法律声明

- 本课件包括演示文稿、示例、代码、题库、视频和声音等内容，小象学院和主讲老师拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意及内容，我们保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

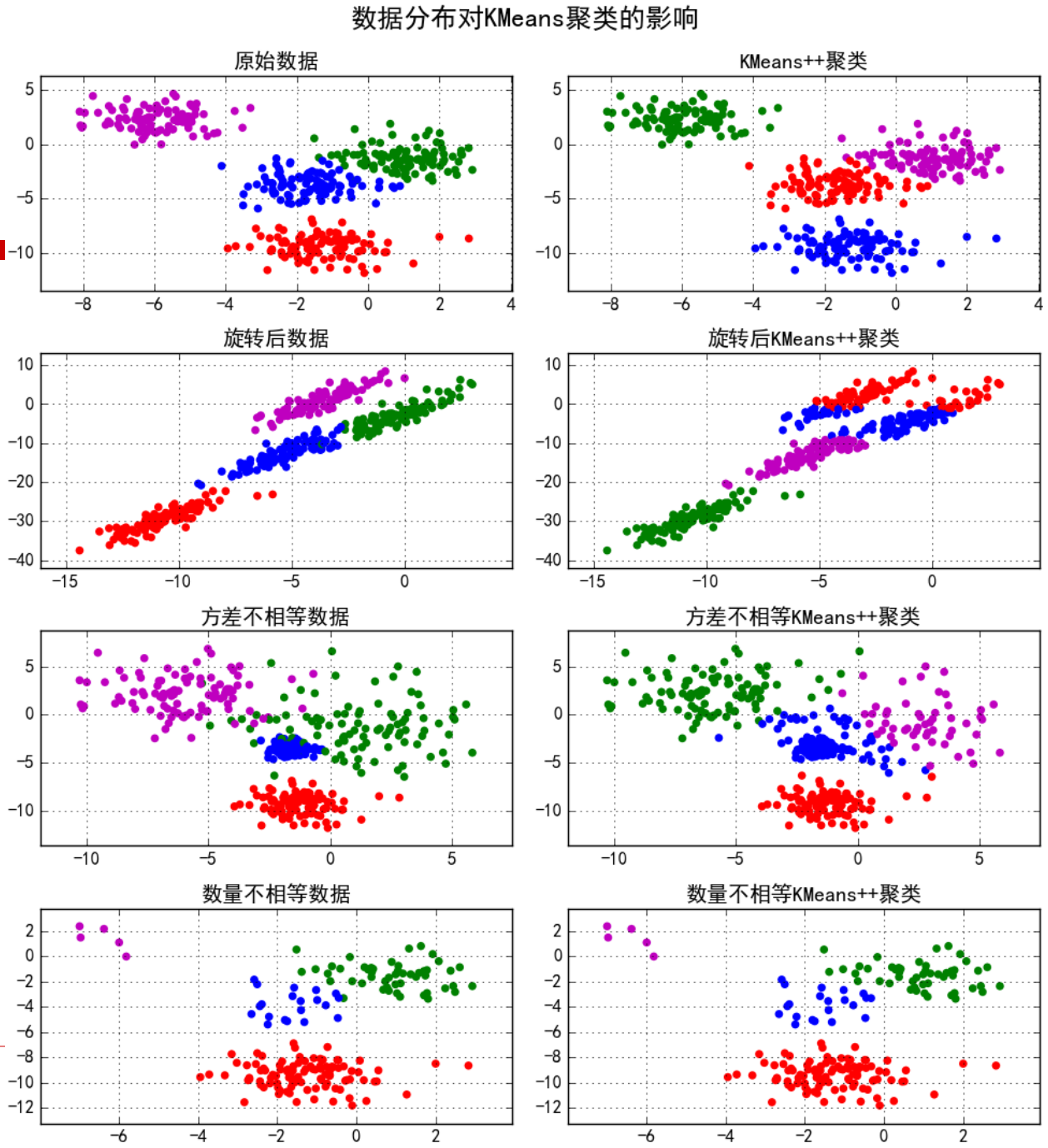
聚类实践

邹博

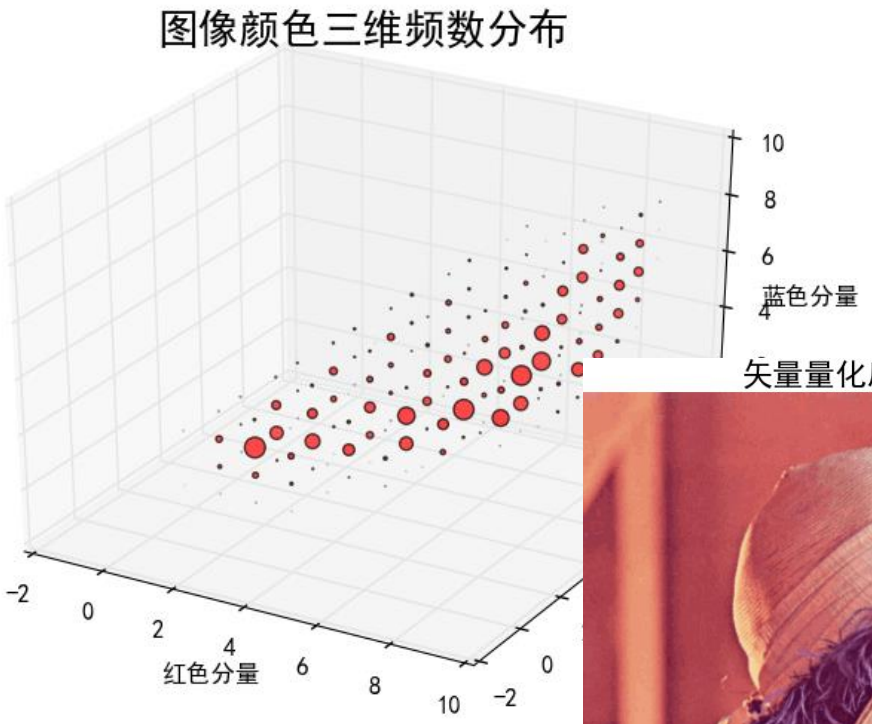
本次目标

