Forward School

Program Code: J620-002-4:2020

Program Name: FRONT-END SOFTWARE DEVELOPMENT

Title : Exe20 - Decision Tree Exercise 2

Name: Chong Mun Chen

IC Number: 960327-07-5097

Date : 18/7/2023

Introduction : Practising on supervised machine learning with decision tree classification.

Conclusion : Found the accuracy score for both decision trees using either entropy or gini with the same training and testing sets.

Decision Tree

In [1]:
```
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier
import pandas as pd

iris = load_iris()
X = iris.data[:, 2:] # petal length and width
y = iris.target

df = pd.DataFrame(data = iris.data, columns = iris.feature_names)

df
```

Out[1]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...	...	...	...	...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

DecisionTree Modeling

In [2]:
```
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn import metrics, tree

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando
```

Build decision tree in both entropy and GINI

In [3]:
```
# entropy
clf_entropy = DecisionTreeClassifier(criterion='entropy')
clf_entropy = clf_entropy.fit(X_train,y_train)

clf_entropy.predict(X_test)
```

Out[3]:
```
array([0, 1, 1, 0, 2, 1, 2, 0, 0, 2, 1, 0, 2, 1, 1, 0, 1, 1, 0, 0, 1, 1,
       2, 0, 2, 1, 0, 0, 1, 2, 1, 2, 1, 2, 2, 0, 1, 0, 1, 2, 2, 0, 1, 2,
       1])
```

In [4]:
```
# gini
clf = DecisionTreeClassifier()
clf = clf.fit(X_train,y_train)

clf.predict(X_test)
```

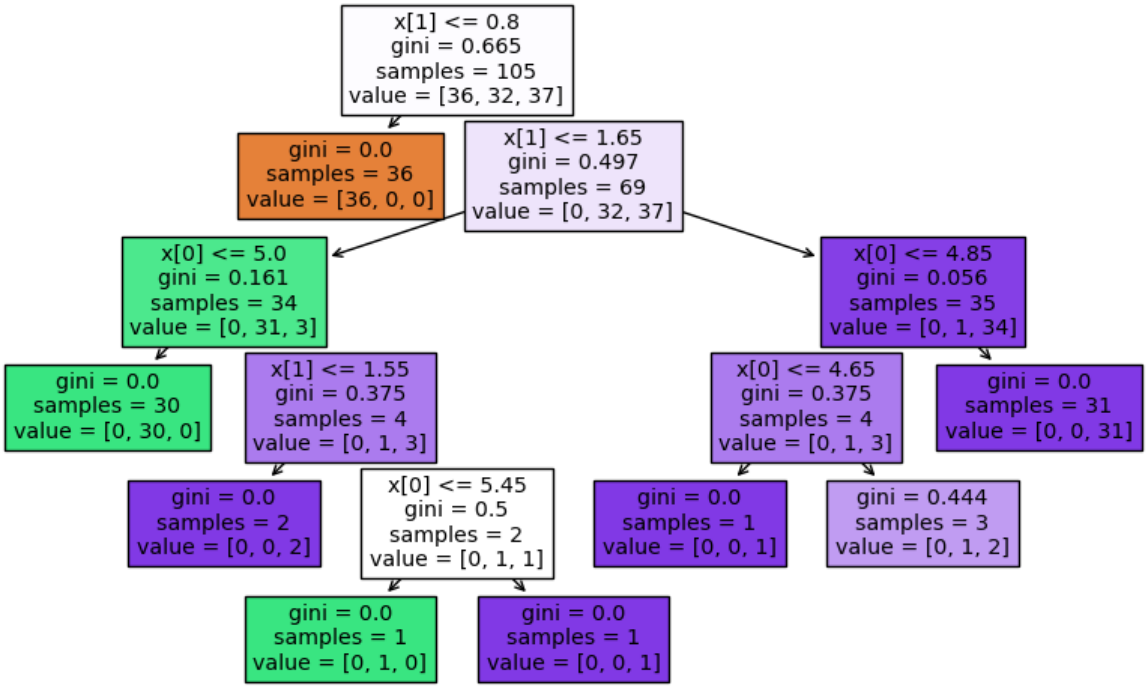
Out[4]:
```
array([0, 1, 1, 0, 2, 1, 2, 0, 0, 2, 1, 0, 2, 1, 1, 0, 1, 1, 0, 0, 1, 1,
       2, 0, 2, 1, 0, 0, 1, 2, 1, 2, 1, 2, 2, 0, 1, 0, 1, 2, 2, 0, 1, 2,
       1])
```

