PCA

May 23, 2022

[1]:
```
# lOAD IRIS DATASET
import pandas as pd
df = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/
 ↪iris/iris.data",
                 names=['sepal length','sepal width','petal length','petal␣
 ↪width','target'])
```

[2]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal length  150 non-null    float64
 1   sepal width   150 non-null    float64
 2   petal length  150 non-null    float64
 3   petal width   150 non-null    float64
 4   target        150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

[3]:
```
df.describe()
```

[3]:

	sepal length	sepal width	petal length	petal width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

[4]:
```
from sklearn.preprocessing import StandardScaler
```

[5]:
```
sc = StandardScaler()
X = df[['sepal length','sepal width','petal length','petal width']]
y = df["target"]
X = sc.fit_transform(X)
```

[6]:
```
X
```

[6]:
```
array([[-9.00681170e-01,  1.03205722e+00, -1.34127240e+00,
        -1.31297673e+00],
       [-1.14301691e+00, -1.24957601e-01, -1.34127240e+00,
        -1.31297673e+00],
       [-1.38535265e+00,  3.37848329e-01, -1.39813811e+00,
        -1.31297673e+00],
       [-1.50652052e+00,  1.06445364e-01, -1.28440670e+00,
        -1.31297673e+00],
       [-1.02184904e+00,  1.26346019e+00, -1.34127240e+00,
        -1.31297673e+00],
       [-5.37177559e-01,  1.95766909e+00, -1.17067529e+00,
        -1.05003079e+00],
       [-1.50652052e+00,  8.00654259e-01, -1.34127240e+00,
        -1.18150376e+00],
       [-1.02184904e+00,  8.00654259e-01, -1.28440670e+00,
        -1.31297673e+00],
       [-1.74885626e+00, -3.56360566e-01, -1.34127240e+00,
        -1.31297673e+00],
       [-1.14301691e+00,  1.06445364e-01, -1.28440670e+00,
        -1.44444970e+00],
       [-5.37177559e-01,  1.49486315e+00, -1.28440670e+00,
        -1.31297673e+00],
       [-1.26418478e+00,  8.00654259e-01, -1.22754100e+00,
        -1.31297673e+00],
       [-1.26418478e+00, -1.24957601e-01, -1.34127240e+00,
        -1.44444970e+00],
       [-1.87002413e+00, -1.24957601e-01, -1.51186952e+00,
        -1.44444970e+00],
       [-5.25060772e-02,  2.18907205e+00, -1.45500381e+00,
        -1.31297673e+00],
       [-1.73673948e-01,  3.11468391e+00, -1.28440670e+00,
        -1.05003079e+00],
       [-5.37177559e-01,  1.95766909e+00, -1.39813811e+00,
        -1.05003079e+00],
       [-9.00681170e-01,  1.03205722e+00, -1.34127240e+00,
        -1.18150376e+00],
       [-1.73673948e-01,  1.72626612e+00, -1.17067529e+00,
        -1.18150376e+00],
       [-9.00681170e-01,  1.72626612e+00, -1.28440670e+00,
        -1.18150376e+00],
```

```
[-5.37177559e-01,  8.00654259e-01, -1.17067529e+00,
 -1.31297673e+00],
[-9.00681170e-01,  1.49486315e+00, -1.28440670e+00,
 -1.05003079e+00],
[-1.50652052e+00,  1.26346019e+00, -1.56873522e+00,
 -1.31297673e+00],
[-9.00681170e-01,  5.69251294e-01, -1.17067529e+00,
 -9.18557817e-01],
[-1.26418478e+00,  8.00654259e-01, -1.05694388e+00,
 -1.31297673e+00],
[-1.02184904e+00, -1.24957601e-01, -1.22754100e+00,
