

Mini Report-GMM

贾丁 jiading.biz@outlook.com

文件清单

1. README.pdf
2. Main.cpp
3. Visualization.py (用于可视化结果)
4. Sample文件夹：示例文件

程序描述

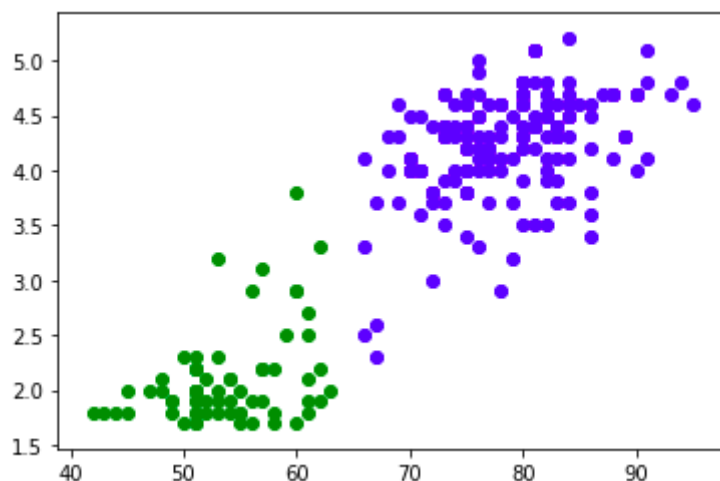
- 面向对象：matrix类、文件读写类、GMM类（除了两个方法没有封装为工具类之外）
- 自己造了matrix的轮子(MyMatrix类)，后来由于性能瓶颈，计算行列式和逆矩阵的部分改换了Eigen库（需要下载并安装），所以私有方法中那两个old方法是不用的
- 理论上更换main方法的MODELNUMBER常数就可以选择GMM中使用的高斯函数的数量，但是只根据样本分布情况测试并使用了使用两个的情况，其他情况可能有bug
- 运行时需要改地址的两个地方：main方法中dataGMM.txt的地址、include语句中Eigen库的安装位置

- GMM模型对初始值很敏感，不同初始值下结果和程序运行时间差别会比较大，程序中设定的初始值为：
 - 两个高斯模型的权重分别为0.5
 - 协方差矩阵初始化为单位矩阵
 - 均值设置为样本中随机选取

