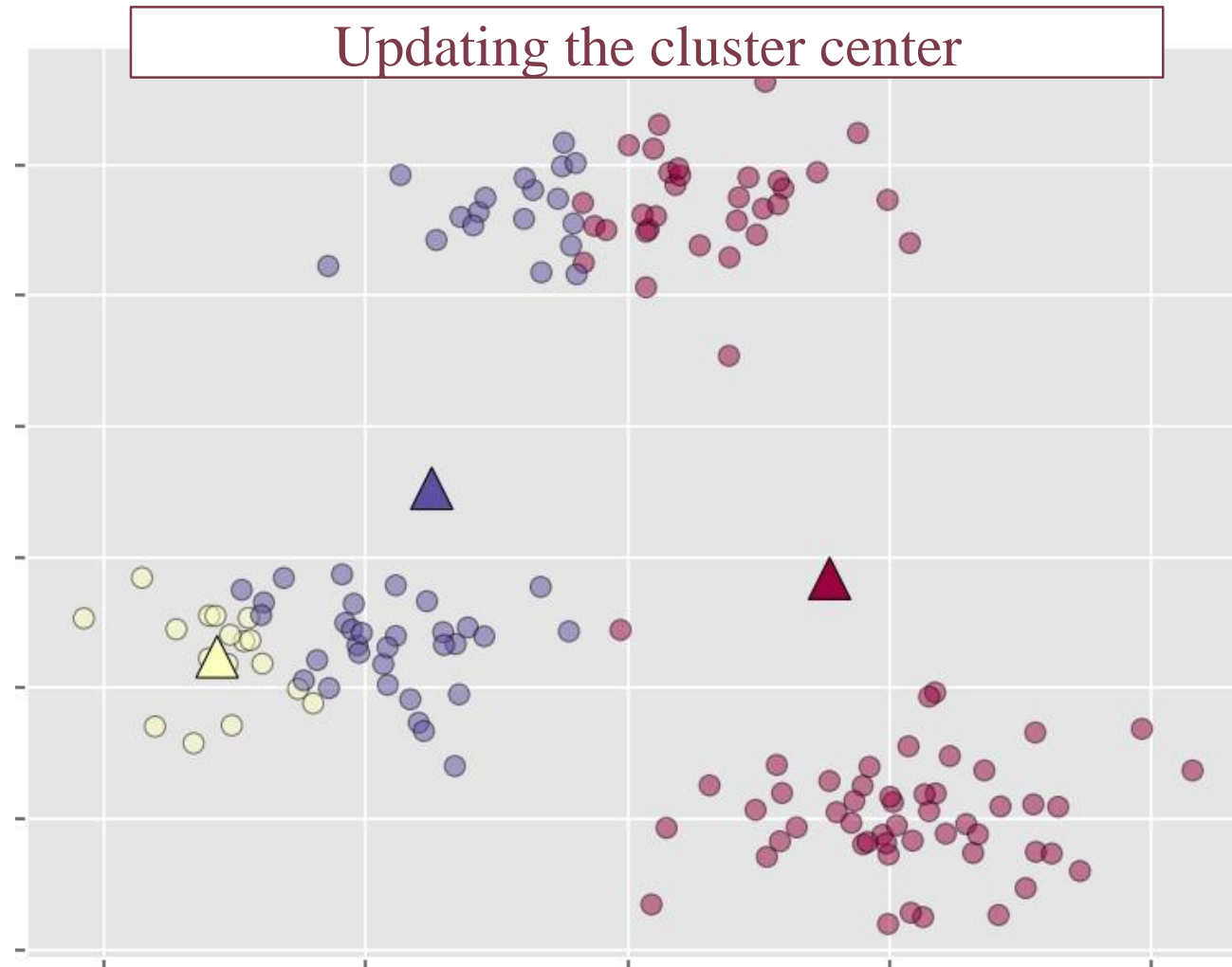
Illustrations of algorithm process

11



Illustrations of algorithm process

12



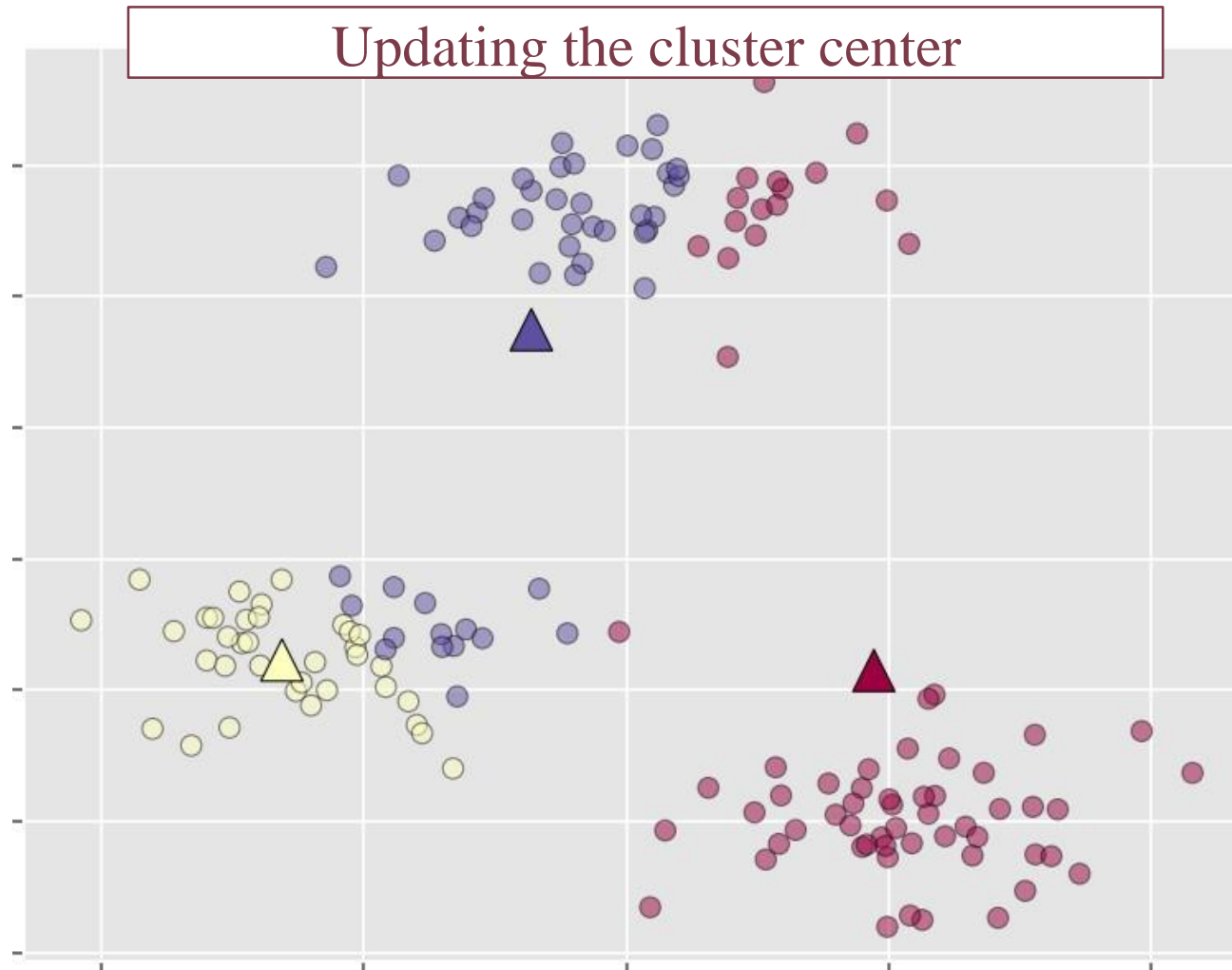
Illustrations of algorithm process

13



Illustrations of algorithm process

14



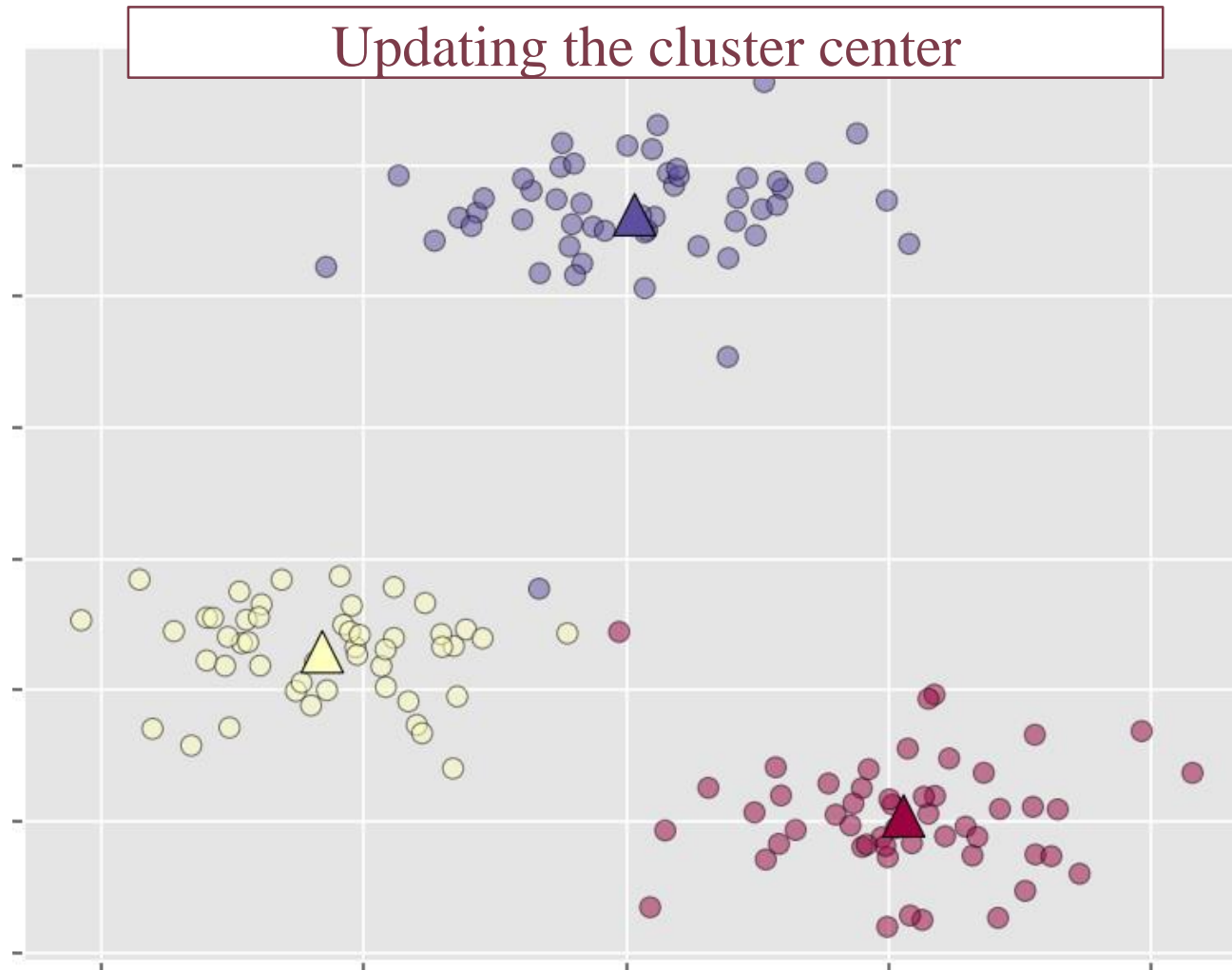
Illustrations of algorithm process

15



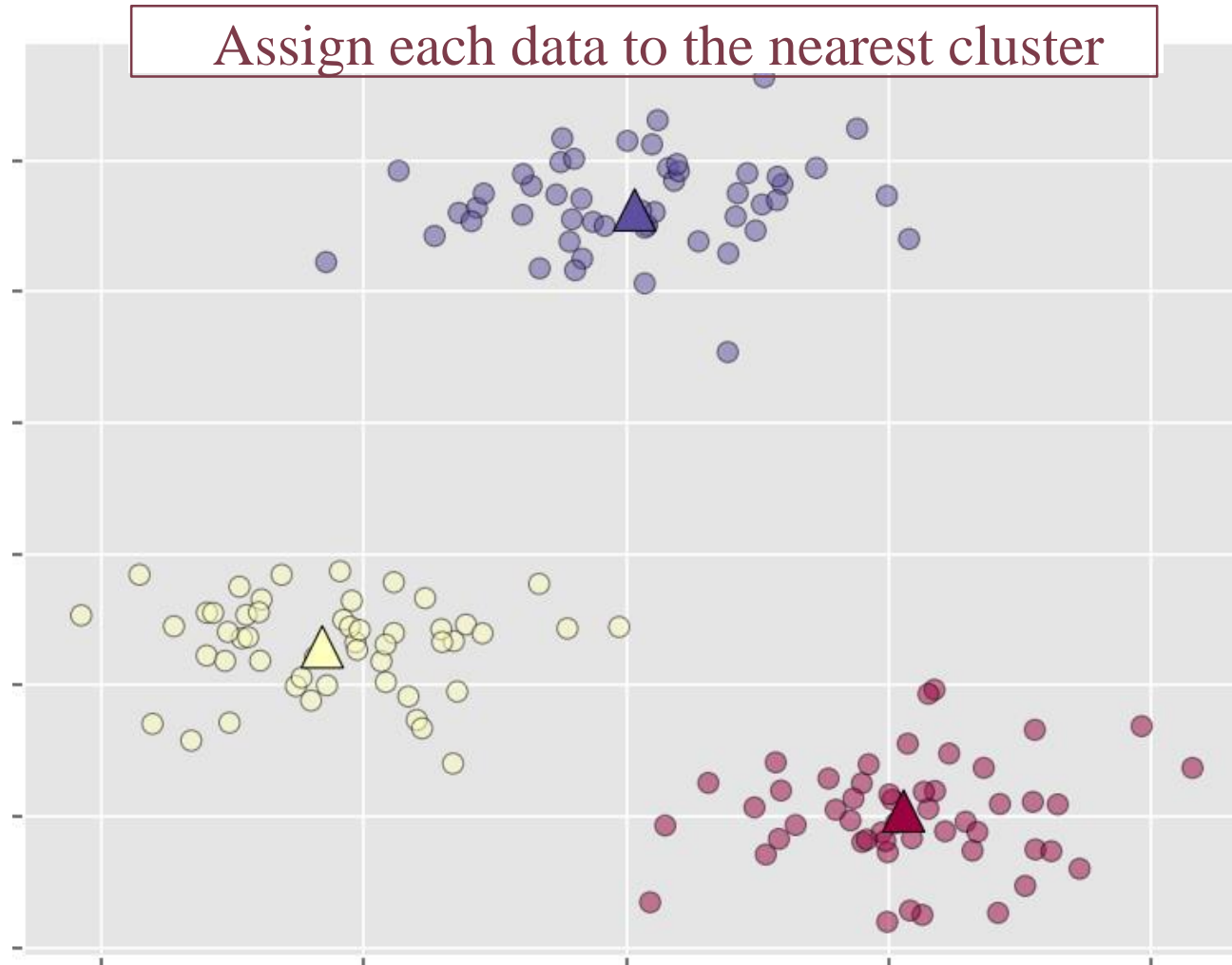
Illustrations of algorithm process

16



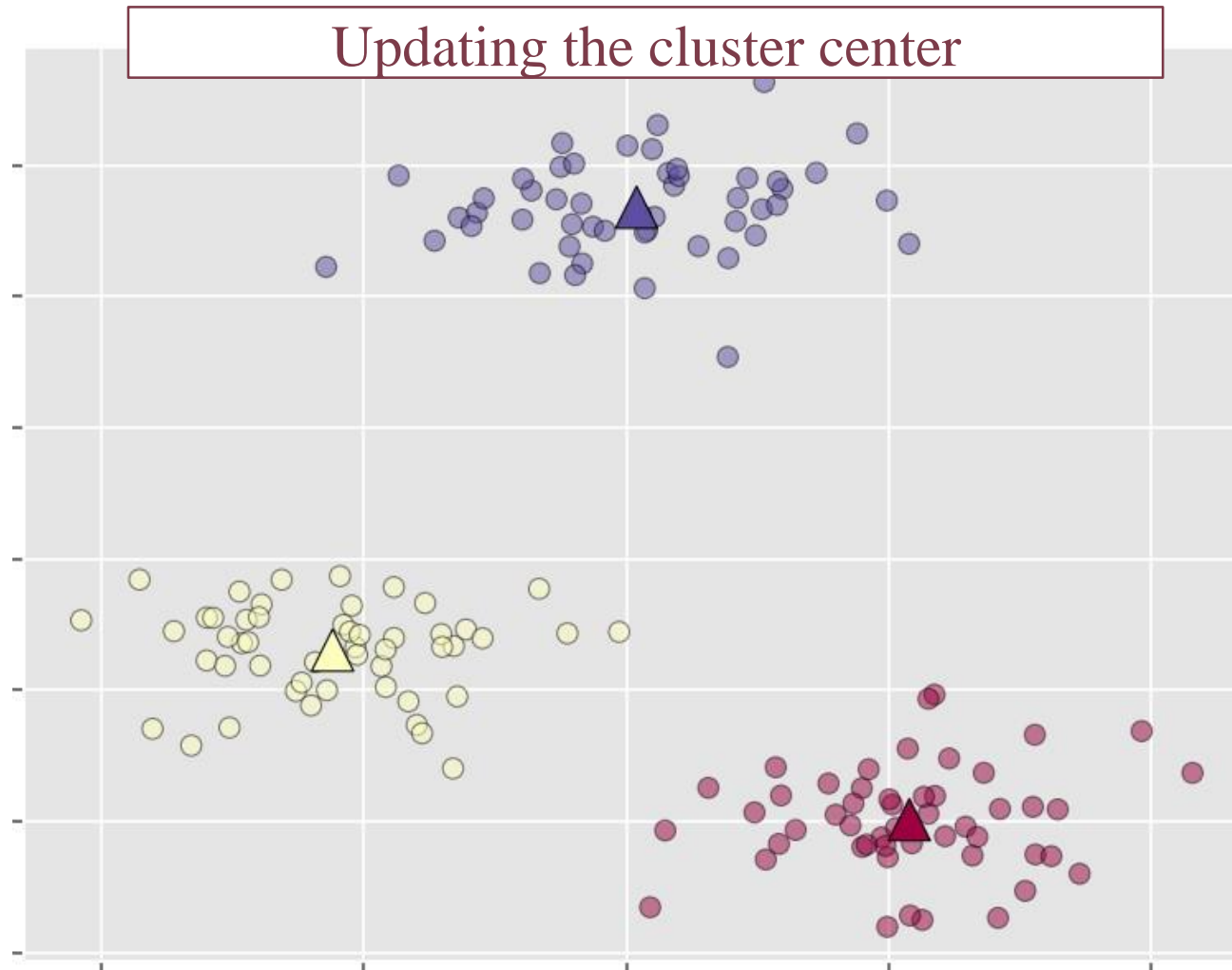
Illustrations of algorithm process

17



Illustrations of algorithm process

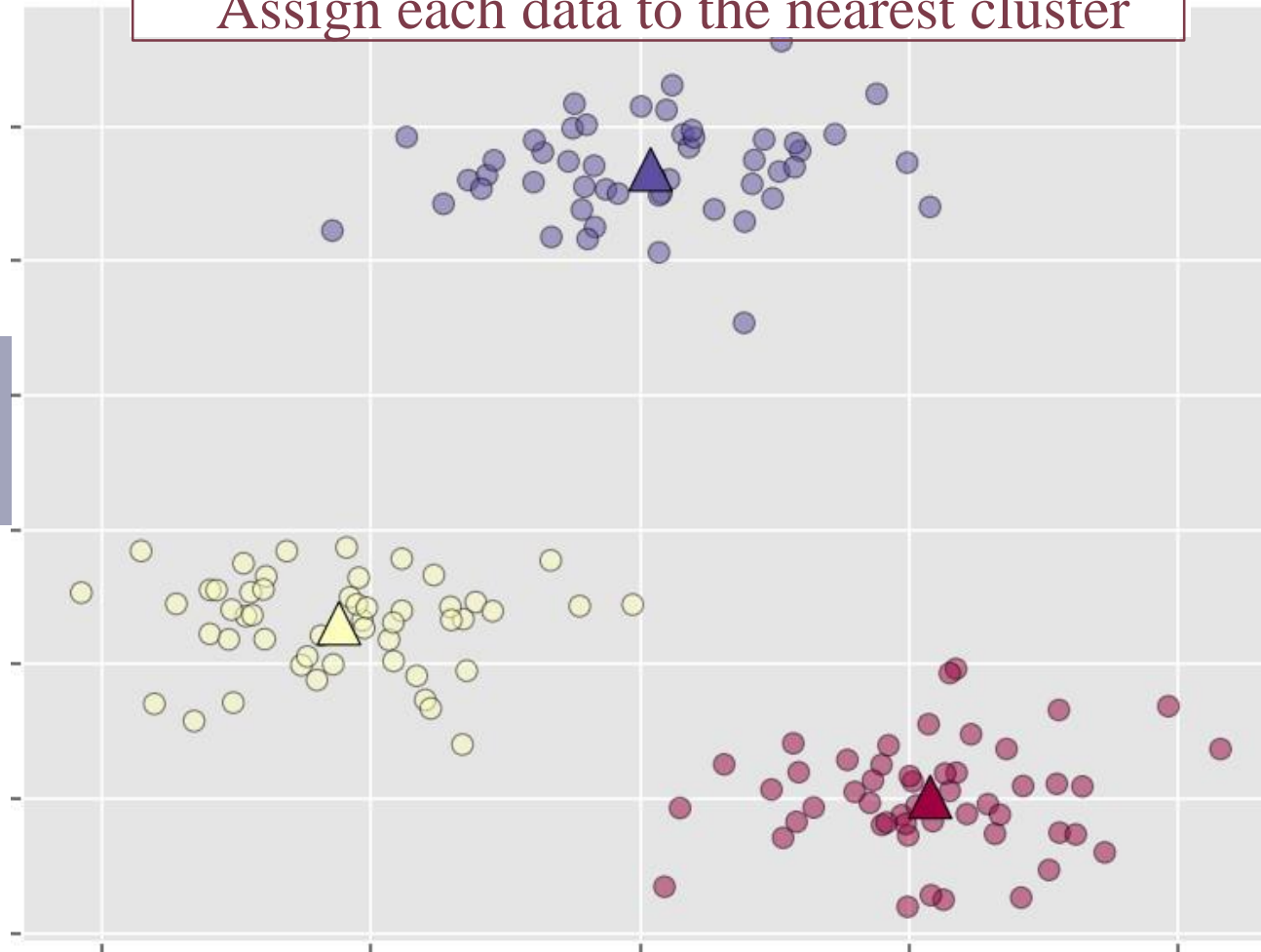
18



