



北京交通大学

# 图像处理与机器学习

Digital Image Processing and Machine Learning

主讲人：彭亚辉

电子信息工程学院



# 图像处理与机器学习

## ➤ 非监督学习

- 监督与非监督学习简介
- 常见的非监督学习算法
- K-means聚类算法简介
- 本讲小结



# 非监督学习

## ➤ 教学目标

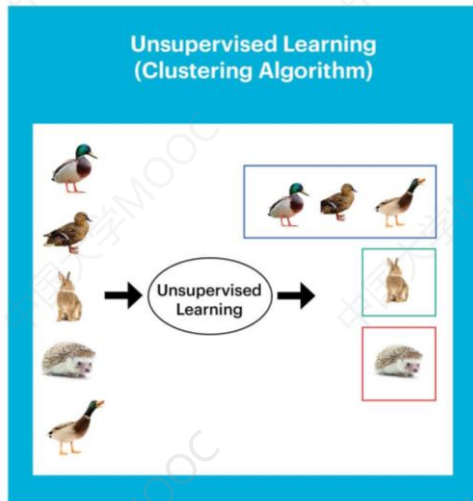
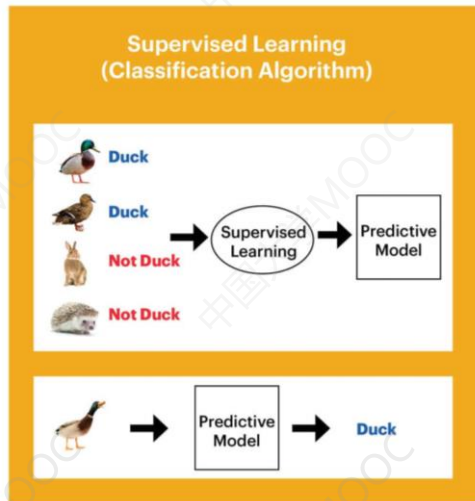
- 掌握非监督学习的概念
- 了解常见非监督算法
- 理解k-means聚类算法



