**第7章 决策树**

**1.**概述**决策树算法**的思路、关键操作（不超过300字，逻辑要通，**要有小标题**，要排版清晰，请不要网络大幅摘抄）。

**决策树（Decision Tree，DT）**

决策树是一种常用的机器学习算法，可用于处理分类和回归问题，既可以用来高效地对未知的数据进行分类，也可以用来做预测。决策树通过学习可以得到一个树形结构模型，它表示的是对象属性与对象值之间的一种映射关系。算法有 ID3、C4.5、CART 等。已经成功被应用于商业、医学、金融分析、分子生物学领域。

**决策树的基本原理**

**基本原理： 决策树学习算法就是一棵树的构造过程，它通过不断地选择最优特征，并根据该特征对训练数据进行分割，即对特征空间进行划分，使得各个子数据集有一个最好的分类过程。一棵决策树包含一个根节点、若干个内部节点和若干个叶节点，根节点包含样本全集，其他每个节点对应一个属性测试，叶节点对应于决策结果，从根节点到各个叶节点的路径对应了一个判定测试序列。**

**2.利用决策树解决问题（回归或分类）。**

附：Sklearn常用数据集

|  |  |  |
| --- | --- | --- |
| 序号 | 数据集 | 学号尾号 |
| 1 | 鸢尾花数据集：load\_iris（） | 0 1 |
| 2 | 手写数字数据集：load\_digits（） | 2 3 |
| 3 | 乳腺癌数据集load-barest-cancer（） | 4 5 |
| 4 | 糖尿病数据集：load-diabetes（） | 6 7 |
| 5 | 波士顿房价数据集：load-boston（） | 8 |
| 6 | 体能训练数据集：load-linnerud（） | 9 |

利用决策树对以上数据集（学号尾号不同采用不同数据集）进行分析，找出最佳参数下的决策树或随机森林更适合本数据集（准确率更高，不同回归、不同参数下的不同准确率要列出）。

要求：

1）针对当前数据、模型，**详叙数据预处理、模型参数调整过程及效果比对**；

2）代码要列出，重点代码加注释说明，特别是自己调试过程中的自我理解；

3）运行结果要截图，结果要文字说明；

4）注意排版

*from* matplotlib.pyplot *import* subplot

*from* numpy.random *import* random

*from* sklearn *import* svm

*import* numpy *as* np

*import* matplotlib.pyplot *as* plt

*import* matplotlib *as* mpl

*from* sklearn.tree *import* plot\_tree

*from* matplotlib *import* colors

*from* sklearn *import* model\_selection

*from* sklearn.tree *import* export\_graphviz

*from* sklearn *import* tree

*import* pydotplus

*from* sklearn.datasets *import* load\_digits

*from* sklearn *import* tree

*import* numpy *as* np

*from* sklearn.preprocessing *import* StandardScaler

def show\_num():

*"""*

*显示手写数字数据集中的前4个数字图像*

*使用matplotlib的matshow函数显示灰度图像*

*"""*

    digits = load\_digits()

    plt.gray()  *# 设置为灰度显示模式*

    plt.matshow(digits.images[0])  *# 显示第一个数字*

    plt.matshow(digits.images[1])  *# 显示第二个数字*

    plt.matshow(digits.images[2])  *# 显示第三个数字*

    plt.matshow(digits.images[3])  *# 显示第四个数字*

    plt.show()

def load\_data():

*"""*

*加载手写数字数据集并进行预处理*

*返回：*

*data\_X: 特征矩阵，每行代表一个数字图像（8x8像素）*

*data\_y: 标签向量，表示每个图像对应的数字（0-9）*

*"""*

    data\_set = []

    data\_X = []

    data\_y = []

    digits = load\_digits()  *# 加载手写数字数据集*

    data\_X.append(digits.data)  *# 添加特征数据*

    data\_y.append(digits.target)  *# 添加标签数据*

    data\_X = np.array(data\_X)

    data\_X = np.reshape(data\_X, (data\_X.shape[1], data\_X.shape[2]))  *# 重塑数组维度*

    data\_y = np.array(data\_y)

    data\_y = np.reshape(data\_y, data\_y.shape[1])  *# 重塑标签数组*

*return* data\_X, data\_y

def split\_data(*data\_X*, *data\_y*):

*"""*

*将数据集分割为训练集和测试集*

*参数：*

*data\_X: 特征矩阵*

*data\_y: 标签向量*

*返回：*

*X\_train: 训练集特征*

*X\_test: 测试集特征*

*y\_train: 训练集标签*

*y\_test: 测试集标签*

*"""*

*# 使用sklearn的train\_test\_split函数分割数据集*

    X\_train, X\_test, y\_train, y\_test = model\_selection.train\_test\_split(*data\_X*, *data\_y*, *test\_size*=0.2, *random\_state*=0)

*# 将特征数据转换为浮点数类型*

*for* index, item *in* enumerate(X\_train):

        X\_train[index] = list(map(float, item))

*for* index, item *in* enumerate(X\_test):

        X\_test[index] = list(map(float, item))

*# 将数据转换为numpy数组*

    X\_train = np.array(X\_train)

    X\_test = np.array(X\_test)

    y\_train = np.array(y\_train)

    y\_test = np.array(y\_test)

*return* X\_train, X\_test, y\_train, y\_test

*# 加载数据*

data\_X, data\_y = load\_data()

*# 分割数据集*

x\_train, x\_test, y\_train, y\_test = split\_data(data\_X, data\_y)

*# 数据标准化处理*

scaler = StandardScaler()

x\_train = scaler.fit\_transform(x\_train)  *# 对训练集进行标准化*

x\_test = scaler.transform(x\_test)  *# 对测试集进行标准化*

*# 创建并训练决策树分类器*

classifier = tree.DecisionTreeClassifier()

classifier.fit(x\_train, y\_train)  *# 训练模型*

*# 评估模型性能*

score = classifier.score(x\_test, y\_test)

print(score)  *# 打印准确率*

*# 将决策树导出为.dot文件，用于可视化*

*with* open("kk.dot", 'w') *as* f:

    f = export\_graphviz(classifier, *out\_file*=f)

