DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment No- 3.2

Student Name: Yash Kumar   **UID: 20BCS9256**
Branch: CSE   **Section/Group: 616-B**
Semester: 5   **Date of Performance: 09-11-22**
Subject Name: Machine Learning   **Subject Code: 20CSP -317**

1. **Aim/Overview of the practical:** Implementing Principal Component Analysis algorithm on any dataset and analyse the accuracy.

2. **Task to be done/ Which logistics used:** Analysing accuracy by implementing the Principal Component Analysis algorithm on any dataset.

3. **Steps of experiment/Code:**

   1. **Importing libraries such as numpy and implementing the function PCA:**

```python
import numpy as np

def PCA(X, num_components):

    #Step-1

    X_meaned = X - np.mean(X, axis = 0)

    #Step-2

    cov_mat = np.cov(X_meaned,  rowvar = False)

    #Step-3

    eigen_values, eigen_vectors = np.linalg.eigh(cov_mat)

    #Step-4

    sorted_index = np.argsort(eigen_values)[::-1]
    sorted_eigenvalue = eigen_values[sorted_index]
    sorted_eigenvectors = eigen_vectors[:,sorted_index]

    #step-5

    eigenvector_subset = sorted_eigenvectors[:, 0:num_components]

    #step-6

    X_reduced = np.dot(eigenvector_subset.transpose(), X_meaned.transpose()).transpose()

    return X_reduced
```

2. **Importing the libraries such as pandas and reading the Iris dataset and implementing Principal Component Analysis:**

```python
import pandas as pd

url= "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
data = pd.read_csv(url, names = ['sepal length', 'sepal width', 'petal length', 'petal width', 'target'])

# prepare the data

x= data.iloc[:, 0:4]

# prepare the target

target = data.iloc[:,4]

# Applying it to a PCA function

mat_reduced= PCA(x, 2)

# Creating a Pandas Dataframe of reduced dataset

principal_df = pd.DataFrame(mat_reduced, columns = ['PC1', 'PC2'])

# Concat it with target variable to create a omplete dataset

principal_df = pd.concat([principal_df, pd.DataFrame(target)], axis =1)
```
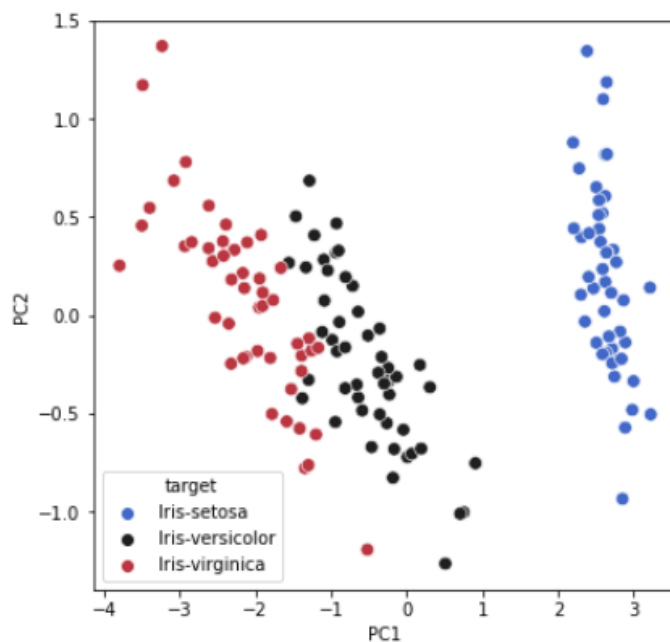
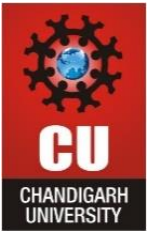**3. Scatter plot using libraries such as seaborn and matplotlib:**

```python
import seaborn as sns

import matplotlib.pyplot as plt

plt.figure(figsize =(6,6))
sns.scatterplot(data = principal_df, x = 'PC1', y = 'PC2' , hue  = 'target', s= 60, palette = 'icefire' )
```

```
<AxesSubplot:xlabel='PC1', ylabel='PC2'>
```

Learning Outcomes (What I have learnt):

1. I have learnt about implementing PCA algorithm on any dataset.

2. I have learnt about assigning few features to one variable and rest to other.

3. I have learnt about various libraries which are supported by python such as sklearn, matplotlib.

4. I have learnt about the various functions provided by various libraries.

5. I have understood the experiment very well.

Evaluation Grid:

	Parameters	Marks Obtained	Maximum Marks
1.	Student Performance (Conduct of experiment) objectives/Outcomes.		12
2.	Viva Voce		10
3.	Submission of Work Sheet (Record)		8
	Total		30