

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



小象学院
ChinaHadoop.cn

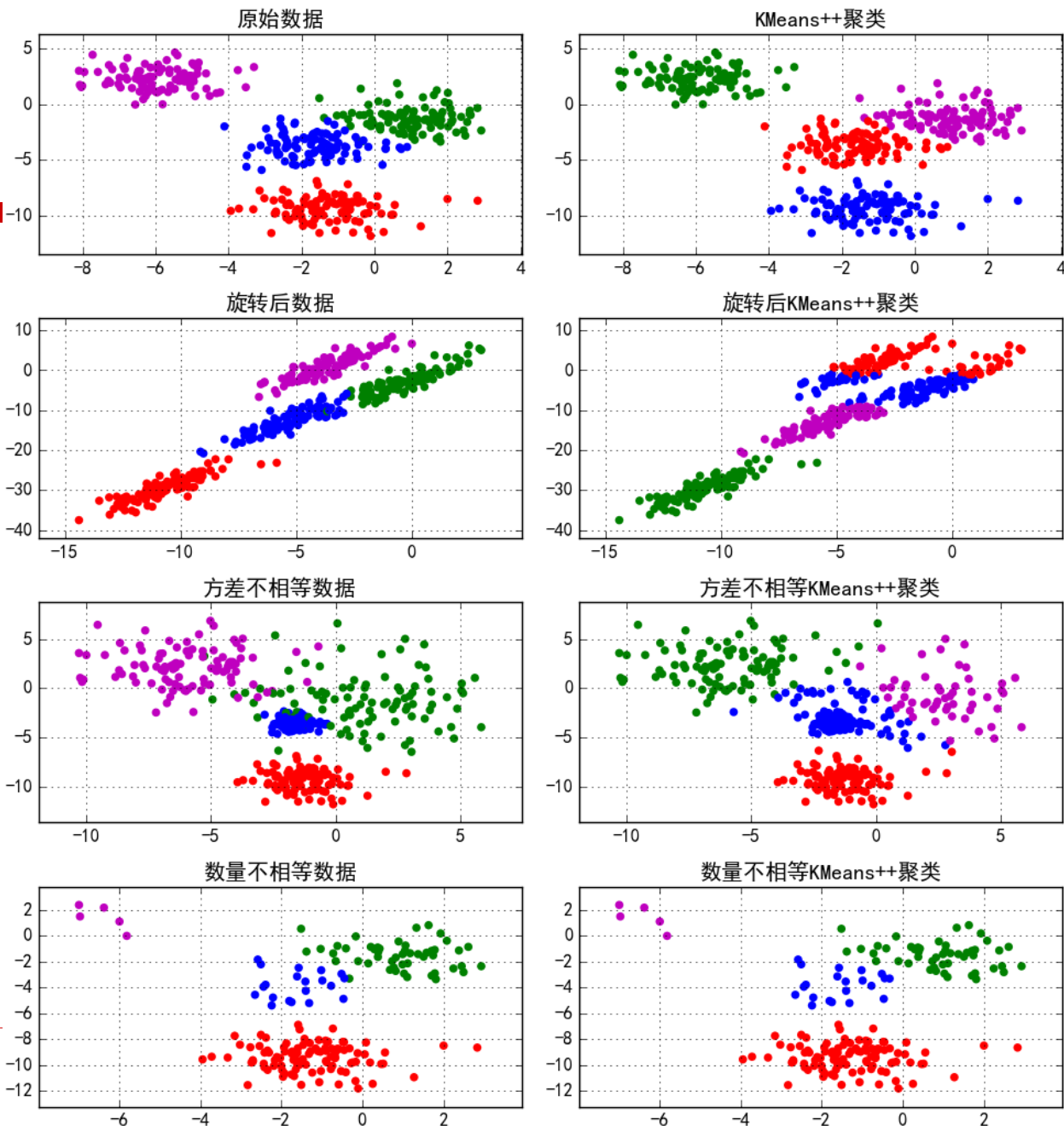
邹博

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

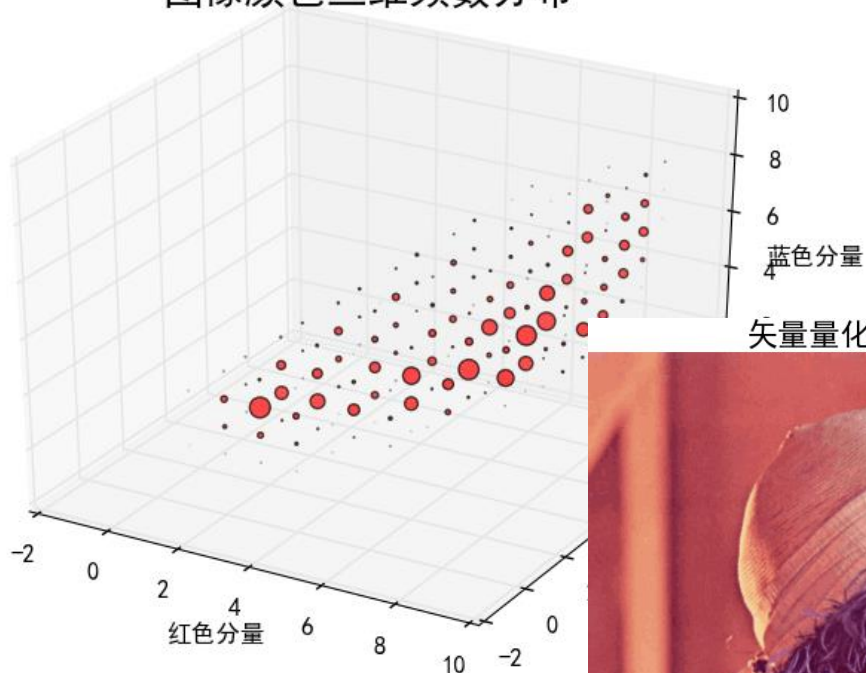
k-Means