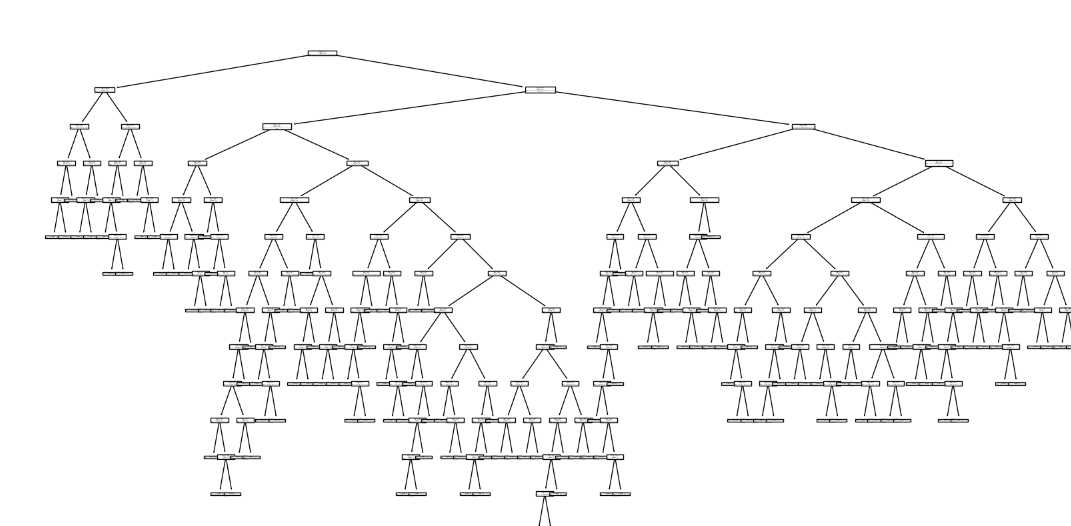
*# 使用matplotlib绘制决策树*

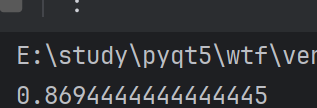
plt.figure(*figsize*=(20, 10))

plot\_tree(classifier)

plt.show()

