In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import load_wine
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [2]:
```
wine = load_wine()
```

In [3]:
```
print(wine.DESCR)
```

```
.. _wine_dataset:

Wine recognition dataset
------------------------

**Data Set Characteristics:**

    :Number of Instances: 178 (50 in each of three classes)
    :Number of Attributes: 13 numeric, predictive attributes and the
class
    :Attribute Information:
            - Alcohol
            - Malic acid
            - Ash
            - Alcalinity of ash
            - Magnesium
            - Total phenols
            - Flavanoids
            - Nonflavanoid phenols
            - Proanthocyanins
            - Color intensity
            - Hue
            - OD280/OD315 of diluted wines
            - Proline

    - class:
            - class_0
            - class_1
            - class_2

    :Summary Statistics:

    ============================= ==== ===== ======= =====
                                   Min   Max   Mean     SD
    ============================= ==== ===== ======= =====
    Alcohol:                      11.0  14.8    13.0   0.8
    Malic Acid:                   0.74  5.80    2.34  1.12
    Ash:                          1.36  3.23    2.36  0.27
```

```
    Alcalinity of Ash:            10.6  30.0    19.5   3.3

    Magnesium:                    70.0 162.0    99.7  14.3
    Total Phenols:                0.98  3.88    2.29  0.63
    Flavanoids:                   0.34  5.08    2.03  1.00
    Nonflavanoid Phenols:         0.13  0.66    0.36  0.12
    Proanthocyanins:              0.41  3.58    1.59  0.57
    Colour Intensity:              1.3  13.0     5.1   2.3
    Hue:                          0.48  1.71    0.96  0.23
    OD280/OD315 of diluted wines: 1.27  4.00    2.61  0.71
    Proline:                       278  1680     746   315
    ============================= ==== ===== ======= =====

    :Missing Attribute Values: None
    :Class Distribution: class_0 (59), class_1 (71), class_2 (48)
    :Creator: R.A. Fisher
    :Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
    :Date: July, 1988

This is a copy of UCI ML Wine recognition datasets.
https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.d
ata (https://archive.ics.uci.edu/ml/machine-learning-databases/wine/w
ine.data)

The data is the results of a chemical analysis of wines grown in the
same
region in Italy by three different cultivators. There are thirteen di
fferent
measurements taken for different constituents found in the three type
s of
wine.

Original Owners:

Forina, M. et al, PARVUS -
An Extendible Package for Data Exploration, Classification and Correl
ation.
Institute of Pharmaceutical and Food Analysis and Technologies,
Via Brigata Salerno, 16147 Genoa, Italy.

Citation:

Lichman, M. (2013). UCI Machine Learning Repository
[https://archive.ics.uci.edu/ml]. Irvine, CA: University of Californi
a,
School of Information and Computer Science.

.. topic:: References

  (1) S. Aeberhard, D. Coomans and O. de Vel,
```

```
  Comparison of Classifiers in High Dimensional Settings,
  Tech. Rep. no. 92-02, (1992), Dept. of Computer Science and Dept. o
f
  Mathematics and Statistics, James Cook University of North Queensla
nd.
  (Also submitted to Technometrics).

  The data was used with many others for comparing various
  classifiers. The classes are separable, though only RDA
  has achieved 100% correct classification.
  (RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN 96.1% (z-transformed data))
  (All results using the leave-one-out technique)

  (2) S. Aeberhard, D. Coomans and O. de Vel,
  "THE CLASSIFICATION PERFORMANCE OF RDA"
  Tech. Rep. no. 92-01, (1992), Dept. of Computer Science and Dept. o
f
  Mathematics and Statistics, James Cook University of North Queensla
nd.
  (Also submitted to Journal of Chemometrics).
```

In [5]:
```
X = wine.data
y = wine.target
print(X.shape,y.shape)
```

```
(178, 13) (178,)
```

In [6]:
```
print(wine.feature_names)
```

```
['alcohol', 'malic_acid', 'ash', 'alcalinity_of_ash', 'magnesium', 't
otal_phenols', 'flavanoids', 'nonflavanoid_phenols', 'proanthocyanins
', 'color_intensity', 'hue', 'od280/od315_of_diluted_wines', 'proline
']
```

In [7]:
```
wine.target_names
```

Out[7]:
```
array(['class_0', 'class_1', 'class_2'], dtype='<U7')
```

### splitting the data

In [8]:
```
from sklearn.model_selection import train_test_split
```

In [9]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.
```

In [10]:
```
sum(y_train == 0)
```
Out[10]: 39

In [11]:
```
sum(y_train == 1)
```
Out[11]: 48

In [12]:
```
sum(y_train == 2)
```
Out[12]: 32

In [13]:
```
48/(39+48+32)
```
Out[13]: 0.40336134453781514

## Logistic Regression

In [15]:
```
lr = LogisticRegression(random_state=0,penalty = 'l2',solver = 'newtor
```

In [16]:
```
pred = lr.predict(X_test)
```

In [17]:
```
accuracy_score(y_test,pred)
```
Out[17]: 0.9152542372881356

In [31]:
```
sum(y_test == 0),sum(y_test == 1),sum(y_test == 2)
```
Out[31]: (20, 23, 16)

In [32]:
```
23/(20+23+16)
```
Out[32]: 0.3898305084745763

## Decision Trees

In [18]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [19]:
```
clf = DecisionTreeClassifier(random_state=0)
clf.fit(X_train,y_train)
```
Out[19]:
```
DecisionTreeClassifier(random_state=0)
```

In [20]:
```
pred = clf.predict(X_test)
```

In [21]:
```
accuracy_score(y_test,pred)
```
Out[21]:
```
0.9491525423728814
```

## Random Forest

In [26]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [30]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.
```

In [37]:
```
clf = RandomForestClassifier()
clf.fit(X_train,y_train)
```
Out[37]:
```
RandomForestClassifier()
```

In [38]:
```
pred = clf.predict(X_test)
```

In [39]:
```
accuracy_score(y_test,pred)
```
Out[39]:
```
0.9830508474576272
```

In [ ]:
```
0.9661016949152542
```