**CHANDIGARH UNIVERSITY**

**UNIVERSITY INSTITUTE OF ENGINEERING**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**



|  |  |
| --- | --- |
| **Submitted By: Kanishk Soni Submitted To: Er. Sudhanshu Sharma** | |
| **Subject Name** | **Machine Learning Lab** |
| **Subject Code** | **20CSP-317** |
| **Branch** | **BE-CSE** |
| **Semester** | **5th** |

**EXPERIMENT-9**

**Student’s Name: Kanishk Soni UID: 20BCS9398**

**Section/Group: 707\_WM\_B Subject Code: 20CSP-317**

**Subject Name: ML Lab Date of performance:9/11/2022**

**Branch: BE CSE Semester:5th**

**Aim:** Implement PCA.

**Objective:** To do PCA on data set.

**Software/Hardware Requirements:** Windows 7 & above version.

**Tools to be used:**

1. Anaconda Jupyter Notebook,
2. numpy, matplotlib, seaborn.

**Introduction to PCA:**

Principal Component Analysis is basically a statistical procedure to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables.

Each of the principal components is chosen in such a way so that it would describe most of them still available variance and all these principal components are orthogonal to each other. In all principal components first principal component has a maximum variance.

**Code:**

--

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

data = pd.read\_csv("iris.csv", 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')

**Output:**