**试验结果：**





digraph Tree {  
node [shape=box, fontname="helvetica"] ;  
edge [fontname="helvetica"] ;  
0 [label="x[36] <= -1.641\ngini = 0.9\nsamples = 1437\nvalue = [151, 147, 141, 154, 151, 142, 137, 140, 135, 139]"] ;  
1 [label="x[42] <= 0.085\ngini = 0.535\nsamples = 227\nvalue = [147, 0, 3, 1, 5, 19, 5, 0, 3, 44]"] ;  
0 -> 1 [labeldistance=2.5, labelangle=45, headlabel="True"] ;  
2 [label="x[21] <= -0.135\ngini = 0.437\nsamples = 62\nvalue = [1, 0, 1, 1, 0, 15, 0, 0, 0, 44]"] ;  
1 -> 2 ;  
3 [label="x[46] <= -0.563\ngini = 0.227\nsamples = 16\nvalue = [1, 0, 1, 0, 0, 14, 0, 0, 0, 0]"] ;  
2 -> 3 ;  
4 [label="x[51] <= 0.746\ngini = 0.5\nsamples = 2\nvalue = [1, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
3 -> 4 ;  
5 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
4 -> 5 ;  
6 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
4 -> 6 ;  
7 [label="gini = 0.0\nsamples = 14\nvalue = [0, 0, 0, 0, 0, 14, 0, 0, 0, 0]"] ;  
3 -> 7 ;  
8 [label="x[33] <= -0.246\ngini = 0.084\nsamples = 46\nvalue = [0, 0, 0, 1, 0, 1, 0, 0, 0, 44]"] ;  
2 -> 8 ;  
9 [label="x[18] <= -1.388\ngini = 0.043\nsamples = 45\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 44]"] ;  
8 -> 9 ;  
10 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
9 -> 10 ;  
11 [label="gini = 0.0\nsamples = 44\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 44]"] ;  
9 -> 11 ;  
12 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
8 -> 12 ;  
13 [label="x[21] <= -1.182\ngini = 0.214\nsamples = 165\nvalue = [146, 0, 2, 0, 5, 4, 5, 0, 3, 0]"] ;  
1 -> 13 ;  
14 [label="x[30] <= 0.042\ngini = 0.729\nsamples = 15\nvalue = [0, 0, 2, 0, 5, 4, 4, 0, 0, 0]"] ;  
13 -> 14 ;  
15 [label="x[27] <= -0.308\ngini = 0.64\nsamples = 10\nvalue = [0, 0, 2, 0, 0, 4, 4, 0, 0, 0]"] ;  
14 -> 15 ;  
16 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 4, 0, 0, 0, 0]"] ;  
15 -> 16 ;  
17 [label="x[20] <= -0.26\ngini = 0.444\nsamples = 6\nvalue = [0, 0, 2, 0, 0, 0, 4, 0, 0, 0]"] ;  
15 -> 17 ;  
18 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 0, 4, 0, 0, 0]"] ;  
17 -> 18 ;  
19 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 2, 0, 0, 0, 0, 0, 0, 0]"] ;  
17 -> 19 ;  
20 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 5, 0, 0, 0, 0, 0]"] ;  
14 -> 20 ;  
21 [label="x[28] <= -0.879\ngini = 0.052\nsamples = 150\nvalue = [146, 0, 0, 0, 0, 0, 1, 0, 3, 0]"] ;  
13 -> 21 ;  
22 [label="gini = 0.0\nsamples = 146\nvalue = [146, 0, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
21 -> 22 ;  
23 [label="x[34] <= 1.248\ngini = 0.375\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 3, 0]"] ;  
21 -> 23 ;  
24 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 3, 0]"] ;  
23 -> 24 ;  
25 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
23 -> 25 ;  
26 [label="x[26] <= 0.064\ngini = 0.888\nsamples = 1210\nvalue = [4.0, 147.0, 138.0, 153.0, 146.0, 123.0, 132.0, 140.0\n132.0, 95.0]"] ;  
0 -> 26 [labeldistance=2.5, labelangle=-45, headlabel="False"] ;  
27 [label="x[53] <= -1.377\ngini = 0.812\nsamples = 589\nvalue = [0.0, 74.0, 131.0, 148.0, 6.0, 8.0, 8.0, 123.0, 70.0\n21.0]"] ;  
26 -> 27 ;  
28 [label="x[19] <= 0.604\ngini = 0.412\nsamples = 154\nvalue = [0, 26, 0, 3, 0, 3, 0, 115, 6, 1]"] ;  
27 -> 28 ;  
29 [label="x[37] <= -1.398\ngini = 0.126\nsamples = 121\nvalue = [0, 0, 0, 3, 0, 0, 0, 113, 4, 1]"] ;  
28 -> 29 ;  
30 [label="x[42] <= -0.679\ngini = 0.444\nsamples = 6\nvalue = [0, 0, 0, 2, 0, 0, 0, 0, 4, 0]"] ;  
29 -> 30 ;  
31 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 2, 0, 0, 0, 0, 0, 0]"] ;  
30 -> 31 ;  
32 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 4, 0]"] ;  
30 -> 32 ;  
33 [label="x[55] <= 1.391\ngini = 0.034\nsamples = 115\nvalue = [0, 0, 0, 1, 0, 0, 0, 113, 0, 1]"] ;  
29 -> 33 ;  
34 [label="gini = 0.0\nsamples = 113\nvalue = [0, 0, 0, 0, 0, 0, 0, 113, 0, 0]"] ;  
33 -> 34 ;  
35 [label="x[52] <= -1.429\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 1]"] ;  
33 -> 35 ;  
36 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
35 -> 36 ;  
37 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
35 -> 37 ;  
38 [label="x[2] <= 0.058\ngini = 0.364\nsamples = 33\nvalue = [0, 26, 0, 0, 0, 3, 0, 2, 2, 0]"] ;  
28 -> 38 ;  
39 [label="gini = 0.0\nsamples = 26\nvalue = [0, 26, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
38 -> 39 ;  
40 [label="x[21] <= -1.021\ngini = 0.653\nsamples = 7\nvalue = [0, 0, 0, 0, 0, 3, 0, 2, 2, 0]"] ;  
38 -> 40 ;  
41 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 3, 0, 0, 0, 0]"] ;  
40 -> 41 ;  
42 [label="x[59] <= -0.012\ngini = 0.5\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 0, 0, 2, 2, 0]"] ;  
40 -> 42 ;  
43 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 2, 0, 0]"] ;  
42 -> 43 ;  
44 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
42 -> 44 ;  
45 [label="x[43] <= -0.889\ngini = 0.761\nsamples = 435\nvalue = [0, 48, 131, 145, 6, 5, 8, 8, 64, 20]"] ;  
27 -> 45 ;  
46 [label="x[29] <= 1.025\ngini = 0.434\nsamples = 173\nvalue = [0.0, 11.0, 2.0, 128.0, 0.0, 3.0, 0.0, 2.0, 9.0, 18.0]"] ;  
45 -> 46 ;  
47 [label="x[34] <= 0.218\ngini = 0.246\nsamples = 141\nvalue = [0, 5, 1, 122, 0, 2, 0, 0, 9, 2]"] ;  
46 -> 47 ;  
48 [label="x[19] <= 1.122\ngini = 0.158\nsamples = 132\nvalue = [0, 5, 1, 121, 0, 2, 0, 0, 1, 2]"] ;  
47 -> 48 ;  
49 [label="x[63] <= 1.848\ngini = 0.063\nsamples = 124\nvalue = [0, 0, 1, 120, 0, 1, 0, 0, 1, 1]"] ;  
48 -> 49 ;  
50 [label="x[18] <= 0.998\ngini = 0.048\nsamples = 123\nvalue = [0, 0, 0, 120, 0, 1, 0, 0, 1, 1]"] ;  
49 -> 50 ;  
51 [label="x[42] <= 1.003\ngini = 0.032\nsamples = 122\nvalue = [0, 0, 0, 120, 0, 0, 0, 0, 1, 1]"] ;  
50 -> 51 ;  
52 [label="x[18] <= 0.644\ngini = 0.017\nsamples = 120\nvalue = [0, 0, 0, 119, 0, 0, 0, 0, 0, 1]"] ;  