 -1.31297673e+00],
[-1.02184904e+00,  8.00654259e-01, -1.22754100e+00,
 -1.05003079e+00],
[-7.79513300e-01,  1.03205722e+00, -1.28440670e+00,
 -1.31297673e+00],
[-7.79513300e-01,  8.00654259e-01, -1.34127240e+00,
 -1.31297673e+00],
[-1.38535265e+00,  3.37848329e-01, -1.22754100e+00,
 -1.31297673e+00],
[-1.26418478e+00,  1.06445364e-01, -1.22754100e+00,
 -1.31297673e+00],
[-5.37177559e-01,  8.00654259e-01, -1.28440670e+00,
 -1.05003079e+00],
[-7.79513300e-01,  2.42047502e+00, -1.28440670e+00,
 -1.44444970e+00],
[-4.16009689e-01,  2.65187798e+00, -1.34127240e+00,
 -1.31297673e+00],
[-1.14301691e+00,  1.06445364e-01, -1.28440670e+00,
 -1.44444970e+00],
[-1.02184904e+00,  3.37848329e-01, -1.45500381e+00,
 -1.31297673e+00],
[-4.16009689e-01,  1.03205722e+00, -1.39813811e+00,
 -1.31297673e+00],
[-1.14301691e+00,  1.06445364e-01, -1.28440670e+00,
 -1.44444970e+00],
[-1.74885626e+00, -1.24957601e-01, -1.39813811e+00,
 -1.31297673e+00],
[-9.00681170e-01,  8.00654259e-01, -1.28440670e+00,
 -1.31297673e+00],
[-1.02184904e+00,  1.03205722e+00, -1.39813811e+00,
 -1.18150376e+00],
[-1.62768839e+00, -1.74477836e+00, -1.39813811e+00,
 -1.18150376e+00],
[-1.74885626e+00,  3.37848329e-01, -1.39813811e+00,
 -1.31297673e+00],
[-1.02184904e+00,  1.03205722e+00, -1.22754100e+00,
```

```
 -7.87084847e-01],
[-9.00681170e-01,  1.72626612e+00, -1.05694388e+00,
 -1.05003079e+00],
[-1.26418478e+00, -1.24957601e-01, -1.34127240e+00,
 -1.18150376e+00],
[-9.00681170e-01,  1.72626612e+00, -1.22754100e+00,
 -1.31297673e+00],
[-1.50652052e+00,  3.37848329e-01, -1.34127240e+00,
 -1.31297673e+00],
[-6.58345429e-01,  1.49486315e+00, -1.28440670e+00,
 -1.31297673e+00],
[-1.02184904e+00,  5.69251294e-01, -1.34127240e+00,
 -1.31297673e+00],
[ 1.40150837e+00,  3.37848329e-01,  5.35295827e-01,
  2.64698913e-01],
[ 6.74501145e-01,  3.37848329e-01,  4.21564419e-01,
  3.96171883e-01],
[ 1.28034050e+00,  1.06445364e-01,  6.49027235e-01,
  3.96171883e-01],
[-4.16009689e-01, -1.74477836e+00,  1.37235899e-01,
  1.33225943e-01],
[ 7.95669016e-01, -5.87763531e-01,  4.78430123e-01,
  3.96171883e-01],
[-1.73673948e-01, -5.87763531e-01,  4.21564419e-01,
  1.33225943e-01],
[ 5.53333275e-01,  5.69251294e-01,  5.35295827e-01,
  5.27644853e-01],
[-1.14301691e+00, -1.51337539e+00, -2.60824029e-01,
 -2.61192967e-01],
[ 9.16836886e-01, -3.56360566e-01,  4.78430123e-01,
  1.33225943e-01],
[-7.79513300e-01, -8.19166497e-01,  8.03701950e-02,
  2.64698913e-01],
[-1.02184904e+00, -2.43898725e+00, -1.47092621e-01,
 -2.61192967e-01],
[ 6.86617933e-02, -1.24957601e-01,  2.50967307e-01,
  3.96171883e-01],
[ 1.89829664e-01, -1.97618132e+00,  1.37235899e-01,
 -2.61192967e-01],
[ 3.10997534e-01, -3.56360566e-01,  5.35295827e-01,
  2.64698913e-01],
[-2.94841818e-01, -3.56360566e-01, -9.02269170e-02,
  1.33225943e-01],
[ 1.03800476e+00,  1.06445364e-01,  3.64698715e-01,
  2.64698913e-01],
[-2.94841818e-01, -1.24957601e-01,  4.21564419e-01,
  3.96171883e-01],
```

```
[-5.25060772e-02, -8.19166497e-01,  1.94101603e-01,
