Principal Component Analysis

MoMA Project

Details:

Chirag Bhuvaneshwara

2571703

Principal Component Analysis Intro:

Principal component analysis (PCA) uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of distinct principal components is equal to the smaller of the number of original variables or the number of observations minus one. This transformation is defined in such a way that the first principal component has the largest possible variance(that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set.

PCA and SVD:

The singular value decomposition method can be used to decompose a matrix of any order into three components: the matrix of left singular vectors, matrix of singular values and the matrix of right singular vectors. By performing matrix multiplication of these matrices with respect to their orders, we can reconstruct the original matrix.

SVD can be used to perform principal component analysis and 99% of the variance can be retained during reconstruction by using the following formula:

$$\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{m} S_{ii}} \geq 0.99$$

Where S_{ii} is the singular value in the i^{th} position along the diagonal of matrix of singular values

k is the number of principal components being considered

m is the total number of singular values

0.99 is the amount of variance

Usually, the amount of variance is set to 99% because that would make the reconstruction retain 99% of the unique information while still significantly reducing the size of the reconstructed matrix.

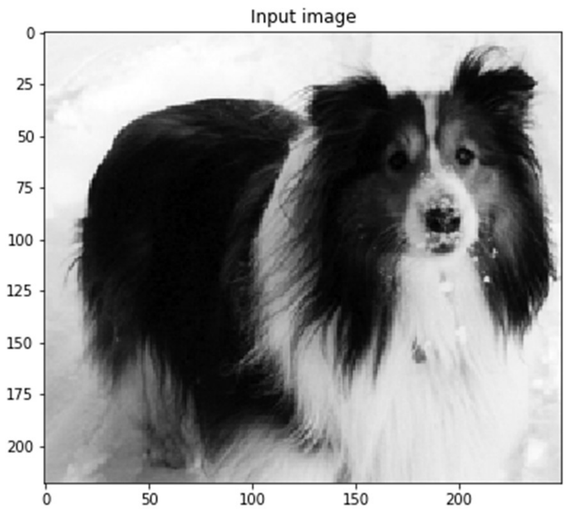
Instead of 0.99 the variance can be set to whatever value we wish, but that might not produce good results in most cases.

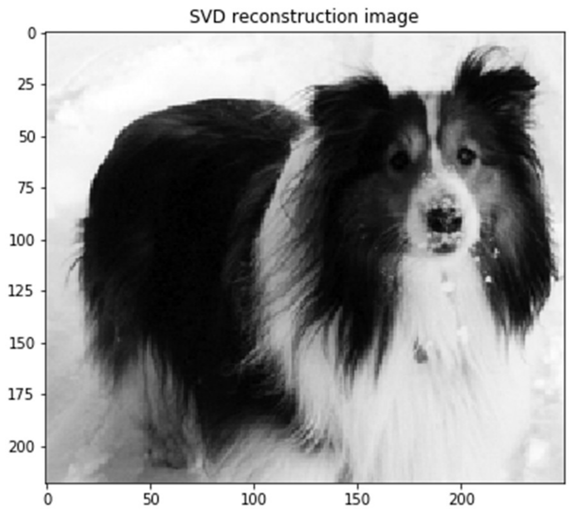
Steps to compute PCA

1. Centre the mean of the matrix
2. Centre the data by subtracting the mean from each element
3. Perform Singular Value Decomposition of the above matrix
4. From the matrix of singular values, use singular values to determine the number of principal components "k" required to retain the specified amount of variance
5. Update the matrix of singular values to contain only "k" number of singular values
6. Perform reconstruction with the updated matrix