# 非监督学习

## ➤ 监督学习与非监督学习

-- 学习过程中是否提供了正确答案





# 非监督学习

## ➤ 什么是“非监督学习”

- 一类不依赖样本标签的机器学习技术
- 对数据本身进行有价值和有意义的整合
- 可以对数据进行分组（分类）
- 也可以对数据进行关联分析等等



# 非监督学习

## ➤ “非监督学习” 的特点

- 不需要“标签”，降低了学习成本
- 计算代价往往比较高
- 学习的目标存在不确定性
- 学习的结果可能不唯一



# 非监督学习

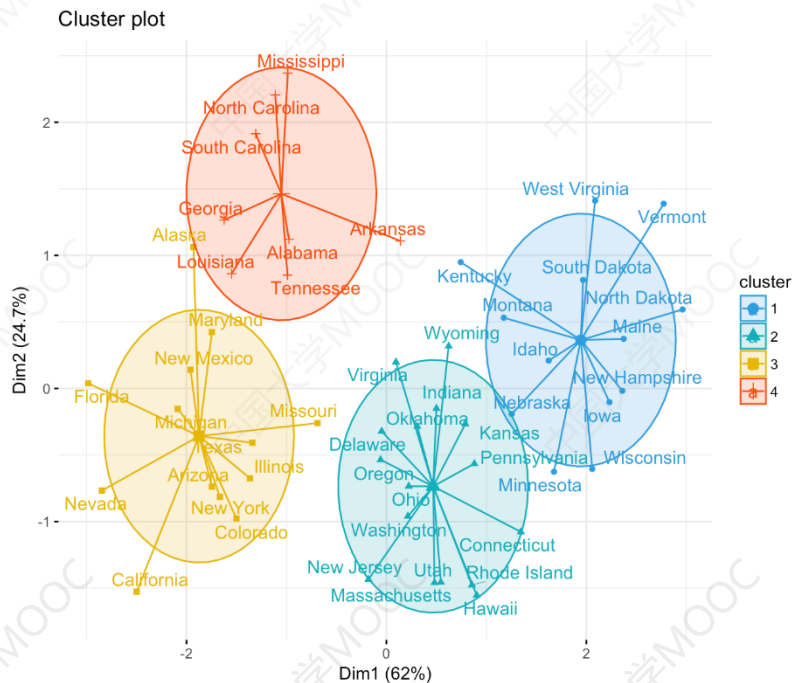
## ➤ 非监督学习概念小结

- “非监督”等价于无“标签”
- 学习的结果只由输入数据本身决定
- 反映的是输入数据的内在结构



## 常见的非监督学习

## -- 聚类





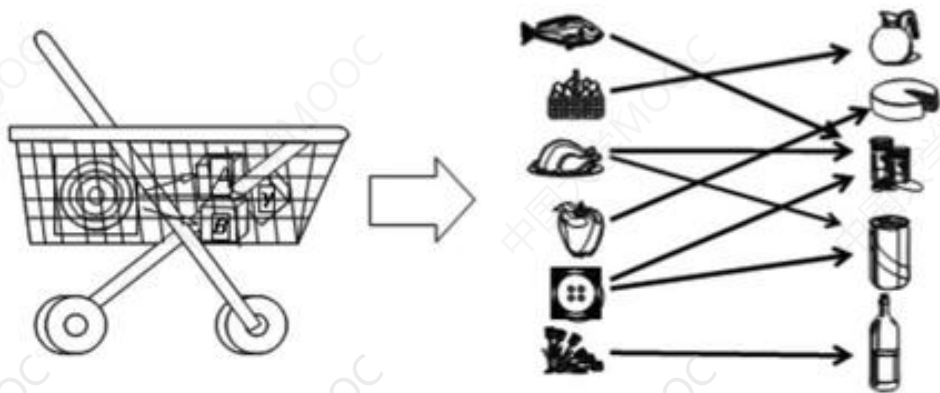


# 非监督学习

## 常见的非监督学习

### -- 关联 (Association) 分析

#### MARKET BASKET ANALYSIS



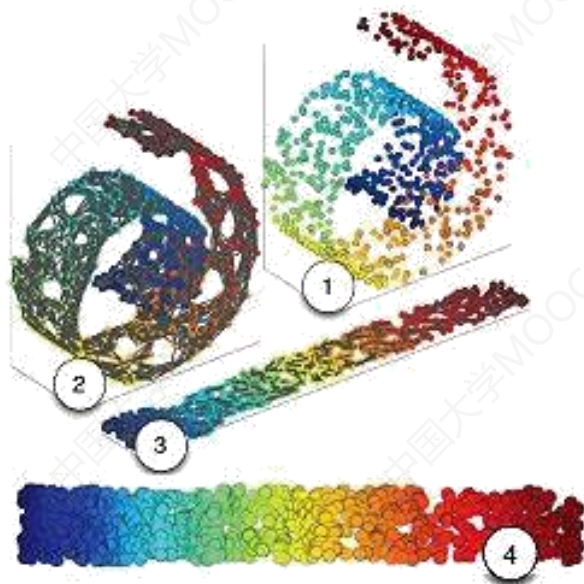
98% of people who purchased items A and B  
also purchased item C