数据分布对KMeans聚类的影响

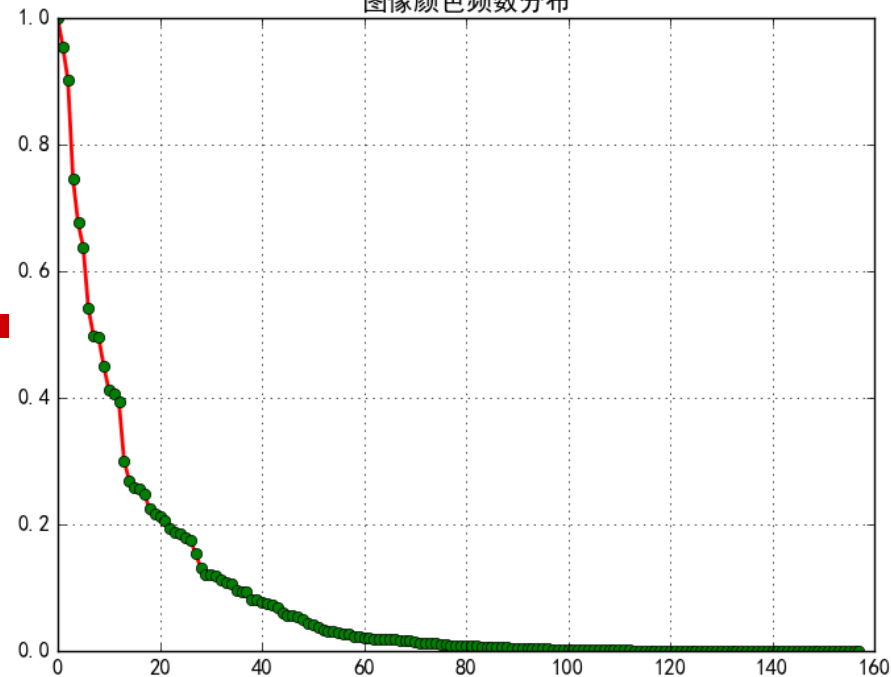


Vector Quantization

图像颜色三维频数分布



图像颜色频数分布

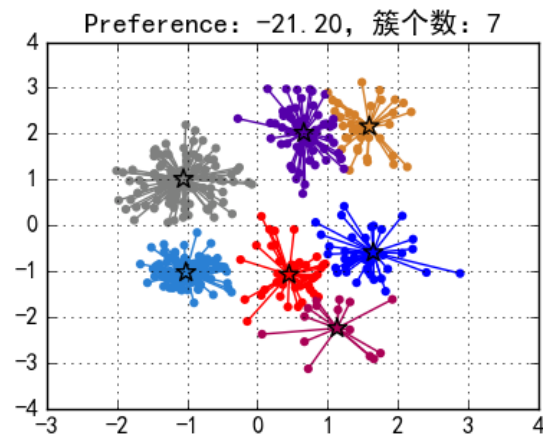
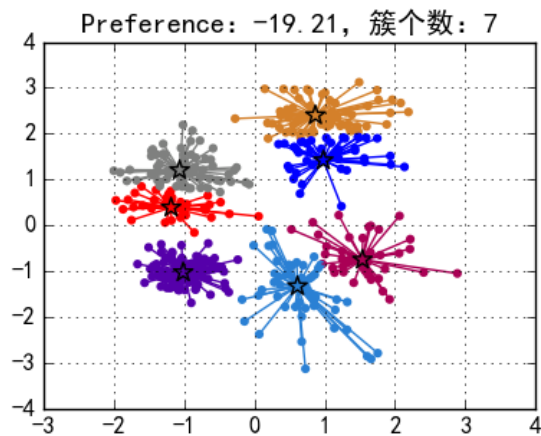
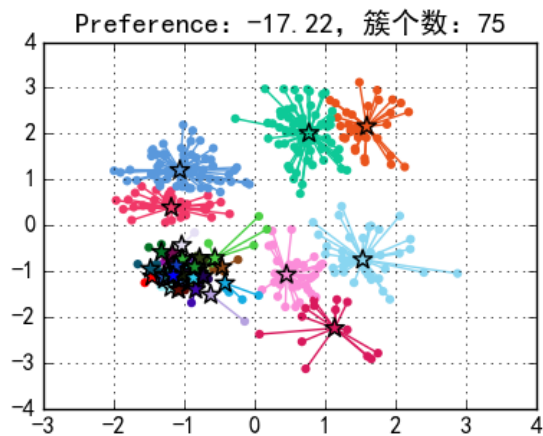
