Lab-10

PCA-part-2

**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. Choose KNN classifier and train on all (ii) Dataset with class label, observer the performance (training and testing) against various performance measures (efficiency, ROC, precision recall etc)

2. Create principle components using (i) Dataset without class label.

3. Choose KNN classifier and train on all components (from step 2, need to include class labels), observer the performance (training and testing) against various performance measures (efficiency, ROC, precision recall etc)

4. Extract top first PCA (from step 2), i.e against highest eigen value, repeat step 3.

5. Extract top two PCA (from step 2), i.e against highest eigen value, repeat step 3.

6. Analyze/observe change in the KNN model performance (step 3, step4, step5). Represent it as table form.