51 -> 52 ;  
53 [label="gini = 0.0\nsamples = 116\nvalue = [0, 0, 0, 116, 0, 0, 0, 0, 0, 0]"] ;  
52 -> 53 ;  
54 [label="x[27] <= 0.626\ngini = 0.375\nsamples = 4\nvalue = [0, 0, 0, 3, 0, 0, 0, 0, 0, 1]"] ;  
52 -> 54 ;  
55 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 3, 0, 0, 0, 0, 0, 0]"] ;  
54 -> 55 ;  
56 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
54 -> 56 ;  
57 [label="x[58] <= 0.667\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 1, 0]"] ;  
51 -> 57 ;  
58 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
57 -> 58 ;  
59 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
57 -> 59 ;  
60 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
50 -> 60 ;  
61 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
49 -> 61 ;  
62 [label="x[61] <= 0.901\ngini = 0.562\nsamples = 8\nvalue = [0, 5, 0, 1, 0, 1, 0, 0, 0, 1]"] ;  
48 -> 62 ;  
63 [label="x[52] <= 0.564\ngini = 0.667\nsamples = 3\nvalue = [0, 0, 0, 1, 0, 1, 0, 0, 0, 1]"] ;  
62 -> 63 ;  
64 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
63 -> 64 ;  
65 [label="x[42] <= -0.297\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 1]"] ;  
63 -> 65 ;  
66 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
65 -> 66 ;  
67 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
65 -> 67 ;  
68 [label="gini = 0.0\nsamples = 5\nvalue = [0, 5, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
62 -> 68 ;  
69 [label="x[44] <= -0.352\ngini = 0.198\nsamples = 9\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 8, 0]"] ;  
47 -> 69 ;  
70 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
69 -> 70 ;  
71 [label="gini = 0.0\nsamples = 8\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 8, 0]"] ;  
69 -> 71 ;  
72 [label="x[3] <= -1.931\ngini = 0.674\nsamples = 32\nvalue = [0, 6, 1, 6, 0, 1, 0, 2, 0, 16]"] ;  
46 -> 72 ;  
73 [label="gini = 0.0\nsamples = 6\nvalue = [0, 6, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
72 -> 73 ;  
74 [label="x[26] <= -1.301\ngini = 0.559\nsamples = 26\nvalue = [0, 0, 1, 6, 0, 1, 0, 2, 0, 16]"] ;  
72 -> 74 ;  
75 [label="x[20] <= -1.065\ngini = 0.406\nsamples = 8\nvalue = [0, 0, 1, 6, 0, 0, 0, 1, 0, 0]"] ;  
74 -> 75 ;  
76 [label="x[36] <= -0.049\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 1, 0, 0, 0, 0, 1, 0, 0]"] ;  
75 -> 76 ;  
77 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
76 -> 77 ;  
78 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
76 -> 78 ;  
79 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 0, 6, 0, 0, 0, 0, 0, 0]"] ;  
75 -> 79 ;  
80 [label="x[44] <= 0.446\ngini = 0.204\nsamples = 18\nvalue = [0, 0, 0, 0, 0, 1, 0, 1, 0, 16]"] ;  
74 -> 80 ;  
81 [label="x[42] <= 0.162\ngini = 0.111\nsamples = 17\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 16]"] ;  
80 -> 81 ;  
82 [label="gini = 0.0\nsamples = 16\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 16]"] ;  
81 -> 82 ;  
83 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
81 -> 83 ;  
84 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
80 -> 84 ;  
85 [label="x[27] <= 0.117\ngini = 0.687\nsamples = 262\nvalue = [0, 37, 129, 17, 6, 2, 8, 6, 55, 2]"] ;  
45 -> 85 ;  
86 [label="x[38] <= -0.694\ngini = 0.237\nsamples = 116\nvalue = [0, 1, 101, 4, 0, 0, 3, 6, 0, 1]"] ;  
85 -> 86 ;  
87 [label="x[25] <= 1.45\ngini = 0.057\nsamples = 103\nvalue = [0.0, 1.0, 100.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0]"] ;  
86 -> 87 ;  
88 [label="x[50] <= -1.247\ngini = 0.039\nsamples = 102\nvalue = [0, 1, 100, 1, 0, 0, 0, 0, 0, 0]"] ;  
87 -> 88 ;  
89 [label="x[37] <= -1.144\ngini = 0.667\nsamples = 3\nvalue = [0, 1, 1, 1, 0, 0, 0, 0, 0, 0]"] ;  
88 -> 89 ;  
90 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
89 -> 90 ;  
91 [label="x[43] <= 1.049\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
89 -> 91 ;  
92 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
91 -> 92 ;  
93 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
91 -> 93 ;  
94 [label="gini = 0.0\nsamples = 99\nvalue = [0, 0, 99, 0, 0, 0, 0, 0, 0, 0]"] ;  
88 -> 94 ;  
95 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
87 -> 95 ;  
96 [label="x[53] <= -0.381\ngini = 0.698\nsamples = 13\nvalue = [0, 0, 1, 3, 0, 0, 2, 6, 0, 1]"] ;  
86 -> 96 ;  
97 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 0, 0, 0, 0, 0, 6, 0, 0]"] ;  
96 -> 97 ;  
98 [label="x[18] <= 0.203\ngini = 0.694\nsamples = 7\nvalue = [0, 0, 1, 3, 0, 0, 2, 0, 0, 1]"] ;  
96 -> 98 ;  
99 [label="x[5] <= -0.767\ngini = 0.625\nsamples = 4\nvalue = [0, 0, 1, 0, 0, 0, 2, 0, 0, 1]"] ;  
98 -> 99 ;  
100 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 2, 0, 0, 0]"] ;  
99 -> 100 ;  
101 [label="x[54] <= 0.594\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 1]"] ;  
99 -> 101 ;  
102 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
101 -> 102 ;  
103 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
101 -> 103 ;  
104 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 3, 0, 0, 0, 0, 0, 0]"] ;  
98 -> 104 ;  
105 [label="x[36] <= -0.132\ngini = 0.749\nsamples = 146\nvalue = [0, 36, 28, 13, 6, 2, 5, 0, 55, 1]"] ;  
85 -> 105 ;  
106 [label="x[5] <= 0.469\ngini = 0.231\nsamples = 30\nvalue = [0, 0, 26, 4, 0, 0, 0, 0, 0, 0]"] ;  
105 -> 106 ;  
107 [label="gini = 0.0\nsamples = 26\nvalue = [0, 0, 26, 0, 0, 0, 0, 0, 0, 0]"] ;  
106 -> 107 ;  
108 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 4, 0, 0, 0, 0, 0, 0]"] ;  
106 -> 108 ;  
109 [label="x[44] <= 1.084\ngini = 0.668\nsamples = 116\nvalue = [0, 36, 2, 9, 6, 2, 5, 0, 55, 1]"] ;  
105 -> 109 ;  
110 [label="x[21] <= -0.538\ngini = 0.433\nsamples = 62\nvalue = [0, 4, 2, 5, 0, 0, 4, 0, 46, 1]"] ;  
