Lab-8

PCA

**Download** the iris data set from given link:

(i) Dataset without class label

<https://drive.google.com/open?id=1DTVhMzFLcDiudohym6P125UQn2CNlo2n>

(ii) Dataset with class label

<https://drive.google.com/open?id=1KO8t9bos_sDAxEVmZdoFc5eyctWaALgG>

**Description of dataset:**

The iris dataset contains measurements for 150 iris flowers from three different species.

The three classes in the Iris dataset are:

1. Iris-setosa (n=50)
2. Iris-versicolor (n=50)
3. Iris-virginica (n=50)

And the four features of in Iris dataset are:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

**Note:** for today’s lab, use (i) Dataset without class label.

**Tasks:**

1. Calculate covariance matrix of data

2. Calculate correlation matrix of data

3. Standardize the data (mean=0 and suing SD)

4. Again calculate covariance matrix of data

5. Observer results of step 2 and 4 is same? Think…

6. Calculate the eigen vector and print it.

7. Calculate the eigen value and print it.

8. Extract top two pca, i.e against highest eigen value

9. Project the data into new feature space.

10. Now, You may use package (PCA) for the same and see the results which are obtained by previous steps.