

2018机器学习导论 作业说明

总体安排

- 作业分为个人作业和团队作业两部分，个人作业占平时成绩30%，团队作业占平时成绩70%
- 个人作业：三选一
- 团队作业：二选一

个人作业

- 汽车购买意向预测
- 通过声音进行性别预测
- 葡萄酒质量分析

汽车购买意向预测

- 背景：在购买汽车时有许多的因素会作为购买意向的考量。本题要求根据某种车型的各项相关数据，预测用户对该车型的购买意向
- 数据说明：
 1. 购买价格，如vhigh、high、med等。
 2. 维护价格，如vhigh、high、med等
 3. 车门的数量，从2到4甚至更多
 4. 载人数量，从2到4甚至更多
 5. 车后备箱大小，包括big、med、small等
 6. 安全性能，从high、med到low等
 7. 购买意向，包括了unacc无法接受、acc可以接受、good好等。
- 数据大小：训练集包含1210条数据，测试集包含518条数据，数据格式为csv文件
- 评分标准：预测结果正确的样本占有所有样本的百分比

通过声音进行性别预测

- 背景：男性和女性的声音存在许多的不同地方，本题要求根据声音的各项特征数据，预测一个声音是属于男性或者女性
- 数据说明：
 1. meanfreq: 平均频率（以kHz为单位）
 2. sd: 频率的标准偏差
 3. skew: 偏度
 4. kurt: 峰度
 5. sp.ent: 谱熵
 6. sfm: 频率谱平坦度
 7. mode: 模式频率
 8. centroid: 频率质心
 9. 等
- 数据大小：训练集包含2050条数据，测试集包含878条数据，数据格式为csv文件
- 评分标准：预测结果正确的样本占有所有样本的百分比

葡萄酒质量分析

- 背景：葡萄酒可以分为白葡萄酒和红葡萄酒。本题要求根据葡萄酒中各项成分的数据，预测该葡萄酒是白葡萄酒还是红葡萄酒
- 数据说明：
 1. Fixed acidity: 非挥发性酸度
 2. Volatile acidity: 挥发性酸度
 3. Citric acid: 柠檬酸
 4. Residual sugar: 剩余糖分
 5. Chlorides: 氯化物
 6. Free sulfur dioxide: 游离二氧化硫含量
 7. Total sulfur dioxide: 总二氧化硫含量
 8. 等
- 数据大小：训练集包含3568条数据，测试集包含1529条数据，数据格式为csv文件
- 评分标准：预测结果正确的样本占有所有样本的百分比

个人作业要求

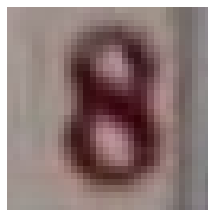
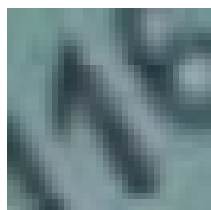
- 独立完成
- 提交报告
- 报告中应当体现你使用的方法、遇到的问题以及你是如何解决的
- 鼓励使用多种方法，会有加分

团队作业

- 街景数字识别
- 动漫角色检测

街景数字识别

- 背景：在街头的门牌号、交通指示牌等，都会频繁出现数字符号，这些数字符号的角度、大小、颜色都会有所不同。本题要求从截取的街景数字图片中识别出对应的数字
- 数据说明：训练集为4200张大小为32*32的图片，train_labels.csv中第一个标签对应训练集中的0.jpg中的数字，第二个标签对应训练集中的1.jpg中的数字，以此类推；测试集为1800张大小为32*32的图片。需要注意的是图像中可能出现不止一个数字，要求只识别最靠近中间的那个数字
- 评分标准：预测结果正确的样本占所有样本的百分比



动漫角色检测

- 背景：人脸识别是如今火热机器学习的运用方向。而动漫角色因为出自画师笔下，可能存在的作画失误而导致同一个角色容貌出现差异。本题要求检测出图像中出现的角色以及在图片中的位置
- 数据说明：训练集为1634张包含动漫角色的图像，测试集为814张包含动漫角色的图像。需要检测的角色一共有四个，需要注意的是是一张图像中可能出现不止一个角色，但需要检测的角色只会出现一个。得到目标角色的检测框后，假设检测框的左下点坐标 $(x1, y1)$ ，右上点坐标 $(x2, y2)$ ，则提交的csv文件中按照 $(x1, y1, x2, y2)$ 的顺序填写每张测试集图片的检测结果。
- 评分标准：每个样本中提交的目标检测框面积与标准答案的面积IOU的累加和。