 -2.61192967e-01],
[ 4.32165405e-01, -1.97618132e+00,  4.21564419e-01,
  3.96171883e-01],
[-2.94841818e-01, -1.28197243e+00,  8.03701950e-02,
 -1.29719997e-01],
[ 6.86617933e-02,  3.37848329e-01,  5.92161531e-01,
  7.90590793e-01],
[ 3.10997534e-01, -5.87763531e-01,  1.37235899e-01,
  1.33225943e-01],
[ 5.53333275e-01, -1.28197243e+00,  6.49027235e-01,
  3.96171883e-01],
[ 3.10997534e-01, -5.87763531e-01,  5.35295827e-01,
  1.75297293e-03],
[ 6.74501145e-01, -3.56360566e-01,  3.07833011e-01,
  1.33225943e-01],
[ 9.16836886e-01, -1.24957601e-01,  3.64698715e-01,
  2.64698913e-01],
[ 1.15917263e+00, -5.87763531e-01,  5.92161531e-01,
  2.64698913e-01],
[ 1.03800476e+00, -1.24957601e-01,  7.05892939e-01,
  6.59117823e-01],
[ 1.89829664e-01, -3.56360566e-01,  4.21564419e-01,
  3.96171883e-01],
[-1.73673948e-01, -1.05056946e+00, -1.47092621e-01,
 -2.61192967e-01],
[-4.16009689e-01, -1.51337539e+00,  2.35044910e-02,
 -1.29719997e-01],
[-4.16009689e-01, -1.51337539e+00, -3.33612130e-02,
 -2.61192967e-01],
[-5.25060772e-02, -8.19166497e-01,  8.03701950e-02,
  1.75297293e-03],
[ 1.89829664e-01, -8.19166497e-01,  7.62758643e-01,
  5.27644853e-01],
[-5.37177559e-01, -1.24957601e-01,  4.21564419e-01,
  3.96171883e-01],
[ 1.89829664e-01,  8.00654259e-01,  4.21564419e-01,
  5.27644853e-01],
[ 1.03800476e+00,  1.06445364e-01,  5.35295827e-01,
  3.96171883e-01],
[ 5.53333275e-01, -1.74477836e+00,  3.64698715e-01,
  1.33225943e-01],
[-2.94841818e-01, -1.24957601e-01,  1.94101603e-01,
  1.33225943e-01],
[-4.16009689e-01, -1.28197243e+00,  1.37235899e-01,
  1.33225943e-01],
[-4.16009689e-01, -1.05056946e+00,  3.64698715e-01,
```

```
  1.75297293e-03],
[ 3.10997534e-01, -1.24957601e-01,  4.78430123e-01,
  2.64698913e-01],
[-5.25060772e-02, -1.05056946e+00,  1.37235899e-01,
  1.75297293e-03],
[-1.02184904e+00, -1.74477836e+00, -2.60824029e-01,
 -2.61192967e-01],
[-2.94841818e-01, -8.19166497e-01,  2.50967307e-01,
  1.33225943e-01],
[-1.73673948e-01, -1.24957601e-01,  2.50967307e-01,
  1.75297293e-03],
[-1.73673948e-01, -3.56360566e-01,  2.50967307e-01,
  1.33225943e-01],
[ 4.32165405e-01, -3.56360566e-01,  3.07833011e-01,
  1.33225943e-01],
[-9.00681170e-01, -1.28197243e+00, -4.31421141e-01,
 -1.29719997e-01],
[-1.73673948e-01, -5.87763531e-01,  1.94101603e-01,
  1.33225943e-01],
[ 5.53333275e-01,  5.69251294e-01,  1.27454998e+00,
  1.71090158e+00],
[-5.25060772e-02, -8.19166497e-01,  7.62758643e-01,
  9.22063763e-01],
[ 1.52267624e+00, -1.24957601e-01,  1.21768427e+00,
  1.18500970e+00],
[ 5.53333275e-01, -3.56360566e-01,  1.04708716e+00,
  7.90590793e-01],
[ 7.95669016e-01, -1.24957601e-01,  1.16081857e+00,
  1.31648267e+00],
[ 2.12851559e+00, -1.24957601e-01,  1.61574420e+00,
  1.18500970e+00],
[-1.14301691e+00, -1.28197243e+00,  4.21564419e-01,
  6.59117823e-01],
[ 1.76501198e+00, -3.56360566e-01,  1.44514709e+00,
  7.90590793e-01],
[ 1.03800476e+00, -1.28197243e+00,  1.16081857e+00,
  7.90590793e-01],
[ 1.64384411e+00,  1.26346019e+00,  1.33141568e+00,
  1.71090158e+00],