# 非监督学习

## ➤ 常见的非监督学习

### -- 数据降维





## -- 聚类



## -- 数据降维



## MARKET BASKET ANALYSIS





# 聚类

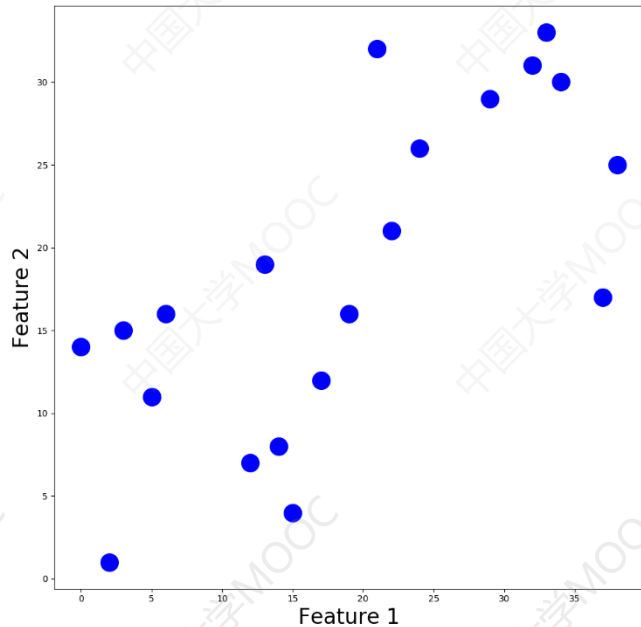
## ➤ K-means算法

- 目标：将 $n$ 个数据点分为 $k(<n)$ 组
- 算法：
  0. (随机) 设定每组数据的中心点
  1. 将数据归为最近中心点代表的组别
  2. 重新计算中心点
  3. 循环执行步骤1和2，直至收敛
- 最终稳定的数据归属即聚类结果