Illustrations of algorithm process

19

Assign each data to the nearest cluster



Convergence :
None of the data clusters were
changed .

K-mean Algorithm

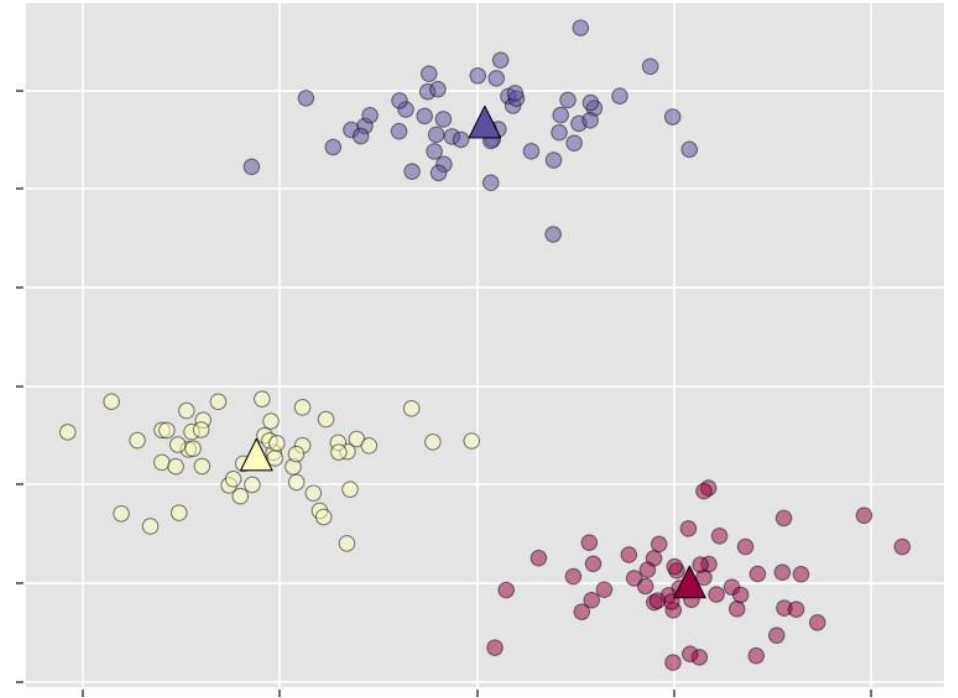
20

□ Inputs.

- Number Of Clusters : K
- Training set

$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$$

- **Note:** In the training set, no label is assigned to the data.
- **Note:** There is no need to add the feature $x_0 = 1$ in the clustering.



K-mean Algorithm

21

randomly initialize K cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$

repeat

{

for $i = 1$ **to** m