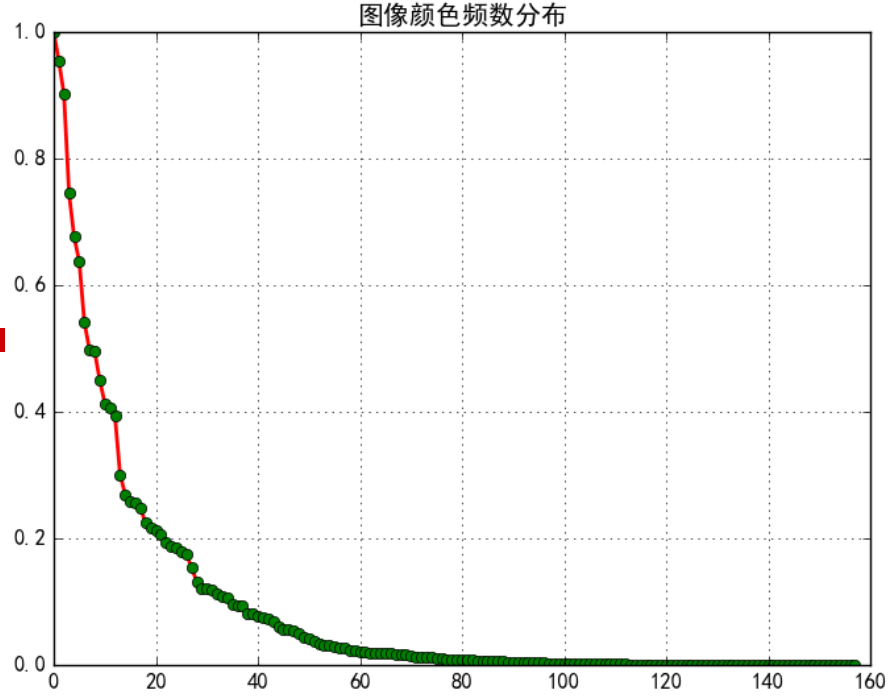
- k-Means算法及其适用范围
- VQ算法及其应用
  - 图像压缩
- AP算法
- MeanShift算法
- DBSCAN算法
- 谱聚类SC

k-Means

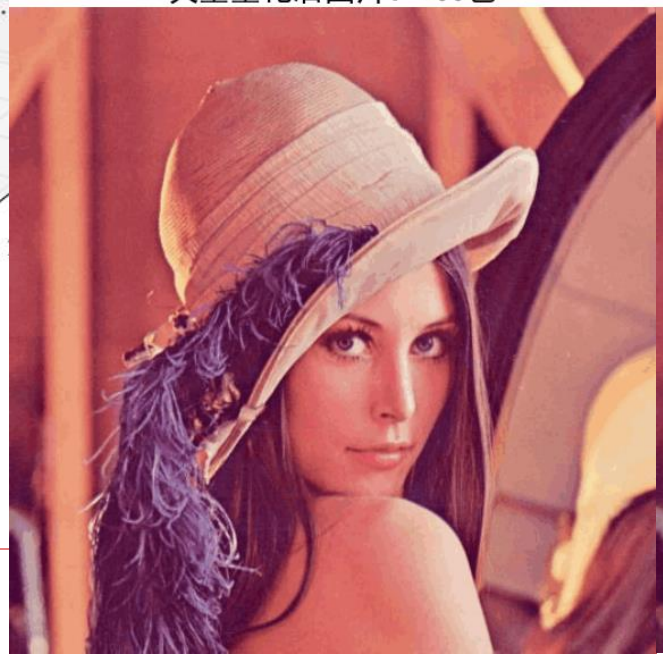

数据分布对KMeans聚类的影响
原始数据
KMeans++聚类
旋转后数据
旋转后KMeans++聚类
方差不相等数据
方差不相等KMeans++聚类
数量不相等数据
数量不相等KMeans++聚类