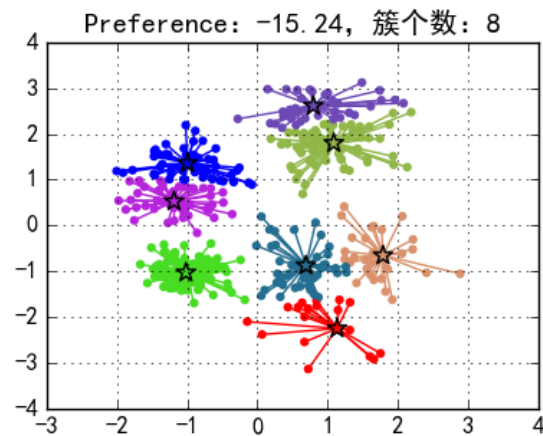
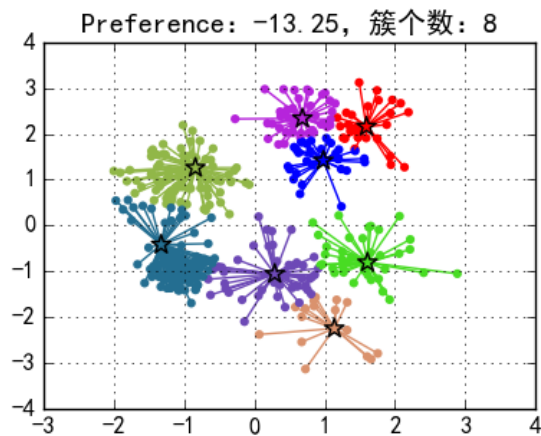
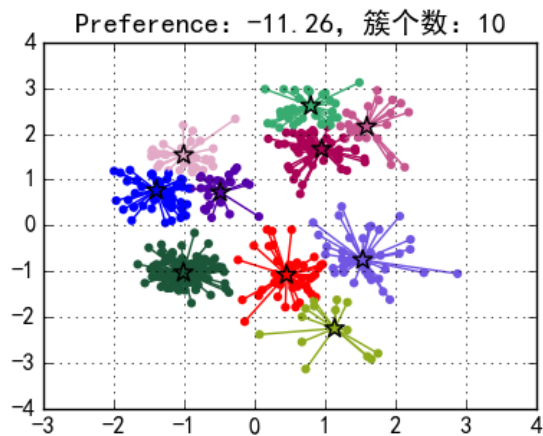
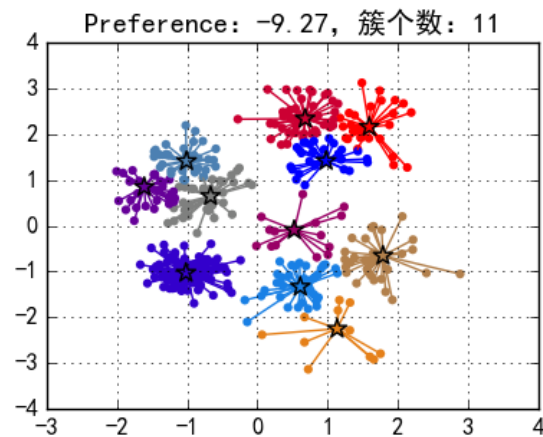
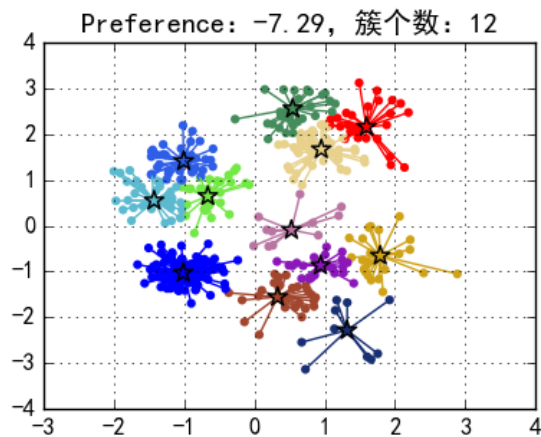
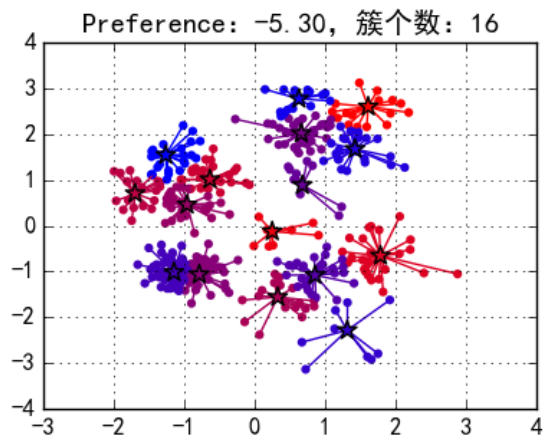


