XGBoost实验报告

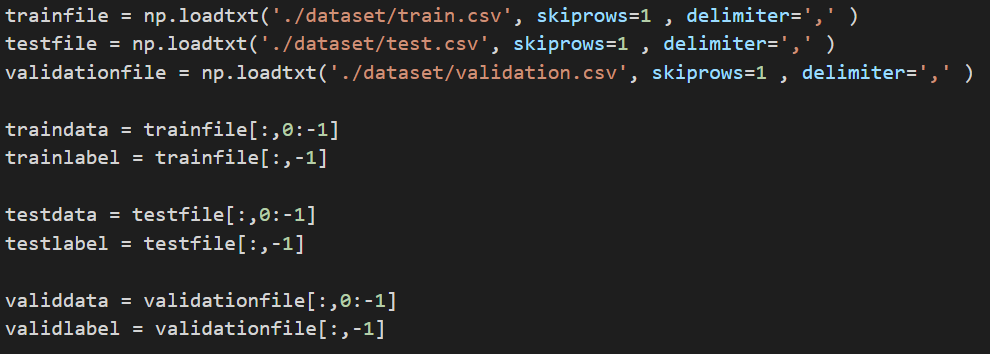
Kotoricyann

项目GitHub：<https://github.com/Kotoricyann/RUC-AI-Course>

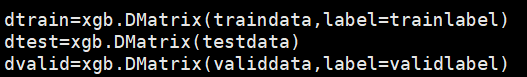
XGBoost是GBDT的一个工程实现，在GBDT的基础上做了多方面优化，包括：在使用CART作为基分类器时，XGBoost显式地加入了正则项来控制模 型的复杂度，有利于防止过拟合，从而提高模型的泛化能力；XGBoost对代 价函数进行二阶泰勒展开，可以同时使用一阶和二阶导数；XGBoost支持多种类型的基分类 器，比如线性分类器；XGBoost采用了与随机 森林相似的策略，支持对数据进行采样；XGBoost能够自动学习出缺 失值的处理策略。

实验通过对一个具有162个feature的数据进行二分类。具体流程如下：

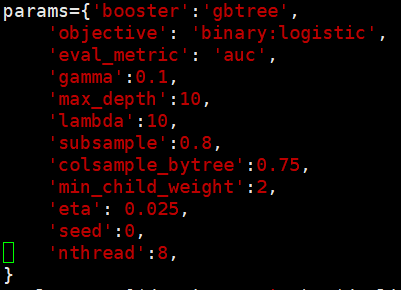
读入数据：



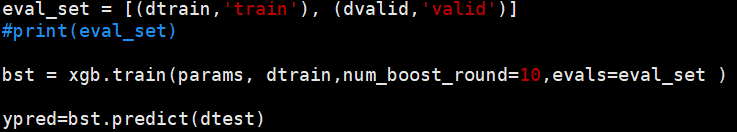
将数据转化为xgb的数据格式：



设置参数：



训练模型并利用模型预测测试集分类：



得到结果：

