### 法律声明

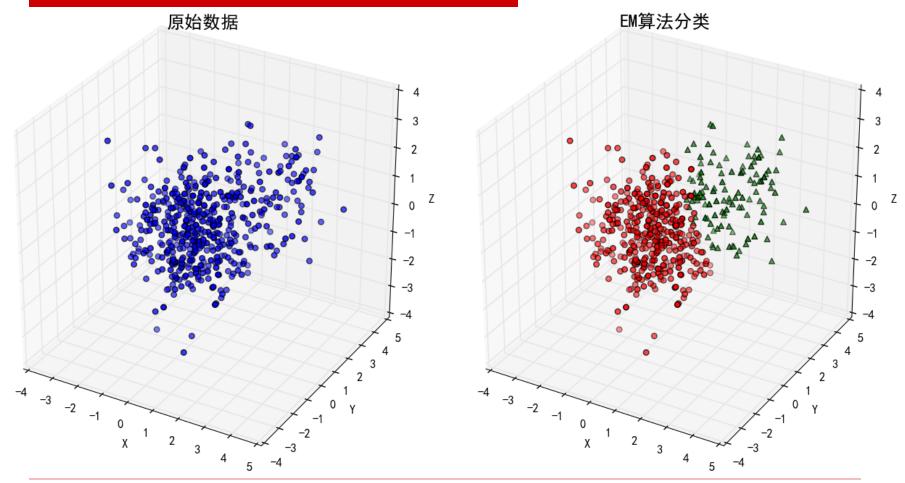
- □本课件包括演示文稿、示例、代码、题库、视频和声音等内容,小象学院和主讲老师拥有完全知识产权的权利;只限于善意学习者在本课程使用,不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意及内容,我们保留一切通过法律手段追究违反者的权利。
- □ 课程详情请咨询
  - 微信公众号:小象
  - 新浪微博: ChinaHadoop



