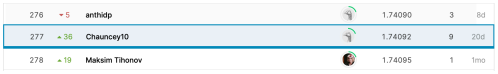
《机器学习导论》赛题三报告

181240035，刘春旭，181240035@smail.nju.edu.cn

2021 年 7 月 1 日

1 队伍得分与排名

队伍名称为 Chauncey10，共 1171 支队伍参加，排名为 277. 截图如下：



(a) 排名

(b) 队伍总数

图 1: Private Leaderboard

2 建模思路与方法

使用 Amazon 开发出的Autogluon完成此次的实验，这是一个自动调参、挑选模型的 AutoML 包，完成本次竞赛全部代码量仅 11 行. 项目地址位于这里下面进行详细说明：

2.1 模型训练

以下部分可以让 Autogluon 进行自主训练

```
from autogluon.tabular import TabularDataset, TabularPredictor
import numpy as np
import pandas as pd
# 导入必要的包

train_data = TabularDataset('train.csv')
id, label = 'id', 'target'
# 获得训练数据

metric = 'log_loss' # 使用log loss作为评估指标
predictor = TabularPredictor(label=label,
    eval_metric=metric).fit(train_data.drop(columns=[id]), presets='best_quality')
    # 定义模型训练参数并开始训练
```

训练好的模型参数会储存在当前文件夹中，但是因为该文件夹大小过于巨大（44GB），故本次试验只上传了代码。

2.2 模型预测

通过以下代码生成提交文件 (submission.csv)

```
test_data = TabularDataset('test.csv') # 导入测试数据集
preds = predictor.predict_proba(test_data.drop(columns=[id]), as_pandas=True) #
    开始训练模型
preds.insert(0, id, test_data[id])
preds.to_csv('submission.csv', index=False) # 生成预测结果
```