

# 非監督式學習 - 分群

財務金融系 張子溥 助理教授  
E-mail: [changtp@yuntech.edu.tw](mailto:changtp@yuntech.edu.tw)



# 非監督式學習 - 分群

財務金融系 張子溥 助理教授  
E-mail: [changtp@yuntech.edu.tw](mailto:changtp@yuntech.edu.tw)

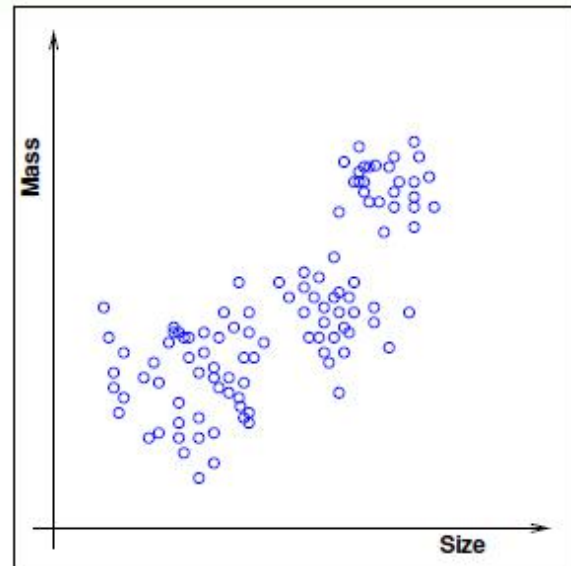
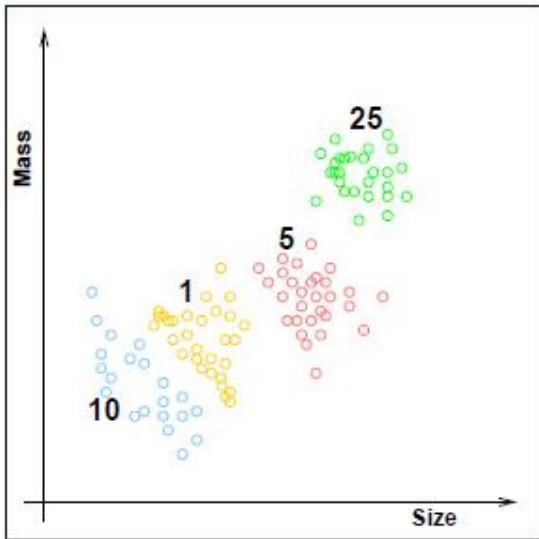


# 非監督式學習 - 分群

財務金融系 張子溥 助理教授  
E-mail: [changtp@yuntech.edu.tw](mailto:changtp@yuntech.edu.tw)



# 辨別硬幣



# 一天9美元，非洲人為矽谷人工智慧打工

- BBC記者Dave Lee撰文講述了人工智慧背後不為人知的貢獻者——生活在肯亞貧民窟的一群人。在非盈利組織Samasource的幫助下，他們為矽谷大型科技公司的人工智慧研究提供資料標註服務。



头条 @品途商业评论





# 非監督式學習

- 集群分析 (Clustering)
- 關聯規則學習
- 資料維度縮減



# 集群分析

- 集群分析可視為多變量分析(multivariate analysis)中精簡資料(data reduction)的一種技術，目的是企圖從一大堆雜亂無章的原始資料中，找出少數幾個較小的群體，使得群體內的分子在某些變項的測量值均很類似，而群體與群體間的分子在該測量值上差異較大，將一大筆資料精簡成少數幾個同質性次群體(homogeneous subgroups)，達到分群的目標。

~國家教育研究院



# 集群分析的兩大類技術

- 集群分析可分成分層法(Hierarchical)、非分層法(Nonhierarchical)
- 分層法有凝聚分層法(Agglomerative)和分離分層法(Divisive)。
- 非分層法最具代表性的為K 組平均法(K-Means)
- 另外也有兩階段法為第一階段分層法分群，決定群組個數，第二階段再以K 組平均法進行群集，移動各群組內的個體，保持全部群組為k 組。



