Project

Due 2021/11/21

Topic 1: 天池大数据竞赛（正式赛）

题目——糖尿病性黄斑水肿 (DME) 患者的Anti-VEGF抗血管内皮生长因子(简称Anti-VEGF)治疗转归预测

(https://tianchi.aliyun.com/competition/entrance/531929/introduction)

初赛时间：2021/09/01 - 2021/10/31

复赛时间：2021/11/02 – 2021/11/19

Topic 2: 天池大数据竞赛（长期赛）

题目——天猫复购预测

(https://tianchi.aliyun.com/competition/entrance/231576/introduction)

Topic 3: 天体光谱智能识别分类

背景：

天文望远镜每晚能采集万余条光谱，使得传统的人工或半人工的利用模板匹配的方式不能很好应对，需要高效而准确的天体光谱智能识别分类算法。

任务描述：

给出18万个天体的光谱数据，要求根据光谱数据，把未知天体分成行星（star），星系（galaxy）和类星体（qso）三类。评价指标即预测准确率。

数据描述

数据为csv文件，每一行代表一个天体，每一列代表这个天体在不同波长上的值。

训练集带有数据标签，即每个天体的分类。star代表恒星，galaxy代表星系，qso代表类星体。测试集要求预测天体的标签。

训练集：

训练集数据集为csv文件，数据包括14万个天体的信息。其中 id列 为天体 ID，FE0列,FE1列,FE2列,…,FE2599列,共计2600列为波长信息，作为数据的输入特征。天体的类别在对应label文件中。

验证集：

验证集数据有4万行，格式和训练集格式相同。（需要注意，由于存入一个文件太大，每2万数据为一个文件存储，共9个文件，同时只有第一个文件有表头信息，训练集是sets\_v1\_0至sets\_v1\_6，验证集是sets\_v1\_7和sets\_v1\_8）

数据下载地址：

地址：https://pan.baidu.com/s/1Di5VMzz0HC0qc--W5NNrig

提取码：muls

大作业验收：

最后验收请提交实验报告来阐述你的实验方案、方法和性能分析。同时，现场会提供测试集文件和label文件，同学需要当场运行程序进行预测并计算准确率。