**一、使用说明（重要！）**

1. 配置环境：

Anaconda可以帮助管理数量繁多的python库，并且当安装新的python库时可以一键安装 – 即会帮助安装所有依赖包，以实现此库的最优运行。

因此我们选择Anaconda虚拟环境运行代码，使用的是Anaconda中自带的python 3.6.7版本。

安装激活Anaconda：

进去Anaconda官网下载windows 64bit版本。

运行.exe进行安装。

进入cmd，执行activate base进入Anaconda虚拟环境。

在虚拟环境base下安装sklearn 和 xgboost：

安装sklearn：conda install scikit-learn

使用在conda库中寻找xgboost的各种版本：anaconda search -t conda xgboost

选择windows 64bit对应的版本安装：conda install -c anaconda py-xgboost

2.参数调节：

我们根据教程PPT中给出的网址教程进行参数调节：

<https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>

先确定一个高学习率：eta = 0.1

调节树相关参数max\_depth和min\_child\_weight

再调整损失函数增益最小值gamma值

再调整subsample和colsample\_bytree值

再调整正则项值lambda和alpha

最后再选择合适的学习率eta

最后我们的参数值为

'eta': 0.025, 'min\_child\_weight': 3, 'max\_depth': 6, 'gamma': 0.1, 'subsample': 0.7, 'colsample\_bytree': 0.7, 'lambda': 2

3. 代码运行过程中输出终端的内容详细表明了程序运行流程。最后的结果也会输出到终端中。并将预测结果输出到predict.txt中。

4. 请将所需的数据文件：iris.date同 代码.py 放在同一目录下，以保证代码正常运行。

5. 祝每天开心！

**二、实验过程与代码**

L1 – L5 导入必须的sklearn包和xgboost包。

L8 – L14 数据读入并标准化代码。

L62 – L78 xgboost训练并预测代码。

L80 – L88 输出结果代码。

**三、实验结果及分析**

因为iris数据集较少，且数据良好，所以处理结果很好。

实验结果截图如下：

