Clustering Method
in
Financial Time series

Chun-Hui,Wu

2017.12.18

Outlines

1. Introduction
2. Data pre-processing
3. Clusterig method
4. Result

1.Introduction

- What is clustering?
  - An unsupervised method to learn “specific structure” in data via Algorithm
  - Hard-clustering : K-means , Hierarchical Clustering
  - Soft-Clustering: Gausian Mixture Model

- Compare
  - time series plot.
  - mean-std plot.

2. Data pre-processing

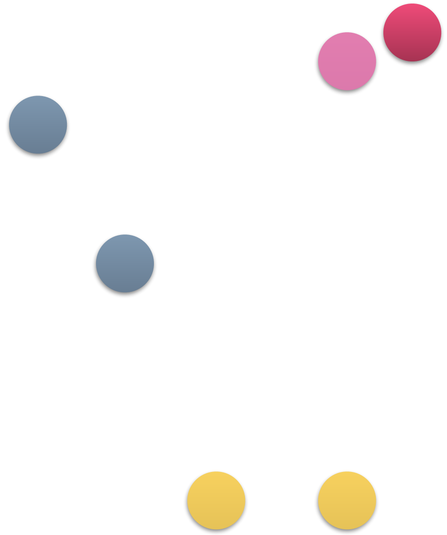
- The data set consists of 77 mutual fund price with different time length .
- step1. Choose the price data between 2014/10/8 and 2017/10/16 .
- step2. Smooth missing values in series.
- step3. All series are normalized to 2014/10/8 , by the following formula $P_t = \frac{p_t - p_1}{p_1}$

3. Clustering Method

- Hierarchical clustering
- DTW distance

Hierachichal Clustering

- In hierarchical clustering, we need to define distance between object , and distance between cluster . (Do not need coordinate!)

