

## Practical 3

### AIM:- Study and implement PCA in python.

#### Principal Component Analysis (PCA)

PCA (Principal Component Analysis) is a dimensionality reduction technique and helps us to reduce the number of features in a dataset while keeping the most important information. It changes complex datasets by transforming correlated features into a smaller set of uncorrelated components.

It helps us to remove redundancy, improve computational efficiency and make data easier to visualize and analyze.

#### Why PCA is used?

- To reduce number of features
- To remove redundancy (correlation)
- To reduce noise
- To improve training speed
- To visualize high-dimensional data
- To avoid overfitting

#### Working of PCA (steps)

1. Standardize the data
2. Compute covariance matrix
3. Find eigenvalues and eigenvectors
4. Sort eigenvectors by eigenvalues
5. Select top k components
6. Project data onto new axes

Example

```
import numpy as np
```

```
data = np.array([  
    [170, 65],  
    [180, 80],  
    [160, 55],
```

```
[175, 70],  
[165, 60]  
])  
print("Original Data:\n", data)  
print("-" * 40)  
m_val = np.mean(data, axis=0)  
X_m = data - m_val  
print("mean centered data:\n", X_m)  
print("-" * 40)  
cov = np.cov(X_m, rowvar=False)  
print("covariance Matrix:\n", cov)  
print("-" * 40)  
Evalues, Evecs = np.linalg.eigh(cov)  
print("eigenvalues:\n", Evalues)  
print("eigenvectors:\n", Evecs)  
print("-" * 40)  
sindex = np.argsort(Evalues)[::-1]  
seigenvectors = Evecs[:, sindex]  
print("sorted eigenvalues:\n", seigenvectors)  
print("principal components \n", seigenvectors)  
print("-" * 40)  
pca_data = np.dot(X_m, seigenvectors)  
print("Final PCA Transformed Data:\n", pca_data)  
print("-" * 40)
```

ouput:-

```
...     Original Data:  
...     [[170  65]  
...      [180  80]  
...      [160  55]  
...      [175  70]  
...      [165  60]]  
-----  
Mean Centered Data:  
[[ 0. -1.]  
[ 10. 14.]  
[-10. -11.]  
[ 5.  4.]  
[ -5. -6.]]  
-----  
Covariance Matrix:  
[[62.5 75. ]  
[75. 92.5]]  
-----  
Eigenvalues:  
[ 1.0147073 153.9852927]  
Eigenvectors:  
[[ -0.77334214  0.63398891]  
[ 0.63398891  0.77334214]]  
-----  
Sorted Eigenvalues:  
[153.9852927  1.0147073]  
Principal Components (Sorted Eigenvectors):  
[[ 0.63398891 -0.77334214]  
[ 0.77334214  0.63398891]]  
-----  
Final PCA Transformed Data:  
[[ -0.77334214 -0.63398891]  
[ 17.16667903  1.14242327]  
[ -14.84665261  0.75954345]  
[  6.26331309 -1.33075508]]
```