

**Assignment (1): PCA Algorithm Implementation**

***Course Name: Neural Networks***

***Course Code: CSE440***

***Submitted to:***

**Dr. Mahmoud Khalil**

**Eng. Amr Yassin**

***Faculty of Engineering, Computer Engineering and Software Systems***

**Student Information**

|  |
| --- |
| **Student Name:** |
| **Ahmed Mamdouh Mohammed** |
|  |

|  |
| --- |
| **Student ID:** |
| **16P6020** |
|  |

**Table of Contents**

**01: Introduction.…………………………………………………….** Page 3

**02: Code Walkthrough.………………………………….,……...** Page 4

**03: Input/Output Scenario.……………………………....…...** Page 7

1. **INTRODUCTION:**

Principal Component Analysis (PCA) is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. In this report, I will discuss how I implemented PCA using Python with NumPy and Matplotlib. I implemented everything from scratch and compared my results to the results of the PCA method provided by the SciKit-Learn library by comparing the produced plots from my output to the produced plots of the SciKit-Learn’s PCA method output. This was done to check if my results are valid.

The input feature vectors have a shape of (N x D) where N is the number of data points and D is the number of dimensions. The output shall be (N x K) where K is the number of dimensions after reduction.

I assumed that the dimensions after reduction (K) will always be equal to D. To be able to plot the output data on a 2D scatter plot, I only took into consideration the two principle components which have the highest variability. I also assumed that the dimension of the input feature vectors (D) can be any value, but I will only use the first two dimensions of features to plot the input data on a 2D scatter plot, this is so that I could be able to have a sense of what the input data may look like. To test my implementation of PCA, I used the Iris dataset which is provided through the SciKit-Learn library.

In the next section, I will explain the main parts of the code without going into too much details as the code itself has a lot of comments to explain all the nitty-gritty details.