In [4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_classification
X,y=make_classification(n_samples=1000,n_features=5,n_clusters_per_class=1,n_classes=2,random_state=2529)
X[0:5]
```

Out[4]:
```
array([[ 1.54701705,  0.84770596, -0.41725021, -0.62356778, -0.19388577],
       [ 0.80633556,  0.40985594, -0.45641095, -0.3052022 ,  0.50935923],
       [ 0.94390268,  0.70041038,  1.11385452, -0.49394417,  1.42305455],
       [ 1.92091517,  0.95815739, -1.2235022 , -0.71578154,  0.66588981],
       [ 1.45270369,  0.69035375, -1.18119669, -0.52009219, -0.22745417]])
```

In [5]:
```
y[0:5]
```

Out[5]:
```
array([0, 0, 1, 0, 0])
```

In [6]:
```
X.shape,y.shape
```

Out[6]:
```
((1000, 5), (1000,))
```

In [7]:
```
from sklearn.model_selection  import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.3, random_state=2529)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

Out[7]:
```
((700, 5), (300, 5), (700,), (300,))
```

In [9]:
```
from sklearn.tree  import DecisionTreeClassifier
model=DecisionTreeClassifier()
model.fit(X_train,y_train)
```

Out[9]:
```
DecisionTreeClassifier()
```

In [10]:
```
y_pred=model.predict(X_test)
y_pred.shape
```

Out[10]:
```
(300,)
```

In [11]:
```
y_pred
```

Out[11]:
```
array([1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1,
       0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0,
       0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0,
       0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1,
       0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0,
       1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1,
       0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1,
       1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0,
       1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1,
       0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0,
       0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1])
```

In [12]:
```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
accuracy_score(y_test,y_pred)
```

Out[12]:
```
0.9866666666666667
```

In [13]:
```
confusion_matrix(y_test,y_pred)
```

Out[13]:
```
array([[156,   1],
       [  3, 140]], dtype=int64)
```

In [16]:
```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       157
           1       0.99      0.98      0.99       143

    accuracy                           0.99       300
   macro avg       0.99      0.99      0.99       300
weighted avg       0.99      0.99      0.99       300
```

In [17]:
```
from sklearn.model_selection import GridSearchCV
parameters={'criterion':['gini','entropy'],'max_depth':[2,3,4,5,6,7,8,9,0,10,11,12,15,20,30,20,40,50,70,90,120,150]}
gridsearch=GridSearchCV(DecisionTreeClassifier(),parameters)
gridsearch.fit(X_train,y_train)
```

```
D:\anaconda\lib\site-packages\sklearn\model_selection\_validation.py:548: FitFailedWarning: Estimator fit failed. The
score on this train-test partition for these parameters will be set to nan. Details:
Traceback (most recent call last):
  File "D:\anaconda\lib\site-packages\sklearn\model_selection\_validation.py", line 531, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "D:\anaconda\lib\site-packages\sklearn\tree\_classes.py", line 890, in fit
    super().fit(
  File "D:\anaconda\lib\site-packages\sklearn\tree\_classes.py", line 277, in fit
    raise ValueError("max_depth must be greater than zero. ")
ValueError: max_depth must be greater than zero.

  warnings.warn("Estimator fit failed. The score on this train-test"
D:\anaconda\lib\site-packages\sklearn\model_selection\_validation.py:548: FitFailedWarning: Estimator fit failed. The
score on this train-test partition for these parameters will be set to nan. Details:
Traceback (most recent call last):
  File "D:\anaconda\lib\site-packages\sklearn\model_selection\_validation.py", line 531, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "D:\anaconda\lib\site-packages\sklearn\tree\_classes.py", line 890, in fit
    super().fit(
```

In [19]:
```
gridsearch.best_params_
```

Out[19]:
```
{'criterion': 'gini', 'max_depth': 3}
```

In [21]:
```
gridsearch.best_score_
```

Out[21]:
```
0.9885714285714287
```

In [22]:
```
gridsearch.best_estimator_
```

Out[22]:
```
DecisionTreeClassifier(max_depth=3)
```

In [23]:
```
gridsearch.best_index_
```

Out[23]:
```
1
```

In [24]:
```
y_pred_grid= gridsearch.predict(X_test)
```

In [29]:
```
confusion_matrix(y_test,y_pred_grid)
```

Out[29]:
```
array([[155,   2],
       [  1, 142]], dtype=int64)
```

In [30]:
```
print(classification_report(y_test,y_pred_grid))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       157
           1       0.99      0.99      0.99       143

    accuracy                           0.99       300
   macro avg       0.99      0.99      0.99       300
weighted avg       0.99      0.99      0.99       300
```

In [ ]: