**Experiment 9**

**Student Name:** Sahil Kaundal **UID:** 21BCS8197

**Branch:** BE CSE (Lateral Entry) **Section/Group:** 616/A

**Semester:** 5th **Date of Performance:** 09/11/2022

**Subject Name:** ML Lab **Subject Code:** 20CSP-317

1. **Aim/Overview of the practical:**

Implement Principle Component Analysis.

1. **Task To Be Done:**

Implement Principle Component Analysis.

**3. Apparatus / Simulator Used:**

1. Windows 7 or above.
2. Google Collab.

**Objectives of PCA:**

♣ It is basically a non-dependent procedure in which it reduces attribute space from a large number of variables to a smaller number of factors.

♣ PCA is basically a dimension reduction process but there is no guarantee that the dimension is interpretable.

♣ Main task in this PCA is to select a subset of variables from a larger set, based on which original variables have the highest correlation with the principal amount. Principal Axis Method: PCA basically search a linear combination of variables so that we can extract maximum variance from the variables. Once this process completes it removes it and search for another linear combination which gives an explanation about the maximum proportion of remaining variance which basically leads to orthogonal factors. In this method, we analyze total variance.

**Principal Component Analyis** is basically a statistical procedure to convert a set of observation 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 the still available variance and all these principal components are orthogonal to each other. In all principal components first principal component has maximum variance.

Uses of PCA:

♣ It is used to find inter-relation between variables in the data.

♣ It is used to interpret and visualize data.

♣ As number of variables are decreasing it makes further analysis simpler.

♣ It‟s often used to visualize genetic distance and relatedness between populations. These are basically performed on square symmetric matrix. It can be a pure sum of squares and cross products matrix or Covariance matrix or Correlation matrix. A correlation matrix is used if the individual variance differs much.

**4. Program / Commands:**

#Sahil Kaundal

#21BCS8197

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)

