Statistical Learning

https://github.com/ggorr/Machine-Learning/tree/master/ISLR

10
Unsupervised Learning

- Unsupervised Learning
  - Observations
    - Features X_1, …, X_p
  - Without or unknown response
- To discover interesting things
  - PCA
  - Clustering

10.2 Principal Components Analysis

- To find highly variable directions
- To find subspaces that are as close as possible to the data cloud

10.2.1 What Are Principal Components?

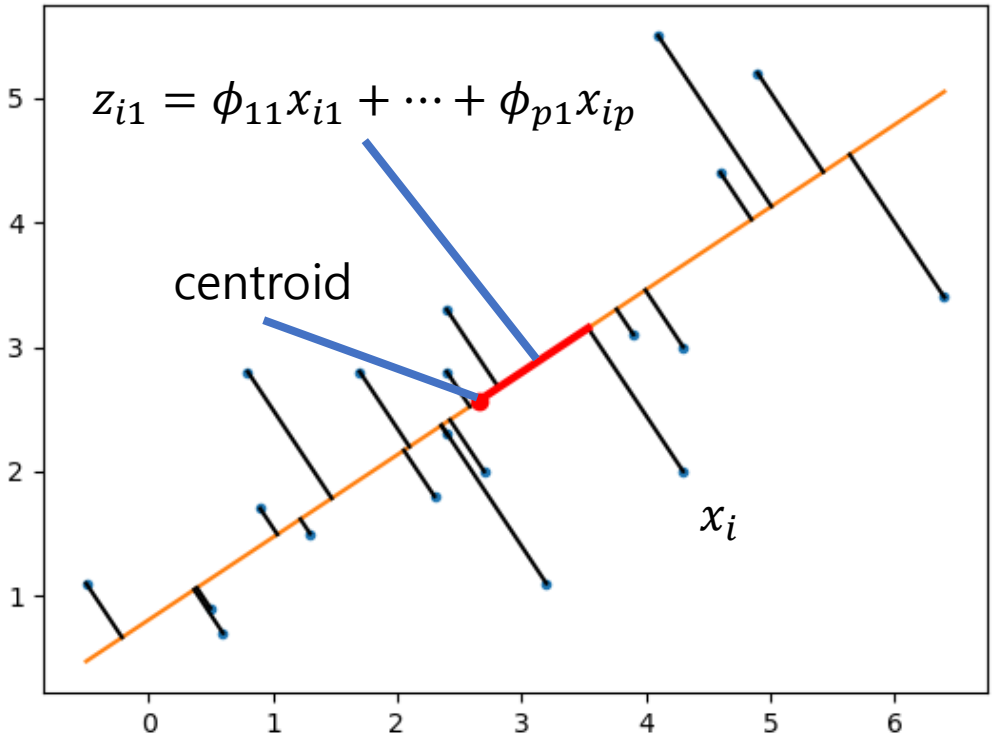
- The first principal component
  - The most variable direction
  - The normalized linear combination

