Practical 5 Train and fine-tune a Decision Tree for the Moons Dataset

May 7, 2022

Practical 5: Train and fine-tune a Decision Tree for the Moons Dataset

[2]:
```
# GridSearchCV to fine-tune a Decision Tree Classifier
import numpy as np
import matplotlib.pyplot as plt
```

[3]:
```
#Visualization
# This function will help in visualization of our dataset.
def plot_dataset(X, y, axes):
    plt.figure(figsize=(10,6))
    plt.plot(X[:, 0][y==0], X[:, 1][y==0], "bs",alpha = 0.5)
    plt.plot(X[:, 0][y==1], X[:, 1][y==1], "g^",alpha = 0.2)
    plt.axis(axes)
    plt.grid(True, which='both')
    plt.xlabel(r"$x_1$", fontsize=20)
    plt.ylabel(r"$x_2$", fontsize=20, rotation=0)
```

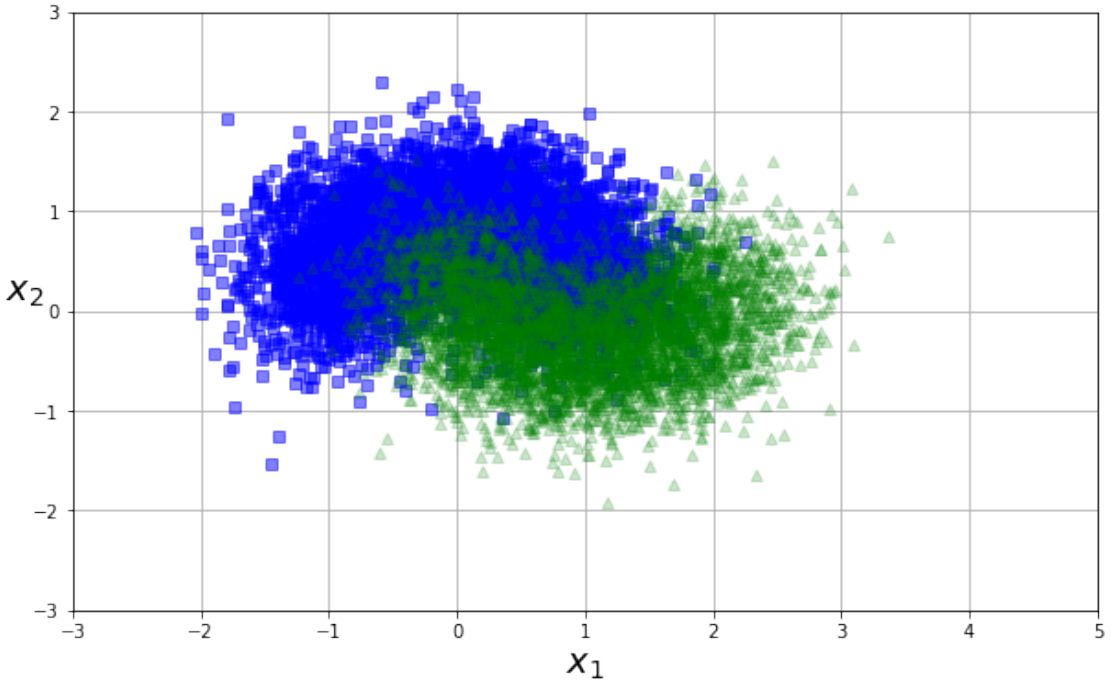
dataset contains two features x1 and x2 , it has ten thousand instances with a Guassian noise of 0.4. it's a binary classification the labels are either 0 or 1.

[4]:
```
from sklearn.datasets import make_moons

X, y = make_moons(n_samples=10000, noise=0.4, random_state=21)
```

Since we have two features it will be easy to visualize the data:

[5]:
```
plot_dataset(X, y, [-3, 5, -3, 3])
```



Splitting the dataset into training and testing set:

[10]:
```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.2)
```

Selecting the model

[11]:
```
from sklearn.tree import DecisionTreeClassifier

tree_clf = DecisionTreeClassifier()
```

We have to select the parameters for our GridSearch:

[12]:
```
from sklearn.model_selection import GridSearchCV

parameter = {
             'criterion' : ["gini", "entropy"],
             'max_leaf_nodes': list(range(2, 50)),
             'min_samples_split': [2, 3, 4]
            }
```

Using a cross validation of 5 and measuring accuracy.