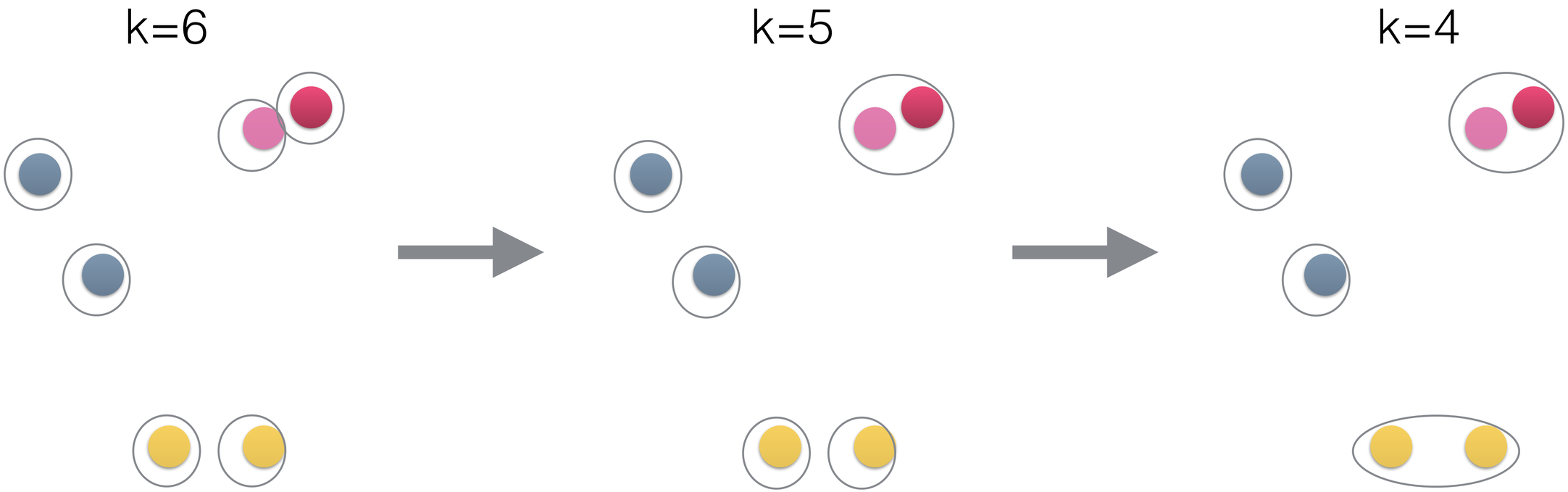


- complete linkage

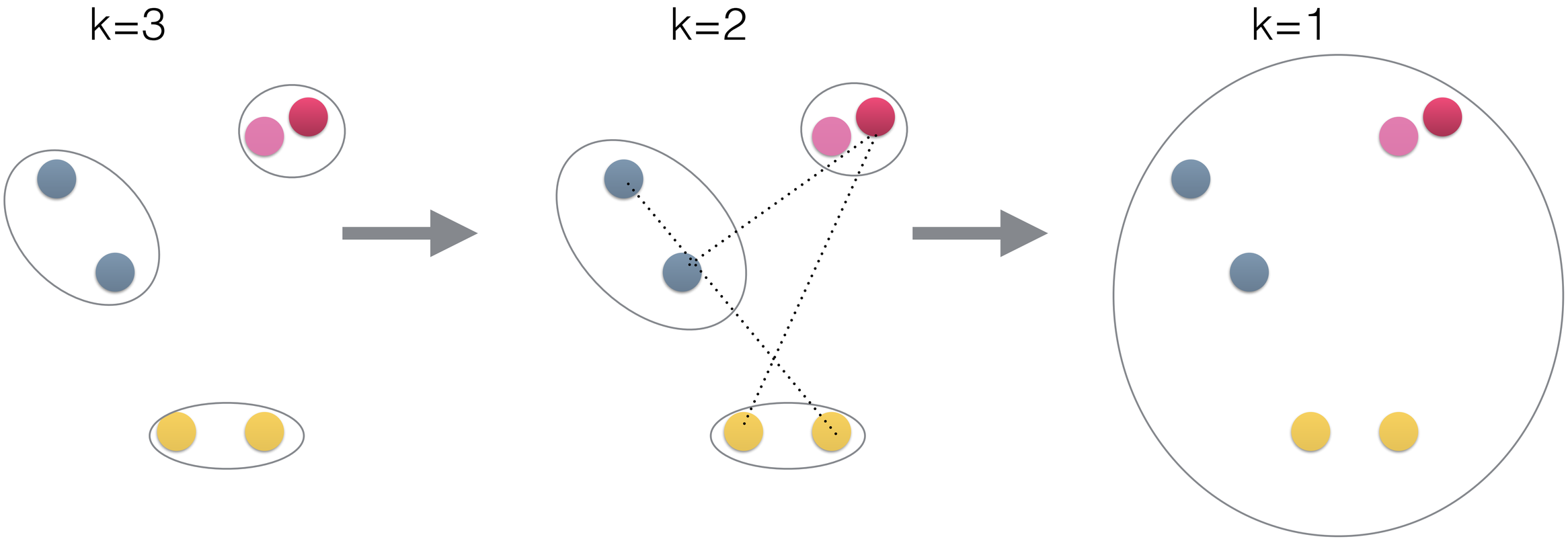
$$D(Clust_i, Clust_j) \equiv \max_{s,k} d(s,k), s \in Clust_i, k \in Clust_j$$

- Agglomerative clustering :
  1. each data is a cluster
  2. merge the closest two clusters (Repeat)

