Requirements:

PIL Library, Numpy

KMEANS:

Tested on various images present in the folder. Optimal inputs being 5 clusters, threshold of 0.01 for "strips.jpg" and 4 clusters, threshold of 0.01 for "4-color.jpg".

Input format:

Expects number of clusters and error threshold. To visualize clusters, input the cluster number (follows zero indexing) when asked

The name of the input can be changed in the code. ("strips.jpg" is left as default)

PCA:

Tested on the two images present in the folder. ("pca.jpg" is left as default)

Fail cases of PCA:

Standard pca can fail when the data is non-linear because it tries to find a linear representation of data in lower dimension.

MLE:

MLE_1.py : Compare with estimator being same distribution

MLE_2.py : Compare with the different distribution(Eg: Gaussian as ground truth and all other types of distribution as estimator)

Number of samples being 1000

Input format: Enter the various parameters of a distribution when prompted. The following is same for both the codes

1 corresponds to Gaussian

2 corresponds to Laplacian

3 corresponds to Exponential

4 corresponds to Poisson

5 corresponds to Binomial

For the case of binomial, the groundtruth was generated from uniform distribution. Also, binomial case expects number of trials as an input.