[13]:
```
clf = GridSearchCV(tree_clf, parameter, cv = 5,scoring =␣
↪"accuracy",return_train_score=True,n_jobs=-1)
```

[14]:
```
clf.fit(X_train, y_train)
```

[14]:
```
GridSearchCV(cv=5, estimator=DecisionTreeClassifier(), n_jobs=-1,
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_leaf_nodes': [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
                                            13, 14, 15, 16, 17, 18, 19, 20, 21,
                                            22, 23, 24, 25, 26, 27, 28, 29, 30,
                                            31, …],
                         'min_samples_split': [2, 3, 4]},
             return_train_score=True, scoring='accuracy')
```

[15]:
```
# Getting the best parameter:
clf.best_params_
```

[15]:
```
{'criterion': 'gini', 'max_leaf_nodes': 23, 'min_samples_split': 2}
```

[16]:
```
#look at the training results:
cvres = clf.cv_results_
for mean_score, params in zip(cvres["mean_train_score"], cvres["params"]):
    print(mean_score, params)
```

```
0.7794062500000001 {'criterion': 'gini', 'max_leaf_nodes': 2,
'min_samples_split': 2}
0.7794062500000001 {'criterion': 'gini', 'max_leaf_nodes': 2,
'min_samples_split': 3}
0.7794062500000001 {'criterion': 'gini', 'max_leaf_nodes': 2,
'min_samples_split': 4}
0.82125 {'criterion': 'gini', 'max_leaf_nodes': 3, 'min_samples_split': 2}
0.82125 {'criterion': 'gini', 'max_leaf_nodes': 3, 'min_samples_split': 3}
0.82125 {'criterion': 'gini', 'max_leaf_nodes': 3, 'min_samples_split': 4}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 4, 'min_samples_split': 2}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 4, 'min_samples_split': 3}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 4, 'min_samples_split': 4}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 5, 'min_samples_split': 2}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 5, 'min_samples_split': 3}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 5, 'min_samples_split': 4}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 6, 'min_samples_split': 2}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 6, 'min_samples_split': 3}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 6, 'min_samples_split': 4}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 7, 'min_samples_split': 2}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 7, 'min_samples_split': 3}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 7, 'min_samples_split': 4}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 8, 'min_samples_split': 2}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 8, 'min_samples_split': 3}
0.8601875 {'criterion': 'gini', 'max_leaf_nodes': 8, 'min_samples_split': 4}
0.8604687500000001 {'criterion': 'gini', 'max_leaf_nodes': 9,
'min_samples_split': 2}
0.8604687500000001 {'criterion': 'gini', 'max_leaf_nodes': 9,
```

```
'min_samples_split': 3}
0.8604687500000001 {'criterion': 'gini', 'max_leaf_nodes': 9,
'min_samples_split': 4}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 10, 'min_samples_split': 2}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 10, 'min_samples_split': 3}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 10, 'min_samples_split': 4}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 11, 'min_samples_split': 2}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 11, 'min_samples_split': 3}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 11, 'min_samples_split': 4}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 12, 'min_samples_split': 2}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 12, 'min_samples_split': 3}
0.8605 {'criterion': 'gini', 'max_leaf_nodes': 12, 'min_samples_split': 4}
0.86071875 {'criterion': 'gini', 'max_leaf_nodes': 13, 'min_samples_split': 2}
0.86071875 {'criterion': 'gini', 'max_leaf_nodes': 13, 'min_samples_split': 3}
0.86071875 {'criterion': 'gini', 'max_leaf_nodes': 13, 'min_samples_split': 4}
0.8617500000000001 {'criterion': 'gini', 'max_leaf_nodes': 14,
'min_samples_split': 2}
0.8617500000000001 {'criterion': 'gini', 'max_leaf_nodes': 14,