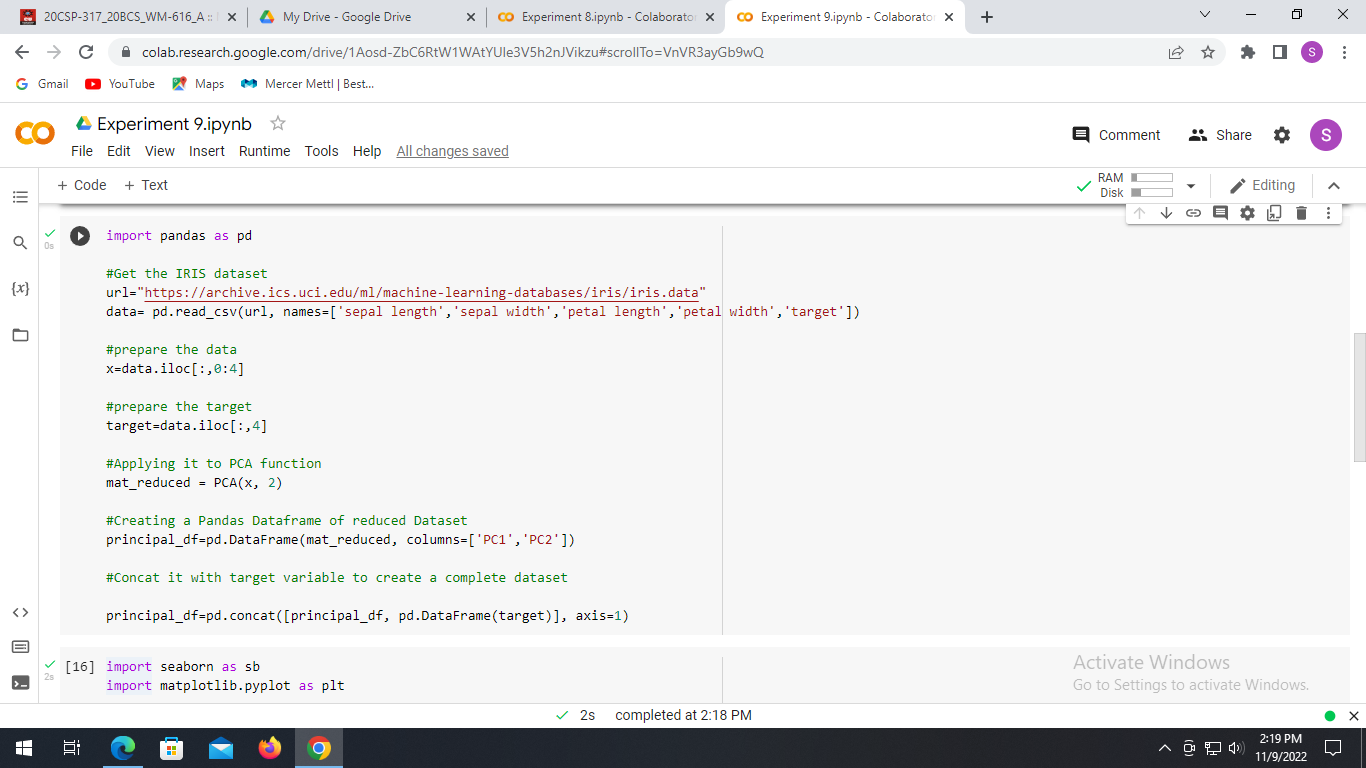
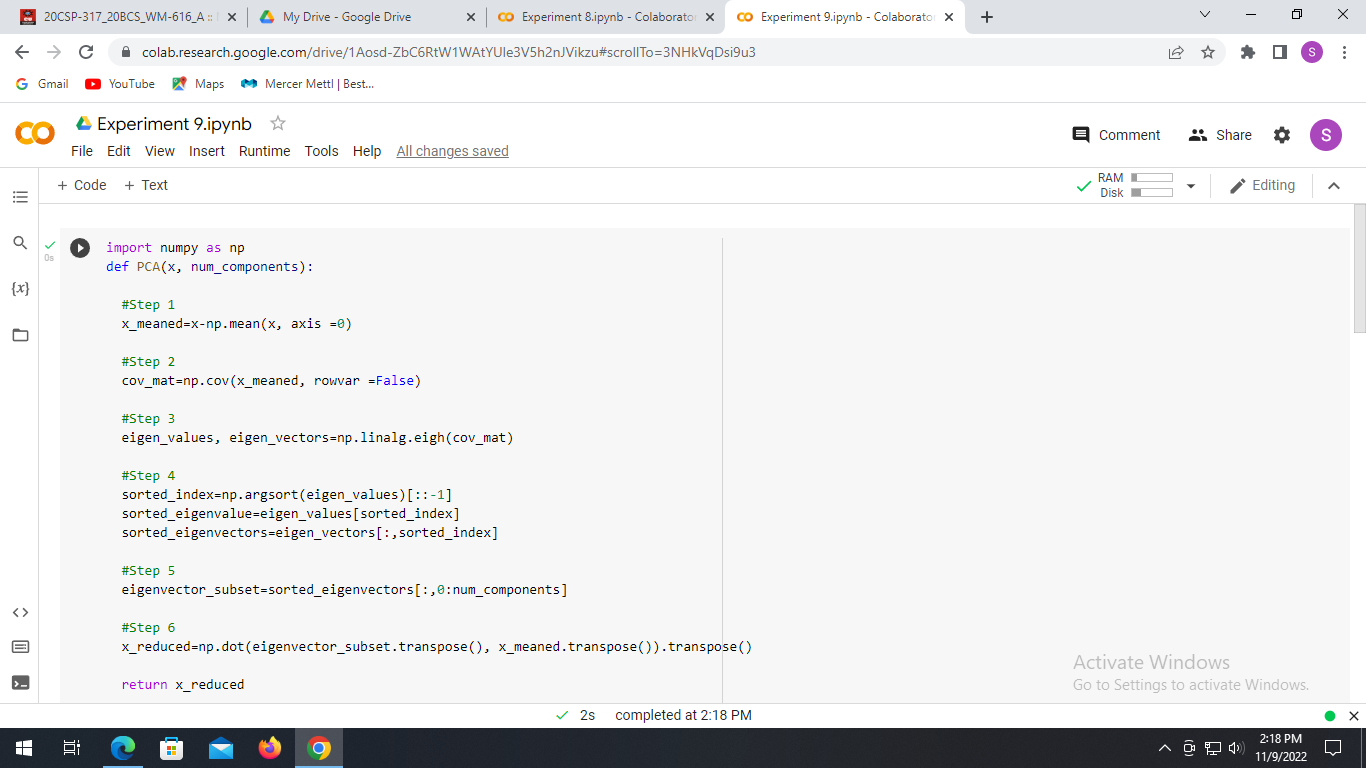
import seaborn as sb