团队作业要求

- 小组合作完成，每组3~5人
- 课堂展示，每组有3分钟展示时间和3分钟提问时间
- 如果使用开源的深度学习网络(如faster rcnn)完成，展示时还应当给出一种非深度学习方法，或者使用自己搭建的神经网络的方法

作业平台

- 地址： <http://contest.mooc.buaa.edu.cn>
- 注册时， 要求用户名为： 学号_姓名， 如： 16061001_张三
- 团队作业需要使用平台的组队功能
- 使用过程中对平台或比赛有任何问题， 及时发邮件联系助教

北航数据工作站

基于高校群体智能促进数据科学研究发展

热门比赛



大数据测试第一阶段

竞赛组织者: roothm

十月 17, 2017
11 位参与者



大数据测试第二阶段

竞赛组织者: roothm

十月 19, 2017
6 位参与者



房价预测问题

竞赛组织者: yqkong

十月 12, 2018
5 位参与者

精选比赛



视频动作识别

竞赛组织者: yqkong

十月 25, 2018
4 位参与者



动漫角色检测

竞赛组织者: KimLee

一月 10, 2018
2 位参与者



汽车购买意向预测

竞赛组织者: KimLee

一月 10, 2018
2 位参与者

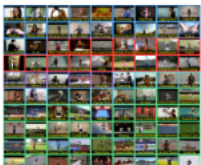
竞赛(我的工作站)

竞赛

我参加的竞赛

我举办的竞赛

我的数据集



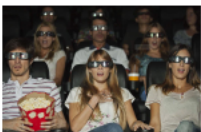
视频动作识别

竞赛组织者: yqkong

在拥有一定数量动作视频的条件下, 对新视频中的动作进行识别

十月 25, 2018- 永不结束

4 位参与者



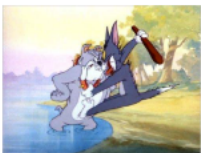
情感分析

竞赛组织者: yqkong

根据观众对电影的评论语句, 分析是正面评论还是负面评论

十月 13, 2018- 永不结束

2 位参与者



猫狗大战阉割版

竞赛组织者: yqkong

在拥有一定数量图片 (已知是猫/狗) 的条件下, 对新的图片进行 (猫/狗) 分类

十月 13, 2018- 永不结束

3 位参与者



泰坦尼克号乘客生还预测

竞赛组织者: yqkong

本次选择的数据是泰坦尼克号上乘客的基本信息, 预测哪些乘客更有可能生还

十月 12, 2018- 永不结束

1 位参与者

公开预测问题

我的竞赛



大数据测试第二阶段

竞赛组织者: roothm - 当前服务器时间: 十一月 5, 2018, 6:32 p.m. 北京时间

最近的

First phase

十月 19, 2017, 8 a.m. 北京时间

结束

竞赛结束

永不结束

比赛细则

阶段

参与比赛

查看成绩

你还没有报名参加这次竞赛

要参加这个比赛, 你必须接受它的具体条款和条件。报名后, 比赛主办方会评审你的申请并在你的申请被批准后通知你

☐ 我接受条款和条件。关于该比赛。

申请参加



参加比赛

- 比赛细则：在对应比赛页面点击比赛细则，可以查看比赛的规则综述、评分标准以及比赛条款。
- 获取数据：在对应比赛页面点击参加比赛-获取数据，点击页面提供的链接即可下载比赛数据。
- 提交结果：在对应比赛页面点击参加比赛-提交结果/查看提交，点击提交按钮，选择你要提交的文件即可完成提交，需要注意**提交文件的命名、格式需要与比赛要求保持一致**。
- 提交成绩：点击你要提交的成绩后的+号，选择submit to leaderboard，即可**将你的成绩提交到排行榜上**。