'min_samples_split': 3}
0.8617500000000001 {'criterion': 'gini', 'max_leaf_nodes': 14,
'min_samples_split': 4}
0.86209375 {'criterion': 'gini', 'max_leaf_nodes': 15, 'min_samples_split': 2}
0.86209375 {'criterion': 'gini', 'max_leaf_nodes': 15, 'min_samples_split': 3}
0.86209375 {'criterion': 'gini', 'max_leaf_nodes': 15, 'min_samples_split': 4}
0.86284375 {'criterion': 'gini', 'max_leaf_nodes': 16, 'min_samples_split': 2}
0.86284375 {'criterion': 'gini', 'max_leaf_nodes': 16, 'min_samples_split': 3}
0.86284375 {'criterion': 'gini', 'max_leaf_nodes': 16, 'min_samples_split': 4}
0.8638125000000001 {'criterion': 'gini', 'max_leaf_nodes': 17,
'min_samples_split': 2}
0.8638125000000001 {'criterion': 'gini', 'max_leaf_nodes': 17,
'min_samples_split': 3}
0.8638125000000001 {'criterion': 'gini', 'max_leaf_nodes': 17,
'min_samples_split': 4}
0.8644999999999999 {'criterion': 'gini', 'max_leaf_nodes': 18,
'min_samples_split': 2}
0.8644999999999999 {'criterion': 'gini', 'max_leaf_nodes': 18,
'min_samples_split': 3}
0.8644999999999999 {'criterion': 'gini', 'max_leaf_nodes': 18,
'min_samples_split': 4}
0.8651249999999999 {'criterion': 'gini', 'max_leaf_nodes': 19,
'min_samples_split': 2}
0.8651249999999999 {'criterion': 'gini', 'max_leaf_nodes': 19,
'min_samples_split': 3}
0.8651249999999999 {'criterion': 'gini', 'max_leaf_nodes': 19,
'min_samples_split': 4}
0.86565625 {'criterion': 'gini', 'max_leaf_nodes': 20, 'min_samples_split': 2}
0.86565625 {'criterion': 'gini', 'max_leaf_nodes': 20, 'min_samples_split': 3}
0.86565625 {'criterion': 'gini', 'max_leaf_nodes': 20, 'min_samples_split': 4}
```

```
0.8674375 {'criterion': 'gini', 'max_leaf_nodes': 21, 'min_samples_split': 2}
0.8674375 {'criterion': 'gini', 'max_leaf_nodes': 21, 'min_samples_split': 3}
0.8674375 {'criterion': 'gini', 'max_leaf_nodes': 21, 'min_samples_split': 4}
0.8676875000000001 {'criterion': 'gini', 'max_leaf_nodes': 22,
'min_samples_split': 2}
0.8676875000000001 {'criterion': 'gini', 'max_leaf_nodes': 22,
'min_samples_split': 3}
0.8676875000000001 {'criterion': 'gini', 'max_leaf_nodes': 22,
'min_samples_split': 4}
0.8685937499999999 {'criterion': 'gini', 'max_leaf_nodes': 23,
'min_samples_split': 2}
0.8685937499999999 {'criterion': 'gini', 'max_leaf_nodes': 23,
'min_samples_split': 3}
0.8685937499999999 {'criterion': 'gini', 'max_leaf_nodes': 23,
'min_samples_split': 4}
0.8690312500000001 {'criterion': 'gini', 'max_leaf_nodes': 24,
'min_samples_split': 2}
0.8690312500000001 {'criterion': 'gini', 'max_leaf_nodes': 24,
'min_samples_split': 3}
0.8690312500000001 {'criterion': 'gini', 'max_leaf_nodes': 24,
'min_samples_split': 4}
0.8695 {'criterion': 'gini', 'max_leaf_nodes': 25, 'min_samples_split': 2}
0.8695 {'criterion': 'gini', 'max_leaf_nodes': 25, 'min_samples_split': 3}
0.8695 {'criterion': 'gini', 'max_leaf_nodes': 25, 'min_samples_split': 4}
0.8696875000000001 {'criterion': 'gini', 'max_leaf_nodes': 26,
'min_samples_split': 2}
0.8696875000000001 {'criterion': 'gini', 'max_leaf_nodes': 26,
'min_samples_split': 3}
0.8696875000000001 {'criterion': 'gini', 'max_leaf_nodes': 26,
'min_samples_split': 4}
0.8703749999999999 {'criterion': 'gini', 'max_leaf_nodes': 27,
'min_samples_split': 2}
0.8703749999999999 {'criterion': 'gini', 'max_leaf_nodes': 27,
'min_samples_split': 3}
0.8703749999999999 {'criterion': 'gini', 'max_leaf_nodes': 27,
'min_samples_split': 4}
0.87084375 {'criterion': 'gini', 'max_leaf_nodes': 28, 'min_samples_split': 2}
0.87084375 {'criterion': 'gini', 'max_leaf_nodes': 28, 'min_samples_split': 3}
0.87084375 {'criterion': 'gini', 'max_leaf_nodes': 28, 'min_samples_split': 4}
0.8712187499999999 {'criterion': 'gini', 'max_leaf_nodes': 29,
'min_samples_split': 2}
0.8712187499999999 {'criterion': 'gini', 'max_leaf_nodes': 29,
'min_samples_split': 3}
0.8712187499999999 {'criterion': 'gini', 'max_leaf_nodes': 29,
'min_samples_split': 4}
0.8714999999999999 {'criterion': 'gini', 'max_leaf_nodes': 30,
'min_samples_split': 2}
0.8714999999999999 {'criterion': 'gini', 'max_leaf_nodes': 30,
```

```
