Homework 5

1. Please give the specific process of the k-means clustering algorithm and analyze its advantages and disadvantages.

2.There are four samples:

$$X_1 = (3,5)^T,\ X_2 = (5,1)^T,\ X_3 = (1,0)^T,\ X_4 = (1,4)^T$$

In beginning, they are divided into two classes: ω_1: $\{X_1,\ X_2\}$ and ω_2: $\{X_3,\ X_4\}$

(1) If we move the sample X_2 to the class ω_2 , please compute the within-class scatter matrix S_w.

(2) If we use the Determinant of S_w as Clustering criteria, please judge whether the movement in (1) is appropriate?

3. There are five samples:

$$X_1 = (0,1,2,1,2,4)^T,$$
$$X_2 = (3,2,3,1,2,1)^T,$$
$$X_3 = (1,0,0,0,1,1)^T,$$
$$X_4 = (2,1,0,2,1,2)^T,$$
$$X_5 = (0,0,1,0,1,0)^T,$$

Please use the hierarchical clustering algorithm to cluster these samples under the minimum distance criterion (distance D can be directly represented by a root number), and give the hierarchical clustering process.

4. Consult the density-based clustering algorithm DBSCAN, and give the pseudo-code representation of the algorithm.