Assignment PCA - IRM2016502

Problem:

To classify river or not river using satellite images

Data:

The satellite images of kolkata having 4 bands is used as the dataset.

Procedure:

- The images are first loaded into a numpy array with 4 rows, each image flattened in each row.
- Mean of the images is found and is subtracted from all of them.
- The covariance of this matrix after subtracting mean is found out.

```
[[ 7.39056118  6.37133815  8.85835449  2.05945495]
 [ 6.37133815  6.60222402  8.85052835  3.57685562]
 [ 8.85835449  8.85052835 15.75011491  3.96733057]
 [ 2.05945495  3.57685562  3.96733057 37.80738645]]
```

- The eigenvalues and eigenvectors are found out of this covariance matrix.
- Of these 2 top eigenvalues and their corresponding eigenvectors are chosen.

  Chosen eigenvectors

```
[[ 0.17494902  0.20630134  0.28671959  0.91903453]
 [-0.45452022 -0.41536811 -0.68303992  0.3928577 ]]
```

- Multiplied them with input to get eigen_images.
- During testing the pixel values from 4 bands are taken, normalised.
- Later multiplied with these principal components and are compared with all other pixels and the class with minimum distance is chosen to classify the pixel as river or not river.
- I've used euclidean distance to find the minimum distant point.