**.py文件说明**

|  |  |  |
| --- | --- | --- |
| **序号** | **文件名** | **说明** |
| 1 | BinaryTree | 包含二叉树类，决策节点类，叶子节点类 |
| 2 | tree\_tools | 其中定义了二分数据方法和测试方法 |
| 3 | plt\_confusion\_matrix | 基于matplotlib绘制混淆矩阵，由DTreeClassifier调用 |
| 4 | DTreeClassifier | 自己实现的分类决策树，包含entropy、gini两种方法 |
| 5 | DTreeRegressor | 自己实现的回归决策树 |
| 6 | prune\_regressor | 回归决策树剪枝方法，由DTreeRegressor调用 |
| 7 | sklearn\_DTreeClassifier | 基于sklearn的分类决策树 |
| 8 | sklearn\_DTreeRegressor | 基于sklearn的回归决策树 |

**figure文件夹内为各个决策树的运行结果**

|  |  |  |  |
| --- | --- | --- | --- |
| 文件 | 说明 | 方法 | 准确率 |
| my\_classifier\_entropy\_confusion\_0.872.png | 自己实现的分类决策树的测试结果混淆矩阵 | entropy | 0.872 |
| my\_classifier\_entropy\_tree\_0.872.gv.pdf | 自己实现的分类决策树的可视化 | entropy | 0.872 |
| my\_classifier\_gini\_confusion\_0.842.png | 自己实现的分类决策树的测试结果混淆矩阵 | gini | 0.842 |
| my\_classifier\_gini\_tree\_0.842.gv.pdf | 自己实现的分类决策树的可视化 | gini | 0.842 |
| my\_regressor\_tree\_SSE\_2595.675\_R2\_0.757.gv.pdf | 自己实现的回归决策树的可视化 | rrs | SSE = 2595.675  = 0.757 |
| sklearn\_classifier\_entropy\_confusion\_0.8444.png | 基于sklearn实现的分类决策树的测试结果混淆矩阵 | entropy | 0.844 |
| sklearn\_classifier\_entropy\_tree.pdf | 基于sklearn实现的分类决策树的可视化 | entropy | 0.844 |
| sklearn\_classifier\_gini\_confusion\_0.8139.png | 基于sklearn实现的分类决策树的测试结果混淆矩阵 | gini | 0.813 |
| sklearn\_classifier\_gini\_tree.pdf | 基于sklearn实现的分类决策树的可视化 | gini | 0.813 |
| sklearn\_regressor\_tree\_SSE\_1933.184\_R2\_0.768.gv.pdf | 基于sklearn实现的回归决策树的可视化 |  | sse = 1933.184  = 0.768 |

**对比**

**分类**

|  |  |  |
| --- | --- | --- |
|  | **我实现的** | **sklearn** |
| **entropy准确率** | 0.872 | 0.844 |
| **entropy混淆矩阵** |  |  |
| **gini准确率** | 0.842 | 0.813 |
| **gini 混淆矩阵** |  |  |

**回归 （树深度同为10）**

|  |  |  |
| --- | --- | --- |
|  | **我实现的（剪枝后）** | **sklearn** |
| **sse** | 2595.675 | 1933.184 |
|  | 0.757 | 0.768 |