109 -> 110 ;  
111 [label="x[58] <= -0.794\ngini = 0.72\nsamples = 10\nvalue = [0, 3, 1, 1, 0, 0, 4, 0, 1, 0]"] ;  
110 -> 111 ;  
112 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 0, 0, 0, 4, 0, 0, 0]"] ;  
111 -> 112 ;  
113 [label="x[26] <= -0.498\ngini = 0.667\nsamples = 6\nvalue = [0, 3, 1, 1, 0, 0, 0, 0, 1, 0]"] ;  
111 -> 113 ;  
114 [label="x[10] <= 0.934\ngini = 0.667\nsamples = 3\nvalue = [0, 0, 1, 1, 0, 0, 0, 0, 1, 0]"] ;  
113 -> 114 ;  
115 [label="x[60] <= -0.99\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 1, 0]"] ;  
114 -> 115 ;  
116 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
115 -> 116 ;  
117 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
115 -> 117 ;  
118 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
114 -> 118 ;  
119 [label="gini = 0.0\nsamples = 3\nvalue = [0, 3, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
113 -> 119 ;  
120 [label="x[18] <= -1.123\ngini = 0.244\nsamples = 52\nvalue = [0, 1, 1, 4, 0, 0, 0, 0, 45, 1]"] ;  
110 -> 120 ;  
121 [label="x[19] <= 0.432\ngini = 0.594\nsamples = 8\nvalue = [0, 1, 0, 4, 0, 0, 0, 0, 3, 0]"] ;  
120 -> 121 ;  
122 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 4, 0, 0, 0, 0, 0, 0]"] ;  
121 -> 122 ;  
123 [label="x[43] <= 0.584\ngini = 0.375\nsamples = 4\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 3, 0]"] ;  
121 -> 123 ;  
124 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 3, 0]"] ;  
123 -> 124 ;  
125 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
123 -> 125 ;  
126 [label="x[43] <= 1.282\ngini = 0.088\nsamples = 44\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 42, 1]"] ;  
120 -> 126 ;  
127 [label="x[28] <= -0.149\ngini = 0.045\nsamples = 43\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 42, 1]"] ;  
126 -> 127 ;  
128 [label="x[3] <= -0.416\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 1]"] ;  
127 -> 128 ;  
129 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
128 -> 129 ;  
130 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
128 -> 130 ;  
131 [label="gini = 0.0\nsamples = 41\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 41, 0]"] ;  
127 -> 131 ;  
132 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0, 0, 0, 0, 0, 0, 0]"] ;  
126 -> 132 ;  
133 [label="x[41] <= 1.63\ngini = 0.602\nsamples = 54\nvalue = [0, 32, 0, 4, 6, 2, 1, 0, 9, 0]"] ;  
109 -> 133 ;  
134 [label="x[10] <= -0.349\ngini = 0.511\nsamples = 48\nvalue = [0, 32, 0, 4, 0, 2, 1, 0, 9, 0]"] ;  
133 -> 134 ;  
135 [label="x[22] <= -0.086\ngini = 0.169\nsamples = 33\nvalue = [0, 30, 0, 0, 0, 0, 1, 0, 2, 0]"] ;  
134 -> 135 ;  
136 [label="x[28] <= -1.122\ngini = 0.064\nsamples = 30\nvalue = [0, 29, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
135 -> 136 ;  
137 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
136 -> 137 ;  
138 [label="gini = 0.0\nsamples = 29\nvalue = [0, 29, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
136 -> 138 ;  
139 [label="x[50] <= 1.308\ngini = 0.444\nsamples = 3\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
135 -> 139 ;  
140 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
139 -> 140 ;  
141 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
139 -> 141 ;  
142 [label="x[42] <= -0.144\ngini = 0.676\nsamples = 15\nvalue = [0, 2, 0, 4, 0, 2, 0, 0, 7, 0]"] ;  
134 -> 142 ;  
143 [label="x[13] <= 1.038\ngini = 0.656\nsamples = 8\nvalue = [0, 1, 0, 4, 0, 2, 0, 0, 1, 0]"] ;  
142 -> 143 ;  
144 [label="x[3] <= 0.75\ngini = 0.625\nsamples = 4\nvalue = [0, 1, 0, 0, 0, 2, 0, 0, 1, 0]"] ;  
143 -> 144 ;  
145 [label="x[52] <= 1.133\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
144 -> 145 ;  
146 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
145 -> 146 ;  
147 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
145 -> 147 ;  
148 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 2, 0, 0, 0, 0]"] ;  
144 -> 148 ;  
149 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 4, 0, 0, 0, 0, 0, 0]"] ;  
143 -> 149 ;  
150 [label="x[29] <= 0.682\ngini = 0.245\nsamples = 7\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 6, 0]"] ;  
142 -> 150 ;  
151 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 6, 0]"] ;  
150 -> 151 ;  
152 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
150 -> 152 ;  
153 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 0, 0, 6, 0, 0, 0, 0, 0]"] ;  
133 -> 153 ;  
154 [label="x[21] <= -1.182\ngini = 0.836\nsamples = 621\nvalue = [4, 73, 7, 5, 140, 115, 124, 17, 62, 74]"] ;  
26 -> 154 ;  
155 [label="x[42] <= 0.238\ngini = 0.652\nsamples = 282\nvalue = [0, 14, 5, 3, 26, 109, 122, 0, 3, 0]"] ;  
154 -> 155 ;  
156 [label="x[5] <= -0.767\ngini = 0.356\nsamples = 133\nvalue = [0, 9, 5, 3, 6, 106, 1, 0, 3, 0]"] ;  
155 -> 156 ;  
157 [label="x[62] <= 3.154\ngini = 0.771\nsamples = 24\nvalue = [0, 9, 5, 2, 4, 1, 1, 0, 2, 0]"] ;  
156 -> 157 ;  
158 [label="x[33] <= 0.184\ngini = 0.797\nsamples = 16\nvalue = [0, 1, 5, 2, 4, 1, 1, 0, 2, 0]"] ;  
157 -> 158 ;  
159 [label="x[27] <= -0.393\ngini = 0.75\nsamples = 12\nvalue = [0, 1, 5, 2, 0, 1, 1, 0, 2, 0]"] ;  
158 -> 159 ;  
160 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 5, 0, 0, 0, 0, 0, 0, 0]"] ;  
159 -> 160 ;  
161 [label="x[18] <= 0.026\ngini = 0.776\nsamples = 7\nvalue = [0, 1, 0, 2, 0, 1, 1, 0, 2, 0]"] ;  
159 -> 161 ;  
162 [label="x[62] <= -0.372\ngini = 0.72\nsamples = 5\nvalue = [0, 1, 0, 2, 0, 1, 1, 0, 0, 0]"] ;  
161 -> 162 ;  
163 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 2, 0, 0, 0, 0, 0, 0]"] ;  
162 -> 163 ;  
164 [label="x[61] <= 1.327\ngini = 0.667\nsamples = 3\nvalue = [0, 1, 0, 0, 0, 1, 1, 0, 0, 0]"] ;  
162 -> 164 ;  
165 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
164 -> 165 ;  
