4. Principal Component Analysis (PCA)

November 15, 2022

1 PRINCIPAL COMPONENT ANALYSIS (PCA)

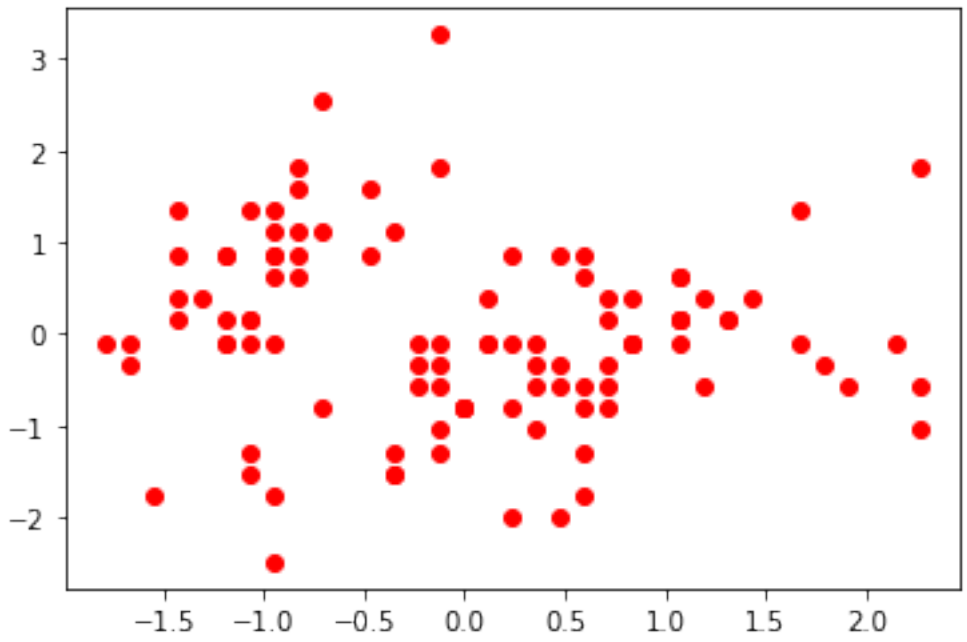
Principal component analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

[2]:
```
import numpy as np
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
```

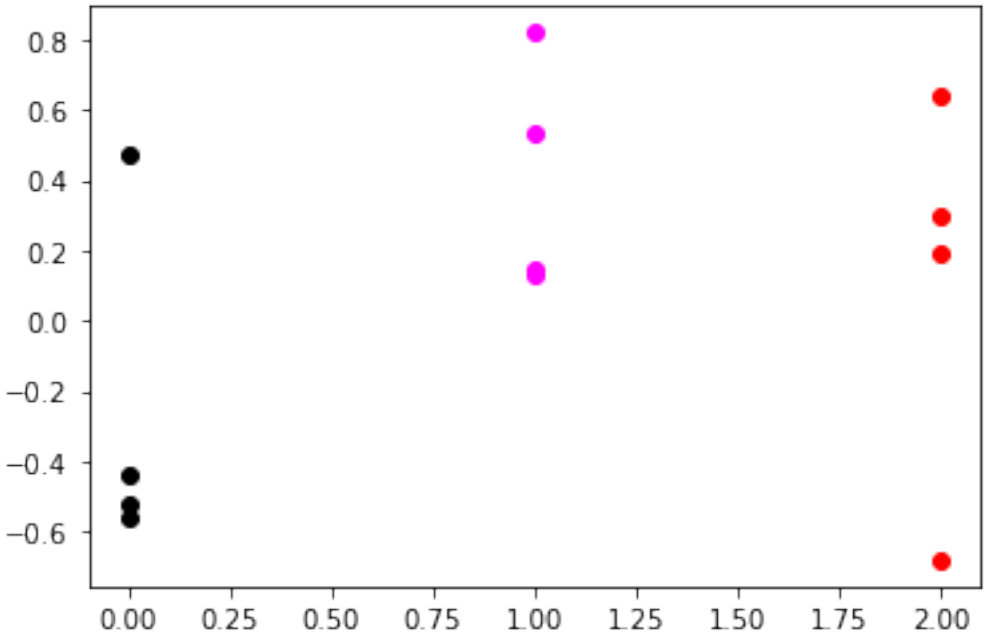
[3]:
```
names = ['sepal-length','sepal-width','petal-length','petal-width','Class']

# Step1: Load Iris dataset
iris = load_iris()

# Step2: Divide dependent and independent variables
X,y = iris.data,iris.target

# Step3: Split train data and test data
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.
 ↪3,random_state=1)

#print(X)

sc = StandardScaler()

X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

print(X_train.shape)
X_1 = X_train

pca = PCA(n_components=3)
```