矢量量化后图片：100色

原始图片

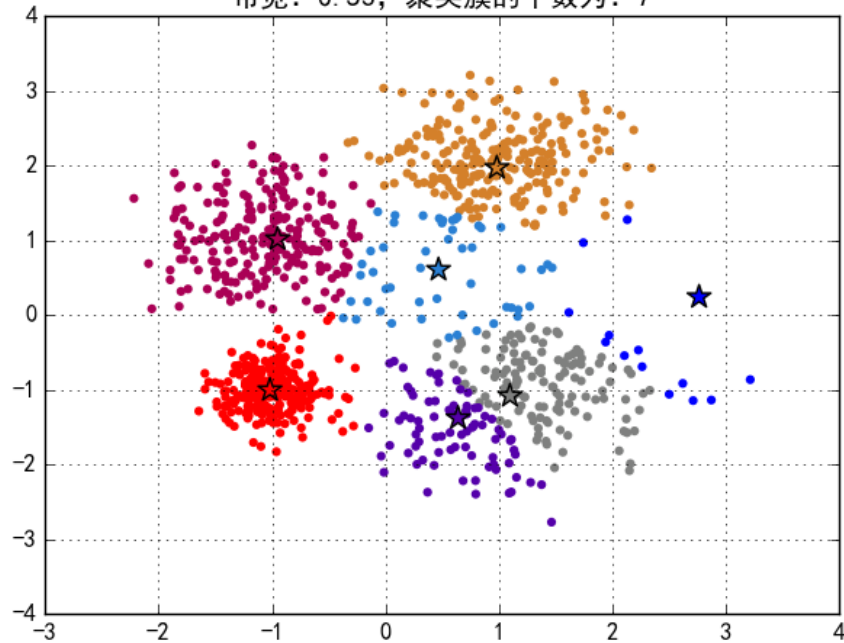


AP聚类

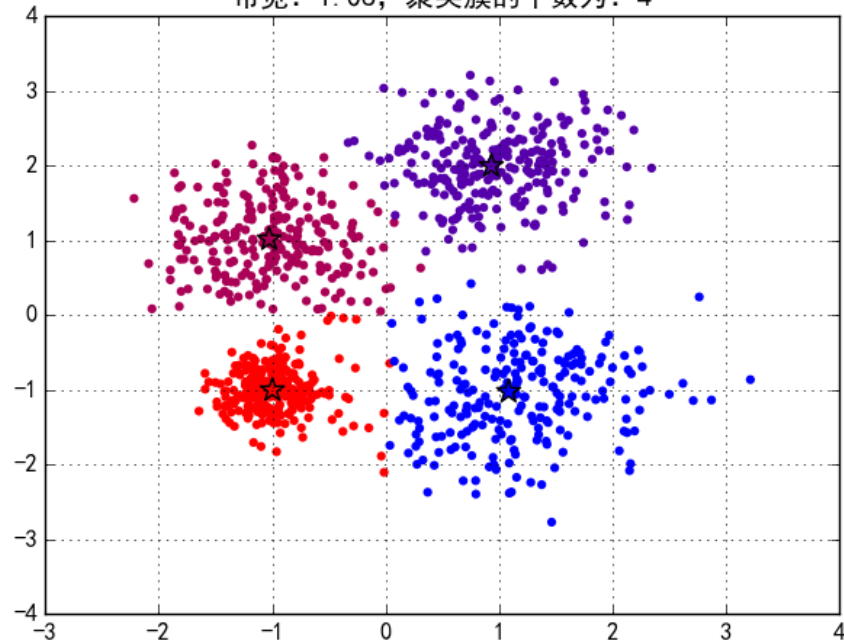


MeanShift聚类

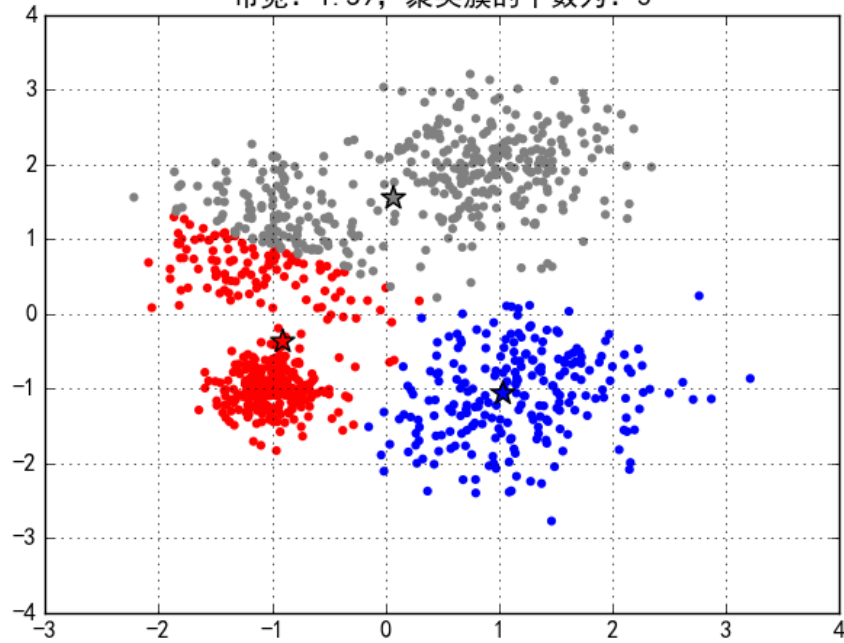
带宽: 0.53, 聚类簇的个数为: 7



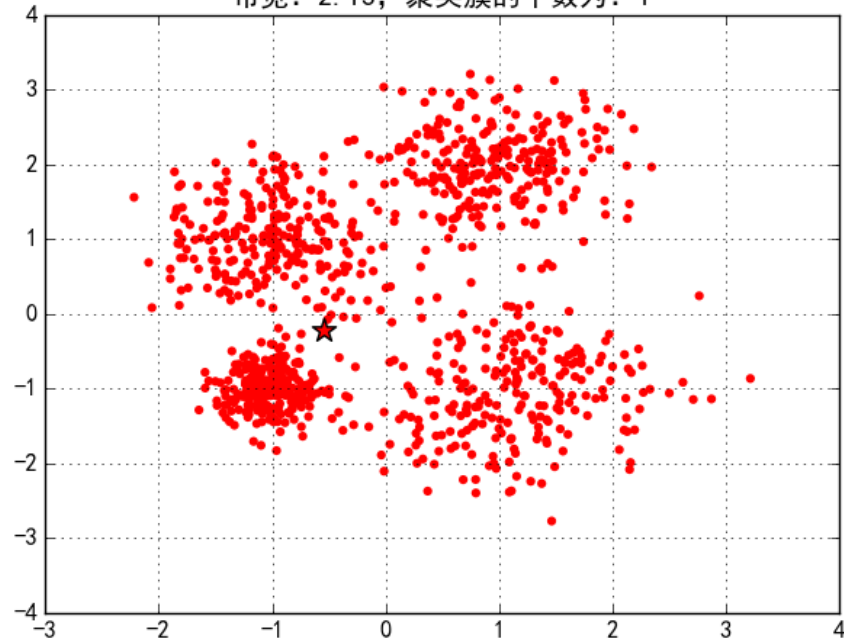
带宽: 1.06, 聚类簇的个数为: 4



带宽: 1.59, 聚类簇的个数为: 3