Vector Quantization





矢量量化后图片：100色

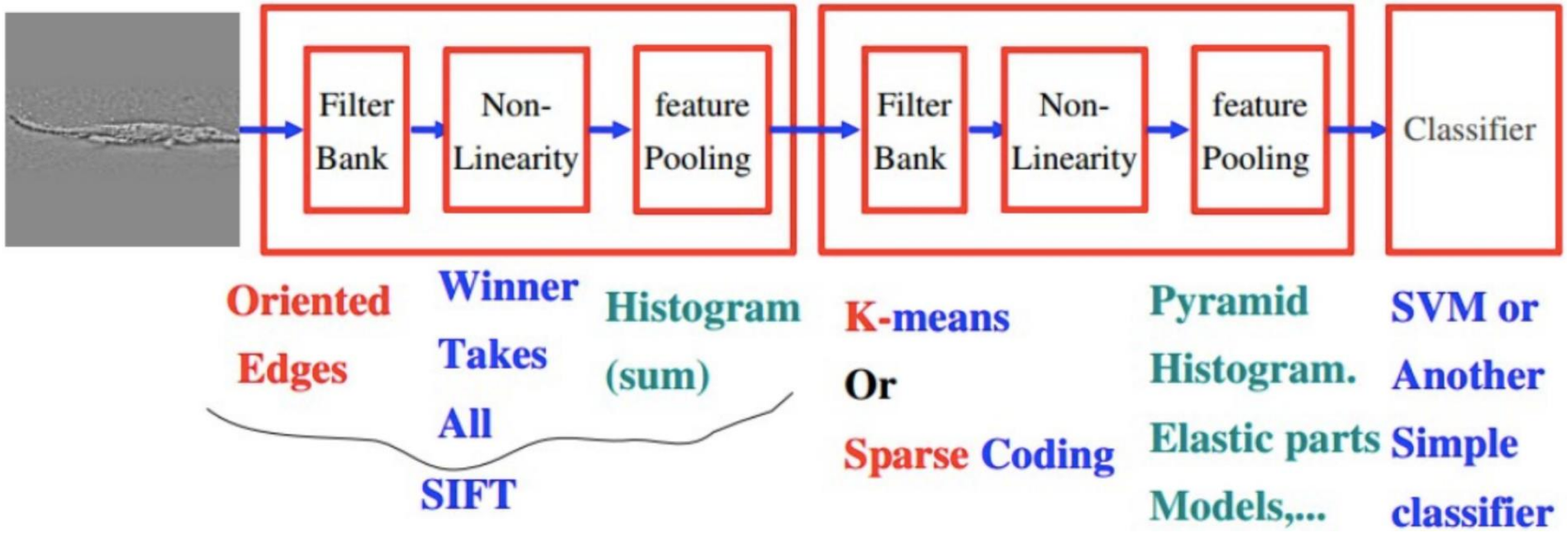


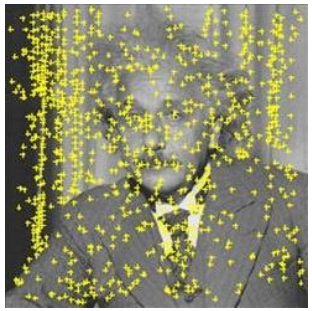
原始图片

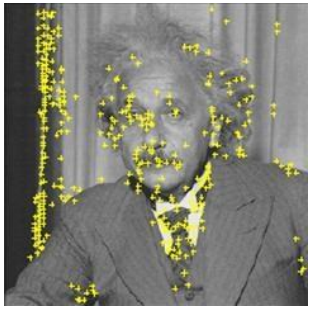
传统模型

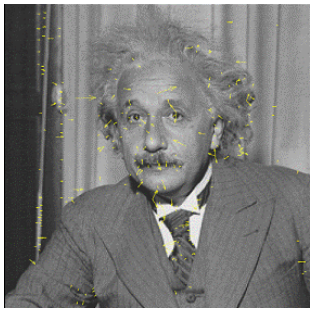
- Fixed Features + unsupervised mid-level features + simple classifier
  - SIFT + Vector Quantization + Pyramid pooling + SVM
    - [Lazebnik et al. CVPR 2006]
  - SIFT + Local Sparse Coding Macrofeatures + Pyramid pooling + SVM
    - [Boureau et al. ICCV 2011]
  - SIFT + Fisher Vectors + Deformable Parts Pooling + SVM
    - [Perronin et al. 2012]

传统图像分类


Filter Bank
Non-Linearity
feature Pooling
Filter Bank
Non-Linearity
feature Pooling
Classifier
Oriented Edges
Winner Takes All
Histogram (sum)
SIFT
K-means Or Sparse Coding
Pyramid Histogram. Elastic parts Models,...
SVM or Another Simple classifier