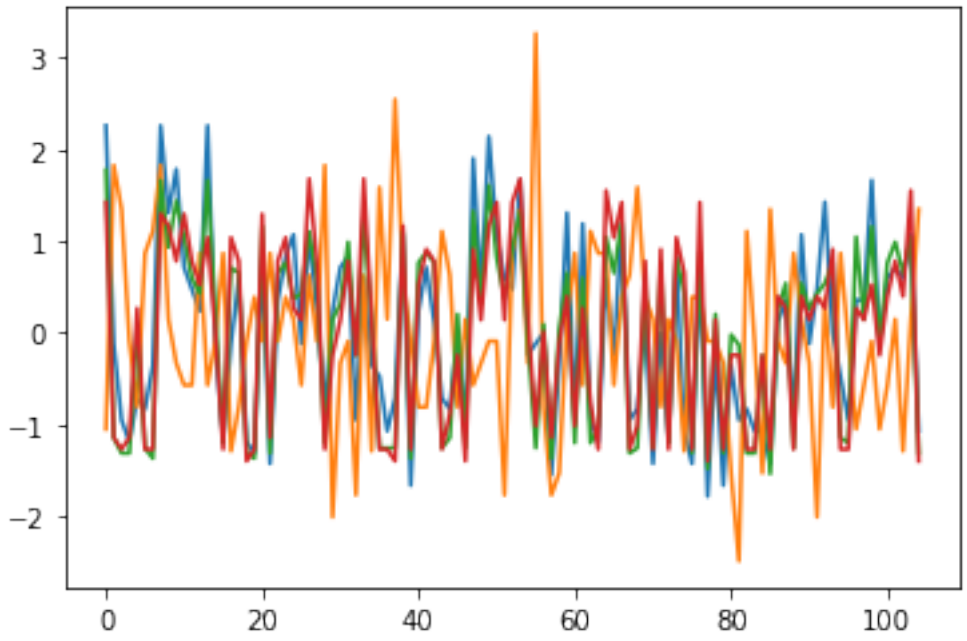
```
X_train = pca.fit_transform(X_train)
X_test = pca.fit_transform(X_test)
print(X_train.shape)

comps = pca.components_

print(type(comps))
plt.scatter(X_1[:,0],X_1[:,1],color='red')
plt.show()

c = ['black','magenta','red','yellow']
for i in range(0,len(comps)):
    print(i,comps[i])
    for j in range(0,len(comps[i])):
        plt.scatter(i,comps[i][j],color=c[i])

plt.show()
plt.plot(X_1)
plt.show()
plt.plot(X_train)
plt.show()
print("covariance is :")
pca.get_covariance()
```

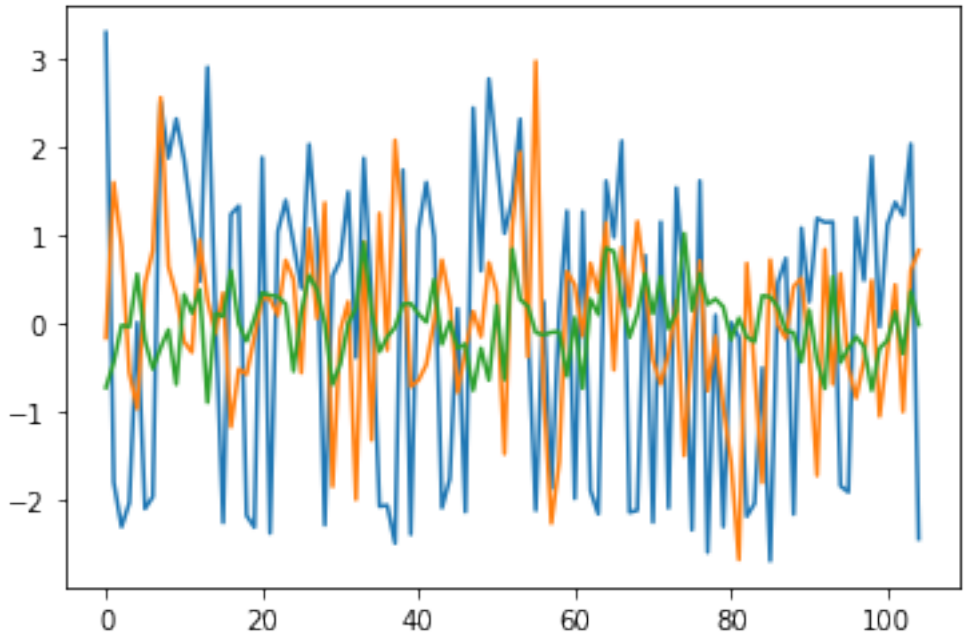
```
(105, 4)
(105, 3)
<class 'numpy.ndarray'>
```



```
0 [-0.43958974  0.46988855 -0.56005943 -0.52182281]
1 [0.53533601 0.82125376 0.13276696 0.14604985]
2 [-0.68182032  0.29986253  0.1911068   0.63928218]
```








```
    covariance is :

[3]: array([[ 0.87923928, -0.19033383,  0.76238719,  0.70121257],
            [-0.19033383,  1.2985837 , -0.64622045, -0.5758761 ],
            [ 0.76238719, -0.64622045,  0.942841  ,  0.86182214],
            [ 0.70121257, -0.5758761 ,  0.86182214,  0.8521541 ]])
```