# 聚类

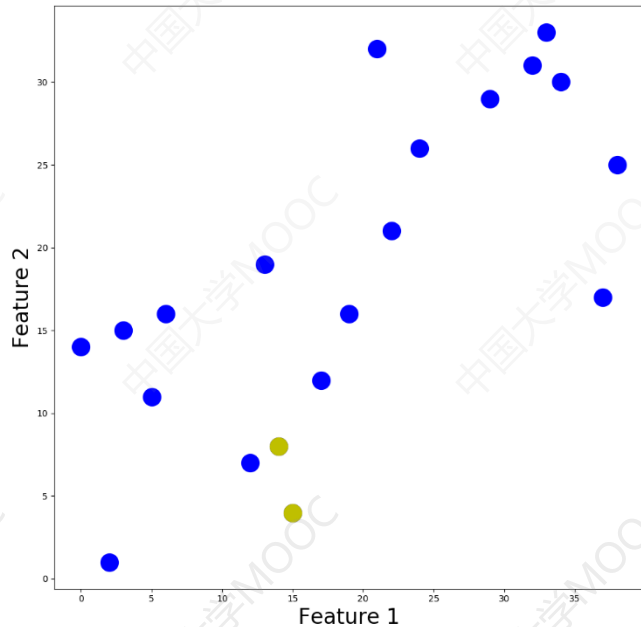
## ➤ K-means算法





# 聚类

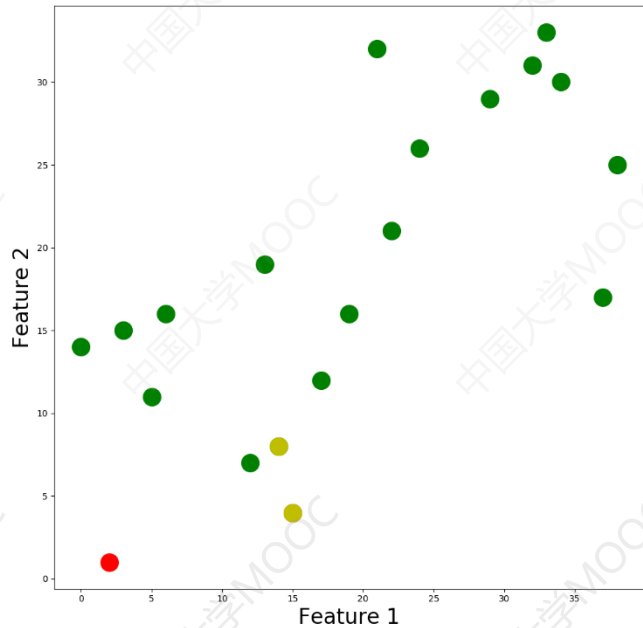
## ➤ K-means算法





# 聚类

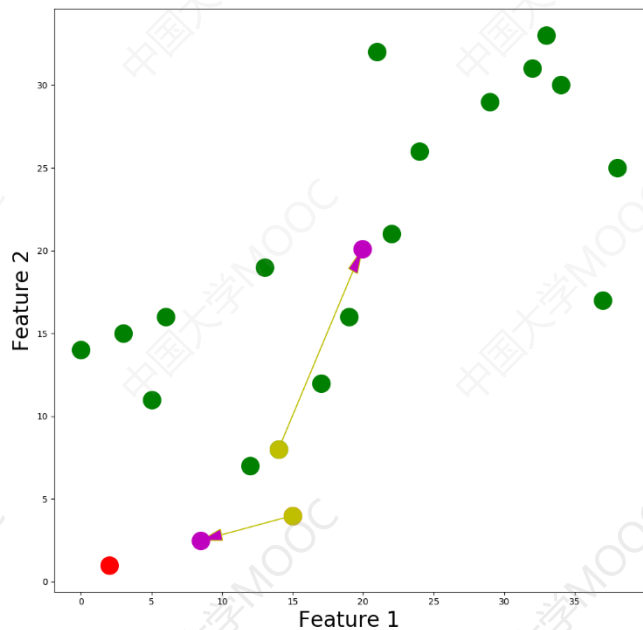
## ➤ K-means算法





# 聚类

## ➤ K-means算法

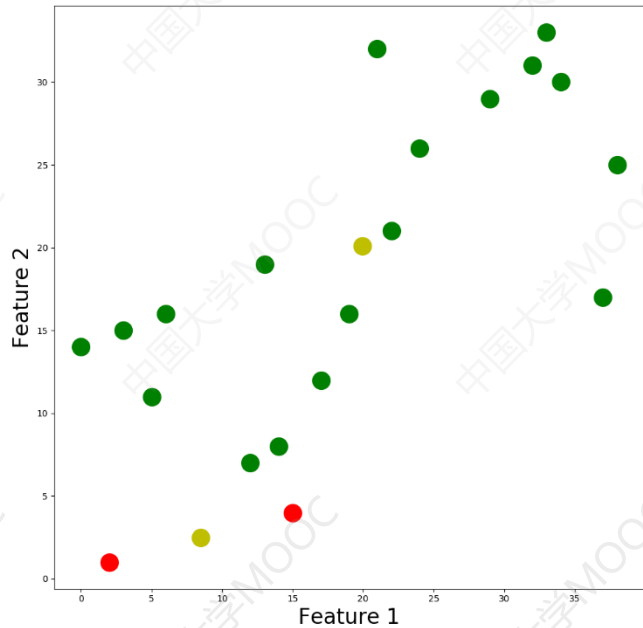