166 [label="x[4] <= -0.909\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
164 -> 166 ;  
167 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
166 -> 167 ;  
168 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
166 -> 168 ;  
169 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
161 -> 169 ;  
170 [label="gini = 0.0\nsamples = 4\nvalue = [0, 0, 0, 0, 4, 0, 0, 0, 0, 0]"] ;  
158 -> 170 ;  
171 [label="gini = 0.0\nsamples = 8\nvalue = [0, 8, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
157 -> 171 ;  
172 [label="x[18] <= -0.946\ngini = 0.072\nsamples = 109\nvalue = [0, 0, 0, 1, 2, 105, 0, 0, 1, 0]"] ;  
156 -> 172 ;  
173 [label="x[15] <= 4.502\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 1, 1, 0, 0, 0, 0, 0]"] ;  
172 -> 173 ;  
174 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
173 -> 174 ;  
175 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 1, 0, 0, 0, 0, 0]"] ;  
173 -> 175 ;  
176 [label="x[16] <= 13.374\ngini = 0.037\nsamples = 107\nvalue = [0.0, 0.0, 0.0, 0.0, 1.0, 105.0, 0.0, 0.0, 1.0, 0.0]"] ;  
172 -> 176 ;  
177 [label="x[24] <= 18.934\ngini = 0.019\nsamples = 106\nvalue = [0, 0, 0, 0, 1, 105, 0, 0, 0, 0]"] ;  
176 -> 177 ;  
178 [label="gini = 0.0\nsamples = 105\nvalue = [0, 0, 0, 0, 0, 105, 0, 0, 0, 0]"] ;  
177 -> 178 ;  
179 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 1, 0, 0, 0, 0, 0]"] ;  
177 -> 179 ;  
180 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
176 -> 180 ;  
181 [label="x[54] <= -0.438\ngini = 0.321\nsamples = 149\nvalue = [0.0, 5.0, 0.0, 0.0, 20.0, 3.0, 121.0, 0.0, 0.0, 0.0]"] ;  
155 -> 181 ;  
182 [label="x[37] <= -0.551\ngini = 0.593\nsamples = 34\nvalue = [0, 5, 0, 0, 20, 3, 6, 0, 0, 0]"] ;  
181 -> 182 ;  
183 [label="x[53] <= -0.464\ngini = 0.579\nsamples = 11\nvalue = [0, 5, 0, 0, 1, 0, 5, 0, 0, 0]"] ;  
182 -> 183 ;  
184 [label="gini = 0.0\nsamples = 5\nvalue = [0, 5, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
183 -> 184 ;  
185 [label="x[41] <= 2.796\ngini = 0.278\nsamples = 6\nvalue = [0, 0, 0, 0, 1, 0, 5, 0, 0, 0]"] ;  
183 -> 185 ;  
186 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 0, 0, 5, 0, 0, 0]"] ;  
185 -> 186 ;  
187 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 1, 0, 0, 0, 0, 0]"] ;  
185 -> 187 ;  
188 [label="x[44] <= -0.033\ngini = 0.299\nsamples = 23\nvalue = [0, 0, 0, 0, 19, 3, 1, 0, 0, 0]"] ;  
182 -> 188 ;  
189 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 3, 0, 0, 0, 0]"] ;  
188 -> 189 ;  
190 [label="x[41] <= -0.203\ngini = 0.095\nsamples = 20\nvalue = [0, 0, 0, 0, 19, 0, 1, 0, 0, 0]"] ;  
188 -> 190 ;  
191 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
190 -> 191 ;  
192 [label="gini = 0.0\nsamples = 19\nvalue = [0, 0, 0, 0, 19, 0, 0, 0, 0, 0]"] ;  
190 -> 192 ;  
193 [label="gini = 0.0\nsamples = 115\nvalue = [0, 0, 0, 0, 0, 0, 115, 0, 0, 0]"] ;  
181 -> 193 ;  
194 [label="x[33] <= 0.327\ngini = 0.776\nsamples = 339\nvalue = [4.0, 59.0, 2.0, 2.0, 114.0, 6.0, 2.0, 17.0, 59.0\n74.0]"] ;  
154 -> 194 ;  
195 [label="x[20] <= 1.349\ngini = 0.748\nsamples = 204\nvalue = [2.0, 41.0, 2.0, 2.0, 7.0, 6.0, 0.0, 15.0, 56.0, 73.0]"] ;  
194 -> 195 ;  
196 [label="x[43] <= -0.579\ngini = 0.665\nsamples = 155\nvalue = [2, 3, 0, 1, 7, 6, 0, 12, 54, 70]"] ;  
195 -> 196 ;  
197 [label="x[28] <= -0.635\ngini = 0.407\nsamples = 83\nvalue = [0, 1, 0, 1, 0, 5, 0, 5, 8, 63]"] ;  
196 -> 197 ;  
198 [label="x[14] <= 1.182\ngini = 0.64\nsamples = 10\nvalue = [0, 1, 0, 1, 0, 0, 0, 5, 3, 0]"] ;  
197 -> 198 ;  
199 [label="x[53] <= 1.112\ngini = 0.56\nsamples = 5\nvalue = [0, 1, 0, 1, 0, 0, 0, 0, 3, 0]"] ;  
198 -> 199 ;  
200 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 3, 0]"] ;  
199 -> 200 ;  
201 [label="x[54] <= 0.387\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
199 -> 201 ;  
202 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
201 -> 202 ;  
203 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
201 -> 203 ;  
204 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 0, 0, 0, 5, 0, 0]"] ;  
198 -> 204 ;  
205 [label="x[42] <= 0.315\ngini = 0.246\nsamples = 73\nvalue = [0, 0, 0, 0, 0, 5, 0, 0, 5, 63]"] ;  
197 -> 205 ;  
206 [label="x[21] <= -0.135\ngini = 0.136\nsamples = 68\nvalue = [0, 0, 0, 0, 0, 5, 0, 0, 0, 63]"] ;  
205 -> 206 ;  
207 [label="x[19] <= -0.604\ngini = 0.278\nsamples = 6\nvalue = [0, 0, 0, 0, 0, 5, 0, 0, 0, 1]"] ;  
206 -> 207 ;  
208 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
207 -> 208 ;  
209 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 0, 5, 0, 0, 0, 0]"] ;  
207 -> 209 ;  
210 [label="gini = 0.0\nsamples = 62\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 62]"] ;  
206 -> 210 ;  
211 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 5, 0]"] ;  
205 -> 211 ;  
212 [label="x[60] <= -0.887\ngini = 0.562\nsamples = 72\nvalue = [2, 2, 0, 0, 7, 1, 0, 7, 46, 7]"] ;  
196 -> 212 ;  
213 [label="x[4] <= -1.026\ngini = 0.72\nsamples = 19\nvalue = [0, 0, 0, 0, 5, 1, 0, 7, 1, 5]"] ;  
212 -> 213 ;  
214 [label="x[52] <= 0.848\ngini = 0.278\nsamples = 6\nvalue = [0, 0, 0, 0, 5, 1, 0, 0, 0, 0]"] ;  
213 -> 214 ;  
215 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 5, 0, 0, 0, 0, 0]"] ;  
214 -> 215 ;  
216 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]"] ;  
214 -> 216 ;  
217 [label="x[17] <= 0.398\ngini = 0.556\nsamples = 13\nvalue = [0, 0, 0, 0, 0, 0, 0, 7, 1, 5]"] ;  
213 -> 217 ;  
218 [label="gini = 0.0\nsamples = 7\nvalue = [0, 0, 0, 0, 0, 0, 0, 7, 0, 0]"] ;  
217 -> 218 ;  
219 [label="x[35] <= 0.791\ngini = 0.278\nsamples = 6\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 5]"] ;  
217 -> 219 ;  
220 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 5]"] ;  
219 -> 220 ;  