我的竞赛



街景数字识别

竞赛组织者: KimLee - 当前服务器时间: 十一月 5, 2018, 6:34 p.m. 北京时间

最近的

test

一月 10, 2018, 8 a.m. 北京时间

结束

竞赛结束

永不结束

比赛细则

阶段

参与比赛

查看成绩

团队



规则综述

评分准则

比赛条款

评分准则

数据表中共有1800条数据, 每条数据所占权重相同。

最后参赛选手的得分为每条评判结果的累加和。

#提交数据的csv文件要命名为submission.csv, 然后压缩成submission.zip的文件上传

| # | 用户名 | 分数 |
|---|--------|--------|
| 1 | KimLee | 100.00 |





街景数字识别

竞赛组织者: KimLee - 当前服务器时间: 十一月 5, 2018, 6:34 p.m. 北京时间

最近的

test

一月 10, 2018, 8 a.m. 北京时间

结束

竞赛结束

永不结束

比赛细则

阶段

参与比赛

查看成绩

团队

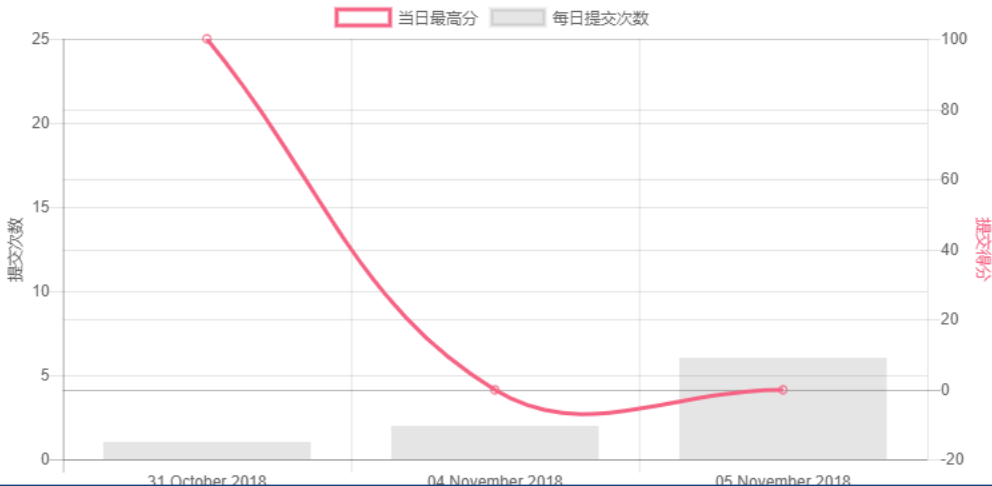
获取数据

训练集和测试集数据

比赛附件

提交结果/查看提交

| # | 用户名 | 分数 |
|---|--------|--------|
| 1 | KimLee | 100.00 |





街景数字识别

竞赛组织者: KimLee - 当前服务器时间: 十一月 5, 2018, 6:40 p.m. 北京时间

最近的

test

一月 10, 2018, 8 a.m. 北京时间

结束

竞赛结束

永不结束

比赛细则

阶段

参加比赛

查看成绩

团队

获取数据

比赛附件

提交结果/查看提交

test

阶段说明

无

每天最大提交数 999

总计最大提交数 100

点击提交按钮上传一个新的提交

可以自愿在此补充一些提交信息

提交

这是目前为止你的提交 (✓ 下载排行榜上的所有提交):

| # | 分数 | 文件名 | 提交日期 | 状态 | ✓ | |
|---|-----|----------------|---------------------|----------|---|---|
| 1 | 0.0 | submission.zip | 11/05/2018 00:47:07 | Finished | | + |
| 2 | 0.0 | submission.zip | 11/05/2018 00:49:50 | Finished | | + |
| 3 | --- | submission.zip | 11/05/2018 09:53:24 | Failed | | + |
| 4 | 0.0 | submission.zip | 11/05/2018 09:54:02 | Finished | | + |