Assigning data to clusters

$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

for $k = 1$ **to** K

Updating the cluster center

$\mu_k =$ average of points assigned to cluster k

}

K-mean Algorithm

22

```
centroids = np.random.random( (K, n) )
```

```
while True:
```

```
    for i in range(m):
```

```
        c[i] = np.argmin(np.linalg.norm(X[i] - centroids, axis=1))
```

```
    for k in range(K):
```

```
        centroids[k] = np.mean(X[c == k], axis=0)
```

Clustering: Objective function

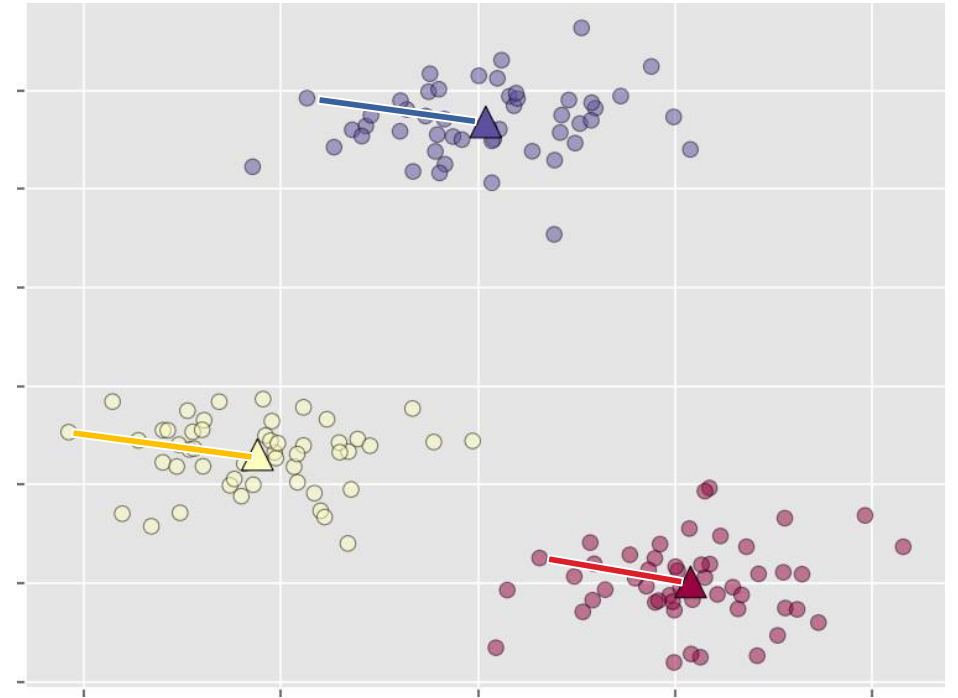
Objective function

24

□ Symbols.

