Homework 4

Name: Mohamed Khaled Ahmed Atyaa

Group No.: 3 Alex

Principal Component Analysis

```
import numpy as np
import pandas as pd
import pprint
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
import matplotlib.pylab as pl
iris = load_iris()
iris_df = pd.DataFrame(iris.data,columns=[iris.feature_names])
X = iris.data
from sklearn.preprocessing import StandardScaler
X_std = StandardScaler().fit_transform(X)
X_covariance_matrix = np.cov(X_std.T)
eig_vals, eig_vecs = np.linalg.eig(X_covariance_matrix)
matrix_w = np.hstack((eig_vecs[:,0].reshape(4,1),
                      eig_vecs[:, 1].reshape(4,1)))
Y = X_std.dot(matrix_w)

pl.figure()
target_names = iris.target_names
y = iris.target
for c, i, target_name in zip("rgb", [0, 1, 2], target_names):
    pl.scatter(Y[y==i,0], Y[y==i,1], c=c, label=target_name)
pl.xlabel('Principal Component 1')
pl.ylabel('Principal Component 2')
pl.legend()
pl.title('PCA of IRIS dataset')
pl.show()
```