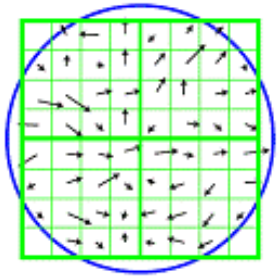
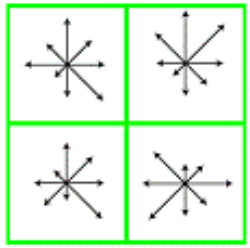
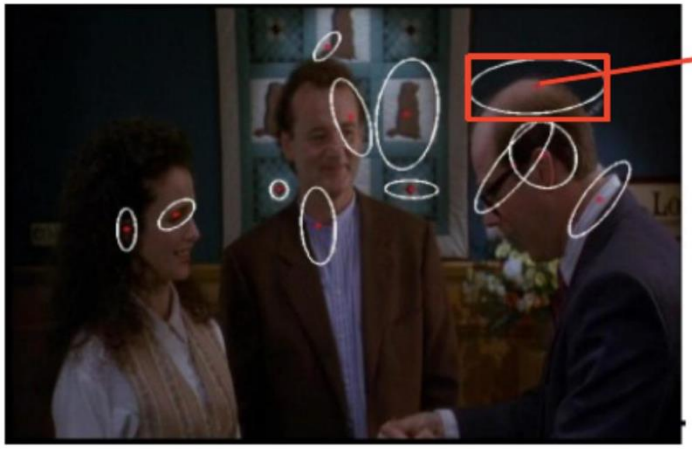