- μ_k : center of the cluster k
- $c^{(i)}$: Cluster number assigned to data $x^{(i)}$
- $\mu_{c^{(i)}}$: Cluster center assigned to data $x^{(i)}$

□ Objective function



$$J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$$

K-mean Algorithm

25

randomly initialize K cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$

repeat

{

for $i = 1$ **to** m

$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

Minimization of the objective with respect to $c^{(i)}$ parameters

for $k = 1$ **to** K

$\mu_k =$ average of points assigned to cluster k

Minimization of the objective with respect to μ parameters

}



Setting initial value for cluster centers

K-mean Algorithm

27

randomly initialize K cluster centroids $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^n$

repeat

{

for $i = 1$ to m

$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

for $k = 1$ to K

$\mu_k =$ average of points assigned to cluster k

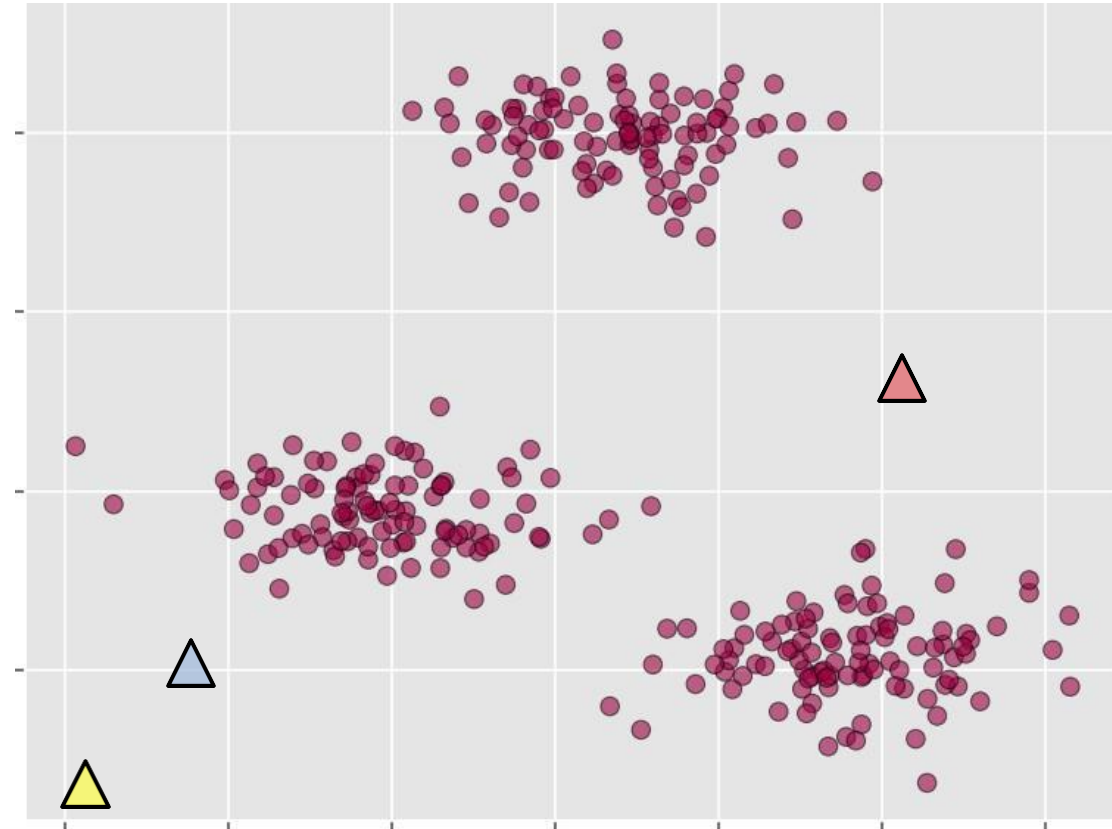
}

Setting initial value for cluster centers

28

- Setting initial value. [$K \leq m$]
 - Selection of K training samples randomly
 - Assigning center of each cluster to K selected samples

A center may be selected in such a way that no data is assigned to it.

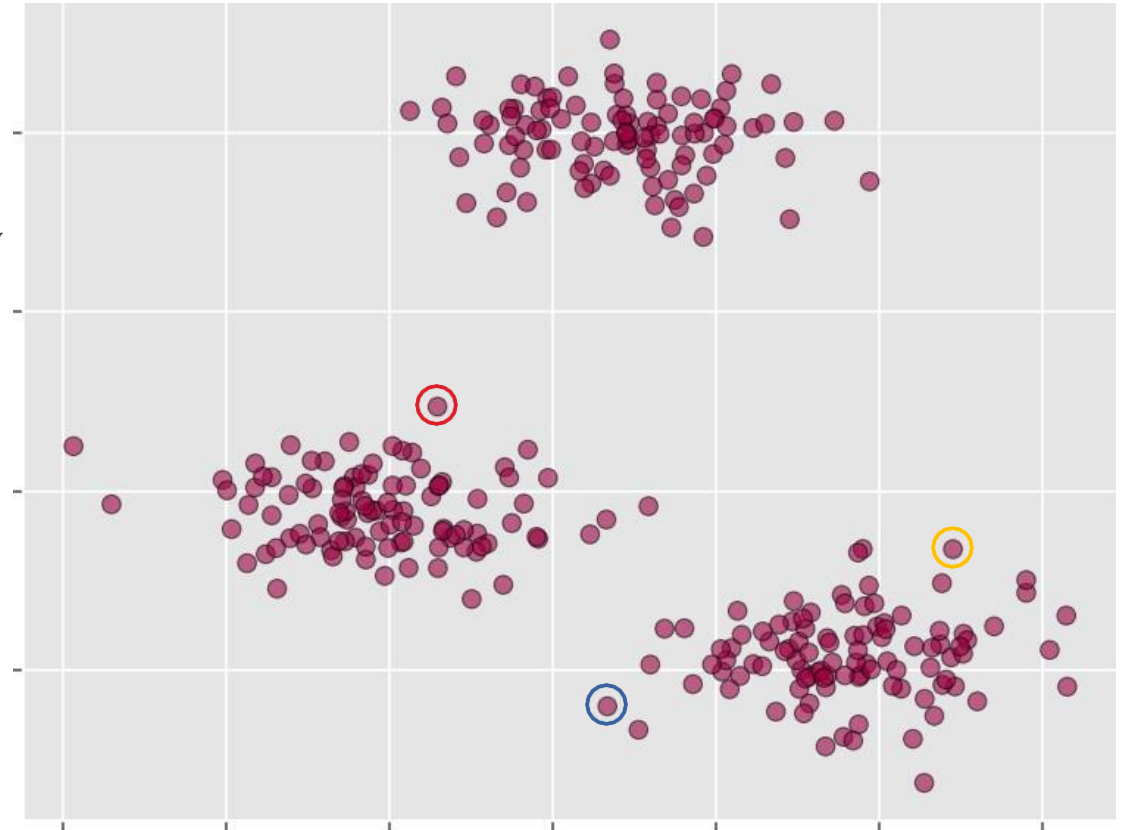


Setting initial value for cluster centers : Better way

29

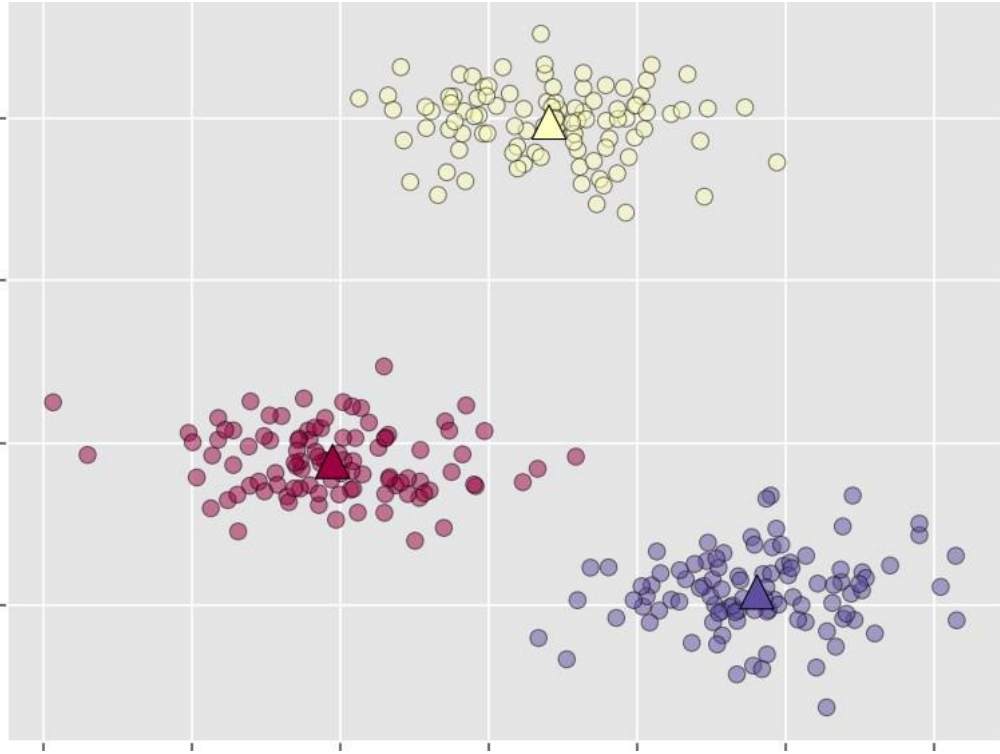
- Setting initial value. [$K \leq m$]
 - Choosing k amount of training samples randomly
 - Assigning center of each cluster to K selected samples

```
C = np.random.permutation(X)[:K]
```

