CL1_01

July 15, 2025

[ ]:
```
'''
NAME:Aher Swami Sandip
ROLL NO.01
COURSE: AI&DS
CLASS: BE
SUB:Computer Laboratory-I (Machine Learning)
'''
```

[ ]:
```
'''
PRACTICAL NO-01:
    To use PCA Algorithm for dimensionality reduction. You have a dataset that␣
 ↪includes
measurements for different variables on wine (alcohol, ash, magnesium, and so␣
 ↪on).
Apply PCA algorithm & transform this data so that most variations in the␣
 ↪measurements
of the variables are captured by a small number of principal components so that␣
 ↪it is
easier to distinguish between red and white wine by inspecting these principal
components.'''
```

[3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
```

[5]:
```
data=pd.read_csv('Wine.csv')
```

[7]:
```
data.head()
```

[7]:

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	\
0	14.23	1.71	2.43	15.6	127	2.80	
1	13.20	1.78	2.14	11.2	100	2.65	
2	13.16	2.36	2.67	18.6	101	2.80	
3	14.37	1.95	2.50	16.8	113	3.85	
4	13.24	2.59	2.87	21.0	118	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  Proline  Customer_Segment
0   3.92     1065                 1
1   3.40     1050                 1
2   3.17     1185                 1
3   3.45     1480                 1
4   2.93      735                 1
```

[9]:
```
data.tail()
```

[9]:
```
     Alcohol  Malic_Acid   Ash  Ash_Alcanity  Magnesium  Total_Phenols  \
173    13.71        5.65  2.45          20.5         95           1.68
174    13.40        3.91  2.48          23.0        102           1.80
175    13.27        4.28  2.26          20.0        120           1.59
176    13.17        2.59  2.37          20.0        120           1.65
177    14.13        4.10  2.74          24.5         96           2.05

     Flavanoids  Nonflavanoid_Phenols  Proanthocyanins  Color_Intensity   Hue  \
173        0.61                  0.52             1.06              7.7  0.64
174        0.75                  0.43             1.41              7.3  0.70
175        0.69                  0.43             1.35             10.2  0.59
176        0.68                  0.53             1.46              9.3  0.60
177        0.76                  0.56             1.35              9.2  0.61

     OD280  Proline  Customer_Segment
173   1.74      740                 3
174   1.56      750                 3
175   1.56      835                 3
176   1.62      840                 3
177   1.60      560                 3
```

[11]:
```
data.shape
```

[11]:
```
(178, 14)
```

[13]:
```
data.describe()
```

[13]:
```
          Alcohol  Malic_Acid         Ash  Ash_Alcanity   Magnesium  \
count  178.000000  178.000000  178.000000    178.000000  178.000000
mean    13.000618    2.336348    2.366517     19.494944   99.741573
std      0.811827    1.117146    0.274344      3.339564   14.282484
```

```
min     11.030000    0.740000    1.360000     10.600000   70.000000
25%     12.362500    1.602500    2.210000     17.200000   88.000000
50%     13.050000    1.865000    2.360000     19.500000   98.000000
75%     13.677500    3.082500    2.557500     21.500000  107.000000
max     14.830000    5.800000    3.230000     30.000000  162.000000

       Total_Phenols  Flavanoids  Nonflavanoid_Phenols  Proanthocyanins  \
count     178.000000  178.000000            178.000000       178.000000
mean        2.295112    2.029270              0.361854         1.590899
std         0.625851    0.998859              0.124453         0.572359
min         0.980000    0.340000              0.130000         0.410000
25%         1.742500    1.205000              0.270000         1.250000
50%         2.355000    2.135000              0.340000         1.555000
75%         2.800000    2.875000              0.437500         1.950000
max         3.880000    5.080000              0.660000         3.580000

       Color_Intensity         Hue       OD280      Proline  Customer_Segment
count       178.000000  178.000000  178.000000   178.000000        178.000000
mean          5.058090    0.957449    2.611685   746.893258          1.938202
std           2.318286    0.228572    0.709990   314.907474          0.775035
min           1.280000    0.480000    1.270000   278.000000          1.000000
25%           3.220000    0.782500    1.937500   500.500000          1.000000
50%           4.690000    0.965000    2.780000   673.500000          2.000000
75%           6.200000    1.120000    3.170000   985.000000          3.000000
max          13.000000    1.710000    4.000000  1680.000000          3.000000
```

[15]:
```
data['Customer_Segment'].unique()
```

[15]:
```
array([1, 2, 3], dtype=int64)
```

[17]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Alcohol               178 non-null    float64
 1   Malic_Acid            178 non-null    float64
 2   Ash                   178 non-null    float64
 3   Ash_Alcanity          178 non-null    float64
 4   Magnesium             178 non-null    int64
 5   Total_Phenols         178 non-null    float64
 6   Flavanoids            178 non-null    float64
 7   Nonflavanoid_Phenols  178 non-null    float64
 8   Proanthocyanins       178 non-null    float64
 9   Color_Intensity       178 non-null    float64
 10  Hue                   178 non-null    float64
```

```
 11  OD280                 178 non-null    float64
 12  Proline               178 non-null    int64
 13  Customer_Segment      178 non-null    int64
dtypes: float64(11), int64(3)
memory usage: 19.6 KB
```

[19]:
```
data.isnull().sum()
```

[19]:
```
Alcohol                 0
Malic_Acid              0
Ash                     0
Ash_Alcanity            0
Magnesium               0
Total_Phenols           0
Flavanoids              0
Nonflavanoid_Phenols    0
Proanthocyanins         0
Color_Intensity         0
Hue                     0
OD280                   0
Proline                 0
Customer_Segment        0
dtype: int64
```

No null values

[22]:
```
x=data.drop('Customer_Segment',axis=1)
y=data['Customer_Segment']
```

[24]:
```
x
```

[24]:

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	\
0	14.23	1.71	2.43	15.6	127	2.80	
1	13.20	1.78	2.14	11.2	100	2.65	
2	13.16	2.36	2.67	18.6	101	2.80	
3	14.37	1.95	2.50	16.8	113	3.85	
4	13.24	2.59	2.87	21.0	118	2.80	
..	…	…	…	…	…	…	
173	13.71	5.65	2.45	20.5	95	1.68	
174	13.40	3.91	2.48	23.0	102	1.80	
175	13.27	4.28	2.26	20.0	120	1.59	
176	13.17	2.59	2.37	20.0	120	1.65	
177	14.13	4.10	2.74	24.5	96	2.05	

	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
..          …                     …                …                 …    …
173        0.61                  0.52             1.06              7.70  0.64
174        0.75                  0.43             1.41              7.30  0.70
175        0.69                  0.43             1.35             10.20  0.59
176        0.68                  0.53             1.46              9.30  0.60
177        0.76                  0.56             1.35              9.20  0.61

     OD280  Proline
0     3.92     1065
1     3.40     1050
2     3.17     1185
3     3.45     1480
4     2.93      735
..     …       …
173   1.74      740
174   1.56      750
175   1.56      835
176   1.62      840
177   1.60      560

[178 rows x 13 columns]
```

[26]:
```
x.shape
```

[26]:
```
(178, 13)
```

[28]:
```
x_standardized = (x - x.mean()) / x.std()
```

[30]:
```
y
```

[30]:
```
0      1
1      1
2      1
3      1
4      1
      ..
173    3
174    3
175    3
176    3
177    3
Name: Customer_Segment, Length: 178, dtype: int64
```

[32]:
```
pca=PCA(n_components=3)
```

[34]:
```
x_pca=pca.fit_transform(x)
```

[35]:
```
x_pca.shape
```

[35]:
```
(178, 3)
```

13 columns got reduced to 3 columns

[39]:
```
pca_df = pd.DataFrame(x_pca, columns = ['pca_col1','pca_col2','pca_col3'])
```

[41]:
```
pca_df
```

[41]:

	pca_col1	pca_col2	pca_col3
0	318.562979	21.492131	-3.130735
1	303.097420	-5.364718	-6.822835
2	438.061133	-6.537309	1.113223
3	733.240139	0.192729	0.917257
4	-11.571428	18.489995	0.554422
..	…	…	…
173	-6.980211	-4.541137	2.474707
174	3.131605	2.335191	4.309931
175	88.458074	18.776285	2.237577
176	93.456242	18.670819	1.788392
177	-186.943190	-0.213331	5.630510

[178 rows x 3 columns]

[43]:
```
pca.explained_variance_ratio_
```

[43]:
```
array([9.98091230e-01, 1.73591562e-03, 9.49589576e-05])
```