Image gradients


Keypoint descriptor

词袋模型bag of words

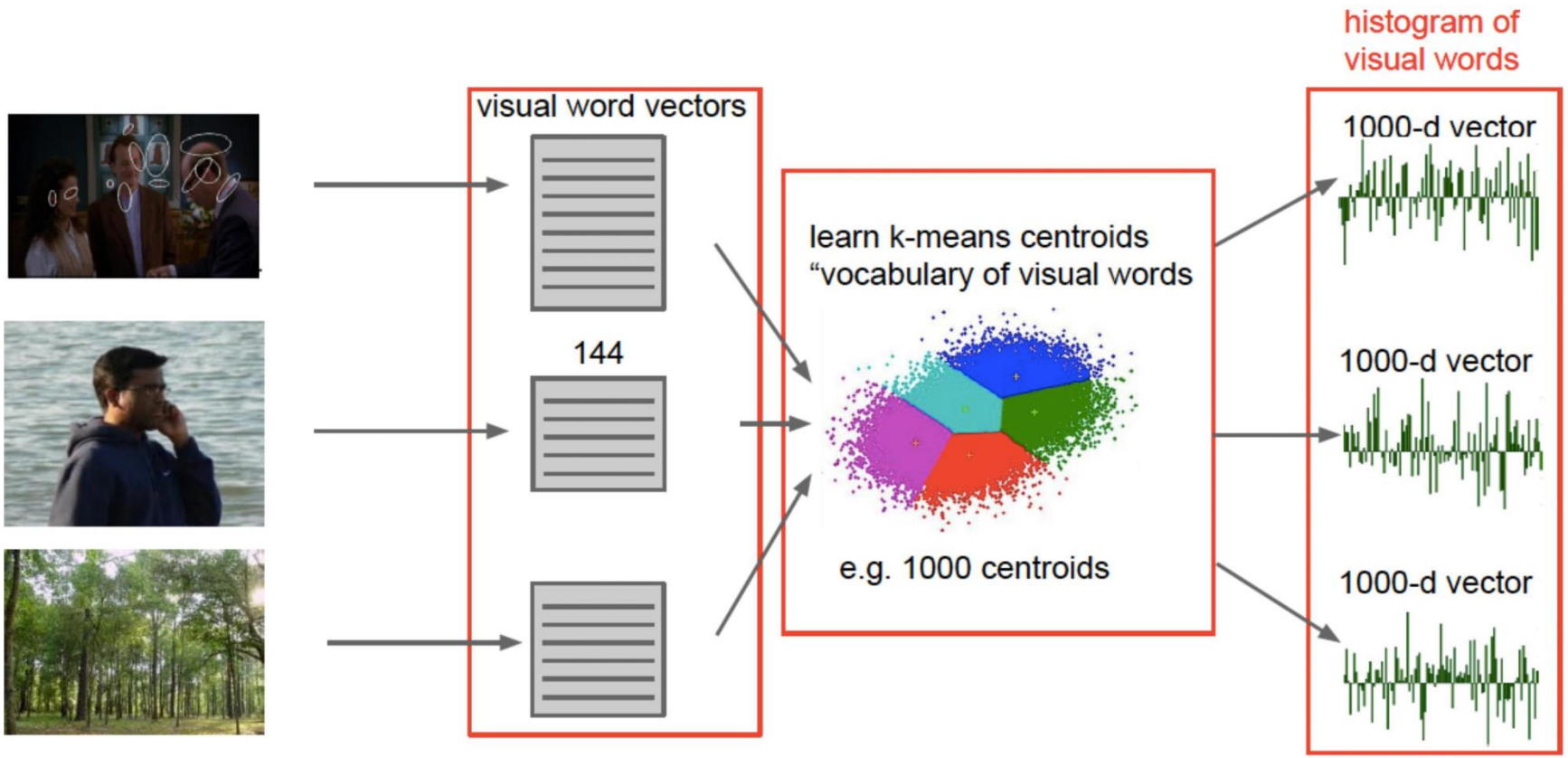


1. Resize patch to a fixed size (e.g. 32x32 pixels)
2. Extract HOG on the patch (get 144 numbers)