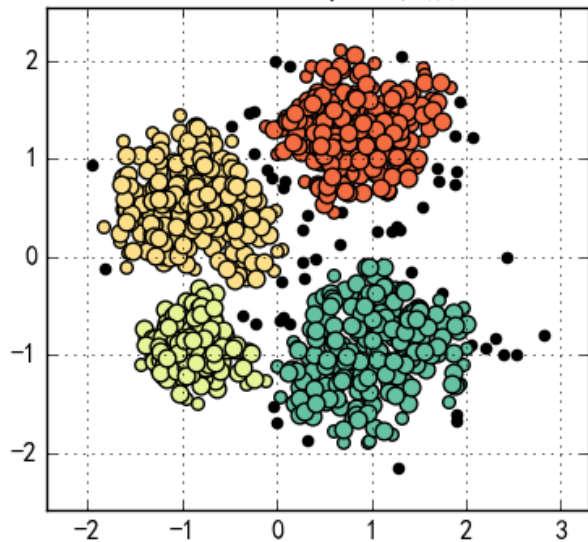


带宽: 2.13, 聚类簇的个数为: 1

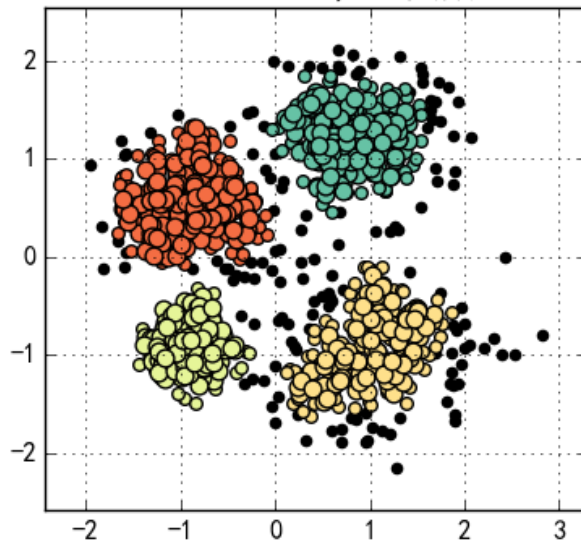


DBSCAN聚类

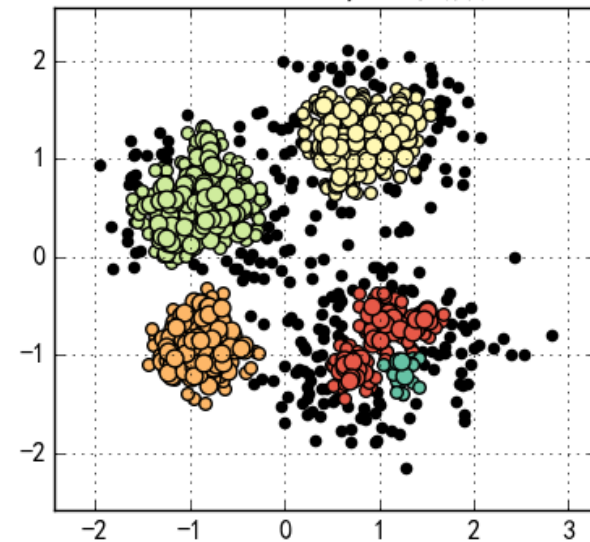
$\epsilon = 0.2$ $m = 5$, 聚类数目: 4



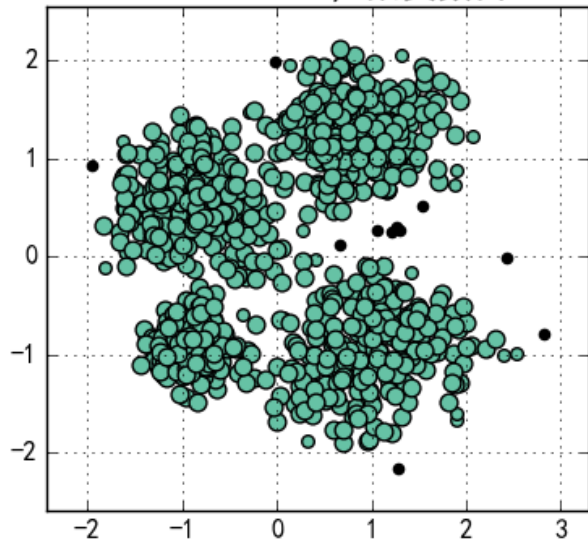
$\epsilon = 0.2$ $m = 10$, 聚类数目: 4