[ 7.95669016e-01,  3.37848329e-01,  7.62758643e-01,
  1.05353673e+00],
[ 6.74501145e-01, -8.19166497e-01,  8.76490051e-01,
  9.22063763e-01],
[ 1.15917263e+00, -1.24957601e-01,  9.90221459e-01,
  1.18500970e+00],
[-1.73673948e-01, -1.28197243e+00,  7.05892939e-01,
  1.05353673e+00],
```

```
[-5.25060772e-02, -5.87763531e-01,  7.62758643e-01,
  1.57942861e+00],
[ 6.74501145e-01,  3.37848329e-01,  8.76490051e-01,
  1.44795564e+00],
[ 7.95669016e-01, -1.24957601e-01,  9.90221459e-01,
  7.90590793e-01],
[ 2.24968346e+00,  1.72626612e+00,  1.67260991e+00,
  1.31648267e+00],
[ 2.24968346e+00, -1.05056946e+00,  1.78634131e+00,
  1.44795564e+00],
[ 1.89829664e-01, -1.97618132e+00,  7.05892939e-01,
  3.96171883e-01],
[ 1.28034050e+00,  3.37848329e-01,  1.10395287e+00,
  1.44795564e+00],
[-2.94841818e-01, -5.87763531e-01,  6.49027235e-01,
  1.05353673e+00],
[ 2.24968346e+00, -5.87763531e-01,  1.67260991e+00,
  1.05353673e+00],
[ 5.53333275e-01, -8.19166497e-01,  6.49027235e-01,
  7.90590793e-01],
[ 1.03800476e+00,  5.69251294e-01,  1.10395287e+00,
  1.18500970e+00],
[ 1.64384411e+00,  3.37848329e-01,  1.27454998e+00,
  7.90590793e-01],
[ 4.32165405e-01, -5.87763531e-01,  5.92161531e-01,
  7.90590793e-01],
[ 3.10997534e-01, -1.24957601e-01,  6.49027235e-01,
  7.90590793e-01],
[ 6.74501145e-01, -5.87763531e-01,  1.04708716e+00,
  1.18500970e+00],
[ 1.64384411e+00, -1.24957601e-01,  1.16081857e+00,
  5.27644853e-01],
[ 1.88617985e+00, -5.87763531e-01,  1.33141568e+00,
  9.22063763e-01],
[ 2.49201920e+00,  1.72626612e+00,  1.50201279e+00,
  1.05353673e+00],
[ 6.74501145e-01, -5.87763531e-01,  1.04708716e+00,
  1.31648267e+00],
[ 5.53333275e-01, -5.87763531e-01,  7.62758643e-01,
  3.96171883e-01],
[ 3.10997534e-01, -1.05056946e+00,  1.04708716e+00,
  2.64698913e-01],
[ 2.24968346e+00, -1.24957601e-01,  1.33141568e+00,
  1.44795564e+00],
[ 5.53333275e-01,  8.00654259e-01,  1.04708716e+00,
  1.57942861e+00],
[ 6.74501145e-01,  1.06445364e-01,  9.90221459e-01,
```

```
        7.90590793e-01],
       [ 1.89829664e-01, -1.24957601e-01,  5.92161531e-01,
        7.90590793e-01],
       [ 1.28034050e+00,  1.06445364e-01,  9.33355755e-01,
        1.18500970e+00],
       [ 1.03800476e+00,  1.06445364e-01,  1.04708716e+00,
        1.57942861e+00],
       [ 1.28034050e+00,  1.06445364e-01,  7.62758643e-01,
        1.44795564e+00],
       [-5.25060772e-02, -8.19166497e-01,  7.62758643e-01,
        9.22063763e-01],
       [ 1.15917263e+00,  3.37848329e-01,  1.21768427e+00,
        1.44795564e+00],
       [ 1.03800476e+00,  5.69251294e-01,  1.10395287e+00,
        1.71090158e+00],
       [ 1.03800476e+00, -1.24957601e-01,  8.19624347e-01,
        1.44795564e+00],
       [ 5.53333275e-01, -1.28197243e+00,  7.05892939e-01,
        9.22063763e-01],
       [ 7.95669016e-01, -1.24957601e-01,  8.19624347e-01,
        1.05353673e+00],
       [ 4.32165405e-01,  8.00654259e-01,  9.33355755e-01,
        1.44795564e+00],
       [ 6.86617933e-02, -1.24957601e-01,  7.62758643e-01,
        7.90590793e-01]])
```