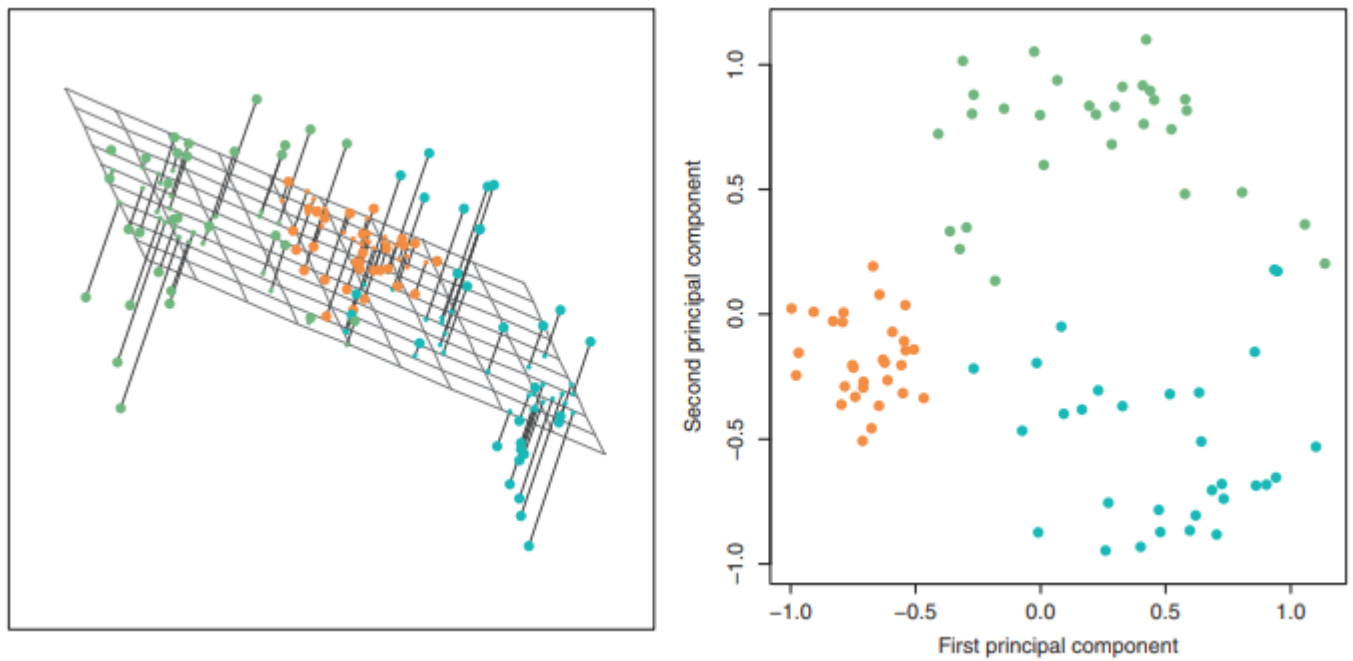
    $$Z_1 = \phi_{11}X_1 + \cdots + \phi_{p1}X_p$$

    where ϕ_{i1}'s are normalized, i.e., $\sum \phi_{i1}^2 = 1$
  - ϕ_{i1}'s are called loadings and