'min_samples_split': 3}
0.8714999999999999 {'criterion': 'gini', 'max_leaf_nodes': 30,
'min_samples_split': 4}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 31,
'min_samples_split': 2}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 31,
'min_samples_split': 3}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 31,
'min_samples_split': 4}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 32,
'min_samples_split': 2}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 32,
'min_samples_split': 3}
0.8717500000000001 {'criterion': 'gini', 'max_leaf_nodes': 32,
'min_samples_split': 4}
0.87253125 {'criterion': 'gini', 'max_leaf_nodes': 33, 'min_samples_split': 2}
0.87253125 {'criterion': 'gini', 'max_leaf_nodes': 33, 'min_samples_split': 3}
0.87253125 {'criterion': 'gini', 'max_leaf_nodes': 33, 'min_samples_split': 4}
0.8729375000000001 {'criterion': 'gini', 'max_leaf_nodes': 34,
'min_samples_split': 2}
0.8729375000000001 {'criterion': 'gini', 'max_leaf_nodes': 34,
'min_samples_split': 3}
0.8729375000000001 {'criterion': 'gini', 'max_leaf_nodes': 34,
'min_samples_split': 4}
0.8733124999999999 {'criterion': 'gini', 'max_leaf_nodes': 35,
'min_samples_split': 2}
0.8733124999999999 {'criterion': 'gini', 'max_leaf_nodes': 35,
'min_samples_split': 3}
0.8733124999999999 {'criterion': 'gini', 'max_leaf_nodes': 35,
'min_samples_split': 4}
0.8735312499999999 {'criterion': 'gini', 'max_leaf_nodes': 36,
'min_samples_split': 2}
0.8735312499999999 {'criterion': 'gini', 'max_leaf_nodes': 36,
'min_samples_split': 3}
0.8735312499999999 {'criterion': 'gini', 'max_leaf_nodes': 36,
'min_samples_split': 4}
0.8739062499999999 {'criterion': 'gini', 'max_leaf_nodes': 37,
'min_samples_split': 2}
0.8739062499999999 {'criterion': 'gini', 'max_leaf_nodes': 37,
'min_samples_split': 3}
0.8739062499999999 {'criterion': 'gini', 'max_leaf_nodes': 37,
'min_samples_split': 4}
0.8739999999999999 {'criterion': 'gini', 'max_leaf_nodes': 38,
'min_samples_split': 2}
0.8739999999999999 {'criterion': 'gini', 'max_leaf_nodes': 38,
'min_samples_split': 3}
0.8739999999999999 {'criterion': 'gini', 'max_leaf_nodes': 38,
'min_samples_split': 4}
```

```
0.874625 {'criterion': 'gini', 'max_leaf_nodes': 39, 'min_samples_split': 2}
0.874625 {'criterion': 'gini', 'max_leaf_nodes': 39, 'min_samples_split': 3}
0.874625 {'criterion': 'gini', 'max_leaf_nodes': 39, 'min_samples_split': 4}
0.8751875 {'criterion': 'gini', 'max_leaf_nodes': 40, 'min_samples_split': 2}
0.8751875 {'criterion': 'gini', 'max_leaf_nodes': 40, 'min_samples_split': 3}
0.8751875 {'criterion': 'gini', 'max_leaf_nodes': 40, 'min_samples_split': 4}
0.8759375 {'criterion': 'gini', 'max_leaf_nodes': 41, 'min_samples_split': 2}
0.8759375 {'criterion': 'gini', 'max_leaf_nodes': 41, 'min_samples_split': 3}
0.8759375 {'criterion': 'gini', 'max_leaf_nodes': 41, 'min_samples_split': 4}
0.8762187499999999 {'criterion': 'gini', 'max_leaf_nodes': 42,
'min_samples_split': 2}
0.8762187499999999 {'criterion': 'gini', 'max_leaf_nodes': 42,
'min_samples_split': 3}
0.8762187499999999 {'criterion': 'gini', 'max_leaf_nodes': 42,
'min_samples_split': 4}
0.8765000000000001 {'criterion': 'gini', 'max_leaf_nodes': 43,
'min_samples_split': 2}
0.8765000000000001 {'criterion': 'gini', 'max_leaf_nodes': 43,
'min_samples_split': 3}
0.8765000000000001 {'criterion': 'gini', 'max_leaf_nodes': 43,
'min_samples_split': 4}
0.8765625 {'criterion': 'gini', 'max_leaf_nodes': 44, 'min_samples_split': 2}
0.8765625 {'criterion': 'gini', 'max_leaf_nodes': 44, 'min_samples_split': 3}
0.8765625 {'criterion': 'gini', 'max_leaf_nodes': 44, 'min_samples_split': 4}
0.876625 {'criterion': 'gini', 'max_leaf_nodes': 45, 'min_samples_split': 2}
0.876625 {'criterion': 'gini', 'max_leaf_nodes': 45, 'min_samples_split': 3}
0.876625 {'criterion': 'gini', 'max_leaf_nodes': 45, 'min_samples_split': 4}
0.8768125000000001 {'criterion': 'gini', 'max_leaf_nodes': 46,
'min_samples_split': 2}
0.8768125000000001 {'criterion': 'gini', 'max_leaf_nodes': 46,
'min_samples_split': 3}
0.8768125000000001 {'criterion': 'gini', 'max_leaf_nodes': 46,
'min_samples_split': 4}
0.87715625 {'criterion': 'gini', 'max_leaf_nodes': 47, 'min_samples_split': 2}
0.87715625 {'criterion': 'gini', 'max_leaf_nodes': 47, 'min_samples_split': 3}
0.87715625 {'criterion': 'gini', 'max_leaf_nodes': 47, 'min_samples_split': 4}
0.87728125 {'criterion': 'gini', 'max_leaf_nodes': 48, 'min_samples_split': 2}
0.87728125 {'criterion': 'gini', 'max_leaf_nodes': 48, 'min_samples_split': 3}
0.87728125 {'criterion': 'gini', 'max_leaf_nodes': 48, 'min_samples_split': 4}
0.8775625 {'criterion': 'gini', 'max_leaf_nodes': 49, 'min_samples_split': 2}
0.8775625 {'criterion': 'gini', 'max_leaf_nodes': 49, 'min_samples_split': 3}
0.8775625 {'criterion': 'gini', 'max_leaf_nodes': 49, 'min_samples_split': 4}
0.7794062500000001 {'criterion': 'entropy', 'max_leaf_nodes': 2,
'min_samples_split': 2}
0.7794062500000001 {'criterion': 'entropy', 'max_leaf_nodes': 2,
'min_samples_split': 3}
0.7794062500000001 {'criterion': 'entropy', 'max_leaf_nodes': 2,
'min_samples_split': 4}
```

```
0.821125 {'criterion': 'entropy', 'max_leaf_nodes': 3, 'min_samples_split': 2}
0.821125 {'criterion': 'entropy', 'max_leaf_nodes': 3, 'min_samples_split': 3}
0.821125 {'criterion': 'entropy', 'max_leaf_nodes': 3, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 4, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 4, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 4, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 5, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 5, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 5, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 6, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 6, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 6, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 7, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 7, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 7, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 8, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 8, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 8, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 9, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 9, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 9, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 10, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 10, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 10, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 11, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 11, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 11, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 12, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 12, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 12, 'min_samples_split': 4}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 13, 'min_samples_split': 2}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 13, 'min_samples_split': 3}
0.8600625 {'criterion': 'entropy', 'max_leaf_nodes': 13, 'min_samples_split': 4}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 14, 'min_samples_split':
2}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 14, 'min_samples_split':
3}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 14, 'min_samples_split':
4}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 15, 'min_samples_split':
2}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 15, 'min_samples_split':
3}
0.86034375 {'criterion': 'entropy', 'max_leaf_nodes': 15, 'min_samples_split':
4}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 16, 'min_samples_split': 2}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 16, 'min_samples_split': 3}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 16, 'min_samples_split': 4}
```

```
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 17, 'min_samples_split': 2}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 17, 'min_samples_split': 3}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 17, 'min_samples_split': 4}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 18, 'min_samples_split': 2}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 18, 'min_samples_split': 3}