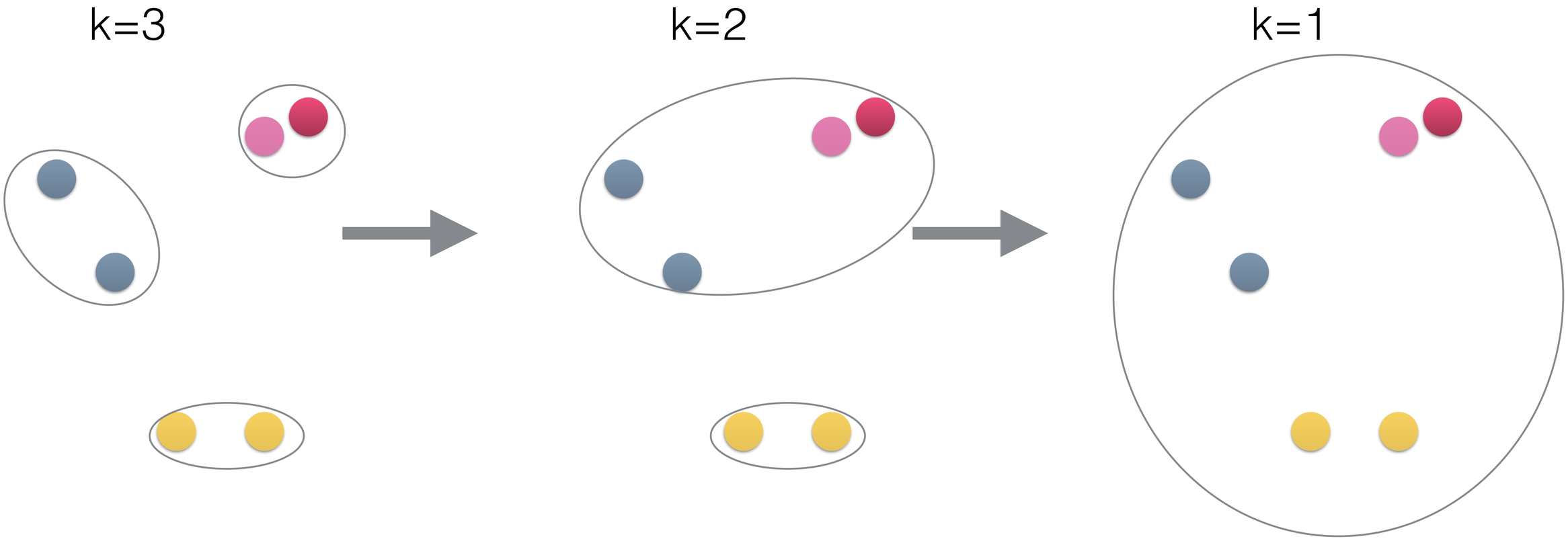
Agglomerative with Complete linkage distance