Observations from the code

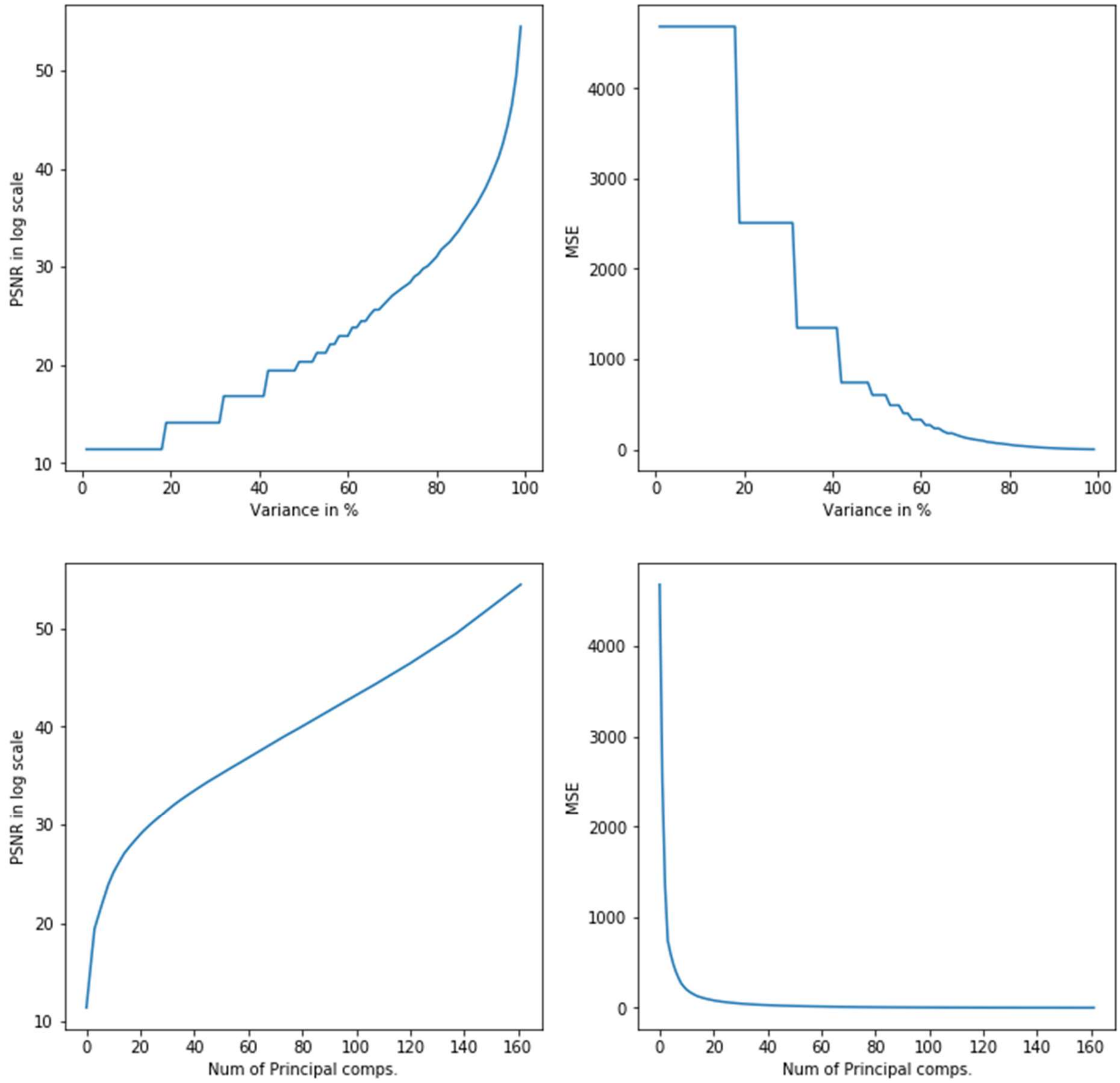
SVD Reconstruction





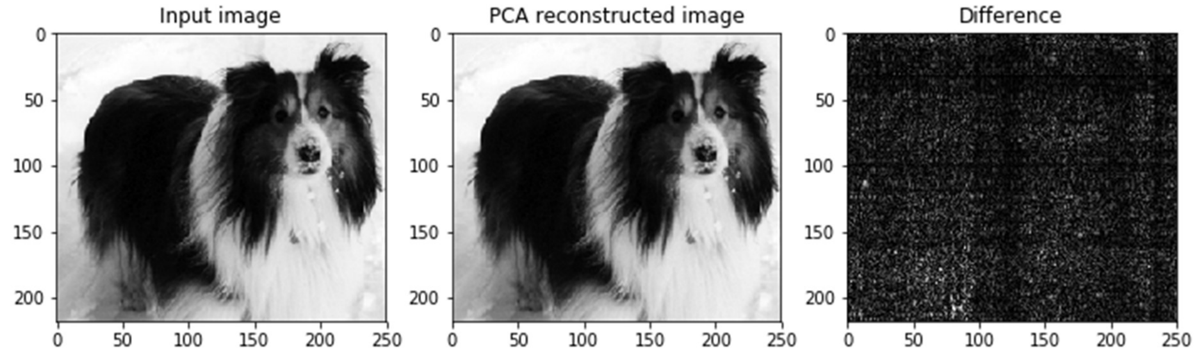
- Above, to the left, is the image of a dog which was fed as the input
- And to the right is the image of the dog obtained by multiplying the SVD components
- As we can see here, no data seems to be lost or not much noise seems to be added while performing reconstruction from the SVD components.

Principal Component Analysis



- The above graphs are obtained for the same image of the dog.
- From the above graphs, we can see that as the variance increases, PSNR also increases and MSE decreases.
- Also, the same result follows with the number of principal components because the variance and the number of principal components are directly proportional.

PCA Reconstruction



- The above reconstruction is obtained for a variance of 99%
- As can be seen, the PCA reconstruction image looks good but only some 50 odd number of principal components contain the actual information.
- The difference image shows how much difference exists between the input and the PCA reconstructed image.

Simialr results have been obtained for 5 other images which are presented in the ipython notebook along with the code.