221 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
219 -> 221 ;  
222 [label="x[21] <= -0.538\ngini = 0.273\nsamples = 53\nvalue = [2, 2, 0, 0, 2, 0, 0, 0, 45, 2]"] ;  
212 -> 222 ;  
223 [label="x[12] <= 0.34\ngini = 0.5\nsamples = 4\nvalue = [0, 2, 0, 0, 2, 0, 0, 0, 0, 0]"] ;  
222 -> 223 ;  
224 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 2, 0, 0, 0, 0, 0]"] ;  
223 -> 224 ;  
225 [label="gini = 0.0\nsamples = 2\nvalue = [0, 2, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
223 -> 225 ;  
226 [label="x[38] <= 0.15\ngini = 0.153\nsamples = 49\nvalue = [2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 45.0, 2.0]"] ;  
222 -> 226 ;  
227 [label="x[28] <= -1.122\ngini = 0.043\nsamples = 46\nvalue = [1, 0, 0, 0, 0, 0, 0, 0, 45, 0]"] ;  
226 -> 227 ;  
228 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
227 -> 228 ;  
229 [label="gini = 0.0\nsamples = 45\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 45, 0]"] ;  
227 -> 229 ;  
230 [label="x[42] <= 0.468\ngini = 0.444\nsamples = 3\nvalue = [1, 0, 0, 0, 0, 0, 0, 0, 0, 2]"] ;  
226 -> 230 ;  
231 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 2]"] ;  
230 -> 231 ;  
232 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
230 -> 232 ;  
233 [label="x[9] <= -0.469\ngini = 0.387\nsamples = 49\nvalue = [0.0, 38.0, 2.0, 1.0, 0.0, 0.0, 0.0, 3.0, 2.0, 3.0]"] ;  
195 -> 233 ;  
234 [label="x[52] <= -1.239\ngini = 0.178\nsamples = 42\nvalue = [0, 38, 0, 0, 0, 0, 0, 2, 1, 1]"] ;  
233 -> 234 ;  
235 [label="x[58] <= 0.375\ngini = 0.444\nsamples = 3\nvalue = [0, 0, 0, 0, 0, 0, 0, 2, 0, 1]"] ;  
234 -> 235 ;  
236 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
235 -> 236 ;  
237 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 2, 0, 0]"] ;  
235 -> 237 ;  
238 [label="x[28] <= -0.473\ngini = 0.05\nsamples = 39\nvalue = [0, 38, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
234 -> 238 ;  
239 [label="x[36] <= -0.887\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
238 -> 239 ;  
240 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
239 -> 240 ;  
241 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
239 -> 241 ;  
242 [label="gini = 0.0\nsamples = 37\nvalue = [0, 37, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
238 -> 242 ;  
243 [label="x[35] <= -0.091\ngini = 0.776\nsamples = 7\nvalue = [0, 0, 2, 1, 0, 0, 0, 1, 1, 2]"] ;  
233 -> 243 ;  
244 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 2]"] ;  
243 -> 244 ;  
245 [label="x[59] <= 0.781\ngini = 0.72\nsamples = 5\nvalue = [0, 0, 2, 1, 0, 0, 0, 1, 1, 0]"] ;  
243 -> 245 ;  
246 [label="x[34] <= 0.535\ngini = 0.667\nsamples = 3\nvalue = [0, 0, 0, 1, 0, 0, 0, 1, 1, 0]"] ;  
245 -> 246 ;  
247 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]"] ;  
246 -> 247 ;  
248 [label="x[30] <= 0.855\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 1, 0]"] ;  
246 -> 248 ;  
249 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
248 -> 249 ;  
250 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
248 -> 250 ;  
251 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 2, 0, 0, 0, 0, 0, 0, 0]"] ;  
245 -> 251 ;  
252 [label="x[27] <= 0.966\ngini = 0.353\nsamples = 135\nvalue = [2, 18, 0, 0, 107, 0, 2, 2, 3, 1]"] ;  
194 -> 252 ;  
253 [label="x[13] <= 0.955\ngini = 0.107\nsamples = 108\nvalue = [2, 0, 0, 0, 102, 0, 1, 1, 1, 1]"] ;  
252 -> 253 ;  
254 [label="x[5] <= 1.086\ngini = 0.039\nsamples = 102\nvalue = [0, 0, 0, 0, 100, 0, 0, 1, 0, 1]"] ;  
253 -> 254 ;  
255 [label="gini = 0.0\nsamples = 100\nvalue = [0, 0, 0, 0, 100, 0, 0, 0, 0, 0]"] ;  
254 -> 255 ;  
256 [label="x[29] <= 1.368\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 1]"] ;  
254 -> 256 ;  
257 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
256 -> 257 ;  
258 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]"] ;  
256 -> 258 ;  
259 [label="x[28] <= -1.527\ngini = 0.722\nsamples = 6\nvalue = [2, 0, 0, 0, 2, 0, 1, 0, 1, 0]"] ;  
253 -> 259 ;  
260 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
259 -> 260 ;  
261 [label="x[51] <= 0.27\ngini = 0.625\nsamples = 4\nvalue = [0, 0, 0, 0, 2, 0, 1, 0, 1, 0]"] ;  
259 -> 261 ;  
262 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 2, 0, 0, 0, 0, 0]"] ;  
261 -> 262 ;  
263 [label="x[29] <= 0.082\ngini = 0.5\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 1, 0]"] ;  
261 -> 263 ;  
264 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
263 -> 264 ;  
265 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0]"] ;  
263 -> 265 ;  
266 [label="x[38] <= -0.554\ngini = 0.513\nsamples = 27\nvalue = [0, 18, 0, 0, 5, 0, 1, 1, 2, 0]"] ;  
252 -> 266 ;  
267 [label="x[12] <= -0.491\ngini = 0.198\nsamples = 18\nvalue = [0, 16, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
266 -> 267 ;  
268 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 0, 0, 0, 0, 0, 0, 2, 0]"] ;  
267 -> 268 ;  
269 [label="gini = 0.0\nsamples = 16\nvalue = [0, 16, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
267 -> 269 ;  
270 [label="x[21] <= 0.59\ngini = 0.617\nsamples = 9\nvalue = [0, 2, 0, 0, 5, 0, 1, 1, 0, 0]"] ;  
266 -> 270 ;  
271 [label="x[4] <= -1.261\ngini = 0.278\nsamples = 6\nvalue = [0, 0, 0, 0, 5, 0, 1, 0, 0, 0]"] ;  
270 -> 271 ;  
272 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 1, 0, 0, 0]"] ;  
271 -> 272 ;  
273 [label="gini = 0.0\nsamples = 5\nvalue = [0, 0, 0, 0, 5, 0, 0, 0, 0, 0]"] ;  
271 -> 273 ;  
274 [label="x[37] <= 0.636\ngini = 0.444\nsamples = 3\nvalue = [0, 2, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
270 -> 274 ;  
275 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0]"] ;  
274 -> 275 ;  
276 [label="gini = 0.0\nsamples = 2\nvalue = [0, 2, 0, 0, 0, 0, 0, 0, 0, 0]"] ;  
274 -> 276 ;  
}