    $$\phi_1 = [\phi_{11} \quad \cdots \quad \phi_{p1}]^T$$

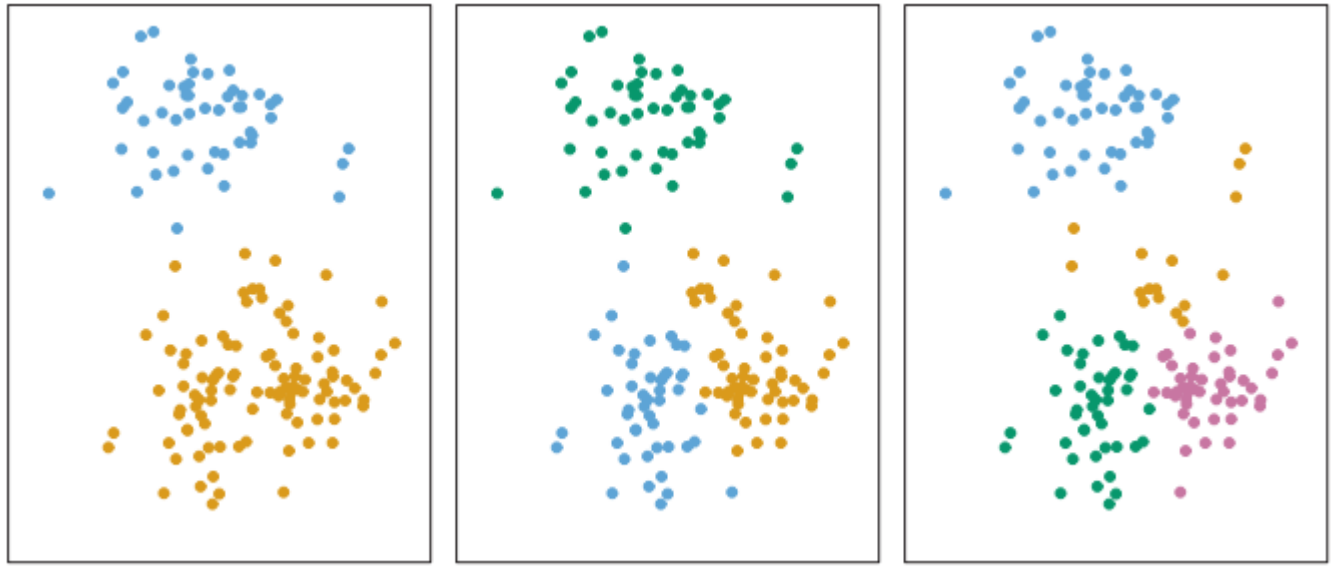
    a loading vector


$z_{i1} = \phi_{11}x_{i1} + \cdots + \phi_{p1}x_{ip}$
centroid
x_i
0
1
2
3
4
5
6
1
2
3
4
5


Second principal component
First principal component
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

10.3 Clustering Methods

- Are observations similar or different?



10.3.1 K-Means Clustering

- To divide observations as K-groups
  1. $C_1 \cup \cdots \cup C_K = \{1, \ldots, n\}$
  2. $C_i \cap C_j = \emptyset$, if $i \neq j$
- To find C_1, …, C_K with

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \sum_k W(C_k)$$

for some measure $W(C_k)$

- Squared Euclidean distance measure

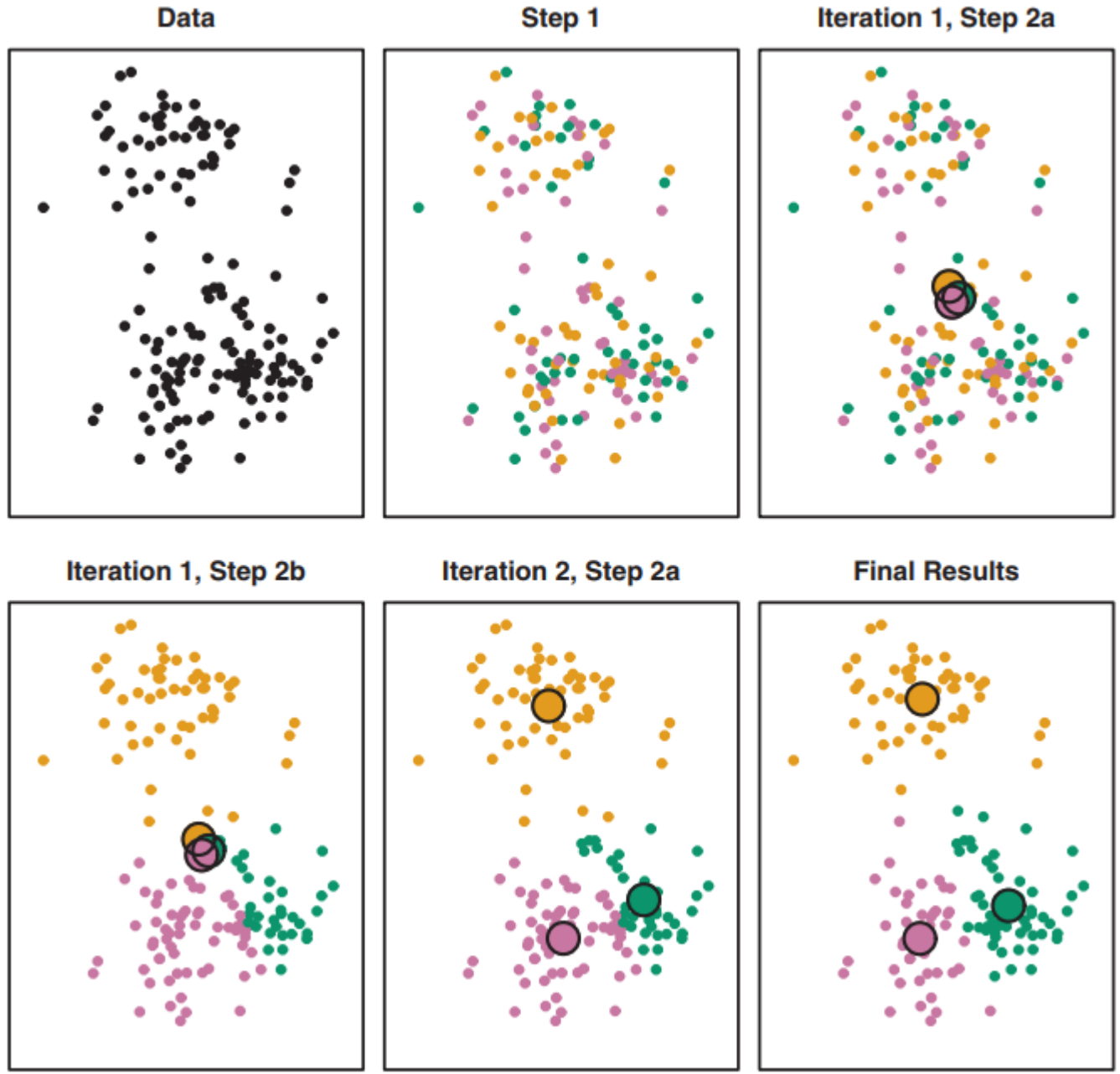
$$W(C_k) = \frac{1}{|C_k|} \sum_{i,j \in C_k} \| x_i - x_j \|^2$$