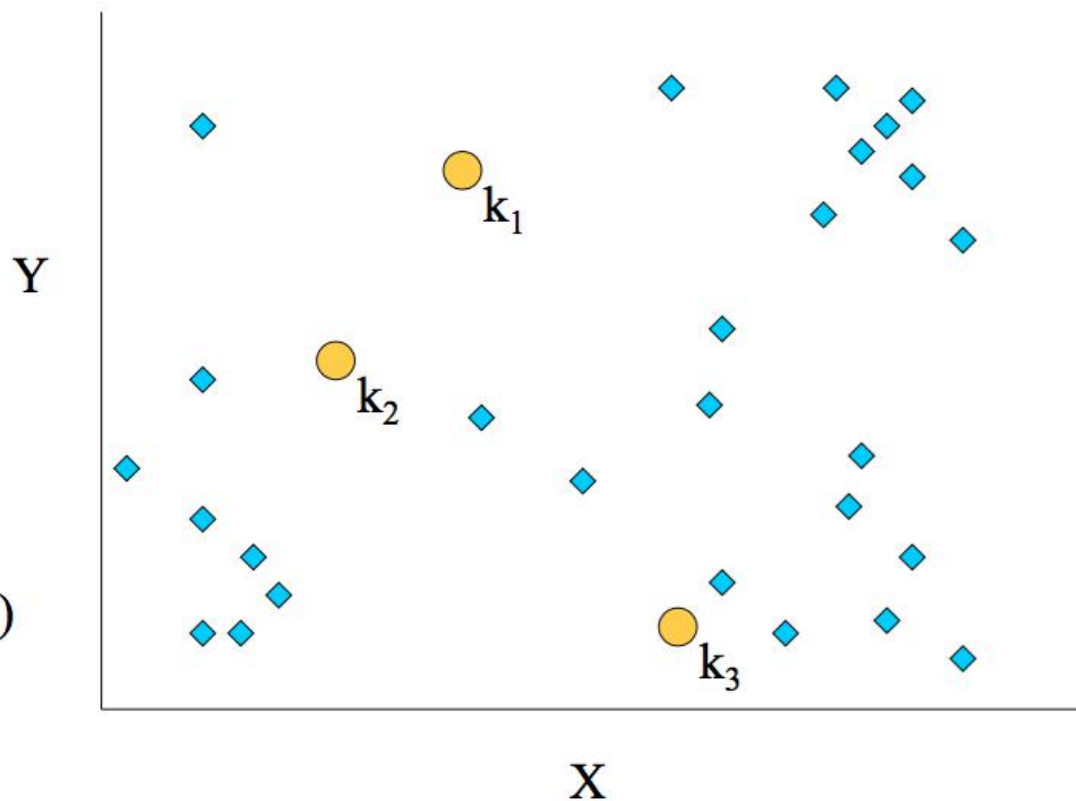


# K-Means

- Given  $k$ , the  $k$ -means algorithm is implemented in four steps:
  - Partition objects into  $k$  nonempty subsets
  - Compute seed points as the centroids of the clusters of the current partition (the centroid is the center, i.e., mean point, of the cluster)
  - Assign each object to the cluster with the nearest seed point
  - Go back to Step 2, stop when no more new assignment