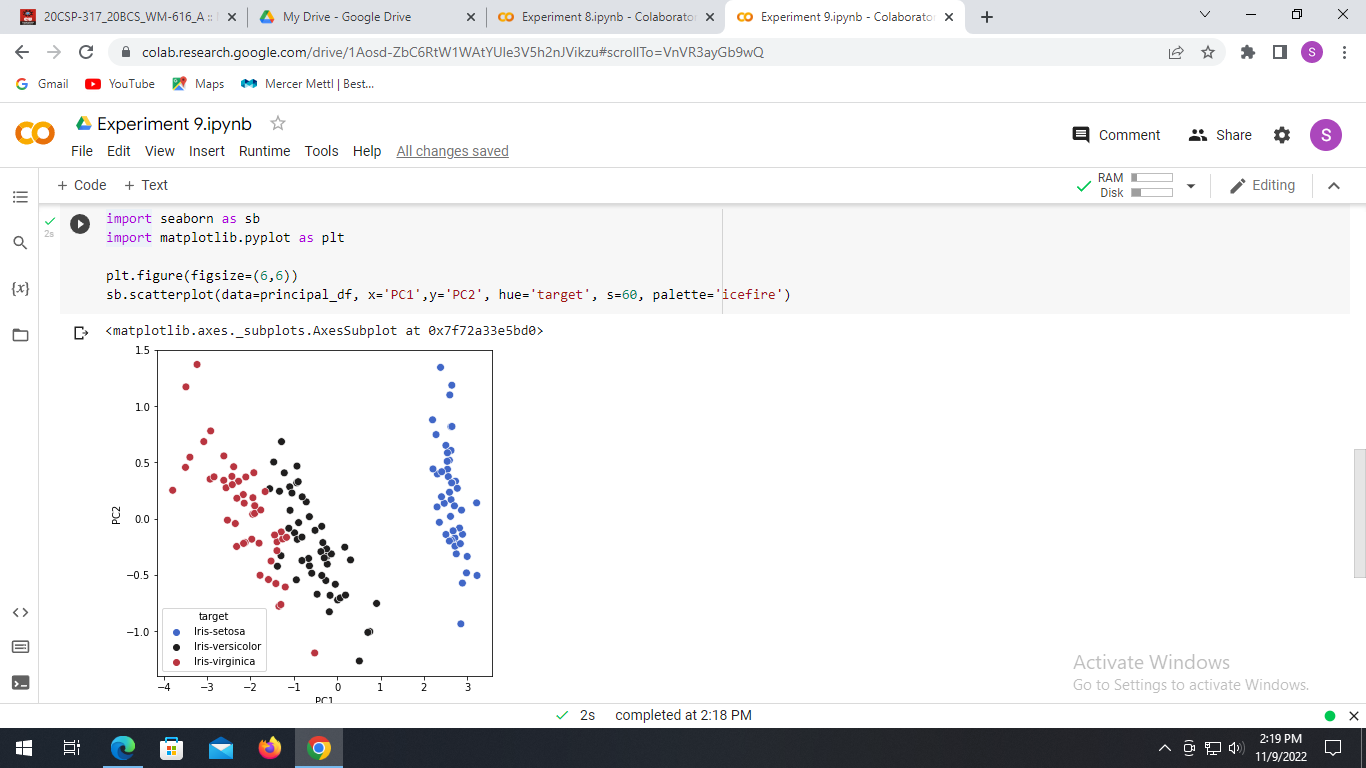
import matplotlib.pyplot as plt

plt.figure(figsize = (6,6))

sb.scatterplot(data = principal\_df , x = 'PC1',y = 'PC2' , hue = 'target' , s = 60 , palette= 'icefire'

1. **Result/Output/Writing Summary:**





**Learning outcomes (What I have learnt):**

* Understood the concept of PCA.
* Learnt how to Covariance Matrix.
* Learnt the separation of eigen value and eigen vectors from CM.
* Plot the graph using seaborn and matplotlib.

**Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Parameters | Marks Obtained | Maximum Marks |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |
|  |  |  |  |