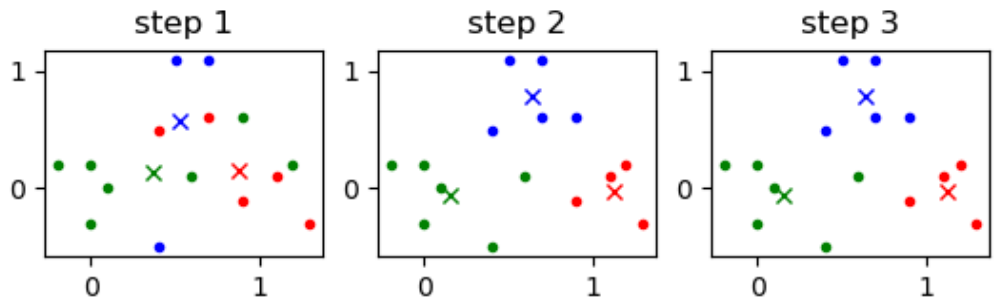
- To find C_1, …, C_K with

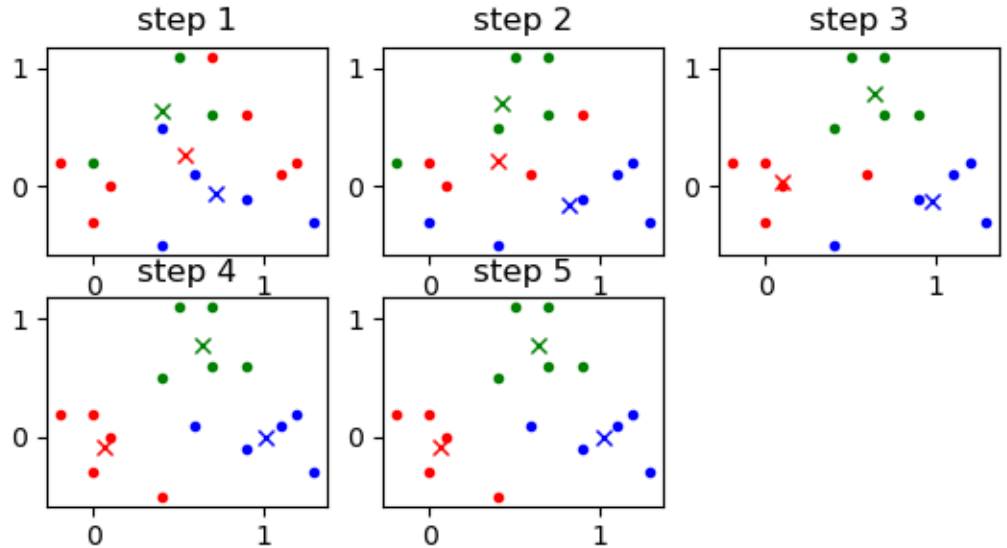
$$\underset{C_1,\ldots,C_K}{\text{minimize}} \sum_k \frac{1}{|C_k|} \sum_{i,j \in C_k} \| x_i - x_j \|^2$$