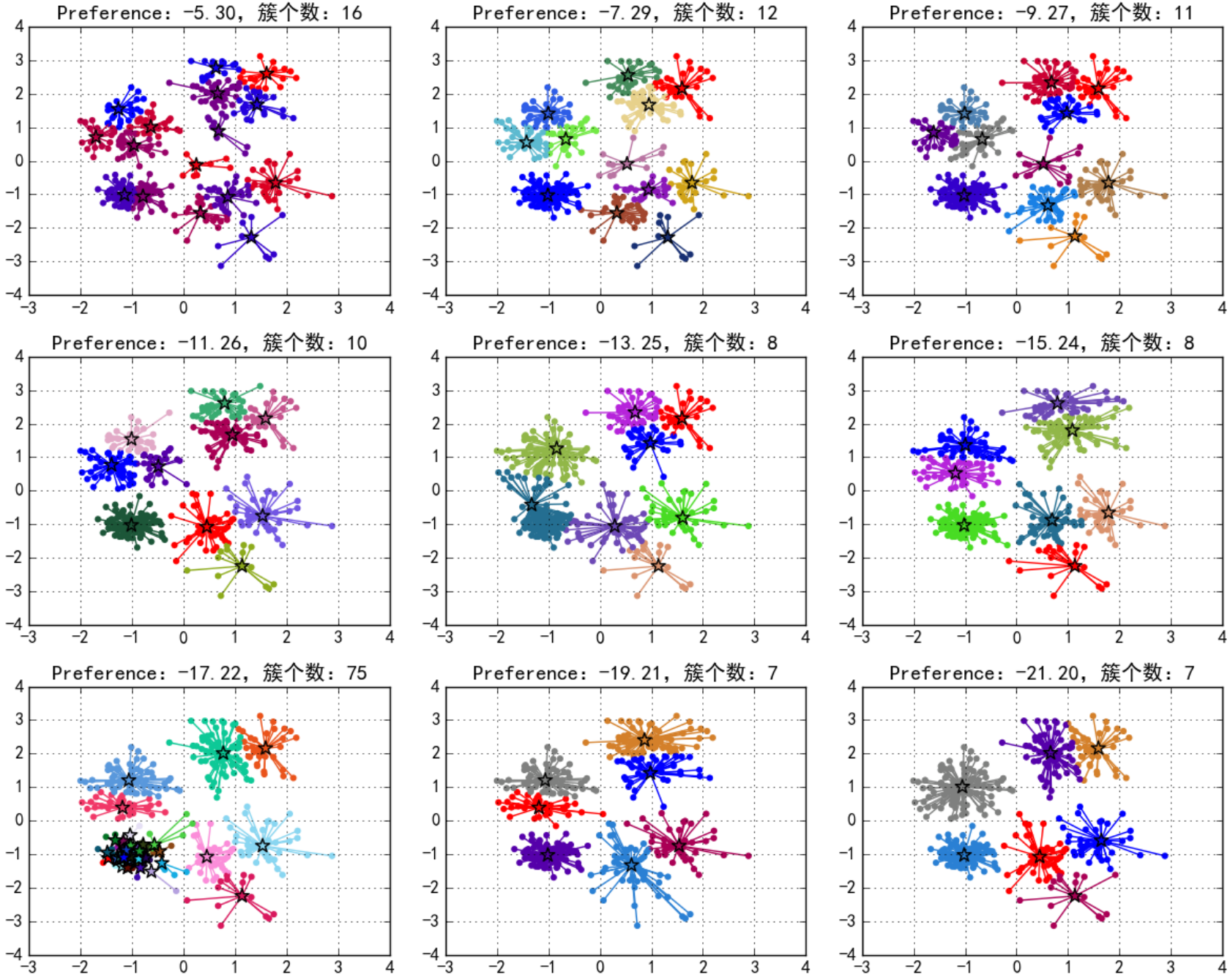
repeat for each detected feature

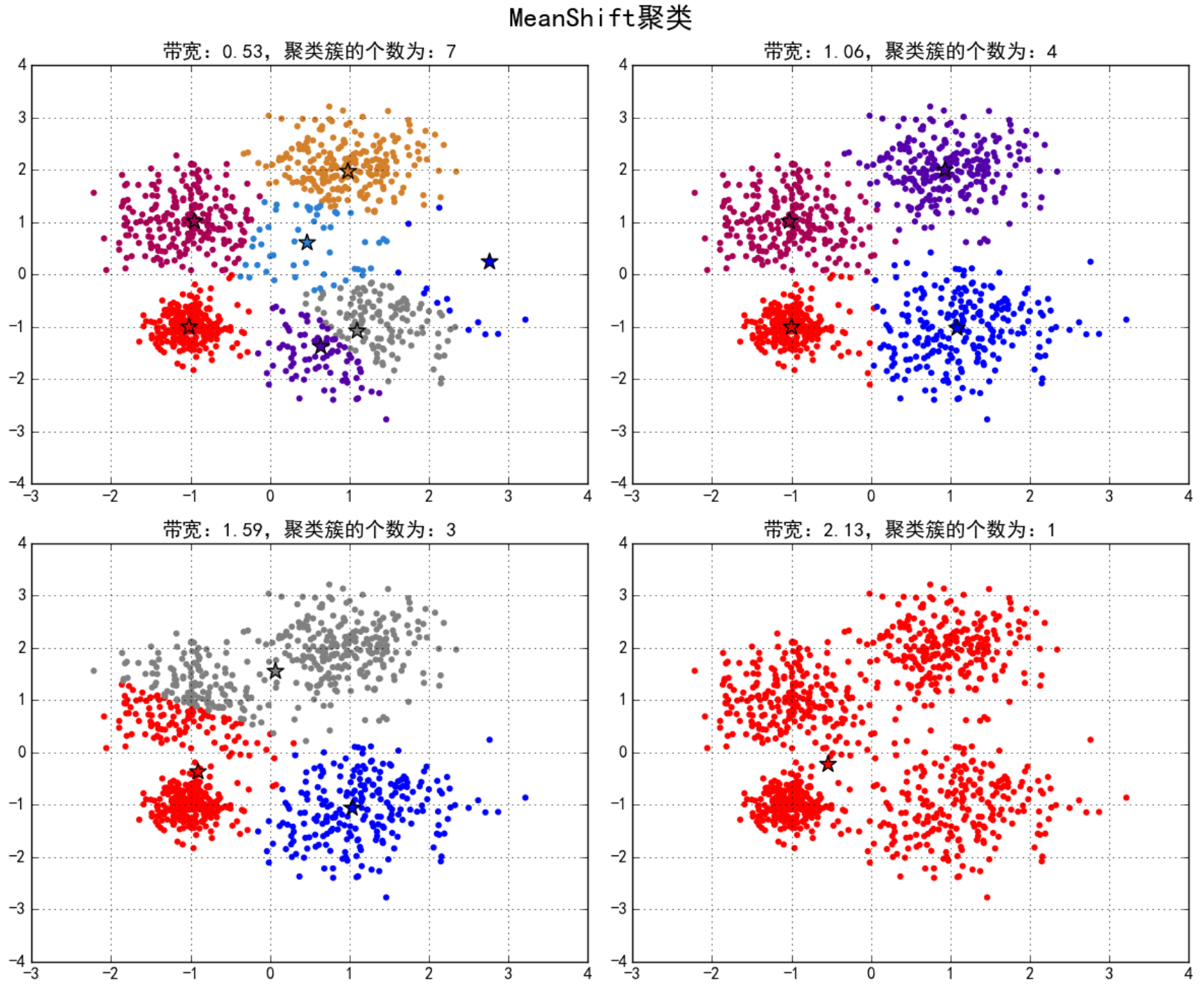
gives a matrix of size
[number_of_features x 144]