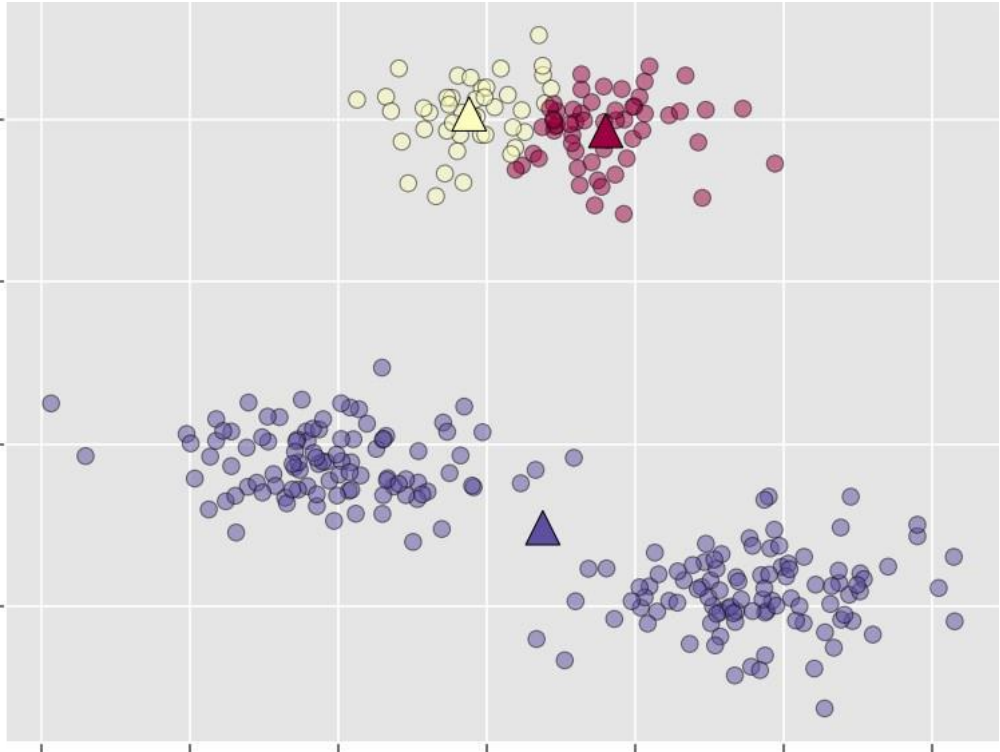


Global optimization and Local optimization

Local optimization



Global optimization



Avoid local optima

31

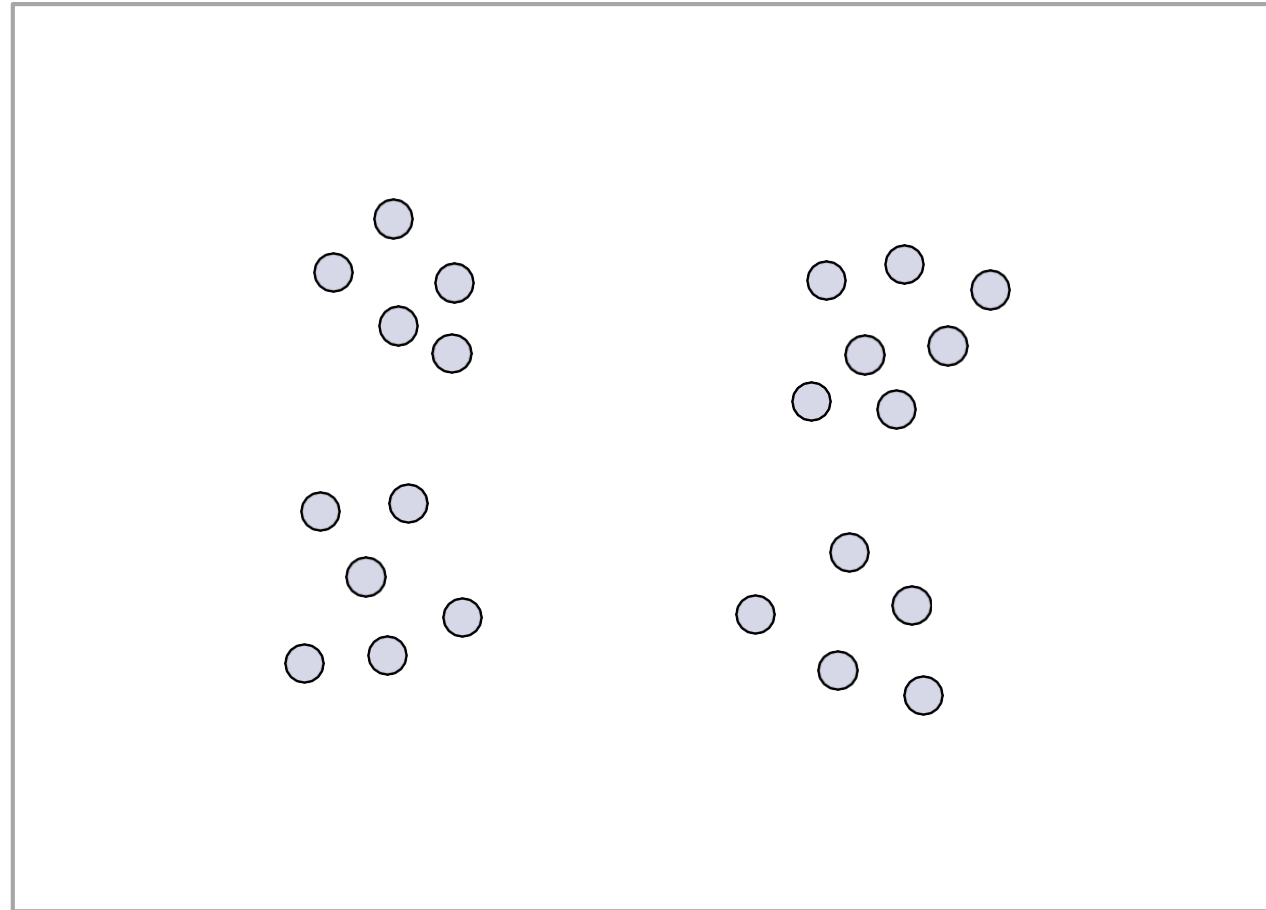
```
for  $t = 1$  to  $MAX$   
{  
  randomly initialize cluster centroids  $\mu_1, \mu_2, \dots, \mu_k$   
  run K-means to get  $c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k$   
  compute cost function  $J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k)$   
}  
pick clustering with minimum cost
```




Determine the number of clusters

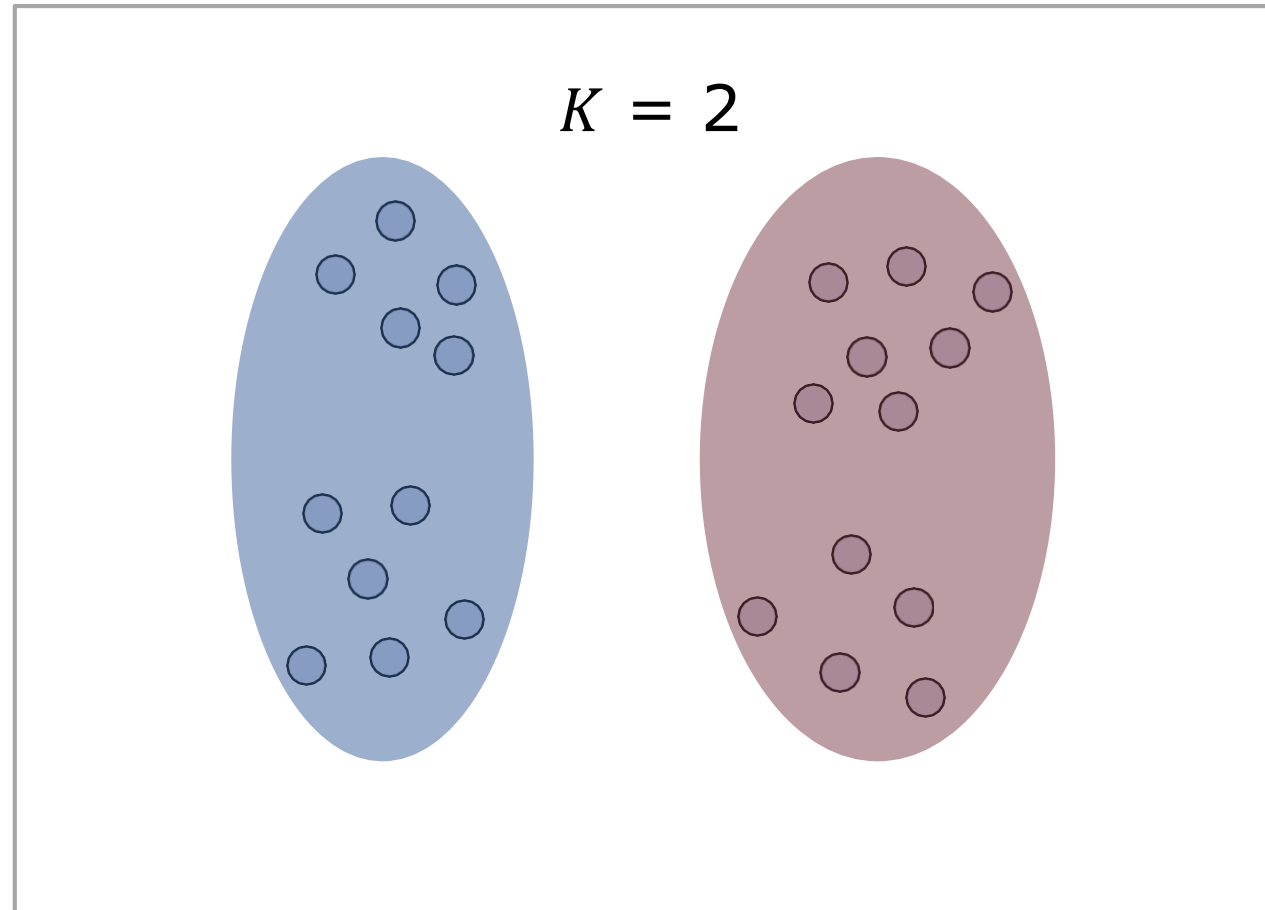
What is the right value for K?