Agglomerative with Complete linkage distance



Agglomerative with Complete linkage distance



Raw based method

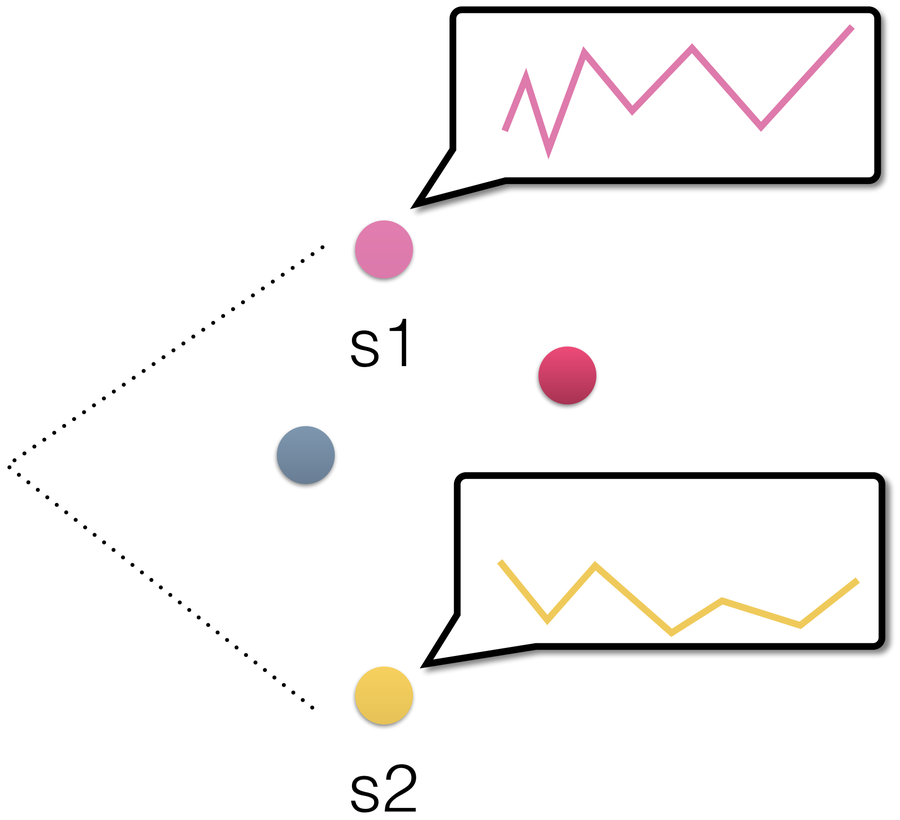
- Define distance between objects

  -> apply clustering method

- Distance Measure is more important than clustering method

- General for almost every domain

Define
d(s1,s2)



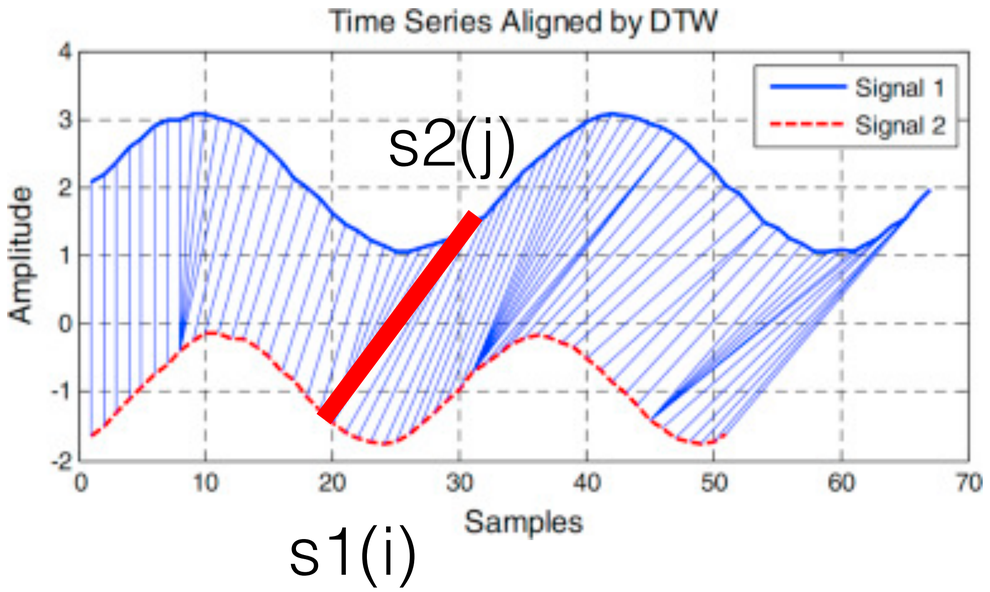
DTW Distance

- Given two time series
  s1 (length m), s2 (length n) ,
  the DTW distance is defined by:

$$D(i,j) = |s1(i) - s2(j)| + min(D(i,j-1), D(i-1,j-1), D(i-1,j))$$

- One should recursively solve the formula above to find DTW distance

- DTW can compare the stretched or compressed time series

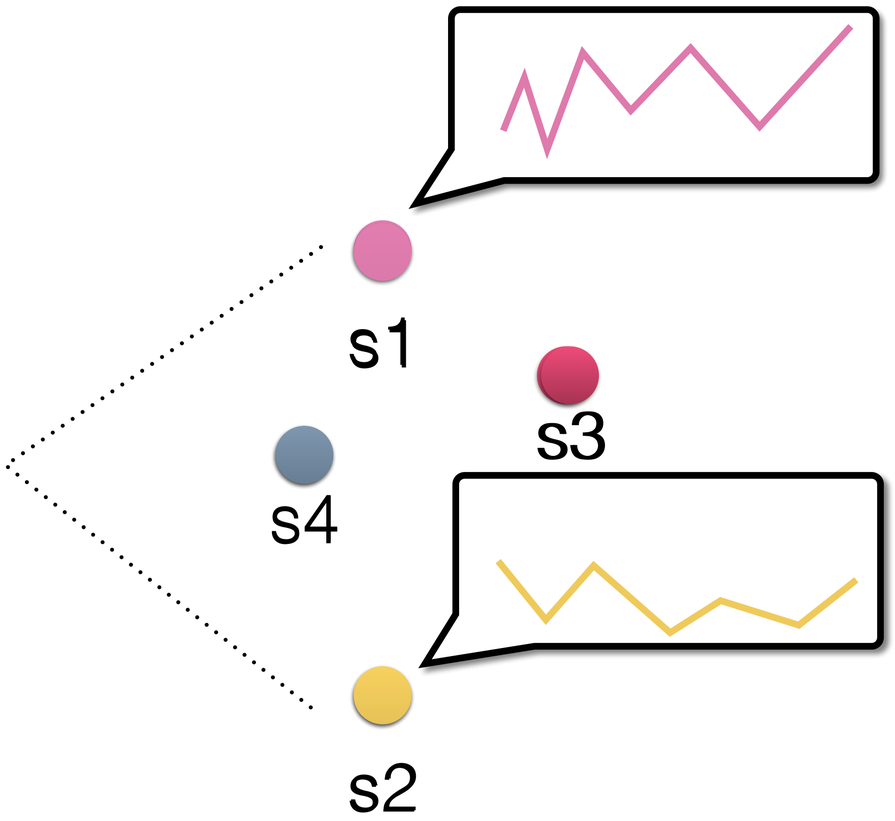


Calculate Dissimilarity Matrix

Define d(si,sj) s.t.

(1) d(si,sj)=d(sj,si)