词袋模型bag of words

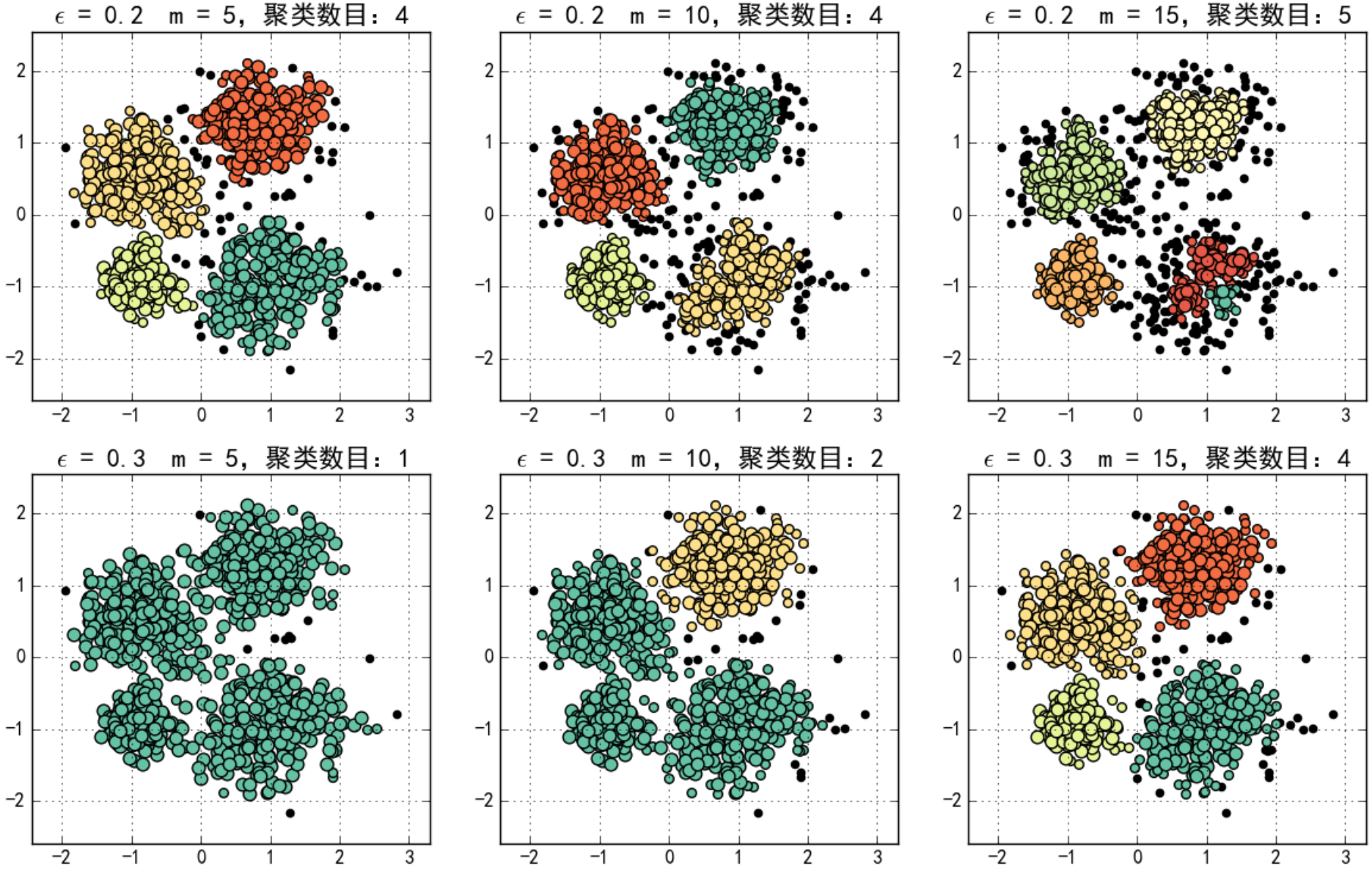


AP聚类

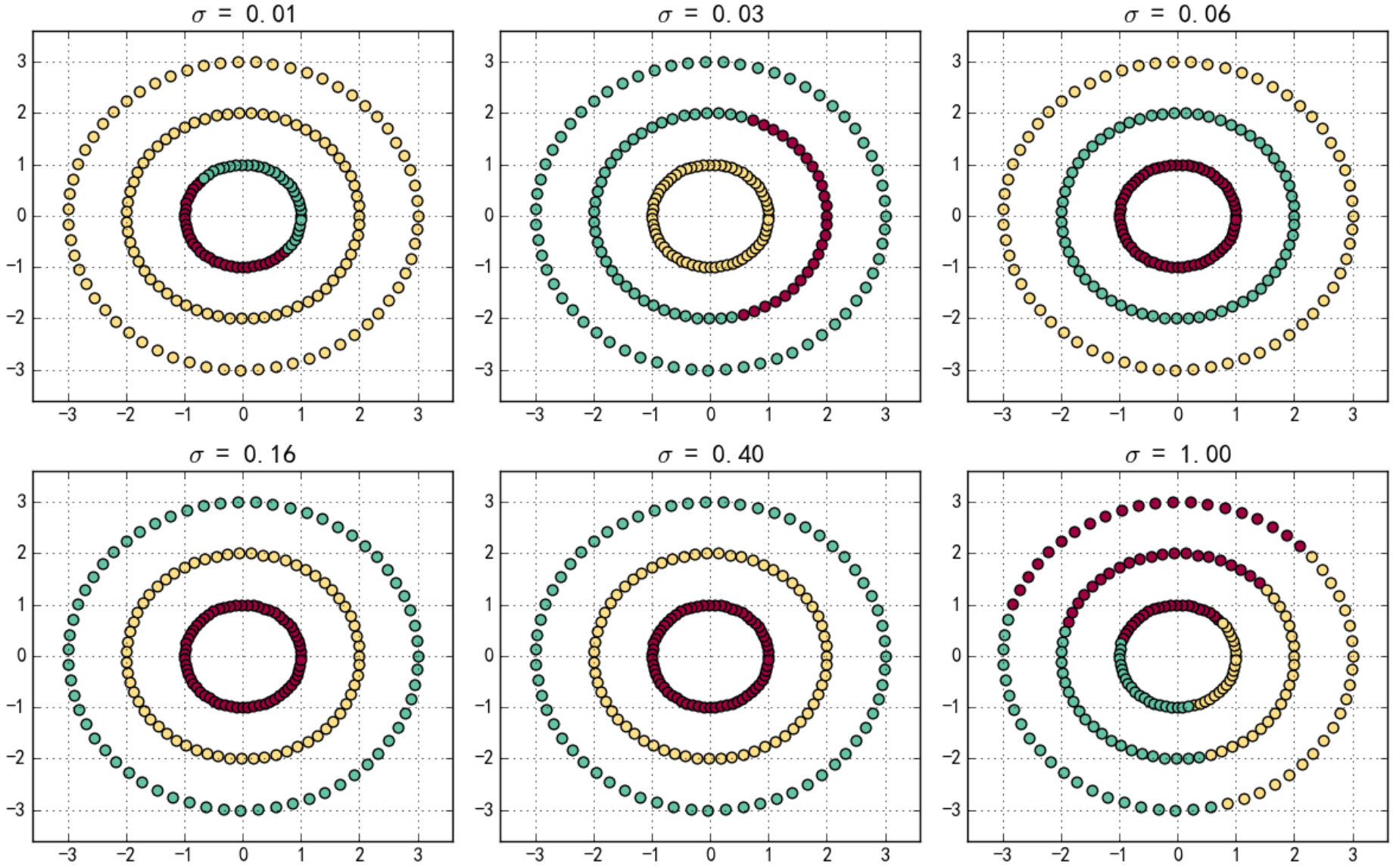

Preference：-5.30，簇个数：16
Preference：-7.29，簇个数：12
Preference：-9.27，簇个数：11
Preference：-11.26，簇个数：10
Preference：-13.25，簇个数：8
Preference：-15.24，簇个数：8
Preference：-17.22，簇个数：75
Preference：-19.21，簇个数：7
Preference：-21.20，簇个数：7