0.861 {'criterion': 'entropy', 'max_leaf_nodes': 18, 'min_samples_split': 4}
0.8615625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 19,
'min_samples_split': 2}
0.8615625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 19,
'min_samples_split': 3}
0.8615625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 19,
'min_samples_split': 4}
0.86403125 {'criterion': 'entropy', 'max_leaf_nodes': 20, 'min_samples_split':
2}
0.86403125 {'criterion': 'entropy', 'max_leaf_nodes': 20, 'min_samples_split':
3}
0.86403125 {'criterion': 'entropy', 'max_leaf_nodes': 20, 'min_samples_split':
4}
0.86428125 {'criterion': 'entropy', 'max_leaf_nodes': 21, 'min_samples_split':
2}
0.86428125 {'criterion': 'entropy', 'max_leaf_nodes': 21, 'min_samples_split':
3}
0.86428125 {'criterion': 'entropy', 'max_leaf_nodes': 21, 'min_samples_split':
4}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 22, 'min_samples_split':
2}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 22, 'min_samples_split':
3}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 22, 'min_samples_split':
4}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 23, 'min_samples_split':
2}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 23, 'min_samples_split':
3}
0.86459375 {'criterion': 'entropy', 'max_leaf_nodes': 23, 'min_samples_split':
4}
0.8650625 {'criterion': 'entropy', 'max_leaf_nodes': 24, 'min_samples_split': 2}
0.8650625 {'criterion': 'entropy', 'max_leaf_nodes': 24, 'min_samples_split': 3}
0.8650625 {'criterion': 'entropy', 'max_leaf_nodes': 24, 'min_samples_split': 4}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 25,
'min_samples_split': 2}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 25,
'min_samples_split': 3}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 25,
'min_samples_split': 4}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 26,
'min_samples_split': 2}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 26,
```

```
'min_samples_split': 3}
0.8662812499999999 {'criterion': 'entropy', 'max_leaf_nodes': 26,
'min_samples_split': 4}
0.8665312500000001 {'criterion': 'entropy', 'max_leaf_nodes': 27,
'min_samples_split': 2}
0.8665312500000001 {'criterion': 'entropy', 'max_leaf_nodes': 27,
'min_samples_split': 3}
0.8665312500000001 {'criterion': 'entropy', 'max_leaf_nodes': 27,
'min_samples_split': 4}
0.8673125 {'criterion': 'entropy', 'max_leaf_nodes': 28, 'min_samples_split': 2}
0.8673125 {'criterion': 'entropy', 'max_leaf_nodes': 28, 'min_samples_split': 3}
0.8673125 {'criterion': 'entropy', 'max_leaf_nodes': 28, 'min_samples_split': 4}
0.867375 {'criterion': 'entropy', 'max_leaf_nodes': 29, 'min_samples_split': 2}
0.867375 {'criterion': 'entropy', 'max_leaf_nodes': 29, 'min_samples_split': 3}
0.867375 {'criterion': 'entropy', 'max_leaf_nodes': 29, 'min_samples_split': 4}
0.8675937499999999 {'criterion': 'entropy', 'max_leaf_nodes': 30,
'min_samples_split': 2}
0.8675937499999999 {'criterion': 'entropy', 'max_leaf_nodes': 30,
'min_samples_split': 3}
0.8675937499999999 {'criterion': 'entropy', 'max_leaf_nodes': 30,
'min_samples_split': 4}
0.8680312499999999 {'criterion': 'entropy', 'max_leaf_nodes': 31,
'min_samples_split': 2}
0.8680312499999999 {'criterion': 'entropy', 'max_leaf_nodes': 31,
'min_samples_split': 3}
0.8680312499999999 {'criterion': 'entropy', 'max_leaf_nodes': 31,
'min_samples_split': 4}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 32,
'min_samples_split': 2}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 32,
'min_samples_split': 3}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 32,
'min_samples_split': 4}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 33,
'min_samples_split': 2}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 33,
'min_samples_split': 3}
0.8682500000000001 {'criterion': 'entropy', 'max_leaf_nodes': 33,
'min_samples_split': 4}
0.86834375 {'criterion': 'entropy', 'max_leaf_nodes': 34, 'min_samples_split':