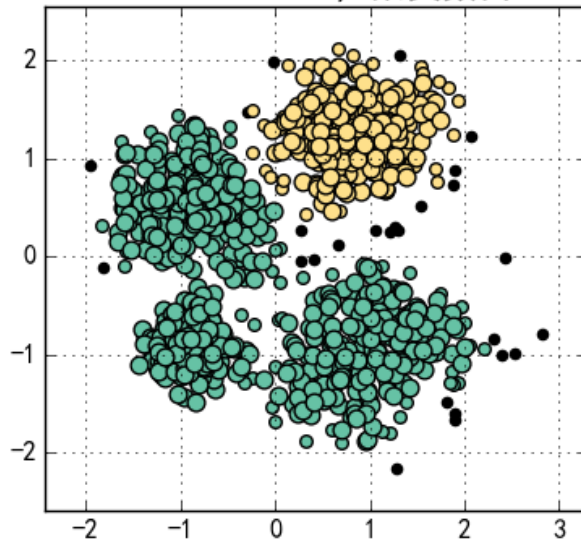
$\epsilon = 0.2$ $m = 15$, 聚类数目: 5



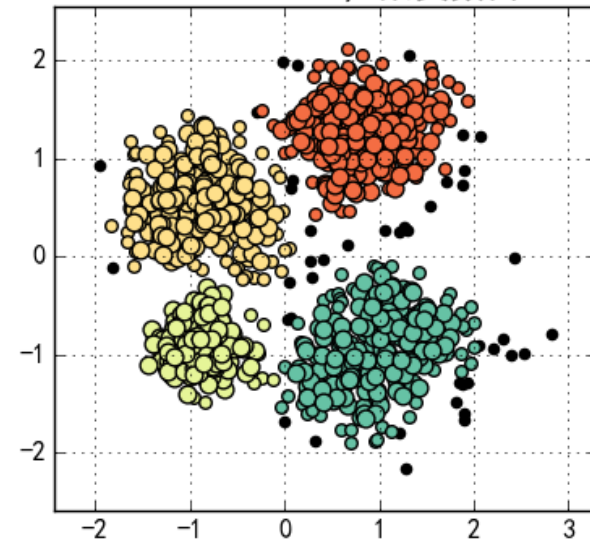
$\epsilon = 0.3$ $m = 5$, 聚类数目: 1



$\epsilon = 0.3$ $m = 10$, 聚类数目: 2

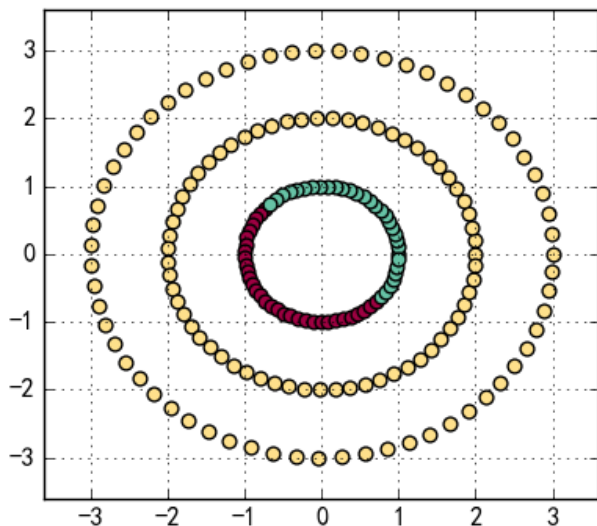


$\epsilon = 0.3$ $m = 15$, 聚类数目: 4