[7]:
```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
```

[8]:
```
import matplotlib.pyplot as plt
```

[9]:
```
X_reduced_components = pca.fit_transform(X)
```

[10]:
```
principalDf = pd.DataFrame(data = X_reduced_components
             , columns = ['principal component 1', 'principal component 2'])
```

[11]:
```
principalDf
```

[11]:

	principal component 1	principal component 2
0	-2.264542	0.505704
1	-2.086426	-0.655405
2	-2.367950	-0.318477
3	-2.304197	-0.575368
4	-2.388777	0.674767
..	…	…
145	1.870522	0.382822
146	1.558492	-0.905314

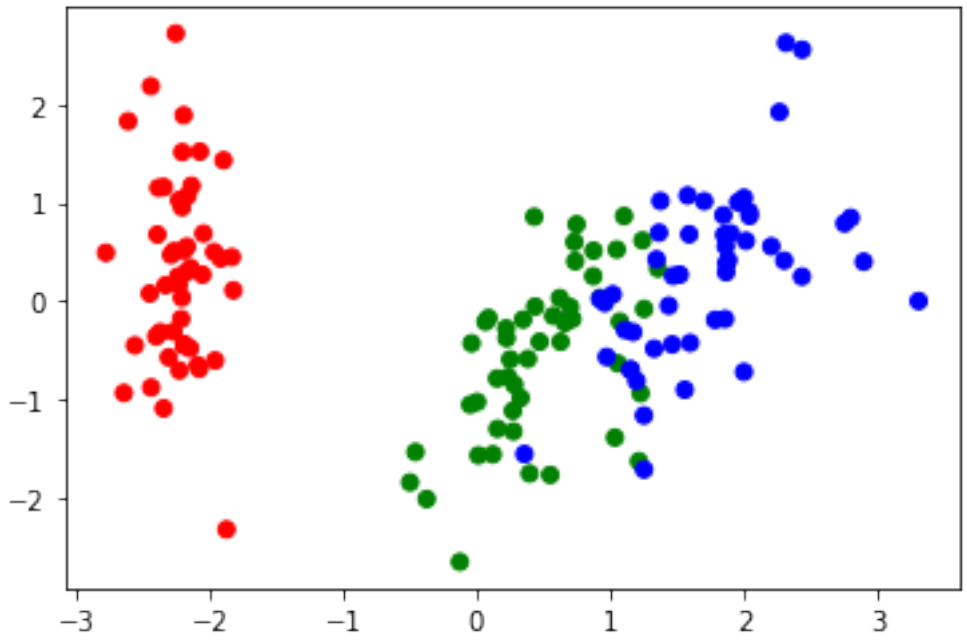
```
147              1.520845              0.266795
148              1.376391              1.016362
149              0.959299             -0.022284

[150 rows x 2 columns]
```

[12]:
```
targets = ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']
colors = ['r', 'g', 'b']
map_colors = dict(zip(targets, colors)) # {'Iris-setosa': 'r',␣
 ↪'Iris-versicolor': 'g', 'Iris-virginica': 'b'}
map_colors = y.apply(lambda x: map_colors[x])
```

[13]:
```
plt.scatter(principalDf['principal component 1'], principalDf['principal␣
 ↪component 2'], c=map_colors)
```

[13]:
```
<matplotlib.collections.PathCollection at 0x7ff8ab96f4c0>
```



[14]:
```
pca.explained_variance_ratio_
```

[14]:
```
array([0.72770452, 0.23030523])
```

[15]:
```
pca.components_
```

[15]:
```
array([[ 0.52237162, -0.26335492,  0.58125401,  0.56561105],
       [ 0.37231836,  0.92555649,  0.02109478,  0.06541577]])
```

[16]:
```
import numpy as np
np.cumsum(pca.explained_variance_ratio_)
# Both PC explain 95% of the variance. This match the goods results we can see␣
 ↪in the plot
```

[16]:
```
array([0.72770452, 0.95800975])
```

[ ]:
```

```