

2018 Dian AI Assignment

本次Assignment包含两道题目，大家可以根据自己的兴趣任选其一完成，如果都完成，算作加分项。

请在规定时间内完成相应任务，做好任务记录，如每次训练的时间，模型的准确度，使用多种机器学习评价标准对模型进行评价，同时请保留模型参数。此外为了方便我们的检查，请写明代码的使用方式及接口，写好注释。

3周内请将所有代码、任务记录、jupyter文件、测试结果、任务报告等私发给我们。测试结果的展示全部通过Jupyter Notebook 完成，请没有了解过的同学提前了解一下Jupyter的使用。

DDL: 5月12日夜12時

Task1: 真假咸鱼

【任务】

1. 自己实现逻辑回归算法
2. 使用所给数据集对逻辑回归进行评测，并尝试提升训练效果，与 Scikit Learn 的逻辑回归做对比测试(可尝试使用 AUC、F-score 等评价标准)。

【加分项】

1. 实现并尝试各种优化算法(梯度下降，牛顿法，各种自适应优化算法如Adagrad, Adadelta等)，熟悉其优缺点。
2. 自己手动实现AUC等评价标准
3. 优异的测试效果

【数据集】

有很多人都叫自己咸鱼，但是有很多是假的咸鱼，为了分辨到底是真的咸鱼还是假的咸鱼，我们搜集了 salted fish 数据集: 链接: <https://share.weiyun.com/54uELEp> 密码: Dian18

前面 8 列是他们的特征，最后一列是他们的标签，0 是假的咸鱼，1 是真的咸鱼，还有一些人在我们统计的时候并没有告诉我们某些特征(数据有缺失)，我们的数据集已经分为两个部分一个是训练集，一个是测试集。你需要使用逻辑回归建立一个模型，来根据特征判断他们是不是咸鱼。

【提交】

1. 完整代码
2. 测试展示用的Jupyter Notebook

Task2: “苟”年之“苟”的识别

【任务】

- 利用机器学习或者深度学习，实现狗的品种识别 (一共30类)。
- 输入:狗的图片输出:输出排名前五的类别及其置信度，即(Top-5)。

【要求】

1. 可以使用 PyTorch / TensorFlow / MXNet / Scikit-Learn 等第三方库。
2. 对图片实施相应的预处理。

3. 自行划分数据集验证集。
4. 提交可以直接测试的 Jupyter Notebook、参数文件、模型文件。(注意!直接测试)。
5. 完成一份模型与结果分析报告(超参数说明, 不同模型差异对比, 模型的优缺点, 结果曲线)。
6. 如果抄袭, 直接取消资格。
7. 不得使用他人任何预训练模型, 包括ImageNet预训练模型。
8. 使用多样化数据增强
9. 一份完整且清晰的报告
10. 合理的训练 tricks
11. 测试效果加分

【提交】

1. 完整代码
2. 报告
3. 测试展示用的Jupyter Notebook

【数据集】 苟

链接: <https://share.weiyun.com/5tUuckq> 密码: Dian18

李泽康, 张中洋

2018.4.21