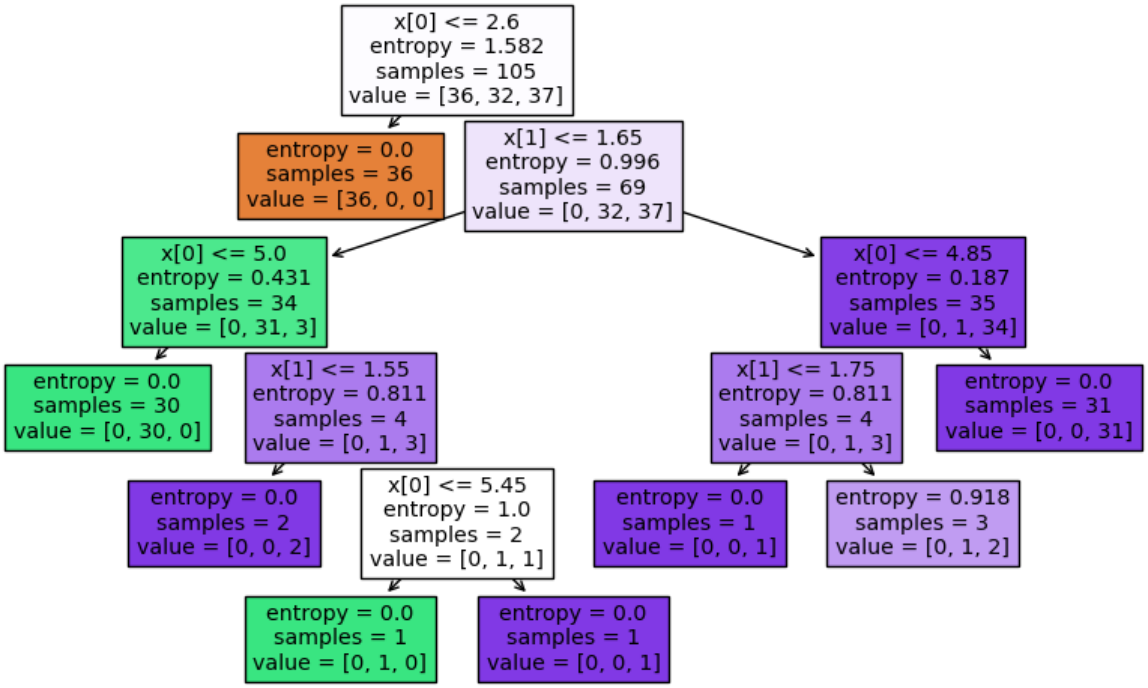
Convert to Decision Tree Diagram

In [5]:
```
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
tree.plot_tree(clf, filled=True)
plt.show()

plt.figure(figsize=(10, 6))
tree.plot_tree(clf_entropy, filled=True)
plt.show()
```




x[0] <= 2.6
entropy = 1.582
samples = 105
value = [36, 32, 37]
entropy = 0.0
samples = 36
value = [36, 0, 0]
x[1] <= 1.65
entropy = 0.996
samples = 69
value = [0, 32, 37]
x[0] <= 5.0
entropy = 0.431
samples = 34
value = [0, 31, 3]
x[0] <= 4.85
entropy = 0.187
samples = 35
value = [0, 1, 34]
entropy = 0.0
samples = 30
value = [0, 30, 0]
x[1] <= 1.55
entropy = 0.811
samples = 4
value = [0, 1, 3]
x[1] <= 1.75
entropy = 0.811
samples = 4
value = [0, 1, 3]
entropy = 0.0
samples = 31
value = [0, 0, 31]
entropy = 0.0
samples = 2
value = [0, 0, 2]
x[0] <= 5.45
entropy = 1.0
samples = 2
value = [0, 1, 1]
entropy = 0.0
samples = 1
value = [0, 0, 1]
entropy = 0.918
samples = 3
value = [0, 1, 2]
entropy = 0.0
samples = 1
value = [0, 1, 0]
entropy = 0.0
samples = 1
value = [0, 0, 1]