lab6

September 5, 2024

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from  sklearn.datasets import load_wine
from scipy import stats
import warnings
warnings.filterwarnings("ignore")
```

6.1 Loading the Wine Dataset

[3]:
```
data=load_wine()
```

[4]:
```
data
```

[4]:
```
{'data': array([[1.423e+01, 1.710e+00, 2.430e+00, …, 1.040e+00, 3.920e+00,
         1.065e+03],
        [1.320e+01, 1.780e+00, 2.140e+00, …, 1.050e+00, 3.400e+00,
         1.050e+03],
        [1.316e+01, 2.360e+00, 2.670e+00, …, 1.030e+00, 3.170e+00,
         1.185e+03],
        …,
        [1.327e+01, 4.280e+00, 2.260e+00, …, 5.900e-01, 1.560e+00,
         8.350e+02],
        [1.317e+01, 2.590e+00, 2.370e+00, …, 6.000e-01, 1.620e+00,
         8.400e+02],
        [1.413e+01, 4.100e+00, 2.740e+00, …, 6.100e-01, 1.600e+00,
         5.600e+02]]),
 'target': array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2,
        2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
        2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
        2, 2]),
 'frame': None,
```

'target_names': array(['class_0', 'class_1', 'class_2'], dtype='<U7'),
'DESCR': '.. _wine_dataset:\n\nWine recognition
dataset\n------------------------\n\n**Data Set Characteristics:**\n\n:Number of
Instances: 178\n:Number of Attributes: 13 numeric, predictive attributes and the
class\n:Attribute Information:\n   - Alcohol\n   - Malic acid\n   - Ash\n
- Alcalinity of ash\n   - Magnesium\n   - Total phenols\n   - Flavanoids\n
- Nonflavanoid phenols\n   - Proanthocyanins\n   - Color intensity\n   -
Hue\n   - OD280/OD315 of diluted wines\n   - Proline\n   - class:\n   -
class_0\n   - class_1\n   - class_2\n\n:Summary
Statistics:\n\n============================= ==== ===== ======= =====\n
Min   Max   Mean   SD\n============================= ==== ===== =======
=====\nAlcohol: 11.0 14.8   13.0   0.8\nMalic Acid:
0.74 5.80   2.34 1.12\nAsh: 1.36 3.23   2.36
0.27\nAlcalinity of Ash: 10.6 30.0   19.5   3.3\nMagnesium:
70.0 162.0   99.7 14.3\nTotal Phenols: 0.98 3.88   2.29
0.63\nFlavanoids: 0.34 5.08   2.03 1.00\nNonflavanoid
Phenols: 0.13 0.66   0.36 0.12\nProanthocyanins: 0.41
3.58   1.59 0.57\nColour Intensity: 1.3 13.0   5.1   2.3\nHue:
0.48 1.71   0.96 0.23\nOD280/OD315 of diluted wines: 1.27 4.00   2.61
0.71\nProline: 278 1680   746
315\n============================= ==== ===== ======= =====\n\n:Missing
Attribute Values: None\n:Class Distribution: class_0 (59), class_1 (71), class_2
(48)\n:Creator: R.A. Fisher\n:Donor: Michael Marshall
(MARSHALL%PLU@io.arc.nasa.gov)\n:Date: July, 1988\n\nThis is a copy of UCI ML
Wine recognition datasets.\nhttps://archive.ics.uci.edu/ml/machine-learning-
databases/wine/wine.data\n\nThe data is the results of a chemical analysis of
wines grown in the same\nregion in Italy by three different cultivators. There
are thirteen different\nmeasurements taken for different constituents found in
the three types of\nwine.\n\nOriginal Owners:\n\nForina, M. et al, PARVUS -\nAn
Extendible Package for Data Exploration, Classification and
Correlation.\nInstitute of Pharmaceutical and Food Analysis and
Technologies,\nVia Brigata Salerno, 16147 Genoa, Italy.\n\nCitation:\n\nLichman,
M. (2013). UCI Machine Learning Repository\n[https://archive.ics.uci.edu/ml].
Irvine, CA: University of California,\nSchool of Information and Computer
Science.\n\n|details-start|\n**References**\n|details-split|\n\n(1) S.
Aeberhard, D. Coomans and O. de Vel,\nComparison of Classifiers in High
Dimensional Settings,\nTech. Rep. no. 92-02, (1992), Dept. of Computer Science
and Dept. of\nMathematics and Statistics, James Cook University of North
Queensland.\n(Also submitted to Technometrics).\n\nThe data was used with many
others for comparing various\nclassifiers. The classes are separable, though
only RDA\nhas achieved 100% correct classification.\n(RDA : 100%, QDA 99.4%, LDA
98.9%, 1NN 96.1% (z-transformed data))\n(All results using the leave-one-out
technique)\n\n(2) S. Aeberhard, D. Coomans and O. de Vel,\n"THE CLASSIFICATION
PERFORMANCE OF RDA"\nTech. Rep. no. 92-01, (1992), Dept. of Computer Science and
Dept. of\nMathematics and Statistics, James Cook University of North
Queensland.\n(Also submitted to Journal of Chemometrics).\n\n|details-end|\n',
'feature_names': ['alcohol',

```
  'malic_acid',
  'ash',
  'alcalinity_of_ash',
  'magnesium',
  'total_phenols',
  'flavanoids',
  'nonflavanoid_phenols',
  'proanthocyanins',
  'color_intensity',
  'hue',
  'od280/od315_of_diluted_wines',
  'proline']}
```

[5]:
```
df=pd.DataFrame(data=data.data,columns=data.feature_names)
```

[6]:
```
df.head()
```

[6]:
```
   alcohol  malic_acid   ash  alcalinity_of_ash  magnesium  total_phenols  \
0    14.23        1.71  2.43               15.6      127.0           2.80
1    13.20        1.78  2.14               11.2      100.0           2.65
2    13.16        2.36  2.67               18.6      101.0           2.80
3    14.37        1.95  2.50               16.8      113.0           3.85
4    13.24        2.59  2.87               21.0      118.0           2.80

   flavanoids  nonflavanoid_phenols  proanthocyanins  color_intensity   hue  \
0        3.06                  0.28             2.29             5.64  1.04
1        2.76                  0.26             1.28             4.38  1.05
2        3.24                  0.30             2.81             5.68  1.03
3        3.49                  0.24             2.18             7.80  0.86
4        2.69                  0.39             1.82             4.32  1.04

   od280/od315_of_diluted_wines  proline
0                          3.92   1065.0
1                          3.40   1050.0
2                          3.17   1185.0
3                          3.45   1480.0
4                          2.93    735.0
```

[7]:
```
df["target"]=data.target
```

[8]:
```
df.sample(3)
```

[8]:
```
     alcohol  malic_acid   ash  alcalinity_of_ash  magnesium  total_phenols  \
108    12.22        1.29  1.94               19.0       92.0           2.36
155    13.17        5.19  2.32               22.0       93.0           1.74
6      14.39        1.87  2.45               14.6       96.0           2.50
```

```
     flavanoids  nonflavanoid_phenols  proanthocyanins  color_intensity   hue  \
108        2.04                  0.39             2.08             2.70  0.86
155        0.63                  0.61             1.55             7.90  0.60
6          2.52                  0.30             1.98             5.25  1.02

     od280/od315_of_diluted_wines  proline  target
108                          3.02    312.0       1
155                          1.48    725.0       2
6                            3.58   1290.0       0
```

6.2 Performing Descriptive Statistics

Mean

[11]:
```python
df.mean()
```

[11]:
```
alcohol                          13.000618
malic_acid                        2.336348
ash                               2.366517
alcalinity_of_ash                19.494944
magnesium                        99.741573
total_phenols                     2.295112
flavanoids                        2.029270
nonflavanoid_phenols              0.361854
proanthocyanins                   1.590899
color_intensity                   5.058090
hue                               0.957449
od280/od315_of_diluted_wines      2.611685
proline                         746.893258
target                            0.938202
dtype: float64
```

Median

[13]:
```python
df.median()
```

[13]:
```
alcohol                          13.050
malic_acid                        1.865
ash                               2.360
alcalinity_of_ash                19.500
magnesium                        98.000
total_phenols                     2.355
flavanoids                        2.135
nonflavanoid_phenols              0.340
proanthocyanins                   1.555
color_intensity                   4.690
hue                               0.965
od280/od315_of_diluted_wines      2.780
proline                         673.500
```

```
target                          1.000
dtype: float64
```

Mode

```
[15]: df.mode().iloc[0]
```

```
[15]: alcohol                        12.37
      malic_acid                      1.73
      ash                             2.28
      alcalinity_of_ash              20.00
      magnesium                      88.00
      total_phenols                   2.20
      flavanoids                      2.65
      nonflavanoid_phenols            0.26
      proanthocyanins                 1.35
      color_intensity                 2.60
      hue                             1.04
      od280/od315_of_diluted_wines    2.87
      proline                       520.00
      target                          1.00
      Name: 0, dtype: float64
```

Standard Deviation

```
[17]: df.std()
```

```
[17]: alcohol                           0.811827
      malic_acid                        1.117146
      ash                               0.274344
      alcalinity_of_ash                 3.339564
      magnesium                        14.282484
      total_phenols                     0.625851
      flavanoids                        0.998859
      nonflavanoid_phenols              0.124453
      proanthocyanins                   0.572359
      color_intensity                   2.318286
      hue                               0.228572
      od280/od315_of_diluted_wines      0.709990
      proline                         314.907474
      target                            0.775035
      dtype: float64
```

Variance

```
[19]: df.var()
```

```
[19]: alcohol                             0.659062
      malic_acid                          1.248015
```

```
ash                             0.075265
alcalinity_of_ash              11.152686
magnesium                     203.989335
total_phenols                   0.391690
flavanoids                      0.997719
nonflavanoid_phenols            0.015489
proanthocyanins                 0.327595
color_intensity                 5.374449
hue                             0.052245
od280/od315_of_diluted_wines    0.504086
proline                     99166.717355
target                          0.600679
dtype: float64
```

Range

[21]:
```
df.max()-df.min()
```

[21]:
```
alcohol                            3.80
malic_acid                         5.06
ash                                1.87
alcalinity_of_ash                 19.40
magnesium                         92.00
total_phenols                      2.90
flavanoids                         4.74
nonflavanoid_phenols               0.53
proanthocyanins                    3.17
color_intensity                   11.72
hue                                1.23
od280/od315_of_diluted_wines       2.73
proline                         1402.00
target                             2.00
dtype: float64
```

Skewness

[23]:
```
df.skew()
```

[23]:
```
alcohol                         -0.051482
malic_acid                       1.039651
ash                             -0.176699
alcalinity_of_ash                0.213047
magnesium                        1.098191
total_phenols                    0.086639
flavanoids                       0.025344
nonflavanoid_phenols             0.450151
proanthocyanins                  0.517137
color_intensity                  0.868585
```

```
hue                             0.021091
od280/od315_of_diluted_wines   -0.307285
proline                         0.767822
target                          0.107431
dtype: float64
```

Kurtosis

[25]:
```
df.kurt()
```

[25]:
```
alcohol                        -0.852500
malic_acid                      0.299207
ash                             1.143978
alcalinity_of_ash               0.487942
magnesium                       2.104991
total_phenols                  -0.835627
flavanoids                     -0.880382
nonflavanoid_phenols           -0.637191
proanthocyanins                 0.554649
color_intensity                 0.381522
hue                            -0.344096
od280/od315_of_diluted_wines   -1.086435
proline                        -0.248403
target                         -1.322787
dtype: float64
```

6.3 Performing Inferential Statistics

[27]:
```
flavanoids  = df['flavanoids']
```

[28]:
```
population_mean = 0.05
```

[29]:
```
t_stat, p_value = stats.ttest_1samp(flavanoids, population_mean)
```

[30]:
```
t_stat
```

[30]:
```
26.436923792243125
```

[31]:
```
p_value
```

[31]:
```
2.3103364017205547e-63
```

6.4 Confidence Intervals

[33]:
```
sample_mean = np.mean(flavanoids)
```

[34]:
```
standard_error = stats.sem(flavanoids)
```

[35]:
```
sample_mean
```

[35]: 2.0292696629213487

[36]:
```
standard_error
```

[36]: 0.07486762372489372

95% Confidence Interval for flavanoids

[38]:
```
confidence_interval_95 = stats.norm.interval(0.95, loc=sample_mean,␣
↪scale=standard_error)
```

[39]:
```
confidence_interval_95
```

[39]: (1.8825318168124605, 2.176007509030237)

99% Confidence Interval for flavanoids Selection deleted

[41]:
```
confidence_interval_99 = stats.norm.interval(0.99, loc=sample_mean,␣
↪scale=standard_error)
```

[42]:
```
confidence_interval_99
```

[42]: (1.8364234438436946, 2.222115881999003)

6.5 Regression Analysis

Logistic Regression

[45]:
```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

[46]:
```
x=df.drop(columns=["target"])
```

[47]:
```
y=df["target"]
```

[48]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
↪1,random_state=42)
```

[49]:
```
x_train.head(3)
```

[49]:
```
     alcohol  malic_acid   ash  alcalinity_of_ash  magnesium  total_phenols  \
9      13.86        1.35  2.27               16.0       98.0           2.98
114    12.08        1.39  2.50               22.5       84.0           2.56
18     14.19        1.59  2.48               16.5      108.0           3.30

     flavanoids  nonflavanoid_phenols  proanthocyanins  color_intensity   hue  \
9          3.15                  0.22             1.85             7.22  1.01
114        2.29                  0.43             1.04             2.90  0.93
18         3.93                  0.32             1.86             8.70  1.23
```

```
     od280/od315_of_diluted_wines  proline
9                            3.55   1045.0
114                          3.19    385.0
18                           2.82   1680.0
```

[50]:
```
y_train.head(3)
```

[50]:
```
9      0
114    1
18     0
Name: target, dtype: int32
```

Standardizing the data

[52]:
```
from sklearn.preprocessing import StandardScaler
```

[53]:
```
scaler=StandardScaler()
```

[54]:
```
x_train_transform=scaler.fit_transform(x_train)
```

[55]:
```
x_test_transform=scaler.fit_transform(x_test)
```

[56]:
```
x_train_transform
```

[56]:
```
array([[ 1.06763294, -0.867326  , -0.32621584, …,  0.22658118,
         1.33106454,  0.95501956],
       [-1.13795143, -0.83169086,  0.51563149, …, -0.12058654,
         0.82742322, -1.14261482],
       [ 1.47653341, -0.65351514,  0.44242737, …,  1.1812924 ,
         0.30979187,  2.9731981 ],
       …,
       [ 1.71196096, -0.40406913,  0.0764068 , …,  1.05110451,
         0.56161253,  2.55049299],
       [-0.3821051 , -0.70696785, -0.3628179 , …,  0.00960136,
        -0.7534509 , -0.79300909],
       [-0.81578742,  0.11264046,  0.36922326, …, -0.68473408,
         1.09323392, -0.97416842]])
```

[57]:
```
x_test_transform
```

[57]:
```
array([[ 7.45264747e-01,  6.31938891e-01,  4.76841934e-01,
        -2.05593023e+00,  1.36541560e+00,  8.22039981e-01,
         1.09028353e+00, -1.85145878e+00,  4.17434297e-01,
         2.58527671e-01,  2.71976217e-02,  9.98624407e-01,
         2.50095060e-01],
       [ 1.43104199e+00,  1.54581975e+00,  2.90251612e-02,
        -4.92859986e-01,  1.00291588e+00,  1.10769094e+00,
         6.58712968e-01, -3.41907022e-01, -4.19702962e-01,
```

```
 3.30356748e-01, -4.13403850e-01,  9.51196335e-01,
 1.01392329e+00],
[-1.08949017e-01,  3.49996924e-01,  9.99294836e-01,
 3.94287989e-01, -8.45832673e-02, -1.38699410e+00,
-1.78306786e+00,  2.32883840e+00, -1.44060206e+00,
 1.99525214e-03, -9.02961040e-01, -6.61358117e-01,
-5.46236503e-01],
[ 8.53545365e-01, -9.23602995e-01,  9.99294836e-01,
 1.02796511e+00,  2.77916450e-01,  1.39334190e+00,
 1.34014018e+00, -4.58026389e-01,  1.88752899e+00,
 5.66366573e-01,  1.15317916e+00, -2.89838218e-02,
 1.68024154e+00],
[-7.82695084e-01, -1.24443351e+00, -1.91151419e+00,
-1.97143994e-01, -1.38958225e+00, -3.01520456e-01,
-7.94998410e-02, -6.90265121e-01, -8.48480582e-01,
 4.30404391e-02,  8.10489126e-01,  1.18833670e+00,
-8.38766464e-01],
[ 1.53932261e+00, -5.15273250e-01,  1.07393096e+00,
-2.81634278e-02,  1.65541537e+00,  1.01247395e+00,
 1.21521186e+00,  6.45107590e-03,  1.05039174e+00,
 8.22898991e-01,  5.65710531e-01, -1.23839966e-01,
 1.66398987e+00],
[-1.22784874e+00,  9.52769405e-01, -1.61296967e+00,
-4.50614844e-01, -7.37082757e-01, -5.10997826e-01,
-4.88356166e-01,  4.70928540e-01,  8.46211917e-01,
-1.70138001e+00, -1.19669535e-01,  5.08534329e-01,
-6.63248487e-01],
[ 4.56516432e-01,  1.41943197e+00,  1.78297419e-01,
 1.23919082e+00,  3.50416393e-01, -8.91865771e-01,
-1.49913986e+00,  1.16764474e+00, -9.30152509e-02,
 1.38727031e+00, -1.24565107e+00, -1.84705992e+00,
-5.86865664e-02],
[-1.69706475e+00, -1.06943505e+00,  9.99294836e-01,
-2.81634278e-02, -2.29583154e-01,  8.98213570e-01,
 9.65355212e-01, -4.58026389e-01,  2.11212680e+00,
-9.98481180e-01,  2.71976217e-02,  8.40530833e-01,
-2.86209870e-01],
[ 4.08391713e-01,  1.06943505e-01, -3.06837418e-01,
-2.81634278e-02, -5.92082871e-01, -1.65360166e+00,
-1.78306786e+00,  4.70928540e-01, -1.66519986e+00,
 5.15060089e-01, -1.24565107e+00, -4.08408399e-01,
 3.88234208e-02],
[ 5.76828230e-01, -6.22216754e-01,  6.63432256e-01,
-2.81634278e-02, -8.45832673e-02,  4.98302228e-01,
 6.13284488e-01, -5.74145755e-01,  4.17434297e-01,
-5.52114772e-01,  8.10489126e-01,  1.72585485e+00,
 2.50095060e-01],
```

```
[ 5.76828230e-01,  6.51383165e-01,  7.00750321e-01,
  1.66164224e+00,  1.87291520e+00, -1.65360166e+00,
 -5.67856007e-01, -1.27086195e+00, -4.19702962e-01,
  2.05425460e+00, -1.78416398e+00, -2.25810321e+00,
 -8.71269793e-01],
[ 4.68547612e-01,  1.35137701e+00, -1.16515290e+00,
 -5.35105128e-01, -5.19582927e-01,  3.45955050e-01,
  6.92784329e-01, -6.90265121e-01,  4.99106225e-02,
 -1.62185496e-01, -2.17580973e-01,  4.29487542e-01,
  8.67658312e-01],
[-3.01447893e-01,  9.52769405e-01, -1.68760580e+00,
 -1.71796909e+00, -1.24458236e+00, -1.21560352e+00,
 -9.31283852e-01,  1.16764474e+00, -1.27725820e+00,
 -6.13682552e-01, -1.24565107e+00, -9.61735907e-01,
 -1.28731241e+00],
[ 7.33233567e-01, -6.22216754e-01,  9.99294836e-01,
 -1.21102739e+00,  1.07541583e+00,  1.10769094e+00,
  9.53998092e-01, -3.41907022e-01,  9.07465863e-03,
  1.38727031e+00,  1.59378063e+00,  2.39775253e-01,
  1.76149986e+00],
[-6.02227387e-01, -1.94442736e-02, -9.78562578e-01,
  3.94287989e-01, -6.64582814e-01,  5.36389022e-01,
  2.27142403e-01, -8.06384487e-01, -4.80956907e-01,
 -1.33197332e+00, -2.66536692e-01,  8.16816797e-02,
 -1.44007805e+00],
[-1.93768834e+00, -1.66248539e+00,  2.52933548e-01,
  3.94287989e-01, -6.64582814e-01,  4.03085242e-01,
 -6.81427209e-02,  1.05152537e+00, -3.17613052e-02,
 -7.77863300e-01,  7.12577688e-01, -6.61358117e-01,
 -1.08579177e+00],
[-1.13159930e+00, -1.28332206e+00,  2.90251612e-01,
  1.66164224e+00, -1.38958225e+00, -5.10997826e-01,
 -5.56498887e-01,  8.19286639e-01, -1.13433233e-01,
 -1.22936036e+00,  1.74064779e+00, -1.31744645e-02,
 -4.48726515e-01]])
```

Predicting the Score

[59]:
```
clf=LogisticRegression()
```

[60]:
```
clf.fit(x_train,y_train)
```

[60]:
```
LogisticRegression()
```

[61]:
```
y_pred=clf.predict(x_test)
```

[62]:
```
from sklearn.metrics import accuracy_score
```

[63]:
```
accuracy_score(y_test,y_pred)
```

[63]: 0.9444444444444444

[ ]:
```
```