2}
0.86834375 {'criterion': 'entropy', 'max_leaf_nodes': 34, 'min_samples_split':
3}
0.86834375 {'criterion': 'entropy', 'max_leaf_nodes': 34, 'min_samples_split':
4}
0.8684375000000001 {'criterion': 'entropy', 'max_leaf_nodes': 35,
'min_samples_split': 2}
0.8684375000000001 {'criterion': 'entropy', 'max_leaf_nodes': 35,
```

```
'min_samples_split': 3}
0.8684375000000001 {'criterion': 'entropy', 'max_leaf_nodes': 35,
'min_samples_split': 4}
0.8684687500000001 {'criterion': 'entropy', 'max_leaf_nodes': 36,
'min_samples_split': 2}
0.8684687500000001 {'criterion': 'entropy', 'max_leaf_nodes': 36,
'min_samples_split': 3}
0.8684687500000001 {'criterion': 'entropy', 'max_leaf_nodes': 36,
'min_samples_split': 4}
0.8686875000000001 {'criterion': 'entropy', 'max_leaf_nodes': 37,
'min_samples_split': 2}
0.8686875000000001 {'criterion': 'entropy', 'max_leaf_nodes': 37,
'min_samples_split': 3}
0.8686875000000001 {'criterion': 'entropy', 'max_leaf_nodes': 37,
'min_samples_split': 4}
0.86878125 {'criterion': 'entropy', 'max_leaf_nodes': 38, 'min_samples_split':
2}
0.86878125 {'criterion': 'entropy', 'max_leaf_nodes': 38, 'min_samples_split':
3}
0.86878125 {'criterion': 'entropy', 'max_leaf_nodes': 38, 'min_samples_split':
4}
0.86890625 {'criterion': 'entropy', 'max_leaf_nodes': 39, 'min_samples_split':
2}
0.86890625 {'criterion': 'entropy', 'max_leaf_nodes': 39, 'min_samples_split':
3}
0.86890625 {'criterion': 'entropy', 'max_leaf_nodes': 39, 'min_samples_split':
4}
0.8689375 {'criterion': 'entropy', 'max_leaf_nodes': 40, 'min_samples_split': 2}
0.8689375 {'criterion': 'entropy', 'max_leaf_nodes': 40, 'min_samples_split': 3}
0.8689375 {'criterion': 'entropy', 'max_leaf_nodes': 40, 'min_samples_split': 4}
0.869 {'criterion': 'entropy', 'max_leaf_nodes': 41, 'min_samples_split': 2}
0.869 {'criterion': 'entropy', 'max_leaf_nodes': 41, 'min_samples_split': 3}
0.869 {'criterion': 'entropy', 'max_leaf_nodes': 41, 'min_samples_split': 4}
0.8690624999999998 {'criterion': 'entropy', 'max_leaf_nodes': 42,
'min_samples_split': 2}
0.8690624999999998 {'criterion': 'entropy', 'max_leaf_nodes': 42,
'min_samples_split': 3}
0.8690624999999998 {'criterion': 'entropy', 'max_leaf_nodes': 42,
'min_samples_split': 4}
0.8693125 {'criterion': 'entropy', 'max_leaf_nodes': 43, 'min_samples_split': 2}
0.8693125 {'criterion': 'entropy', 'max_leaf_nodes': 43, 'min_samples_split': 3}
0.8693125 {'criterion': 'entropy', 'max_leaf_nodes': 43, 'min_samples_split': 4}
0.869375 {'criterion': 'entropy', 'max_leaf_nodes': 44, 'min_samples_split': 2}
0.869375 {'criterion': 'entropy', 'max_leaf_nodes': 44, 'min_samples_split': 3}
0.869375 {'criterion': 'entropy', 'max_leaf_nodes': 44, 'min_samples_split': 4}
0.8694687499999999 {'criterion': 'entropy', 'max_leaf_nodes': 45,
'min_samples_split': 2}
0.8694687499999999 {'criterion': 'entropy', 'max_leaf_nodes': 45,
```

```
'min_samples_split': 3}
0.8694687499999999 {'criterion': 'entropy', 'max_leaf_nodes': 45,
'min_samples_split': 4}
0.8694999999999998 {'criterion': 'entropy', 'max_leaf_nodes': 46,
'min_samples_split': 2}
0.8694999999999998 {'criterion': 'entropy', 'max_leaf_nodes': 46,
'min_samples_split': 3}
0.8694999999999998 {'criterion': 'entropy', 'max_leaf_nodes': 46,
'min_samples_split': 4}
0.8696562499999999 {'criterion': 'entropy', 'max_leaf_nodes': 47,
'min_samples_split': 2}
0.8696562499999999 {'criterion': 'entropy', 'max_leaf_nodes': 47,
'min_samples_split': 3}
0.8696562499999999 {'criterion': 'entropy', 'max_leaf_nodes': 47,
'min_samples_split': 4}
0.86990625 {'criterion': 'entropy', 'max_leaf_nodes': 48, 'min_samples_split':
2}
0.86990625 {'criterion': 'entropy', 'max_leaf_nodes': 48, 'min_samples_split':
3}
0.86990625 {'criterion': 'entropy', 'max_leaf_nodes': 48, 'min_samples_split':
4}
0.8700625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 49,
'min_samples_split': 2}
0.8700625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 49,
'min_samples_split': 3}
0.8700625000000001 {'criterion': 'entropy', 'max_leaf_nodes': 49,
'min_samples_split': 4}
```