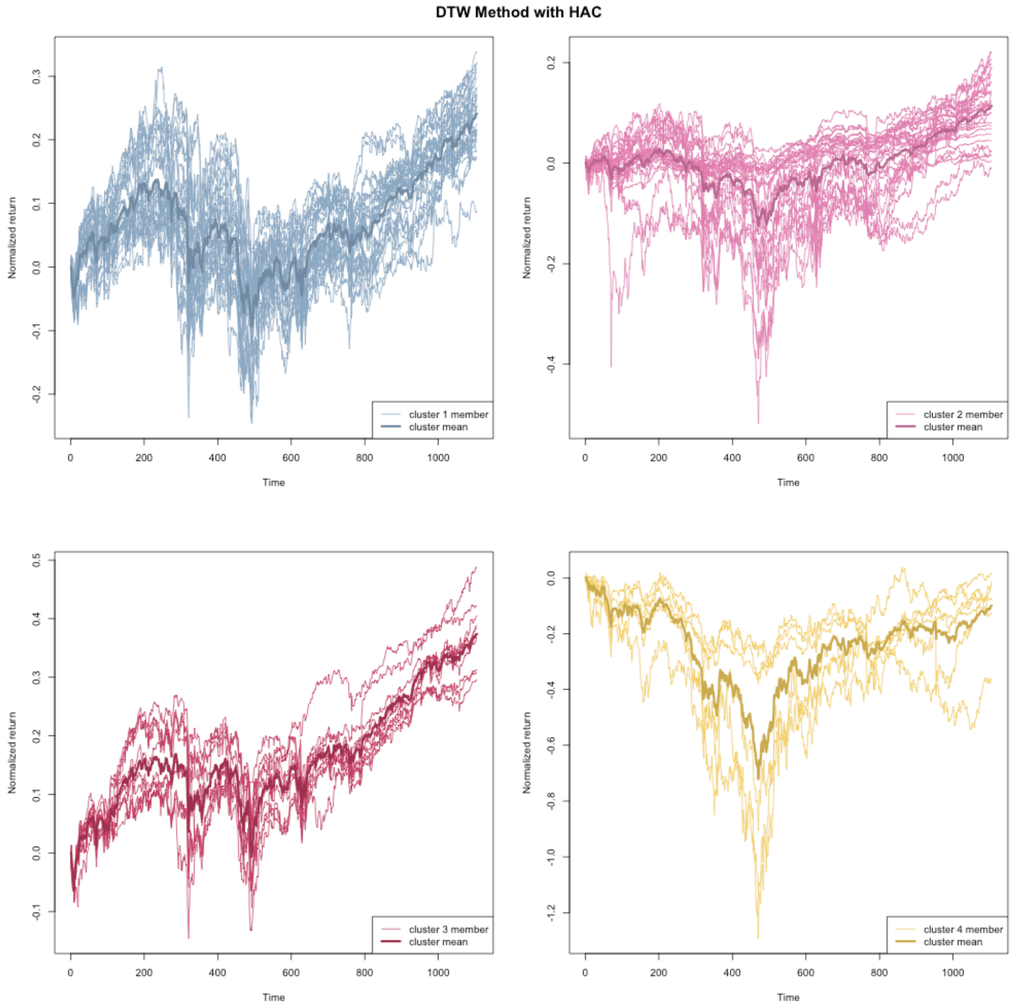
(2) d(si,si)=0

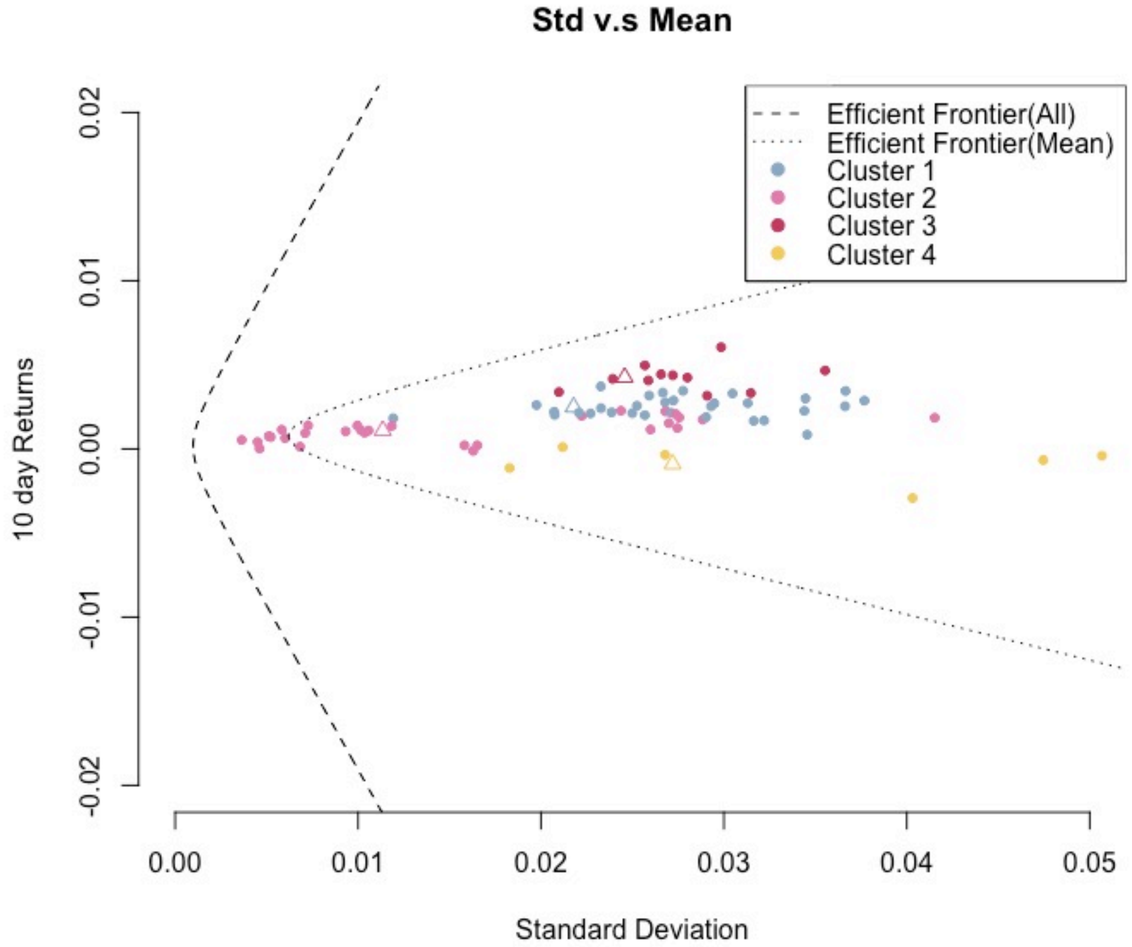


Calculate Dissimilarity Matrix

