

18 级实验班暑假科研作业要求（可自由选择组合，满分 100 分，加满为止）：

1. 仔细阅读 <2020 中国计算机视觉人才调研报告>(<https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/innovation/deloitte-cn-iddc-2020-china-computer-vision-talent-survey-report-zh-210220.pdf>), 结合实验班的学习经验、做过的作业及报告中关于企业对人才的需求, 讨论你未来如何学习计算机视觉技术。(不少于 1000 字, 20 分)
2. 仔细阅读知乎上的《怎么看待近期的人工智能企业上市频繁遇阻或者失败? 》, 并在知乎里检索用户对商汤科技等知名人工智能公司的评论, 讨论你从这些评论中悟出什么。(不少于 1000 字, 20 分)
3. 仔细阅读周志华老师的“深度森林”项目 (<https://deep-forest.readthedocs.io/en/stable/>), 在百度 AI Studio 上新建立一个公开 Notebook 项目 (命名为“深度森林重现-姓名-山大威海”), 用纯 python+飞桨重现深度森林, 并把深度森林和随机森林算法进行对比 (需使用指定 160 个数据集测试, 在 AI Studio/数据集/公开数据集, 搜“深度森林测试数据集”), 使用统计检验的方法论证深度森林是否优于随机森林。具体要求如下:
 - 3.1. Notebook 中详细介绍什么是深度森林及其原理 (20 分)
 - 3.2. Notebook 中用纯 python+飞桨重现深度森林 (20 分)
 - 3.3. Notebook 中分别使用深度森林和随机森林预测不少于 100 个测试数据集的分类准确度 (设定随机种子, 使用) (30 分) 【注意内存溢出, 可以每次只处理若干个数据集】
 - 3.4. 按照这篇论文的方法 (<https://deepai.org/publication/subsampling-bias-and-the-best-discrepancy-systematic-cross-validation>), Notebook 中使用适当的统计检验方法, 回答整体而言, 深度森林是否真正优于随机森林算法 (10 分) 以及在什么样的数据集中会优于或劣于随机森林算法 (20 分)。

截止日前: 2021 年 8 月 15 日 (不能延后, 否则会影响学校推免)

分组: 每组不超过 3 人

提交: 第一题、第二题做成 PDF, 以名字+学号命名, 交给班长。第三题 Notebook 链接汇总给班长 (务必设为公开项目)

视频: B 站视频 (含所有选择的题目内容) 在实验班公共群分享
不允许使用非百度 AI Studio 以外的开发环境。