**第五次上机作业**

**（及计算作业）**

本次作业包含第五周和第六周的教学内容，6月6日（第6周末）交。

从本周开始，请大家注意大作业的探索（最后一节上机课，即第八周周四下午进行展示），如分析过程中有问题，可随时找我讨论。

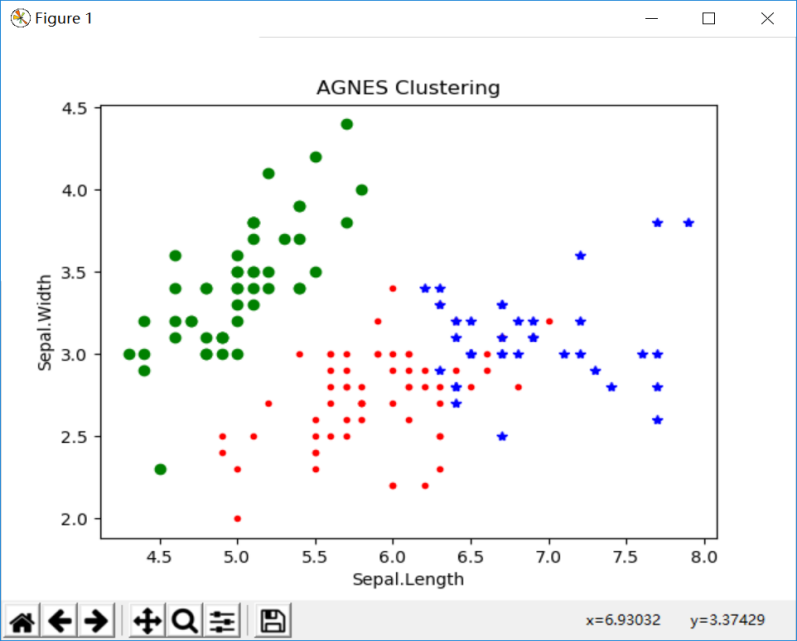
**一、开放练习：根据官网文档，或者参考网络教程，分别用sklearn中的决策树和随机森林方法，对鸢尾花数据集进行分类。**

1. 分类练习，两类或三类均可。
2. 参阅官网文档，为随机森林方法挑选3个感兴趣的参数，改变参数、观察分类结果。自由探索，报告形式自由，并适当写一些心得体会。

**二、开放练习：根据官网文档，或者参考网络教程，利用sklearn对鸢尾花数据集进行聚类。**

1. 对三类样本进行着色，并显示在花萼长度、花萼宽度的二维空间上。

展示结果可以如下图所示（但不要求一样，“能用就好”！）：



1. 用K-means对鸢尾花数据集进行聚类（所有特征都用），并将结果可视化在花萼长度、花萼宽度的二维空间上。将代码运行5次，查看结果是否有变化？
2. 用Gaussian mixtures model对鸢尾花数据集进行聚类（所有特征都用），并将结果可视化在花萼长度、花萼宽度的二维空间上。
3. 用层次聚类对鸢尾花数据集进行聚类（所有特征都用），并绘制聚类树。
4. 自由改变参数进行探索，挑选3个感兴趣的参数，改变参数、观察分类结果，并适当写一些观察体会。自由格式、自由探索。

**三、计算练习：层次聚类**

对如下5个样本（每个样本为一个6维向量），用最小距离准则进行层次聚类。定义距离度量，写清每步聚类的依据，并给出最后的聚类树。

