LAB REPORT

NAME-DEVYANI GORKAR
ROLL NO.-B20ME027

X is the original matrix after converting the **dataset**.**xnew** is the matrix obtained after shifting the actual origin to the mean of the X.**covariance** is the covariance matrix of xnew transpose and it is of shape 745 x 745. Using the Linear algebra library of numpy we get **w** as an array containing the eigenvalues and **v** as an array consisting of eigenvectors with dimension 745 x 745.To sort the eigenvectors along the maximum variance of the data we set the eigenvalues in descending order and sort their corresponding eigenvectors.Finally **values** and **eigenv** are sorted eigenvalues and eigenvectors array.

- pca_5 is the array of first five eigenvectors having maximum variance.X_decomposed1 is the final array after applying PCA with no. of components=5.
- pca_10 is the array of first five eigenvectors having maximum variance.X_decomposed2 is the final array after applying PCA with no. of components=10.
- pca_50 is the array of first five eigenvectors having maximum variance.X_decomposed3 is the final array after applying PCA with no. of components=50.
- pca_100 is the array of first five eigenvectors having maximum variance. X_decomposed4 is the final array after applying PCA with no. of components=100.
- pca_300 is the array of first five eigenvectors having maximum variance. X_decomposed5 is the final array after applying PCA with no. of components=300.
- pca_700 is the array of first five eigenvectors having maximum variance. X_decomposed6 is the final array after applying PCA with no. of components=700.

For every sample image dimensionality reduction is done.In the first case dimensionality is reduced to 700. Images are plotted using original data and reduced data. Both data are subtracted and a new image is plotted.**recovered data** is the data obtained after inversing the pca components and getting back the original data from reduced data .**e1** is the RMS error.Similarly from each reduced dimension data is recovered using inverse pca and stored.For lower dimensions of the data we can see that image gets blurred and unclear.Error is also stored.From the error vs no.components plot we can see that the error minimizes as no. of components increases.This is because as the no. of components increases the information loss between original data and reduced data decreases.So to choose the proper number of pca components we should check how much data loss is bearable so that it does not affect the further model much.