CSL 7020 - Assignment 4
Principal Component Analysis

Vishakh.S (B16CS038)

---

1. DATASET

The peppers image shown below is used for demonstrating dimensionality reduction using principal component analysis (PCA). The image is of size 513 x 383 pixels.

2. APPROACH

Principal Component Analysis (PCA) is an unsupervised, non-parametric statistical technique used frequently for dimensionality reduction in machine learning. High dimensionality means that the data has a large number of features. The numero uno problem associated with high-dimensionality in the machine learning field is model overfitting.

Algorithm :

- If there are 3 input colour channels (R,G,B), separate them and perform the following for each channel.
- Subtract mean from each column vector.
- Compute covariance matrix as *sigma* = $X.X^T$
- Perform an eigenvalue decomposition of *sigma.*

- Suppose *sigma* has N eigenvalues and you want to reconstruct the image using k (<= N) eigenvalues. Then, pick the largest k eigenvalues and eigenvectors corresponding to the largest k eigenvalues and use it for reconstruction.
- Once this is done for each channel, combine the results.

3. EXPERIMENTS

a. SETUP

➔ Setup Python3 virtual environment.
➔ Install numpy and PIL.
➔ Download a suitable image for dimensionality reduction
➔ Implement the algorithm above.
➔ Execute it for various values of the number of principal components to be taken

b. RESULTS

The following images are obtained after reconstruction.

Reconstruction with 20 eigenvalues

Reconstruction with 50 eigenvalues

Reconstruction with 100 eigenvalues

Reconstruction with 150 eigenvalues

Reconstruction with 200 eigenvalues

Reconstruction with 350 eigenvalues

4. ANALYSIS

- As the number of principal components considered for reconstruction increases, the quality of the reconstructed image also increases.