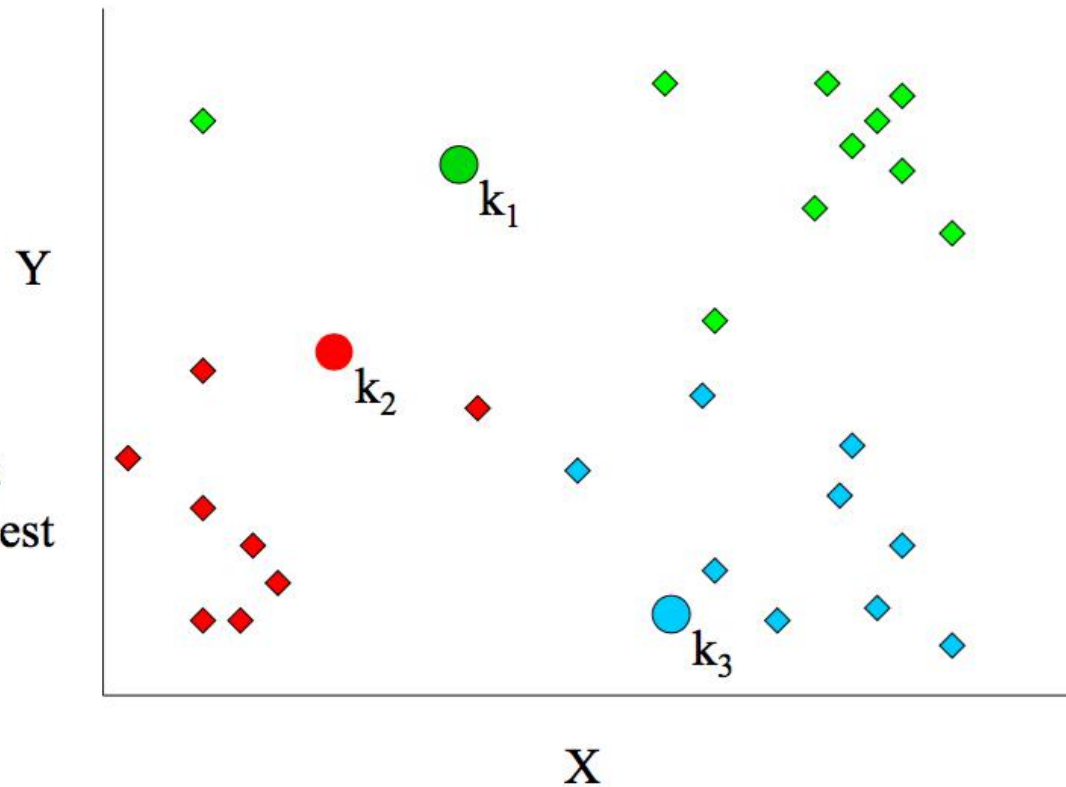


Pick 3  
initial  
cluster  
centers  
(randomly)

