# Week 2 Problem 7

Abhishek Sriram, Dahlia Brown

## Definitions

**Definition 0.1.** ***Principal Component Analysis*** *(PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance.*

**Definition 0.2.** *A **Covariance Matrix** is a type of matrix that is used to represent the covariance values between pairs of elements given in a random vector. The covariance matrix can also be referred to as the variance covariance matrix. This is because the variance of each element is represented along the main diagonal of the matrix. It has the structure below:*

$$\begin{pmatrix} var(X) & cov(X,Y) \\ cov(Y,X) & var(Y) \end{pmatrix}$$

**Definition 0.3.** *An **Eigenvector** or characteristic vector of a linear transformation is a nonzero vector that changes at most by a scalar factor when that linear transformation is applied to it. The corresponding **Eigen value**, often denoted by $\lambda$ , is the factor by which the Eigen vector is scaled.*

## Theory

We know that by definition, **Principal Component Analysis** (PCA) is a way of analyzing multidimensional data. What this means is that we can visualize our weighted data along with its variance in order to eliminate any correlation without data loss or at least minimal data loss.

## Procedure

We begin with this sample set:

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| -1 | 1 | 0 | 2 |
| -2 | 3 | 1 | 0.5 |

In the above sample set, we will be using the steps of Principal Component analysis to analyze the data in order to obtain a covariance matrix. Moreover, from the covariance matrix we will also attain eigenvalues and eigenvectors such that these values represent the orthogonal directions of maximized variance from our given sample data.

### Step 1: Covariance Matrix

Use the variance and covariance formulas to get each entry in the covariance matrix.

$var(x) = \sum \frac{(X-\bar{X})^2}{N-1}$

$cov(X, Y) = \sum \frac{(X-\bar{X})(Y-\bar{Y})}{N-1}$

Once we arrange these values in the structure of the covariance matrix, what then need to achieve is obtaining the eigen values and eigen vectors. This is because the eigen values actually determine the amount of variance it presents in the sample data set.

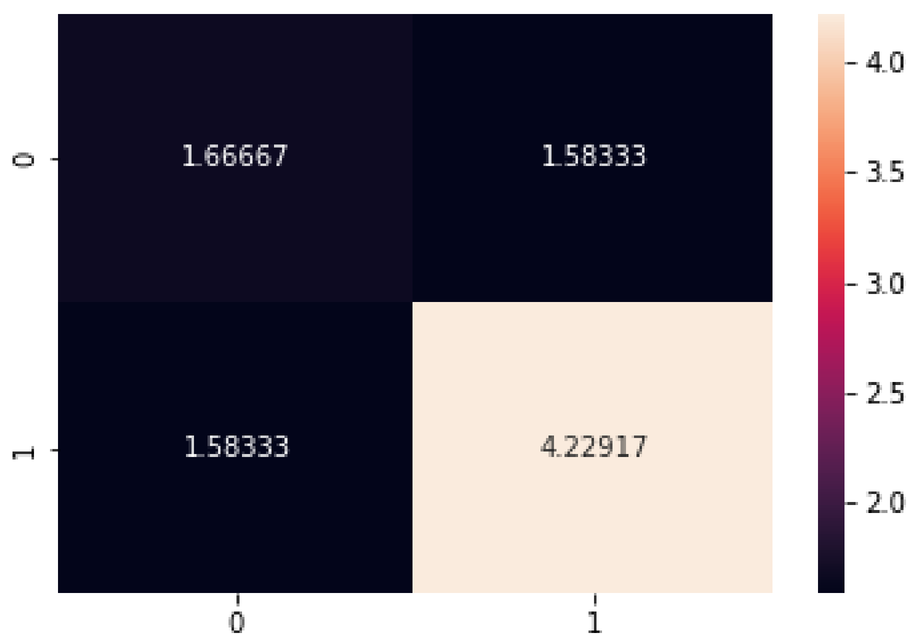
### Step 2: Perform PCA

In this section, we tried our best using resources on how to perform PCA with python as well as graphing. Since we are both novices when it comes to coding, I feel that some of the code we used in the colab may not accurately depict our PCA. As we can see the eigen values we obtained from the covariance matrix are: 0.91111869, 4.98471465. Now from these eigen values we are able to calculate the percentage principal component, where the first eigen value accounts for about 15.45% and the second eigen value accounts for about 84.55%. What this means is that the larger the percentage, the

https://github.com/Milkshake10401/Math_140-weekly-projects/blob/6b627a9e8bf8b8cce03d00c7938cf5159c78c03
Math140_Week_2.ipynb

### Step 3: Heat-Map

In the Heat map below, we can see that the covariance matrix is represented accurately. A Heat Map is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors.




## References

[1] "Principal component analysis: a review and recent developments"
https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202#:~:text=Principal%

20component%20analysis%20(PCA)%20is,variables%20that%20successively%20maximize%20variance

[2] ”How to Calculate Principal Component Analysis (PCA) from Scratch in Python”
https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/

[3] ”sklearn.decomposition.PCA”
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

[4] ”How to draw a screen plot in python?”
https://stats.stackexchange.com/questions/12819/how-to-draw-a-scree-plot-in-python

[5] ”Principal Component Analysis from Scratch in Python”
https://www.askpython.com/python/examples/principal-component-analysis