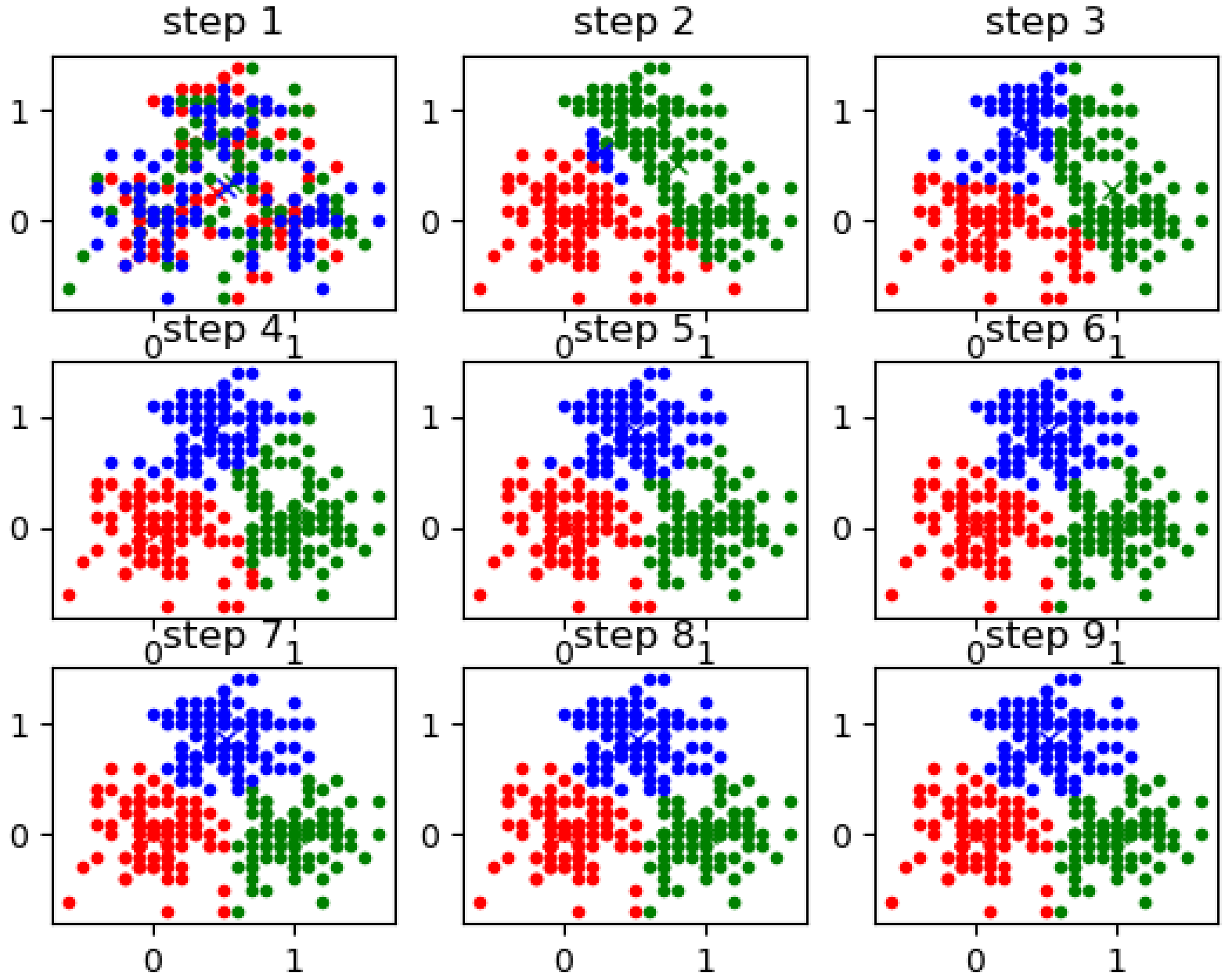
Algorithm 10.1 K-Means Clustering

1. Randomly assign a number, from 1 to K, to each observation
2. Iterate until the cluster assignments stop changing
   a. For each of the K clusters, compute the cluster centroid
   b. Assign each observation to the cluster whose centroid is closest