33



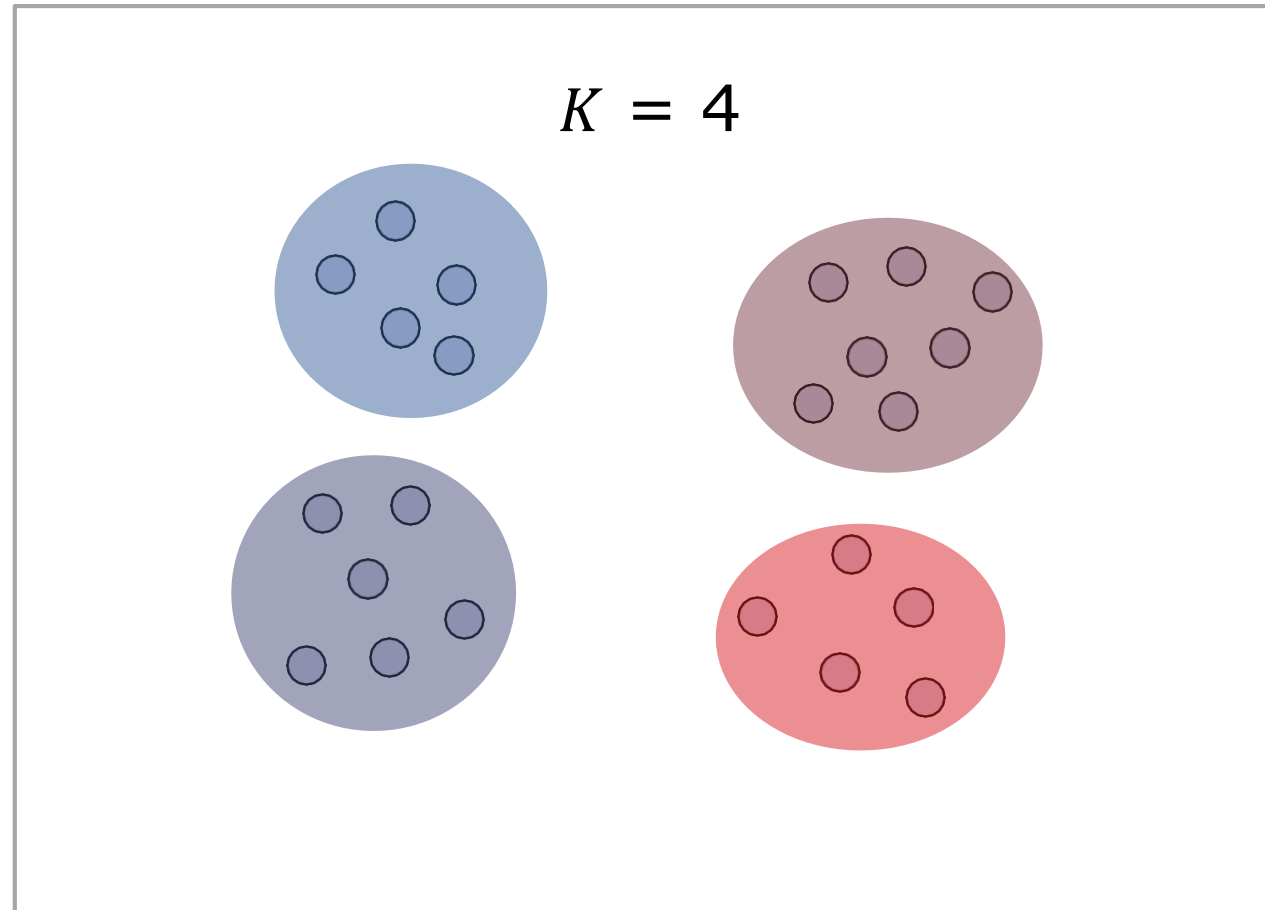
What is the right value for K?

34



What is the right value for K?

35



Determining the appropriate number of clusters: Elbow Method





Improving clustering

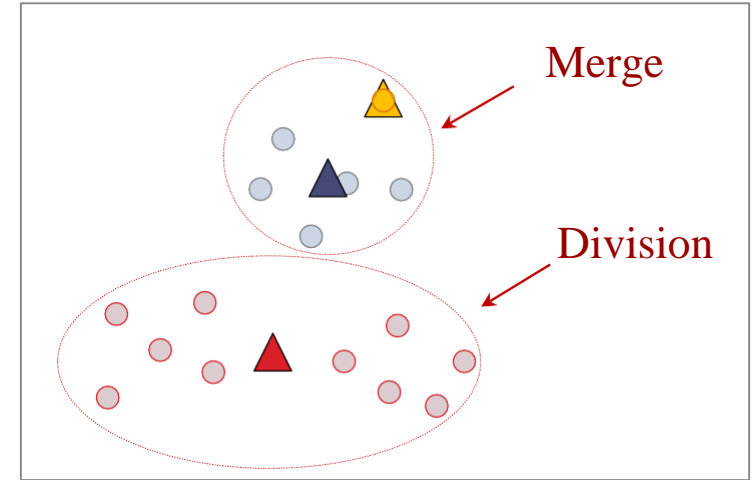
Improving clustering with post-processing of clusters

38

□ Division.

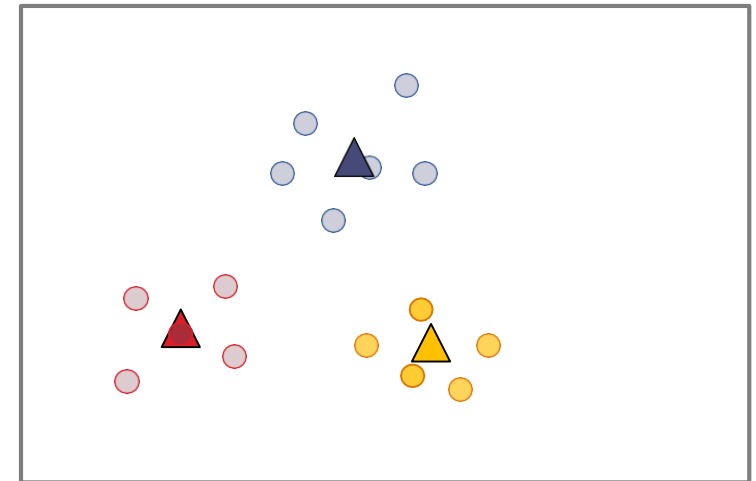
- Splitting a cluster with the highest error into two clusters.

By running the K-means on the data of this cluster with a value of $K = 2$



□ Merge.

- Merge the two closest clusters
- Merging two clusters with minimal increase in total error.



Algorithm of bipartite generator K-Means

39

□ Bipartite algorithm.

- Start with a cluster containing all the data.

- Choose one cluster at a time :
 - Divide the selected cluster into two clusters using K-means algorithm.
 - Calculate the total clustering error.
 - Choose the clustering with the least error.

- Repeat the process above until you reach the desired number of clusters.

Algorithm of bipartite generator K-Means

40

Start with all the points in one cluster

while the number of clusters is less than K

 measure the total error

 for every cluster

 perform K-means clustering with $k=2$ on the given cluster

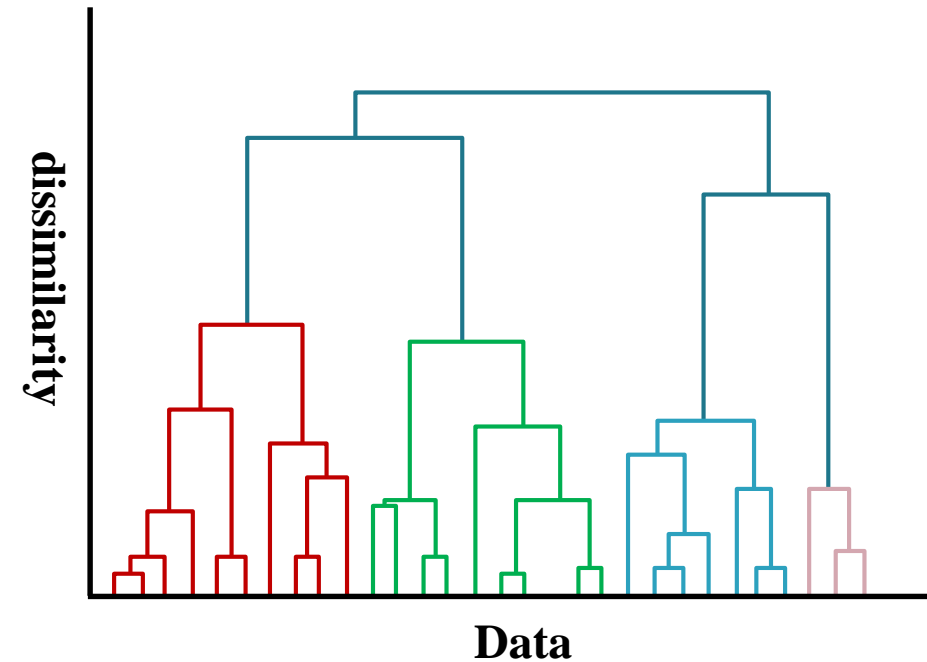
 measure the total error after splitting