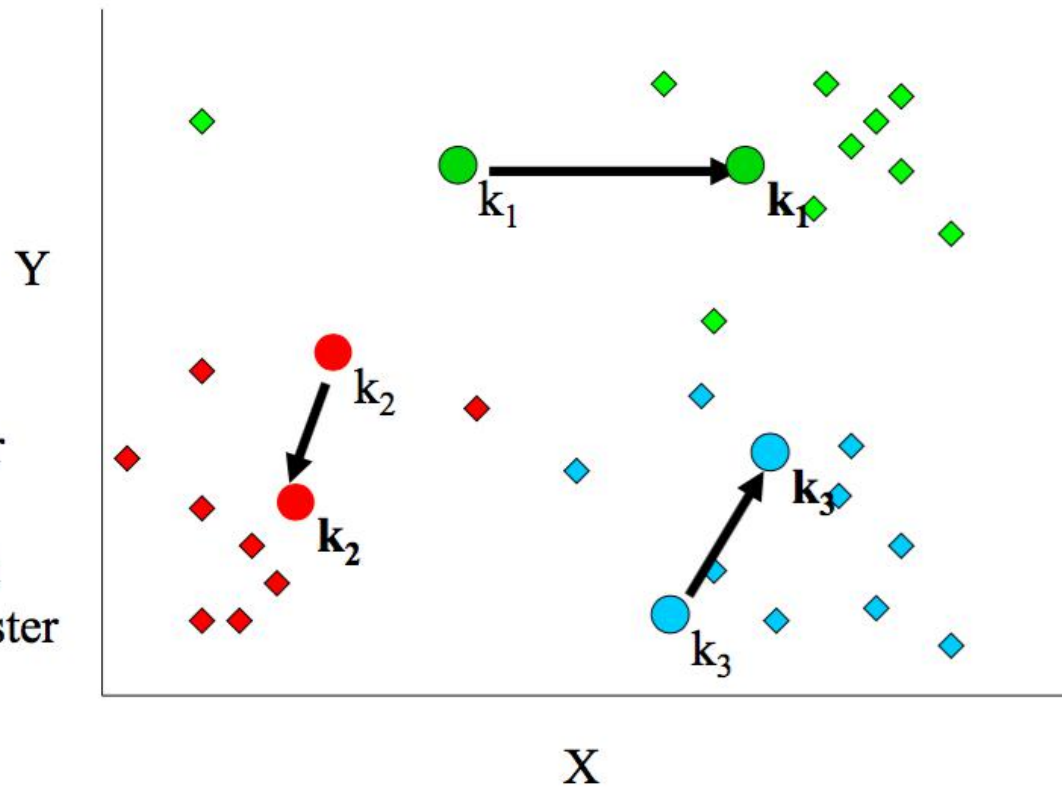




Assign  
each point  
to the closest  
cluster  
center



Update  
each cluster  
center  
to the mean  
of each cluster

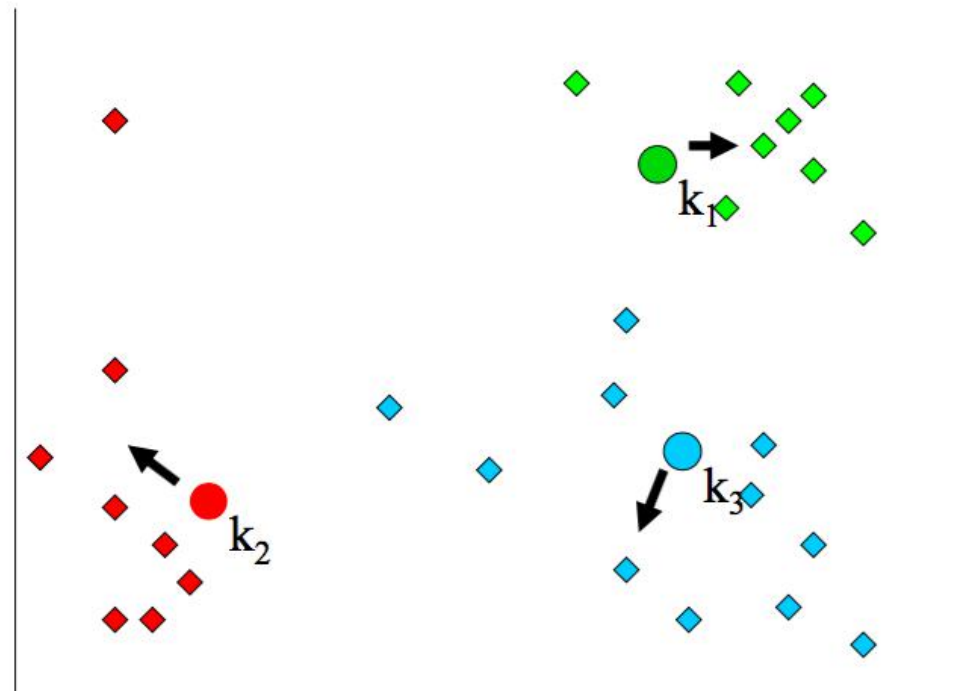






re-compute  
cluster  
means

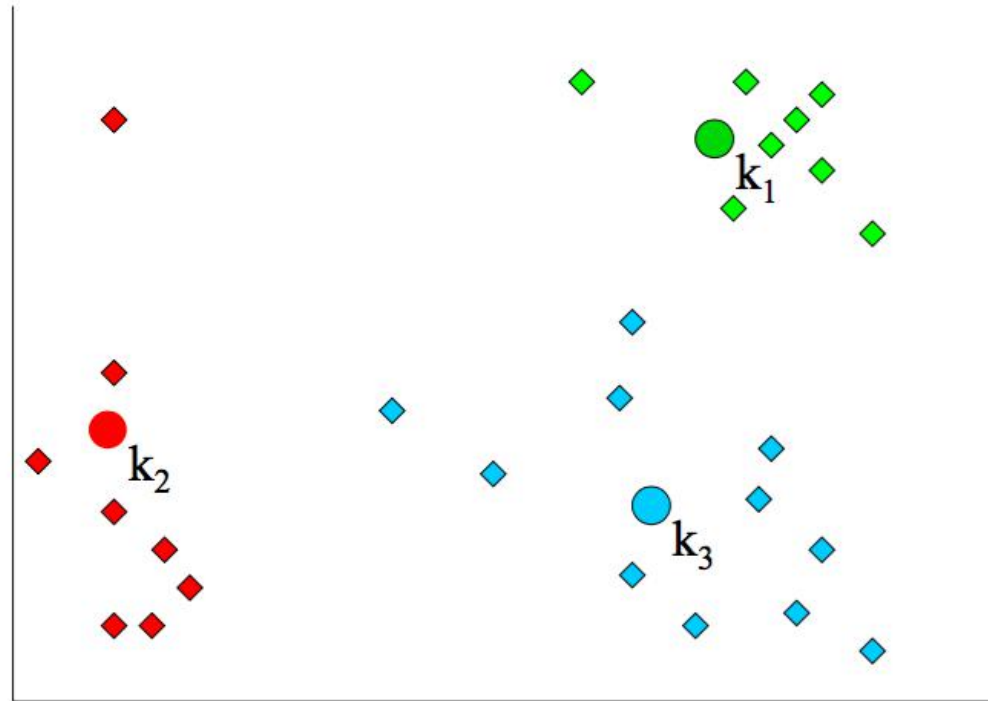
Y



X

update cluster  
centers to  
cluster means

Y



X



# Hierarchical Clustering- Agglomerative

- 開始時每一個體為一群，然後最近的兩個體合成一群，一次結合使群組越變越少，最後所有個體結合成一群。依不同的『群間距離』分為，
  - 最近法(單一聯結法Single Linkage)
  - 最遠法(完全聯結法Complete Linkage)
  - 平均法(Average Linkage)
  - 中心法(Centroid Method)
  - 華德法(華德最小變異法Wards Method)





# 凝聚分層法(Agglomerative)圖示

