

Adaboost 算法实现

Hint

赶时间可以直接转跳到[测试例子](#)

文件目录结构

```
.
├── data
│   ├── data.csv
│   ├── evaluate.py
│   ├── experiments
│   └── targets.csv
├── readme.pdf
├── report
│   └── Adaboost_U202115325_宁毓伟_机器学习结课报告.pdf
├── src
│   ├── Adaboost.py
│   ├── DataReader.py
│   ├── DecisionStump.py
│   ├── LogisticRegressionClassifier.py
│   └── main.py
└── weight
    ├── AdaboostDecisionStumpWeight.npy
    └── AdaboostLogisticRegressionWeight.npy
```

没有给出相应的命令行参数时，相关文件保存的位置与上述目录一致

依赖

```
matplotlib==3.7.1
numpy==1.24.3
pandas==2.0.1
SciencePlots==2.0.1
scikit_learn==1.2.2
tqdm==4.65.0
```

测试方法

```
cd src/
python main.py <args>
```

- 使用 `-t {train, predict}` 指定训练或者预测

- 使用 `-c {DecisionStump, LogisticRegression}` 指定分类器，默认前者
- 使用 `-d <data file>` 指定输入的特征(不包含label)
- 使用 `-l <label file>` 指定输入的label
- 使用 `-o <output dir>` 指定输出的文件夹
- 预测文件形式 `'/base%d_pred.csv' % (n_base_classifier)`
- 使用 `-f <weight file>` 指定权重保存位置

训练例子

```
cd src/  
python main.py -t train -c DecisionStump -d <训练集特征文件> -l <训练集标签文件>  
> -f ../weight/AdaboostDecisionStumpWeight.npy  
# or  
python main.py -t train -c LogisticRegression -d <训练集特征文件> -l <训练集标签文件> -f ../weight/AdaboostLogisticRegressionWeight.npy
```

测试例子

用 `test.csv` 的位置替换 `<新的特征文件>`

本实验已经为您训练了两种权重，分别位于 `weight/AdaboostDecisionStumpWeight.npy` 以及 `weight/AdaboostLogisticRegressionWeight.npy` 中。

```
cd src/  
python main.py -t predict -c DecisionStump -d <新的特征文件> -f  
../weight/AdaboostDecisionStumpWeight.npy  
# or  
python main.py -t predict -c LogisticRegression -d <新的特征文件> -f  
../weight/AdaboostLogisticRegressionWeight.npy
```