# 聚类

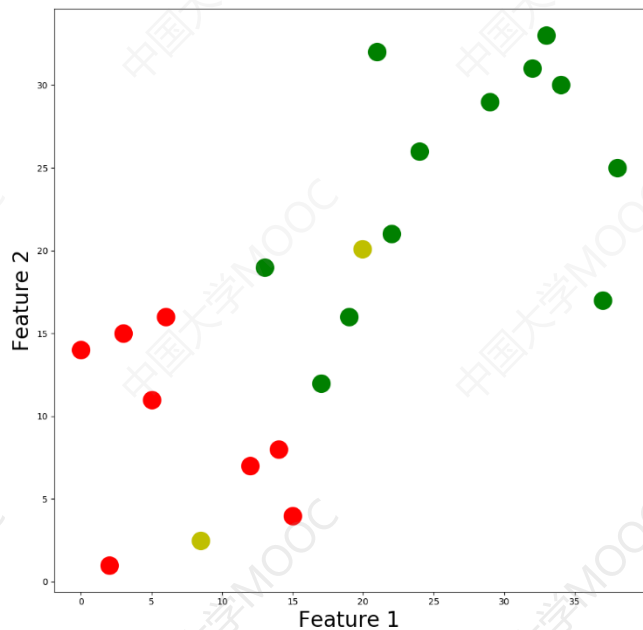
## ➤ K-means算法





# 聚类

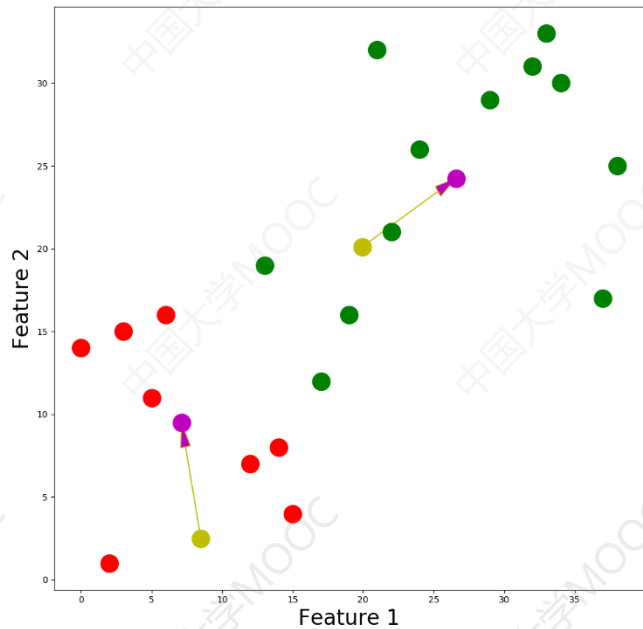
## ➤ K-means算法





# 聚类

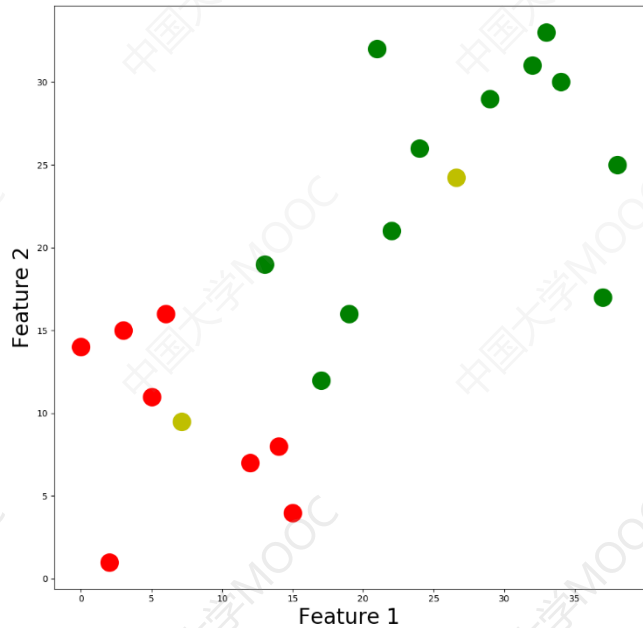
## ➤ K-means算法





# 聚类

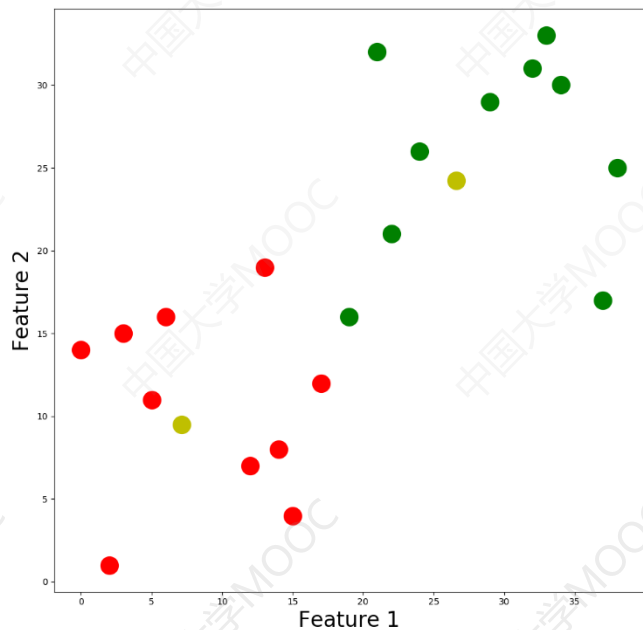
## ➤ K-means算法





# 聚类

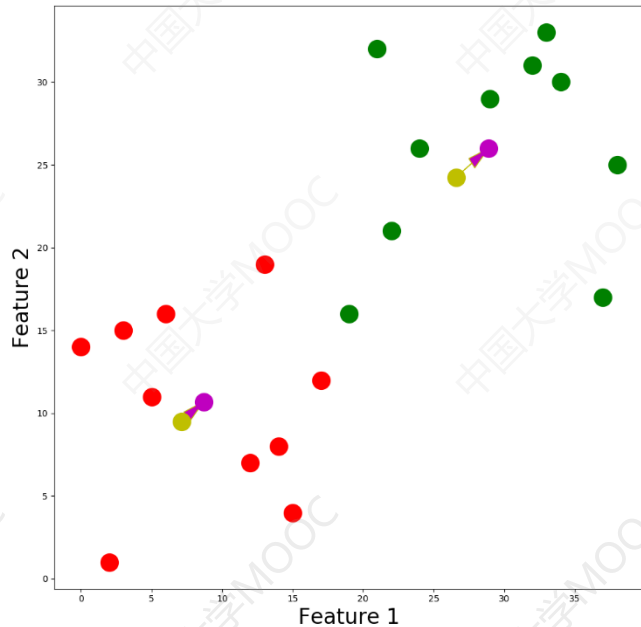
