DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

WORKSHEET – 9

Student Name: Ruchika Raj **UID:** 20BCS9285

Branch: BE-CSE **Section/Group:** 615 B

Semester: 5TH **Date :**05/11/2022

Subject Name: Machine Learning Lab

Subject Code: 20CSP-317

Aim: Implement principle component analysis

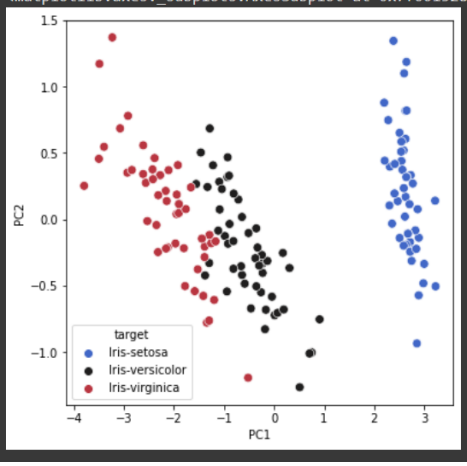
Software used : Google collab

```
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
```

```
import pandas as pd
#Get the IRIS dataset
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 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 complete Dataset
principal_df = pd.concat([principal_df, pd.DataFrame(target)], axis = 1)
```

```
import seaborn as sb
import matplotlib.pyplot as plt

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

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f001328dfd0>
```



Learning Outcomes:

1. In this, we learned about Principal Component analysis.
2. Learned about Principal axis method.