Data
Step 1
Iteration 1, Step 2a
Iteration 1, Step 2b
Iteration 2, Step 2a
Final Results

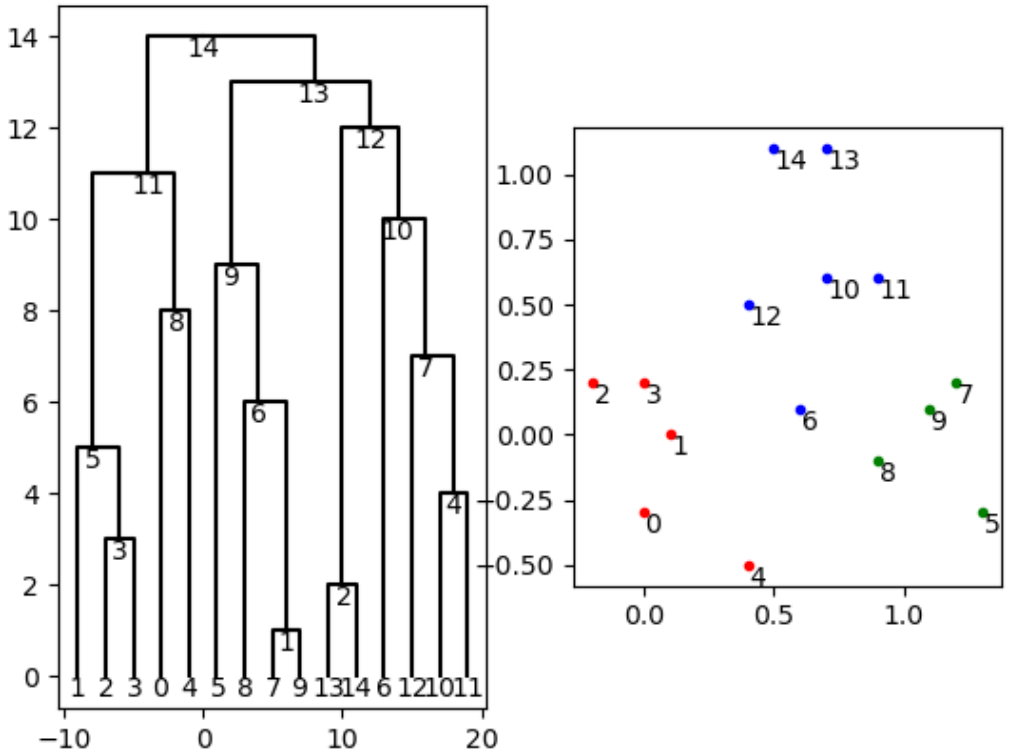

step 1
step 2
step 3
1
0
0
1


step 1
step 2
step 3
step 4
step 5
1
0
0
1


step 1
step 2
step 3
step 4
step 5
step 6
step 7
step 8
step 9
0
1

10.3.2 Hierarchical Clustering

- Hierarchical clustering
  - Tree based clustering