# EM算法实践



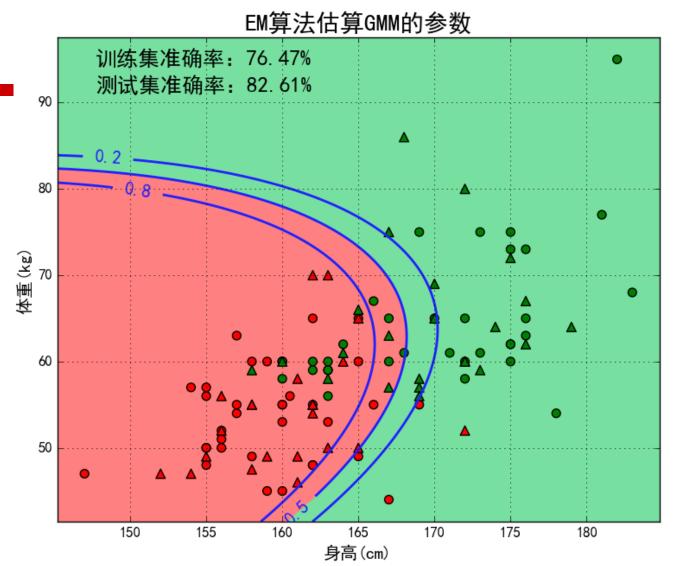
# 多维GMM聚类: EM算法



## EM算法

□副产品

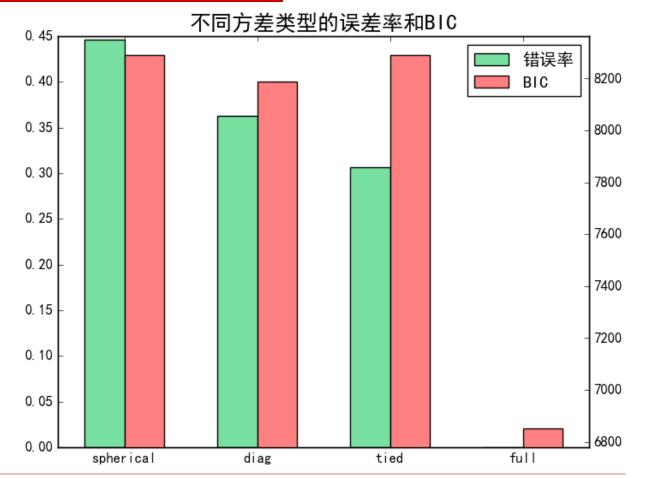
■ 等值线



### GMM调参

□副产品

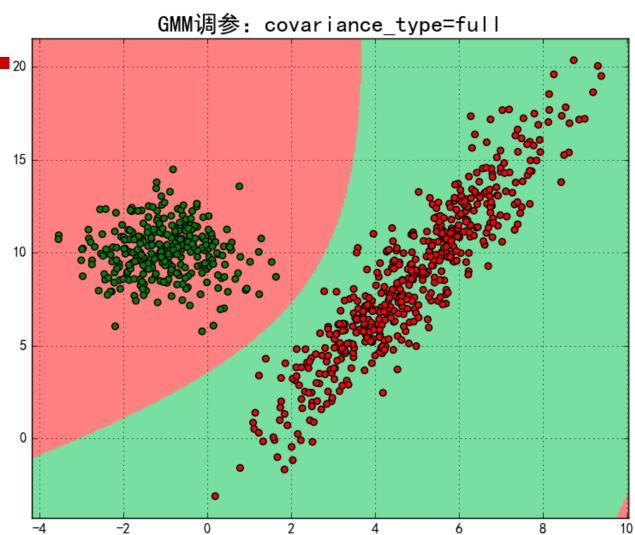
■ 双y轴



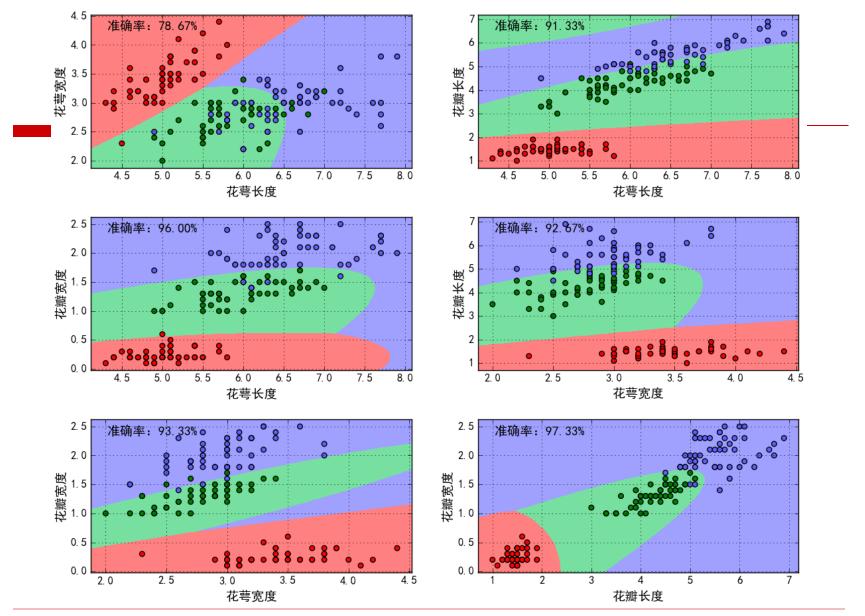
### 模型选择的准则

- □ 记: L为某模型下样本的似然函数值, k为模型中未 知参数的个数(维度), n为样本个数,则:
- $\Box$  AIC =  $-2 \ln L + 2k$ 
  - akaike information criterion
  - 日本统计学家赤池弘次(Akaike)于1973年提出
- $\square BIC = -2\ln L + (\ln n)k$ 
  - Bayesian Information Criterion/Schwarz criterion
  - Akaike于1976年通过改进AIC得到
  - Gideon E. Schwarz在1978年根据Bayesian理论重新发现



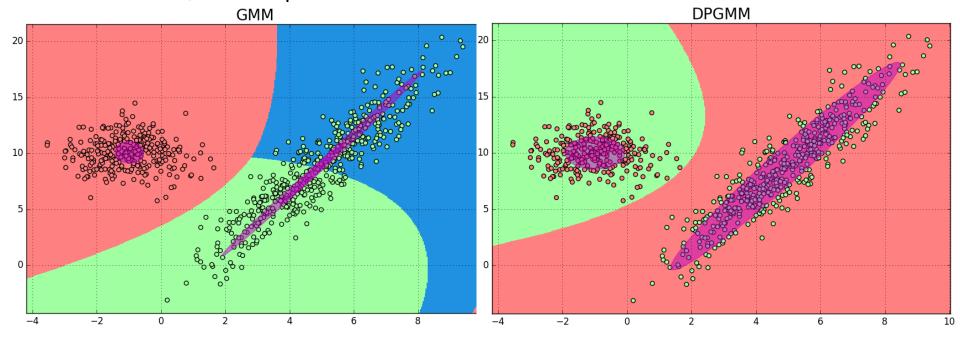


#### EM算法无监督分类鸢尾花数据



### **DPGMM**

- ☐ Dirichlet Process Gaussian Mixture Model
  - 先验分布



### 复习: 二项分布的最大似然估计

- □ 投硬币试验中,进行N次独立试验,n次朝上,N-n次朝下。
- □ 假定朝上的概率为p,使用对数似然函数作 为目标函数:

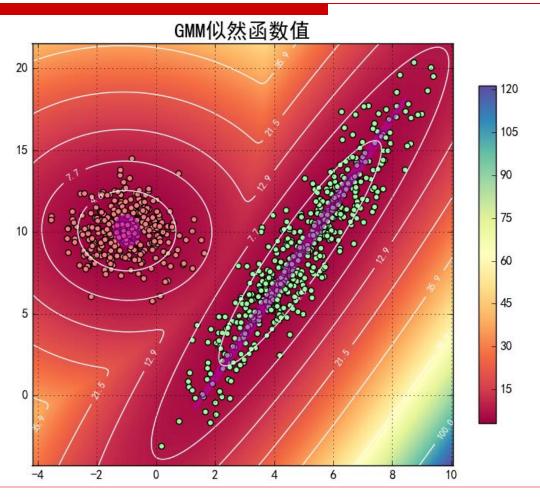
$$f(n \mid p) = \log(p^n(1-p)^{N-n}) \xrightarrow{\Delta} h(p)$$

$$\frac{\partial h(p)}{\partial p} = \frac{n}{p} - \frac{N - n}{1 - p} \xrightarrow{\Delta} 0 \implies p = \frac{n}{N}$$

### 二项分布与先验举例

- - □ 若观察到4个女生和1个男生, 以得出该校女生比例是80%吗?

# 似然函数值:复习Matplotlib的绘图



### 参考文献

https://en.wikipedia.org/wiki/Bayesian\_information\_criterion

### 我们在这里

△ 通知 http://wenda.ChinaHadoop.cn 专题 招聘求职 yarn运行时一直重复这个info...好像没找到资源,应该从哪里检查呢? 大数据行业应用 视频/课程/社区 数据科学 系统与编程 贡献 云计算技术 机器学习 Eric\_Jiang 回复了问题 • 2 人关注 • 1 个回复 • 6 次浏览 • 2016-05-18 13:29 35 微博 贡献 wangxiaolei 回复了问题 • 1 人关注 • 10 个回复 • 47 次浏览 • 2016-05-18 12:04 @ChinaHadoop sqoop把mysql数据导入Hbase报如图错误 贡献 @邹博\_机器学习 kafkaOffsetMonitor打开页面以后无法显示内容? kafka fish 回复了问题 • 4 人关注 • 2 个回复 • 8 次浏览 • □ 微信公众号 markdown公式编辑\$符号不起作用 热门用户 再多 > 贡献 markdown masterwzh 回复了问题 • 3 人关注 • 1 个回复 • 13 次浏览 • 2016-05-18 08:40 小泵 17 个问题, 0 次赞同 找到,进入源码编译之后的目录如图二!这个文件找不到怎么解决呢?是编译没产生? 55 个问题 3 次幣同 **\*\*** ■ 大数据分析挖掘 55 个问题, 12 次營同 opentsdb安装时出现72个warning,是正常的么? 48 个问题, 0 次赞同

► → C wenda.chinahadoop.cn/explore/

贡献 

hiveman 19 个问题, 1 次赞同

opentsdb fish 回复了问题 • 3 人关注 • 5 个回复 • 49 次浏览 • 2016-05-17 18:53

计算机广告 wayaya 回复了问题 • 4 人关注 • 7 个回复 • 108 次浏览 • 2016-05-17 18:26

关于在线广告和个性化推荐区别的一点浅见

# 感谢大家!

恳请大家批评指正!