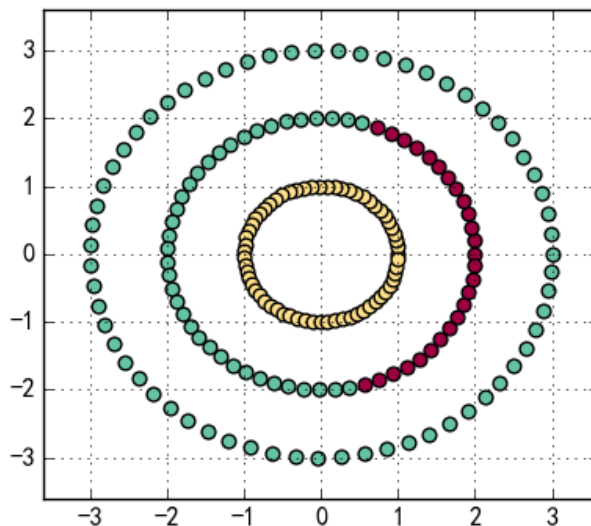


谱聚类

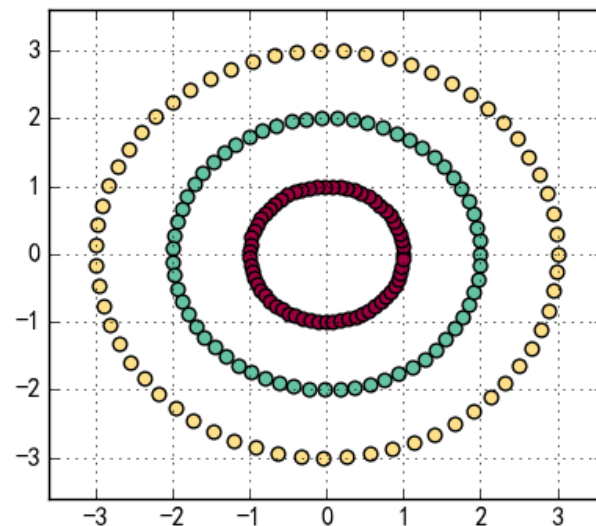
$\sigma = 0.01$



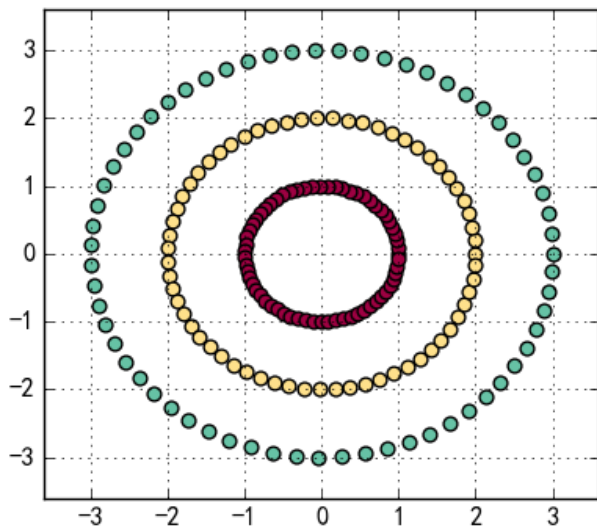
$\sigma = 0.03$



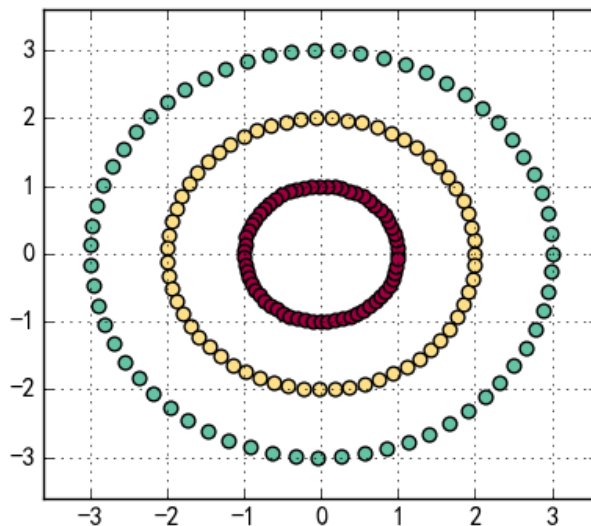
$\sigma = 0.06$



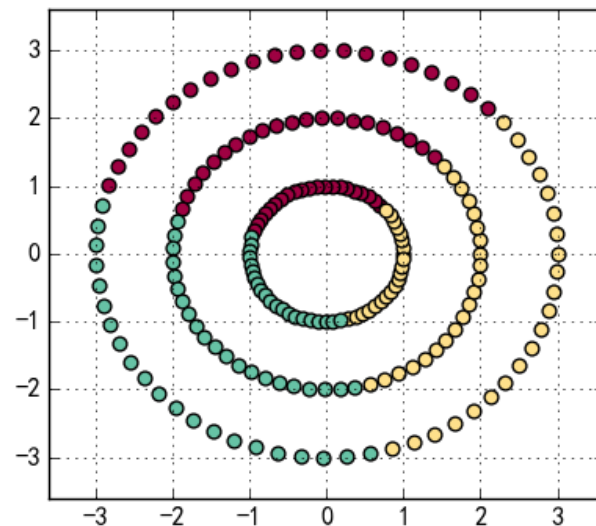