 choose the cluster split that gives the lowest error

Hierarchical clustering

41

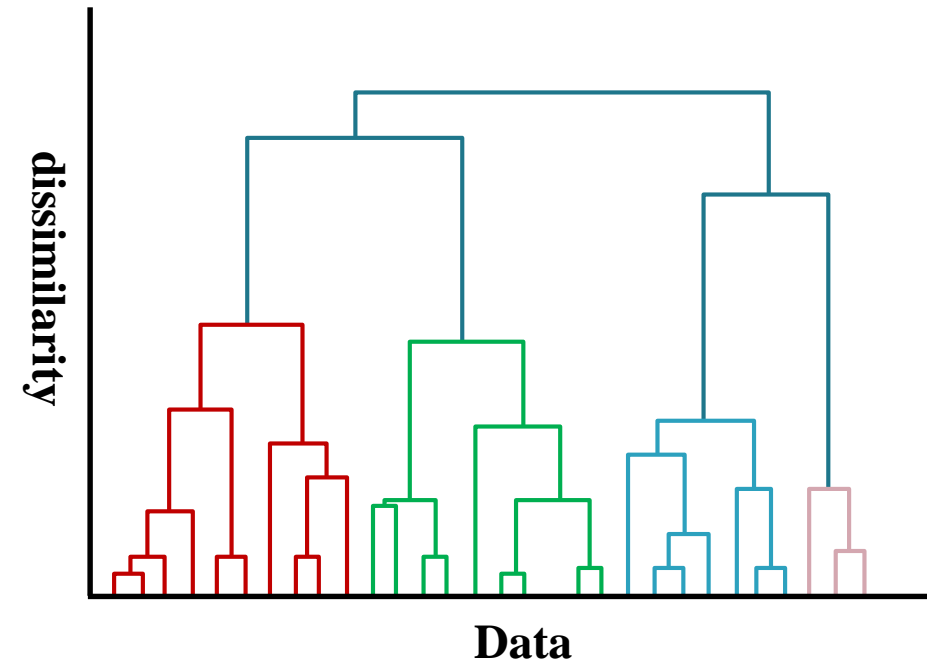
- Hierarchical clustering.
 - First, merge very similar data.
 - Gradually create larger clusters by merging smaller clusters.
- Algorithm.
 - At first, each data represents a cluster.
 - Repeat the following steps:
 - Choose the two **closest** clusters each time.
 - Merge those two clusters into a new cluster.
 - Stop: when only one cluster remains.
- Create a **tree diagram** containing a wide range of clusters.



Hierarchical clustering

42

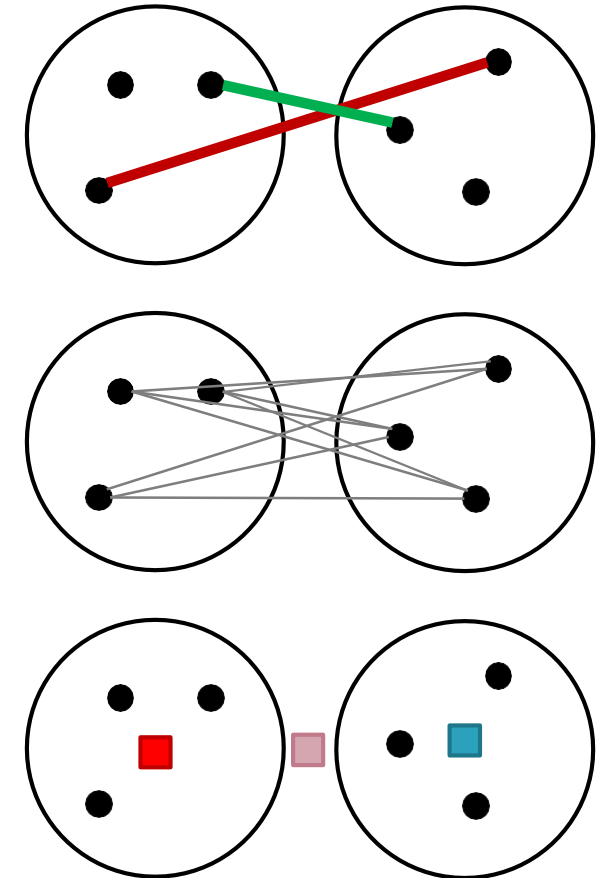
- Hierarchical clustering.
 - First, merge very similar data.
 - Gradually create larger clusters by merging smaller clusters.
- Algorithm.
 - At first, each data represents a cluster.
 - Repeat the following steps:
 - Choose the two **closest** clusters each time.
 - Merge those two clusters into a new cluster.
 - Stop: when only one cluster is left.
- Create a **tree diagram** containing a wide range of clusters.



Hierarchical clustering

43

- How to define the closest two clusters.
- Criteria for determining clusters similarity.
 - **Nearest pair** (one-link clustering)
 - **Farthest pair** (all-link clustering)
 - Average distance of all pairs
 - «Enter» method (least variance, like k-means)
- Different criteria create different clusters.



Clustering of digits

44

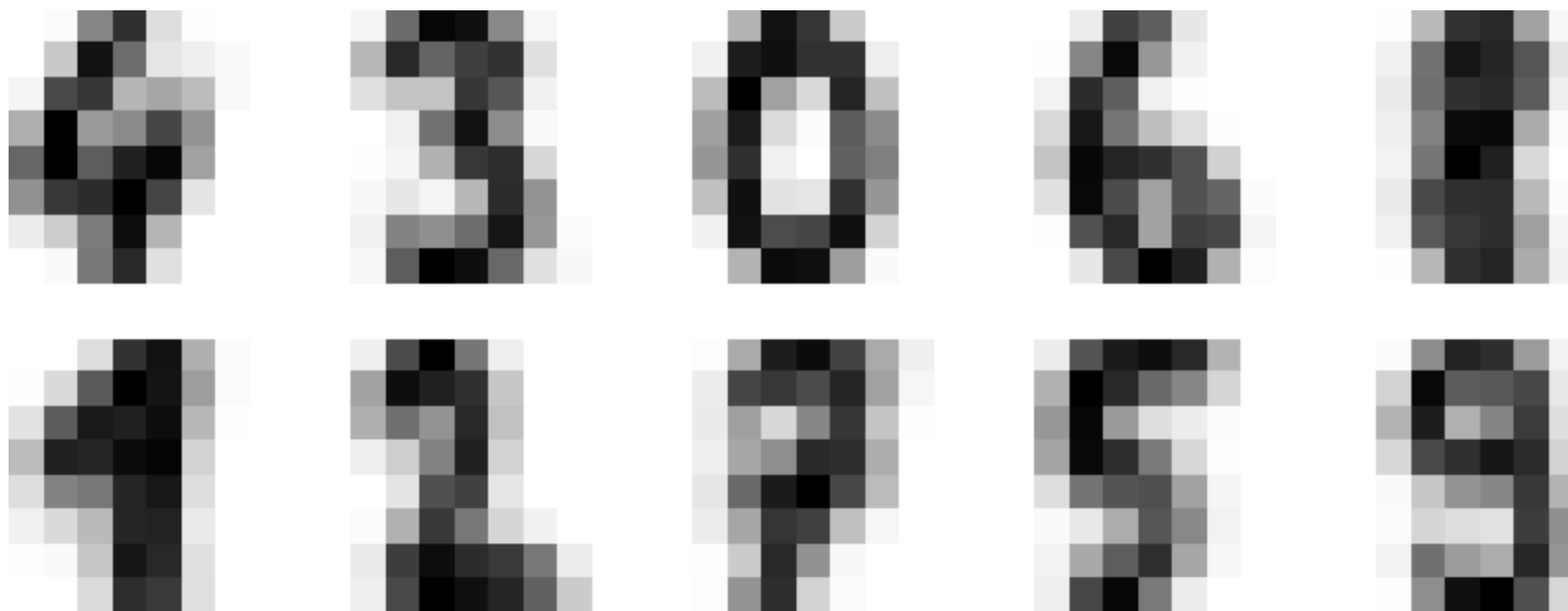
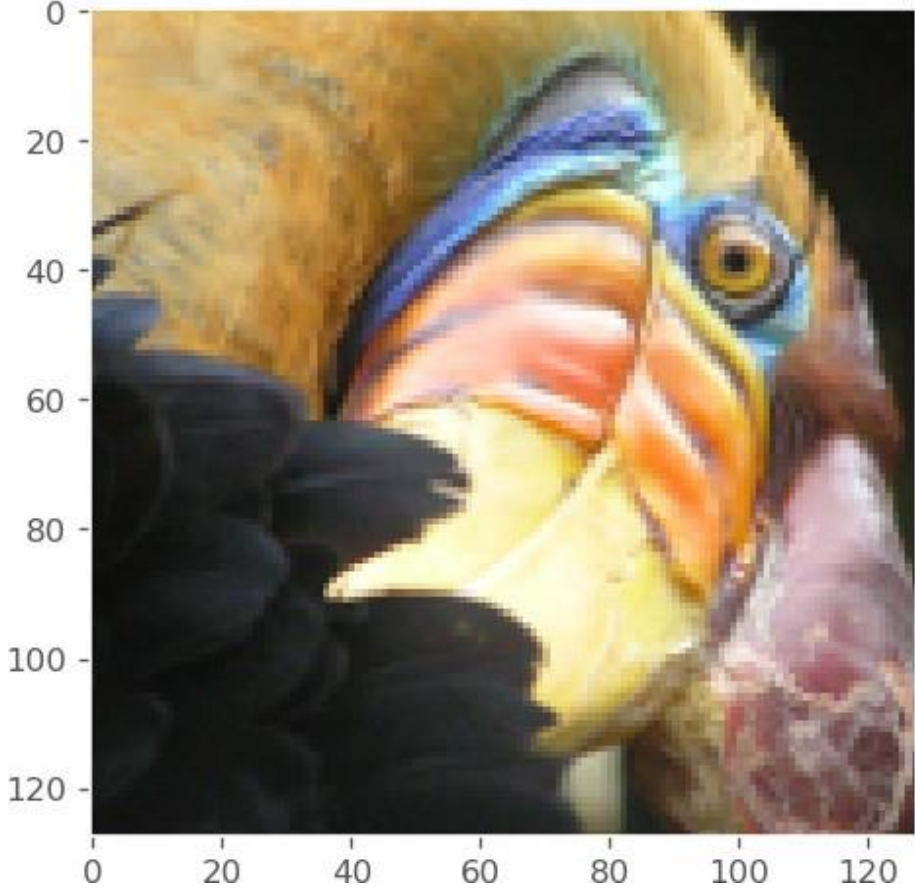