## ➤ K-means算法





# 聚类

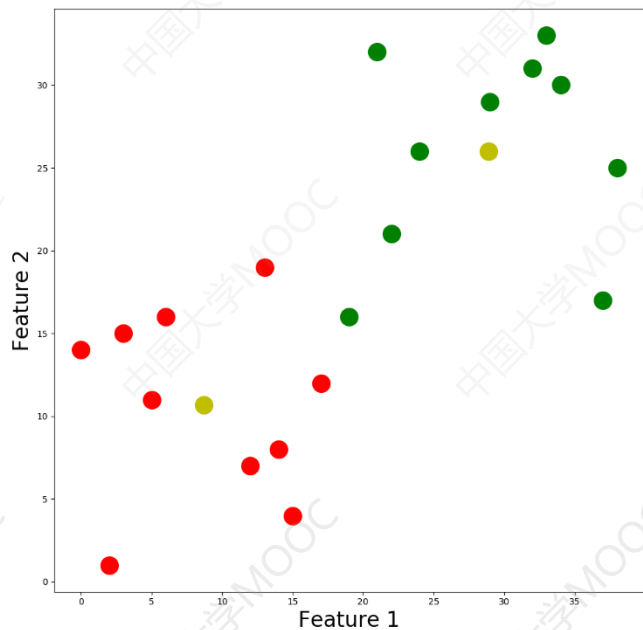
## ➤ K-means算法





# 聚类

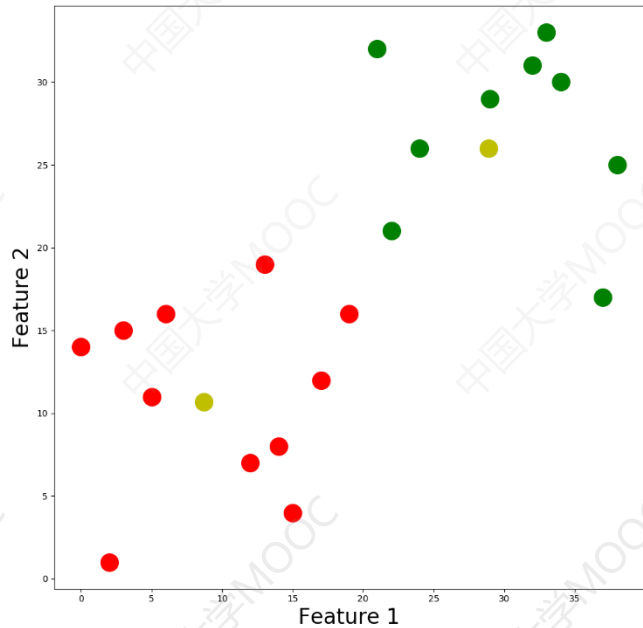
## ➤ K-means算法





# 聚类

## ➤ K-means算法

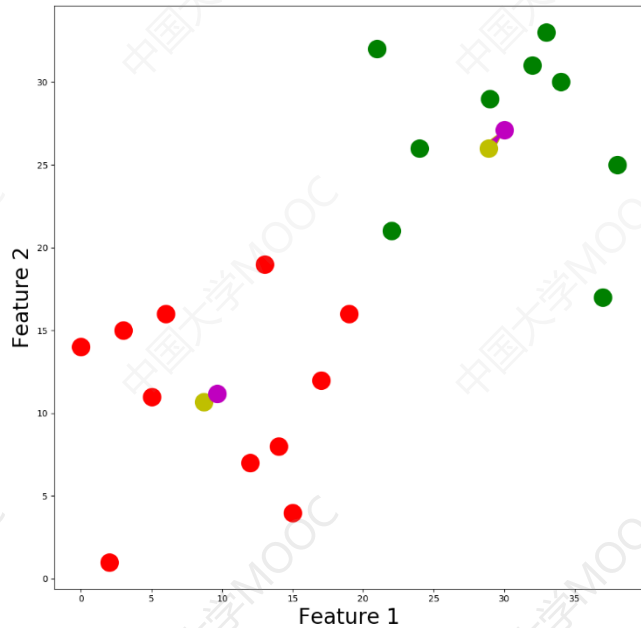






# 聚类

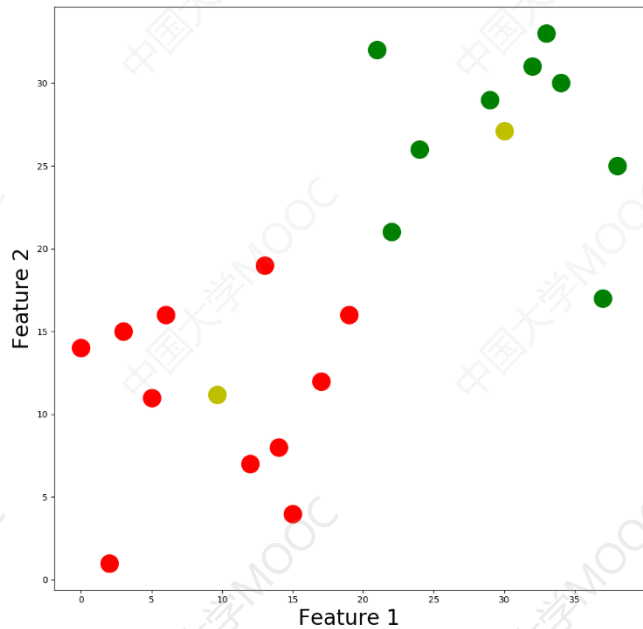
## ➤ K-means算法





# 聚类

## ➤ K-means算法





# 聚类

## ➤ K-means算法

-- 主要优点:

简单

快速

-- 主要缺点:

不一定是最优解

组别数目需要预先确定



# 非监督学习

## ➤ 教学目标

-- 掌握非监督学习的概念

标签／答案

-- 了解常见非监督算法

三类应用

-- 理解k-means聚类算法

中心，归属



# 谢 谢

本课程所引用的一些素材为主讲老师多年的教学积累，来源于多种媒体及同事和同行的交流，难以一一注明出处，特此说明并表示感谢！