$\sigma = 0.16$



$\sigma = 0.40$

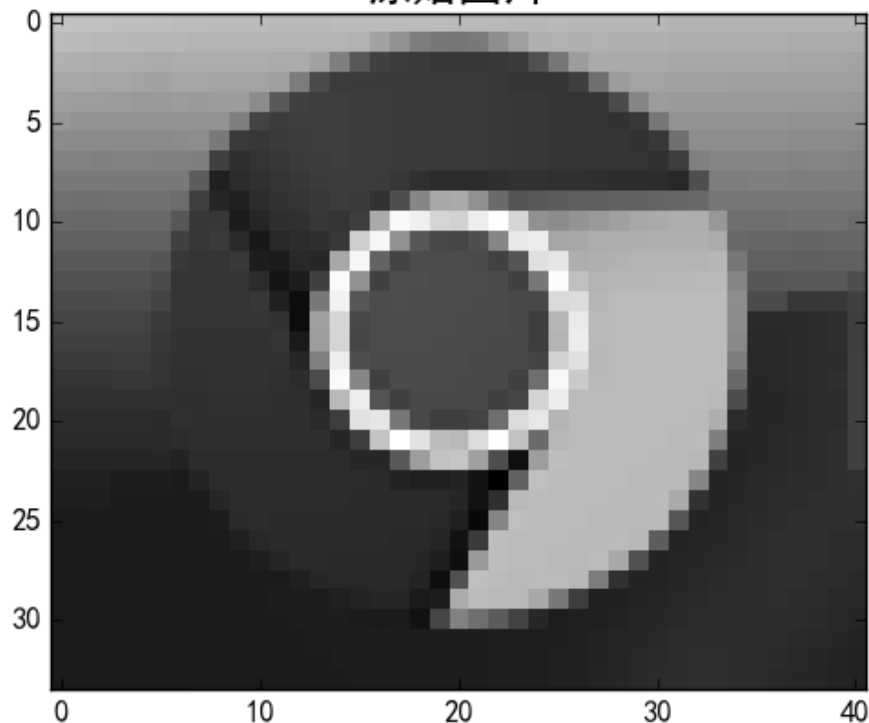


$\sigma = 1.00$

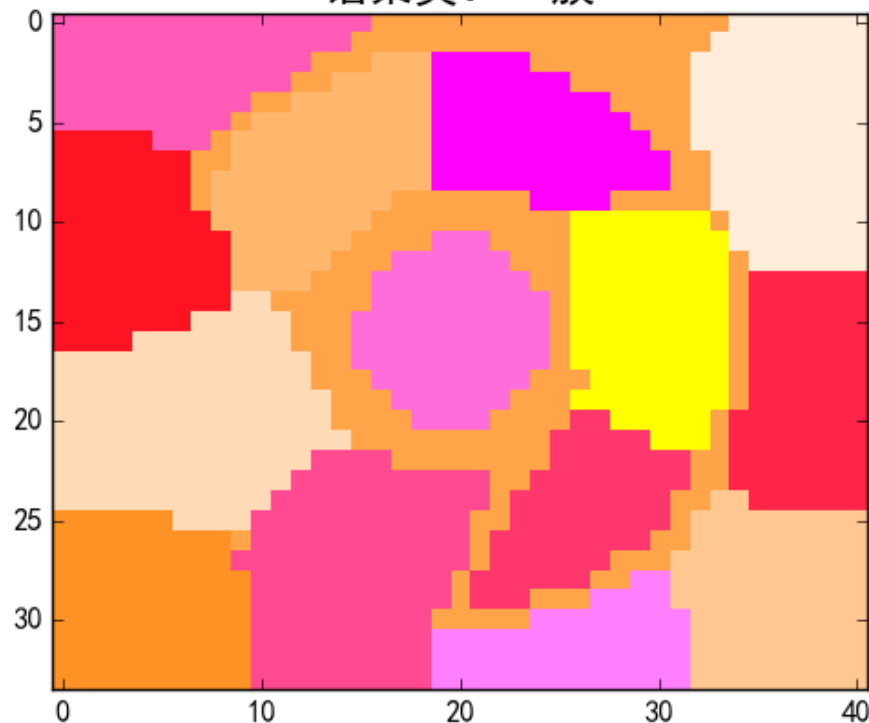


谱聚类与图像切割

原始图片



谱聚类：15簇



Demo

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

■ 大数据分析挖掘



感谢大家！

恳请大家批评指正！