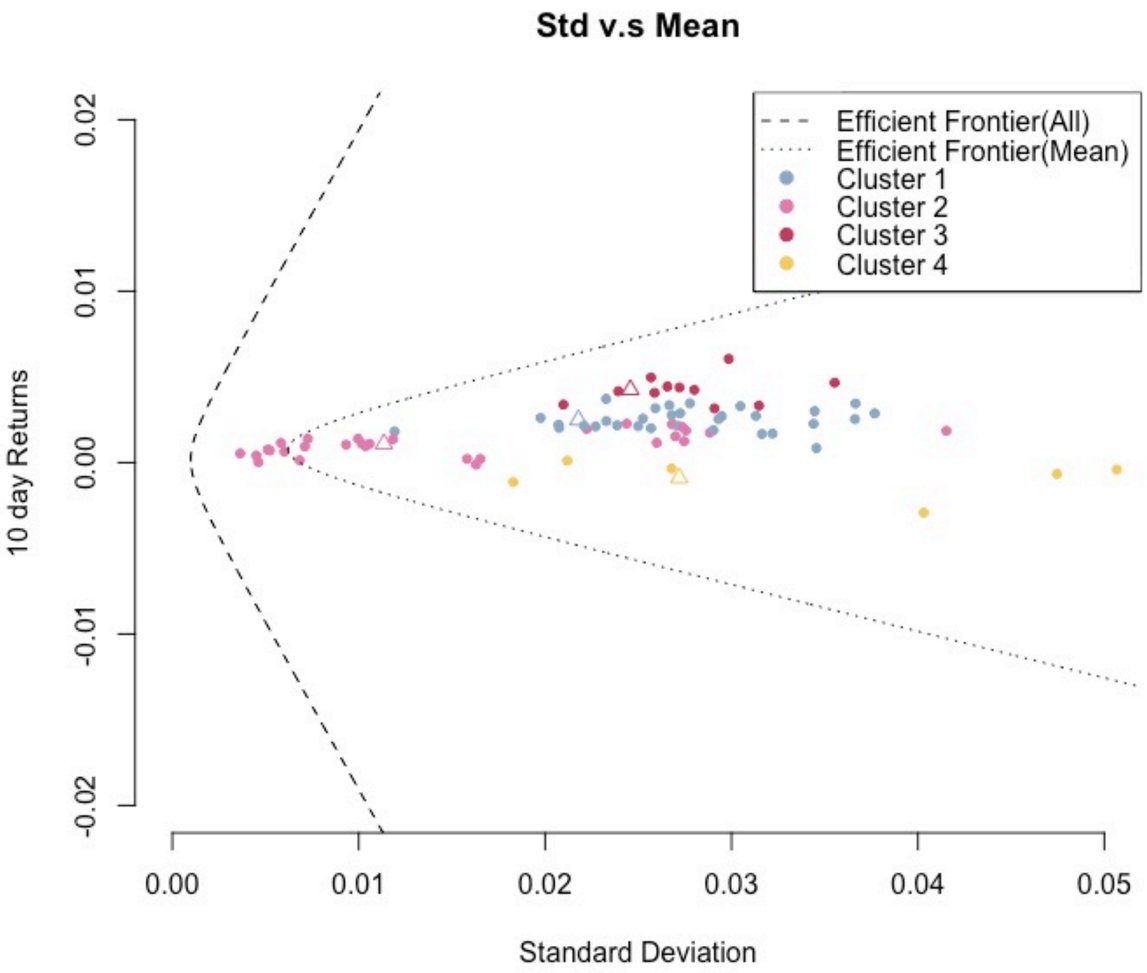
	s1	s2	s3	s4
s1	0	d12	d13	d14
s2	d12	0	d23	d24
s3	d13	d23	0	d34
s4	d14	d24	d34	0