MeanShift聚类
带宽：0.53，聚类簇的个数为：7
带宽：1.06，聚类簇的个数为：4
带宽：1.59，聚类簇的个数为：3
带宽：2.13，聚类簇的个数为：1

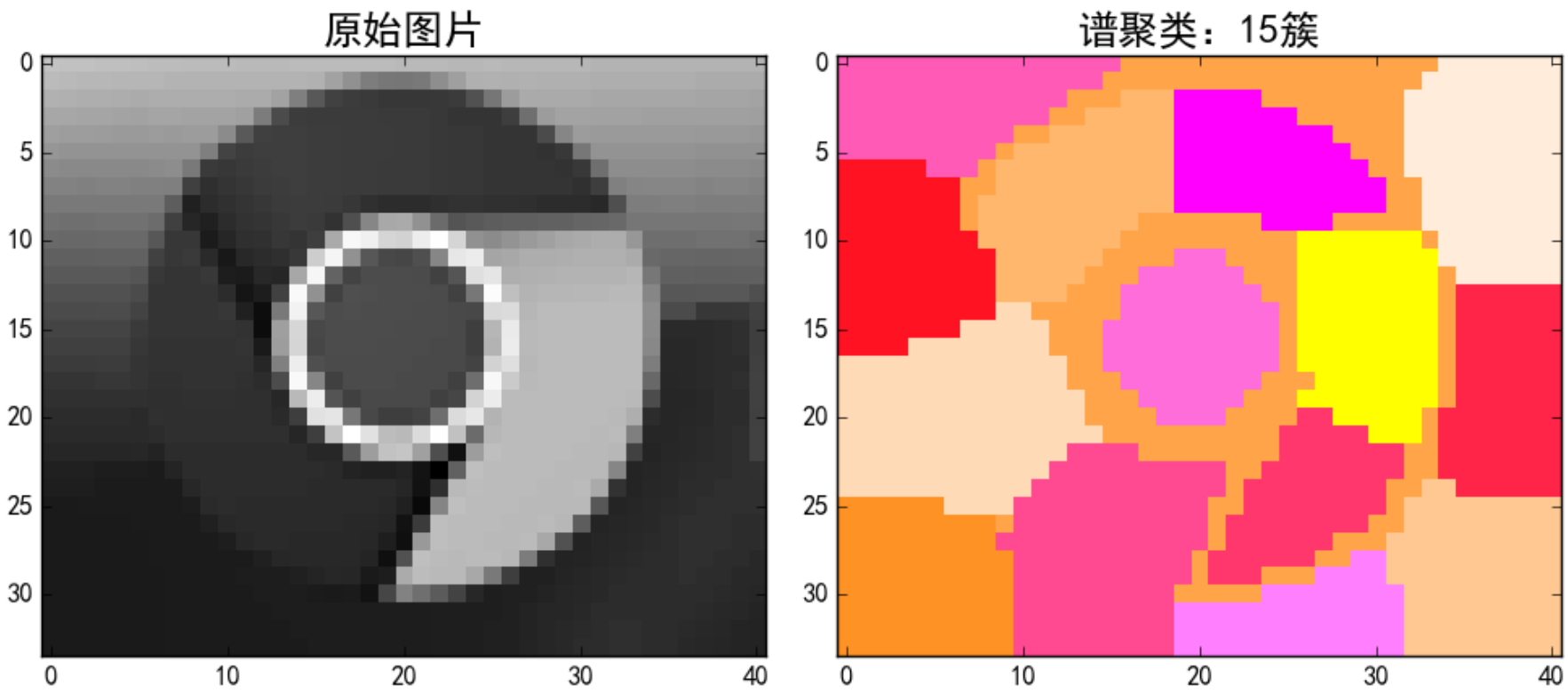
DBSCAN聚类


ϵ = 0.2 m = 5，聚类数目：4
ϵ = 0.2 m = 10，聚类数目：4
ϵ = 0.2 m = 15，聚类数目：5
ϵ = 0.3 m = 5，聚类数目：1
ϵ = 0.3 m = 10，聚类数目：2
ϵ = 0.3 m = 15，聚类数目：4

谱聚类



谱聚类与图像切割



我们在这里

- http://wenda.ChinaHadoop.cn
  - 视频/课程/社区
- 微博
  - @ChinaHadoop
  - @邹博_机器学习
- 微信公众号
  - 小象学院
  - 大数据分析挖掘

感谢大家！

恳请大家批评指正！