**3.[附加题].**利用**决策树**解决**自己遇到的问题**。

[解题要求同题2]

*import* pandas *as* pd

*import* matplotlib.pyplot *as* plt

*from* sklearn.ensemble *import* RandomForestClassifier

*from* sklearn.feature\_extraction *import* DictVectorizer

*from* sklearn.model\_selection *import* train\_test\_split, GridSearchCV, cross\_val\_score

*from* sklearn.tree *import* DecisionTreeClassifier

def decision\_tree():

*# 加载数据*

    train = pd.read\_csv('train.csv')

*# pd.set\_option('display.max\_columns', 10)  # 为最大列设置为10列*

    print(train.head(5))

    print(train.info())

*# 处理数据，找出特征值和目标值*

*# 特征选择*

    features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']

    X\_data = train[['Pclass', 'Age', 'Sex', 'Embarked']]

    X\_data = train[features]

    y\_label = train['Survived']

    print(len(y\_label))

*# 缺失值处理*

    X\_data['Age'].fillna(X\_data['Age'].mean(), *inplace*=True)

    print(X\_data.info())

    labels = X\_data["Embarked"].tolist()

    X\_data['Embarked'] = X\_data['Embarked'].apply(lambda *x*: labels.index(*x*))

    X\_data['Embarked'].fillna(X\_data['Embarked'].mean, *inplace*=True)

    X\_data["Sex"] = (X\_data["Sex"] == "male").astype("int")  *# 先获取bool值，再将bool值转换成0和1*

*# 将数据集分割为训练集和测试集*

    X\_train, X\_test, y\_train, y\_test = train\_test\_split(X\_data, y\_label, *test\_size*=0.25)

    dict = DictVectorizer(*sparse*=False)

    X\_train = dict.fit\_transform(X\_train.to\_dict(*orient*="records"))

    print(X\_train)

    X\_test = dict.transform(X\_test.to\_dict(*orient*="records"))

*# 用决策树进行预测*

    decision\_model = DecisionTreeClassifier()

    decision\_model.fit(X\_train, y\_train)

*# 预测准确率*

    print(decision\_model.predict(X\_test))

    print("预测的准确率：", decision\_model.score(X\_test, y\_test))

    rf = RandomForestClassifier(*n\_jobs*=-1)

    param = {"n\_estimators": [100, 200, 300, 400, 700, 1000], "max\_depth": [6, 9, 18, 26, 35]}

*# 网络搜索与交叉验证*

    grid\_cv = GridSearchCV(rf, *param\_grid*=param, *cv*=2)

    grid\_cv.fit(X\_train, y\_train)

    print("准确率：", grid\_cv.score(X\_test, y\_test))

*return* X\_train, y\_train

def show\_curve(*X*, *y*):

    score\_train = []

    score\_test = []

*for* i *in* range(10):

        dt = DecisionTreeClassifier(*random\_state*=25, *max\_depth*=i + 1, *criterion*="entropy")

        dt = dt.fit(*X*, *y*)  *# 训练过程*

        s\_train = dt.score(*X*, *y*)  *# 训练集得分*

        s\_test = cross\_val\_score(dt, *X*, *y*, *cv*=10).mean()  *# 测试集得分*

        score\_train.append(s\_train)

        score\_test.append(s\_test)

    plt.plot(range(1, 11), score\_train, *color*="red", *label*="train")

    plt.plot(range(1, 11), score\_test, *color*="blue", *label*="test")

    plt.xticks(range(1, 11))

    plt.legend()

    plt.show()

*if* \_\_name\_\_ == '\_\_main\_\_':

    X\_train, y\_label = decision\_tree()

    show\_curve(X\_train, y\_label)

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pclass 891 non-null int64

1 Sex 891 non-null object

2 Age 891 non-null float64

3 SibSp 891 non-null int64

4 Parch 891 non-null int64

5 Fare 891 non-null float64

6 Embarked 889 non-null object