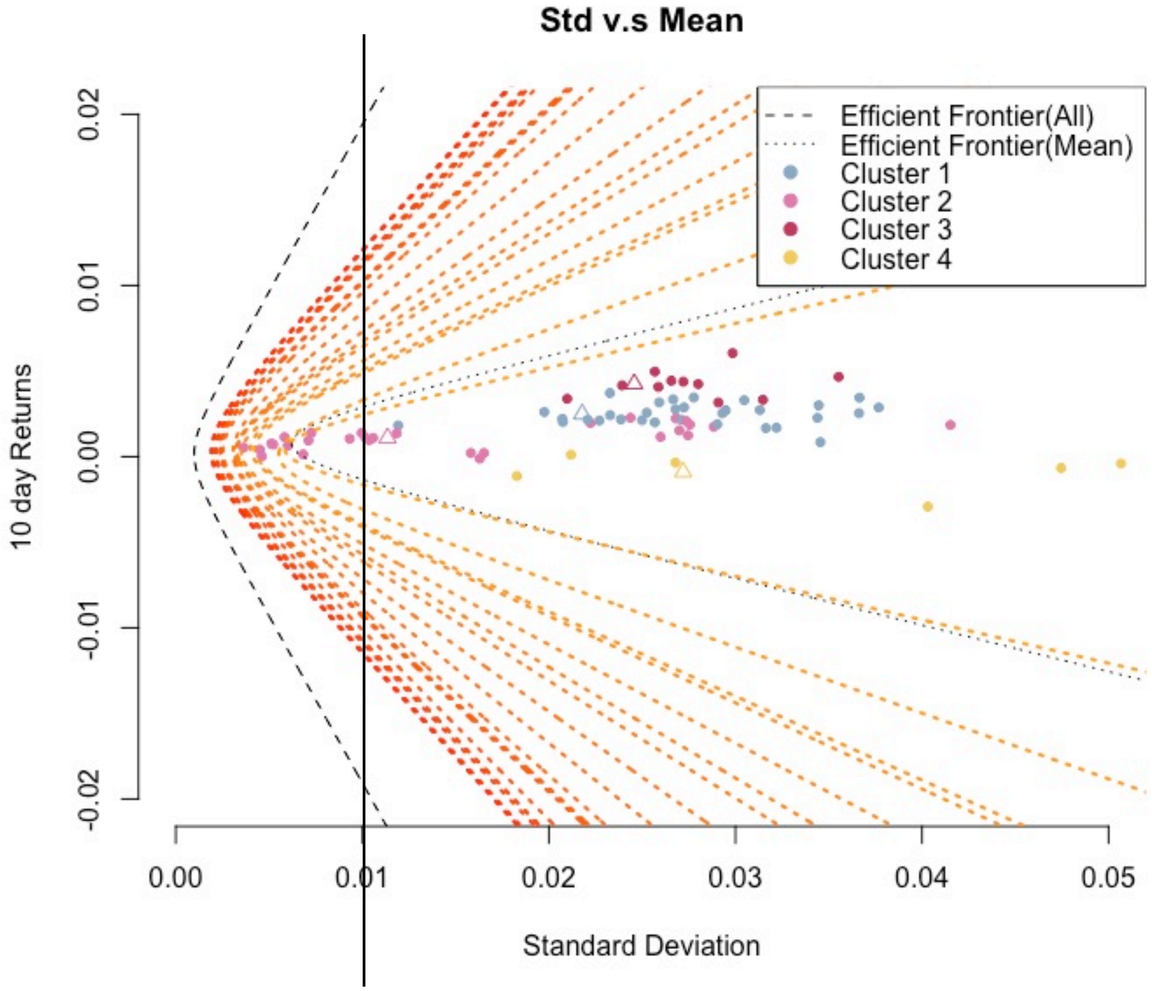
4.Result : DTW-HAC





Application N - Nearest Portfolio





We can consider the Transaction cost between return and standard deviation

Reference

- [1] Distance Measures for Effective Clustering of ARIMA Time-Series
  - Konstantinos Kalpakis

- [2] Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.
  - Toni Giorgino

- [3] K-Shape: Efficient and Accurate Clustering of Time Series
  - John Paparrizos

- [4]《影像學習筆記》：https://dotblogs.com.tw/dragon229/2013/02/04/89919

- [5] Clustering of time series data—a survey
  - T. Warren Liao