[17]:
```
#Getting the training score
clf.score(X_train, y_train)
```

[17]:
```
0.869625
```

We have an accuracy of approximately 87% but accuracy is sometimes not a good measure to use, lets see the confusion matrix.

[18]:
```
from sklearn.metrics import confusion_matrix

pred = clf.predict(X_train)
confusion_matrix(y_train,pred)
```

[18]:
```
array([[3572,  460],
       [ 583, 3385]], dtype=int64)
```

[19]:
```
#from the confusion matrix let's get our precision and recall, which are better␣
 ↪metrics.
from sklearn.metrics import precision_score, recall_score
```

```
pre = precision_score(y_train, pred)
re  = recall_score(y_train, pred)
print(f"Precision: {pre}  Recall:{re}")
```

```
Precision: 0.8803641092327699  Recall:0.8530745967741935
```

[20]:
```
#we have a higher precision than recall but lets combine the two metrics into␣
↪F1  score.
from sklearn.metrics import f1_score

f1_score(y_train, pred)
```

[20]:
```
0.8665045437092026
```

Our F1 Score and accuracy are almost the same.

[21]:
```
#Getting the testing score
clf.score(X_test, y_test)
```

[21]:
```
0.865
```

We have an accuracy of approximately 86% on the testing set.

[ ]:
```

```