

DecisionTreeClassifier_Basic

April 24, 2018

```
In [1]: from sklearn import tree
import numpy as np

In [2]: # lenght, weight, speed
X = [[100, 200, 450], [500, 1000, 200], [150, 250, 150], [150, 300, 300], [200, 300, 250], [300, 450, 500],
      [700, 1200, 270], [650, 1100, 280], [130, 300, 300]]
# sport, bus, jeep assumed
Y = [['SP'], ['BS'], ['JP'], ['SP'], ['JP'], ['SP'], ['BS'], ['BS'], ['SP']]

In [3]: for i, j in zip(X, Y):
print(i, '>', j)

([100, 200, 450], '>', ['SP'])
([500, 1000, 200], '>', ['BS'])
([150, 250, 150], '>', ['JP'])
([150, 300, 300], '>', ['SP'])
([200, 300, 250], '>', ['JP'])
([300, 450, 500], '>', ['SP'])
([700, 1200, 270], '>', ['BS'])
([650, 1100, 280], '>', ['BS'])
([130, 300, 300], '>', ['SP'])

In [4]: dt = tree.DecisionTreeClassifier()
dt.fit(X, Y)

Out[4]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort=False, random_state=None,
splitter='best')

In [5]: # all three category prediction
a = np.array([150, 500, 550]).reshape(1, -1)
print(dt.predict(a))
a = np.array([700, 1300, 290]).reshape(1, -1)
print(dt.predict(a))
a = np.array([270, 300, 250]).reshape(1, -1)
print(dt.predict(a))
```

['SP']
['BS']
['JP']