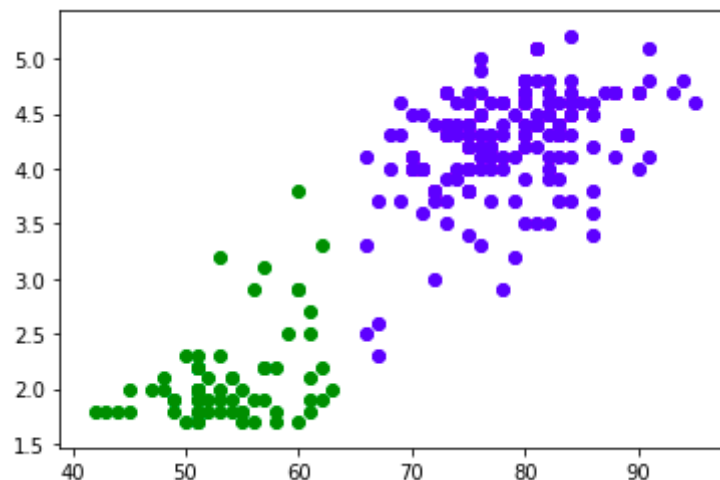
运行结果

我只调整了两个高斯模型的初始权重，效果已经不错，所以没有调其他的

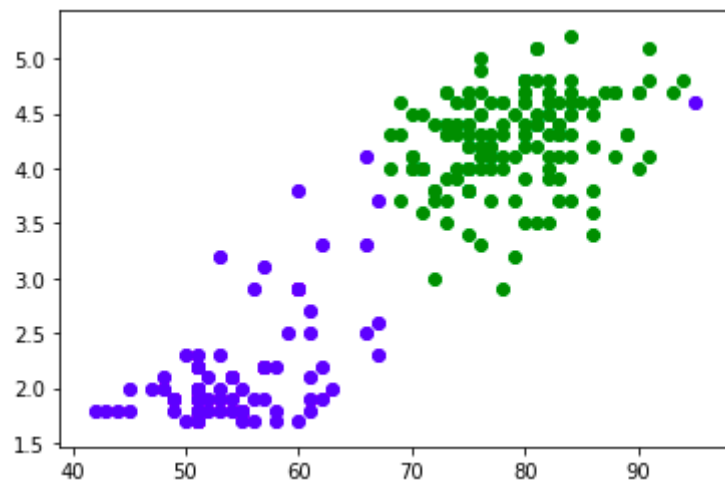
1. 0.1:0.9



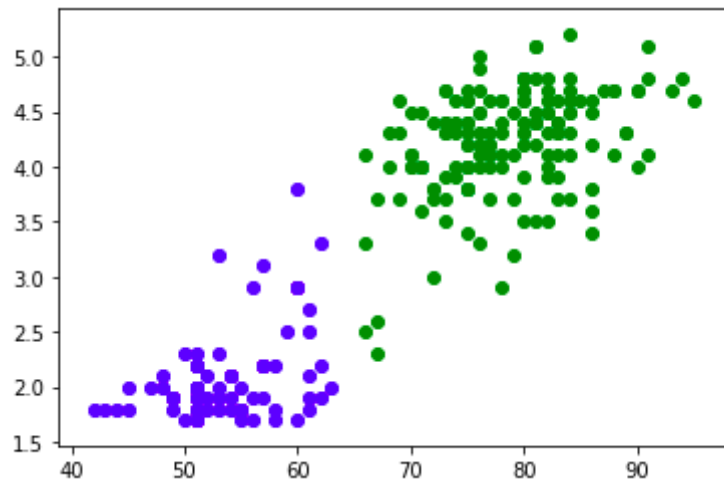
2. 0.2:0.8



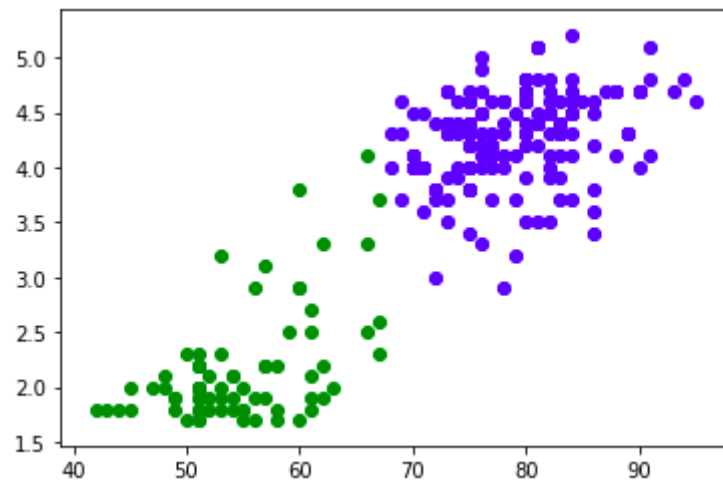
3. 0.3:0.7



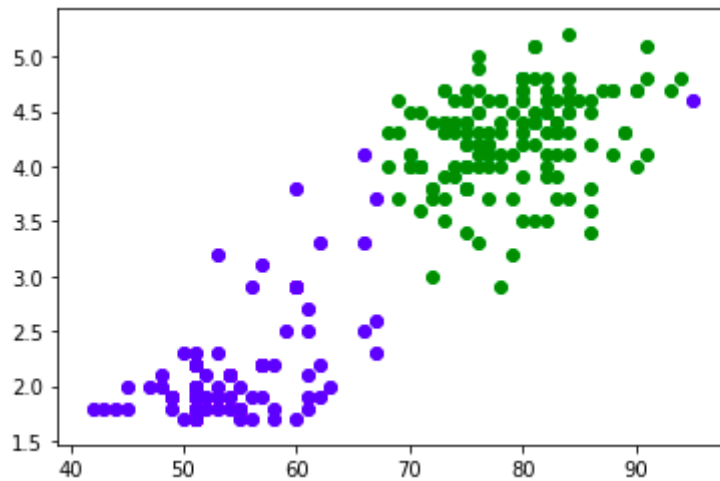
4. 0.4:0.6



5. 0.5:0.5



6. 0.6:0.4



最终由表现选择0.5:0.5的模型，参数输出为：

```
-----report-----  
modelNumber:2  
  
ConverianceList  
32.9107 0.686843  
0.686843 0.18338  
35.2697 1.69223  
1.69223 0.277249  
  
average:  
78.8311 4.25811  
54.8219 2.17945  
  
weight:  
0.666667 0.328829
```

Reference

1. <https://www.cnblogs.com/wxl845235800/p/9027005.html>
2. <https://www.cnblogs.com/luxiaoxun/archive/2013/05/10/3071672.html>
3. <https://brilliant.org/wiki/gaussian-mixture-model/>