Image compression

Original Image



Compressed Image (K = 16)

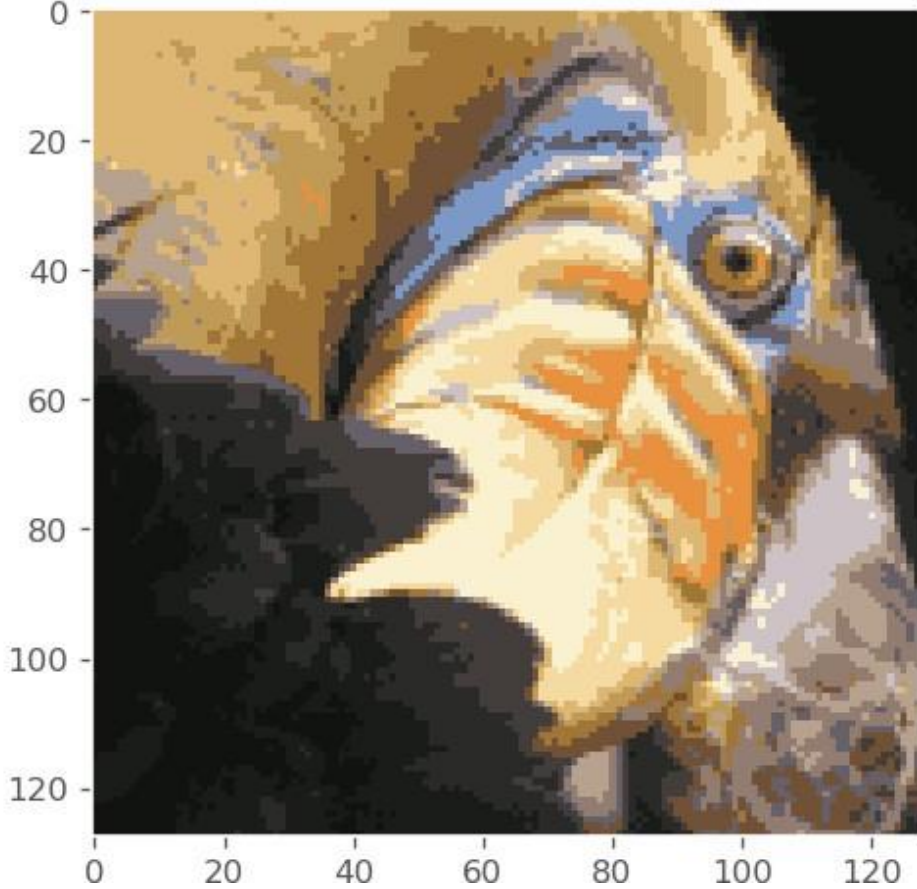
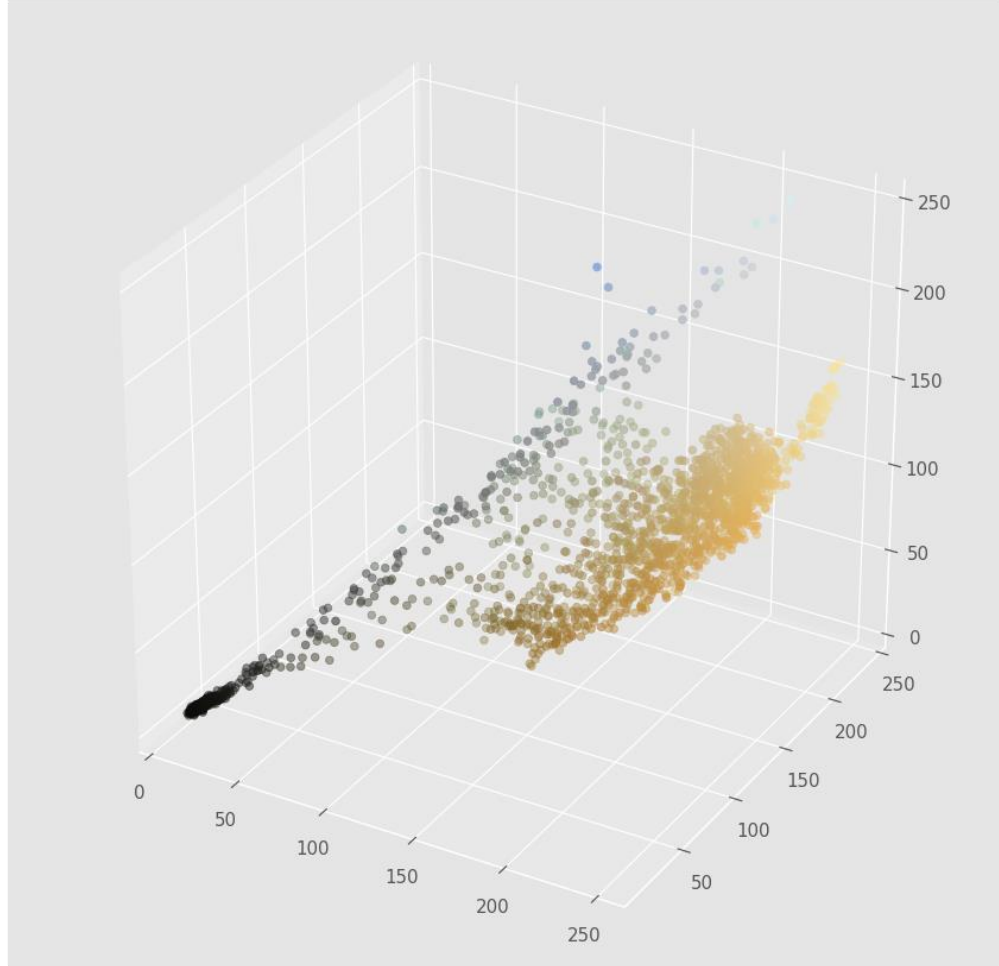
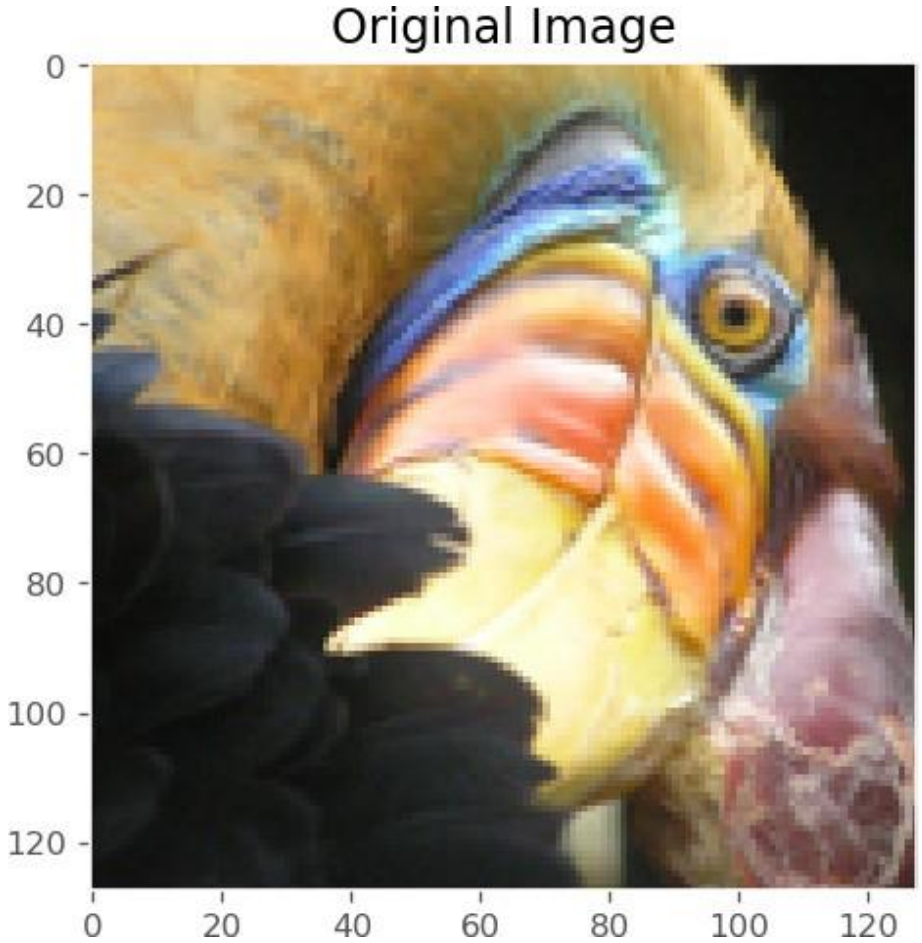


Image compression



Summary

47

- Unsupervised learning, Finding structure in data

- Clustering Grouping similar data
 - K-mean Algorithm
 - Easy implementation
 - Slow for very large data sets
 - Possibility of getting stuck in the local optimum
 - Post-processing of clusters : Splitting and merging clusters
 - Algorithm of bipartite generator K-Means
 - Better clustering than the K-Means algorithm
 - Hierarchical clustering algorithms