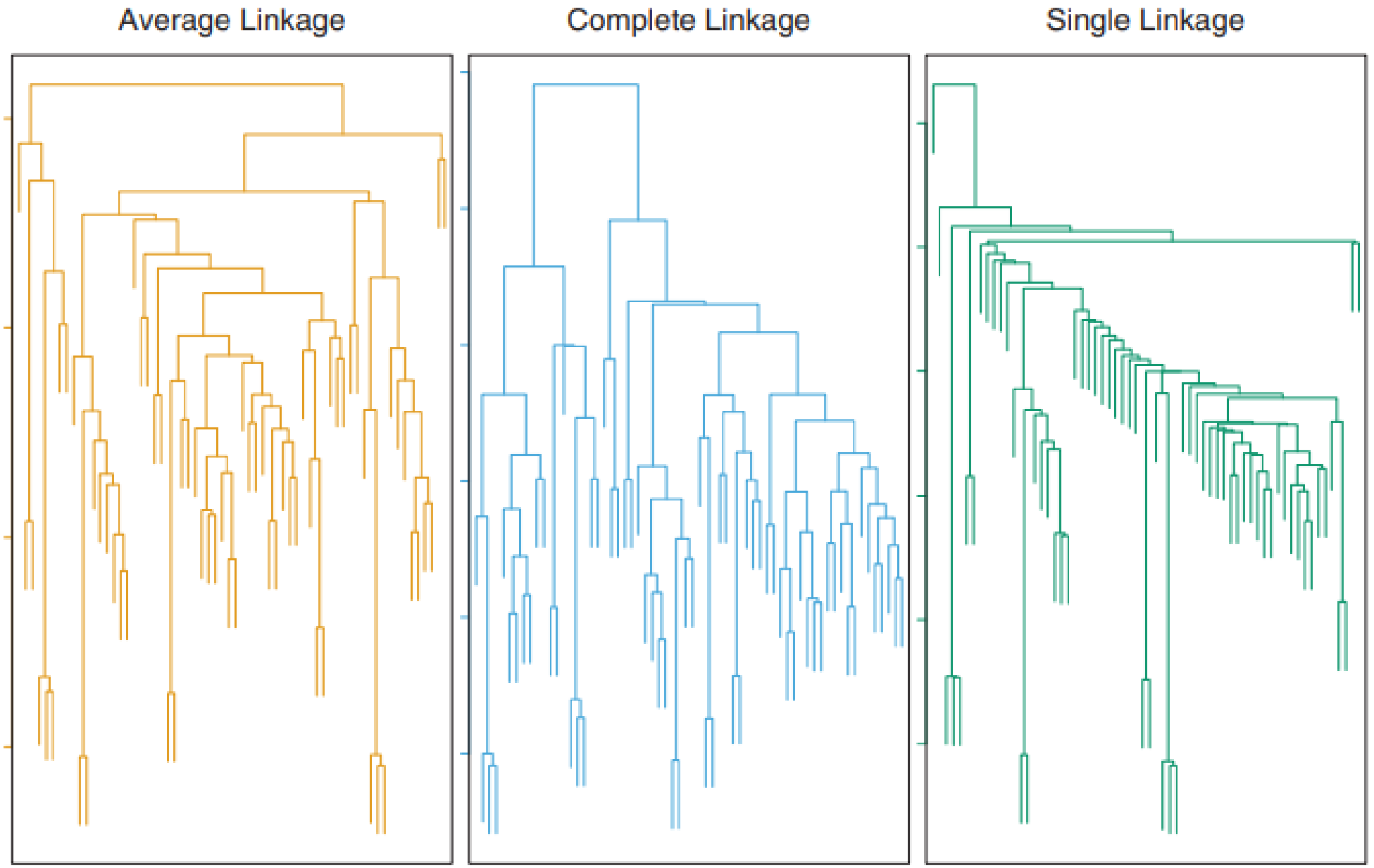


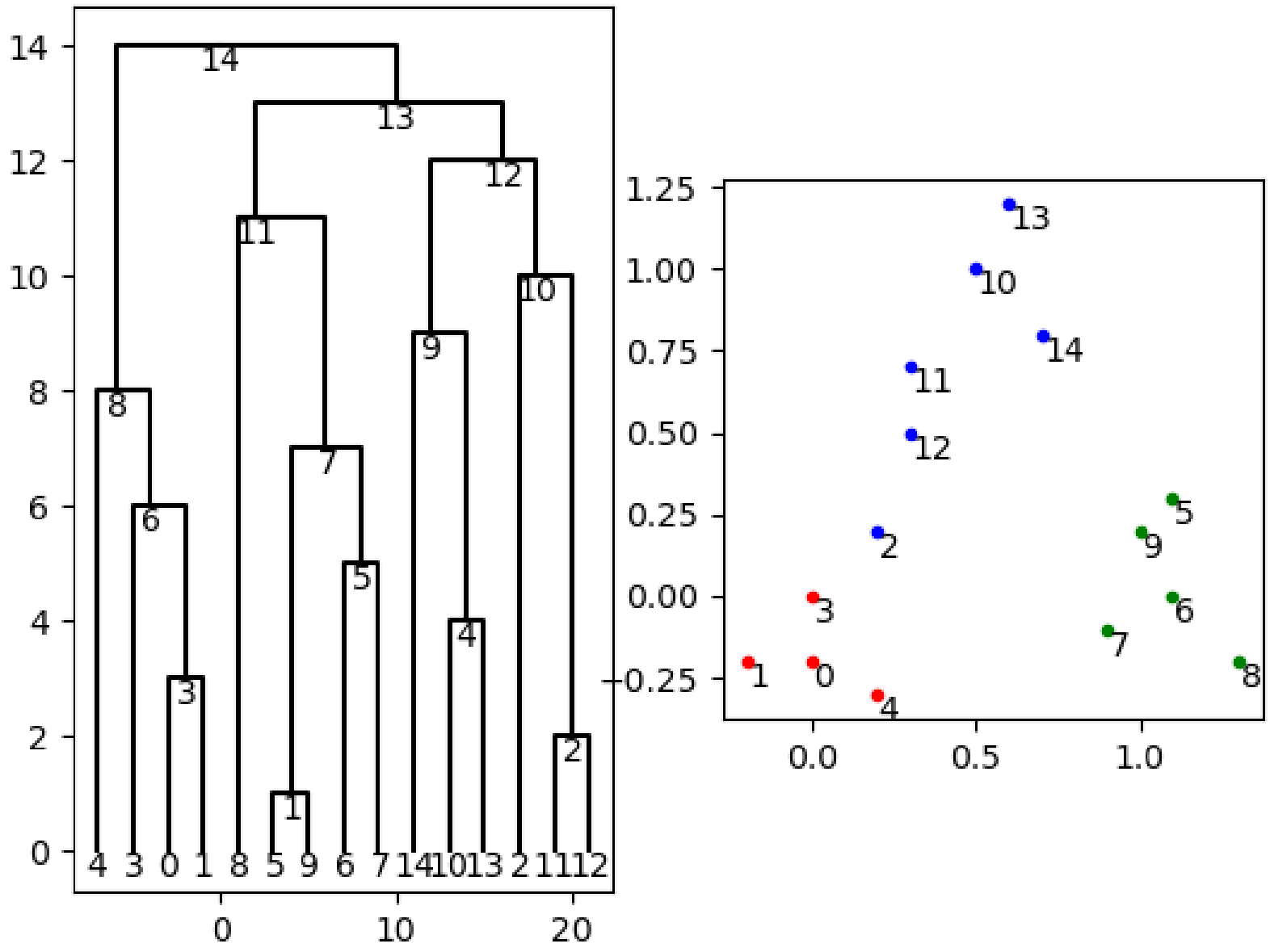
Algorithm 10.2 Hierarchical Clustering

- Begin with n observations as clusters
- Repeat
  - Fuse two clusters that are least dissimilar(most similar)

Dissimility

- Complete linkage
  - Maximal intercluster dissimilarity
- Single linkage
  - Minimal intercluster dissimilarity
- Average linkage
  - Mean intercluster dissimilarity
- Centroid linkage
  - Dissimilarity between centroids


Average Linkage
Complete Linkage
Single Linkage


14
13
12
11
10
9
8
7
6
5
4
3
2
1
4 3 0 1 8 5 9 6 7 14 10 13 2 11 12
0
10
20
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
0.0
0.5
1.0

Hierarchical Clustering with average linkage
600 observations