总计最大提交数 100

点击提交按钮上传一个新的提交

可以自愿在此补充一些提交信息

提交

这是目前为止你的提交 (✓ 下载排行榜上的所有提交):

| # | 分数 | 文件名 | 提交日期 | 状态 | ✓ | |
|---|-----|----------------|---------------------|----------|---|---|
| 1 | 0.0 | submission.zip | 11/05/2018 00:47:07 | Finished | | + |
| 2 | 0.0 | submission.zip | 11/05/2018 00:49:50 | Finished | | + |
| 3 | --- | submission.zip | 11/05/2018 09:53:24 | Failed | | + |
| 4 | 0.0 | submission.zip | 11/05/2018 09:54:02 | Finished | | - |

Description:

更新解释

下载你的提交
浏览评分器读入信息
浏览评分器错误日志
浏览预测的输出日志
浏览预测的错误日志
从预处理阶段下载你代码的运行结果
下载最终的评分结果

Submit to Leaderboard

| # | 用户名 | 分数 |
|---|--------|--------|
| 1 | KimLee | 100.00 |





街景数字识别

竞赛组织者: KimLee - 当前服务器时间: 十一月 6, 2018, 1:08 p.m. 北京时间

最近的

test

一月 10, 2018, 8 a.m. 北京时间

结束

竞赛结束

永不结束

比赛细则

阶段

参与比赛

查看成绩

团队

test

阶段说明

无

每天最大提交数 999

总计最大提交数 100

下载CSV文件

Results

| # | 用户名 | 登录 | 上次登录日期 | 团队名称 | score ▲ |
|---|--------|----|----------|------|------------|
| 1 | KimLee | 10 | 10/31/18 | | 100.00 (1) |

| # | 用户名 | 分数 |
|---|--------|--------|
| 1 | KimLee | 100.00 |



Team Management

[返回竞赛主页](#)

街景数字识别

你不在竞赛的任何一个团队中。你可以自己创建一个新的团队或者向一个已经存在的团队发出登记请求

创建新的团队

来自其他团队的邀请

| 团队 | 消息内容: | 日期 | 操作 |
|----|-------|----|----|
|----|-------|----|----|

给其他团队的请求

| 团队 | 消息内容: | 日期 | 操作 |
|----|-------|----|----|
|----|-------|----|----|



Edit Team

街景数字识别

团队信息

Name:

Description:

Allow requests:



Image:

选择文件

未选择任何文件

提交

machine comprehension

你不在竞赛的任何一个团队中。你可以自己创建一个新的团队或者向一个已经存在的团队发出登记请求



创建新的团队

来自其他团队的邀请

| 团队 | 消息内容: | 日期 | 操作 |
|----|-------|----|----|
|----|-------|----|----|

给其他团队的请求

| 团队 | 消息内容: | 日期 | 操作 |
|----|-------|----|----|
|----|-------|----|----|

| | | |
|---|-----------------------------------|----------------------|
|  | | 十一月 06, 2018 0 成员 |
| 报名请求 | | |
|  | RANCHOR 创建者: xiejinluo test | 十一月 06, 2018 0 成员 |
| 报名请求 | | |

Team Management

[返回竞赛主页](#)

machine comprehension

来自用户的请求

13061002

未处理的请求

状态

通过

原因是:

Process



RANCHOR

创建者: xiejinluo

test

十一月 06, 2018

0 成员

编辑

团队成员

| # | NAME | EMAIL | JOINED | STATUS | ENTRIES |
|---|-----------|-----------------------|------------------------|--------|---------|
| 1 | xiejinluo | xiejinluo@buaa.edu.cn | 十一月 6, 2018, 1:11 p.m. | 拥有者 | 0 |

提交

时间节点

- 11.13 23:00之前，队长将团队成员的学号、姓名以及队长联系方式以excel的形式发送到: xiejinluo@buaa.edu.cn，excel文件命名为队长学号_报名，如:16061001_报名.excel。如果组队存在困难，请在excel中注明你的联系方式和情况，助教会根据情况分配团队。最终分组情况以公邮发布为准。
- 11.26 23:00，个人作业停止提交，同时在此之前，将个人作业报告发送到: xiejinluo@buaa.edu.cn,文件命名为学号_姓名_个人作业.pdf。
- 12.9 23:00之前，队长提交团队展示ppt到xiejinluo@buaa.edu.cn，ppt中需包含团队分工及个人贡献，ppt命名为组号_团队展示.ppt。
- 12.11和12.18 进行大作业展示，按报名顺序的逆序进行展示，展示顺序会提前发送到公邮。