Homework 5

Raka Mandal

April 18, 2016

1 2 Principle Component Analysis

1.1 Question 1

We have the following data points in 2d space (0; 0); (-1; 2); (-3; 6); (1;-2); (3;-6). The following is the plot of the data:-

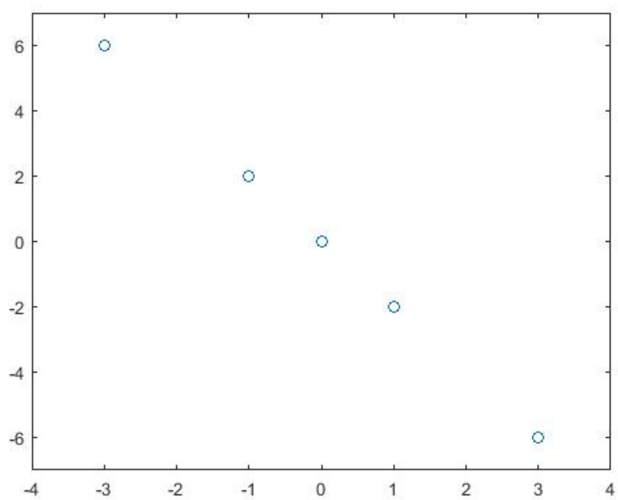

Figure 1: Scatter plot of five data points

The data poitns lie in 1d hyperplane (straight line). The first PC will be in the direction of the straight line $y + 2x = 0$.

The second PC will be zero as S_{2x2} will be singular.

The PC's are found as described below:-
Let $X_{obs} = [x_1, ...x_n]$ be the observed data matrix of size nxd, n being number of observa-

tions, d is dimension of the variable. Define $X = X_{obs} - \bar(X_{obs})$, The centered data matrix. Let $S = \frac{X'X}{n}$.

We want to maximize the variance of data projected onto the direction of the unit vector u_1 subject to $u_1^T u_1 = 1$. i.e maximize $u_1^T S u_1 + \lambda_1(u_1^T u_1 - 1)$

The first Principal component is λ_1 the largest eigenvalue of the matrix S. it's direction is given by the corresponding eigen vector u_1.

Then we want to project the data along the direction of u_2 orthogonal to u_1 and maximize the variance. i.e maximize $u_2^T S u_2 + \lambda_2(u_2^T u_2 - 1) + \phi u_2^T u_1$.

The second Principal component is λ_2 the second largest eigenvalue of the matrix S. it's direction is given by the corresponding eigen vector u_2 and $u_2 \perp u_1$

1.2 Question 2

We are doing PCA on handwritten digit images from the USPS dataset. The matrix A contains all the images of size 16 by 16. Each of the 3000 rows in A corresponds to the image of one handwrit- ten digit (between 0 and 9). We Apply Principal Component Analysis (PCA) to the data using p = 10; 50; 100; 200 principal components. The following pair of images show the original image and image reconstructed after PCA.

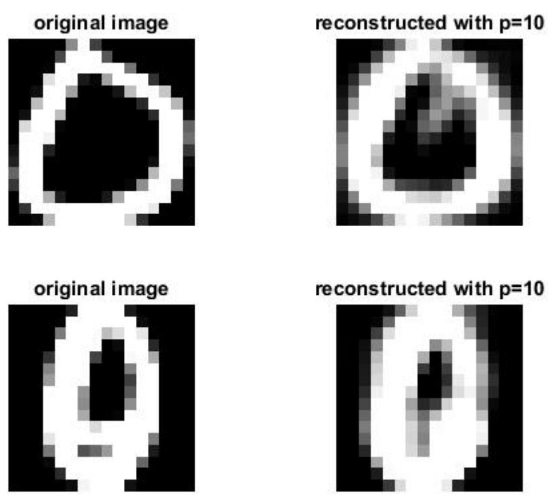

Figure 2: p = 10

We can clearly see the image reconstruction gets better with larger p.

The log of frobenius norm error $e = ||X - X_{pred}||_F$ is plotted for different values of p.It is steadily decreasing with p. The error rate becomes slower as the p gets large.

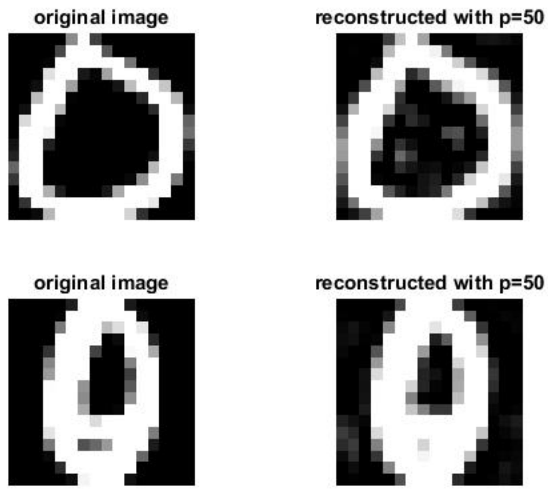

Figure 3: p = 50

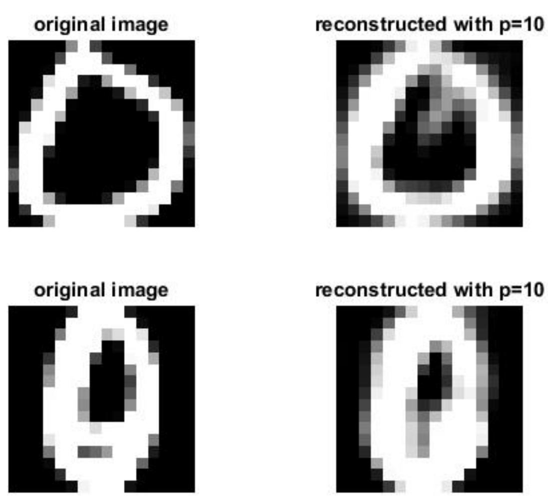

Figure 4: p = 100

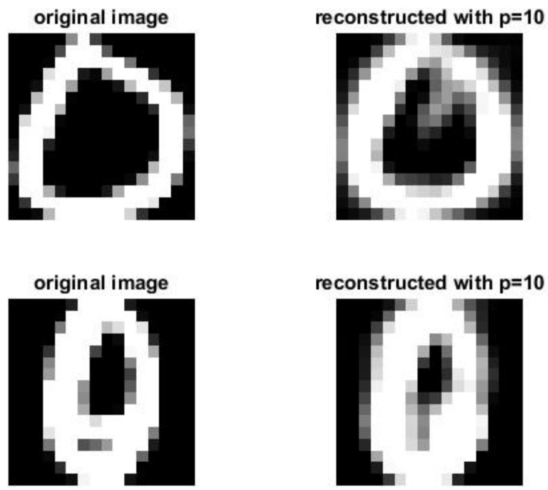

Figure 5: p = 200

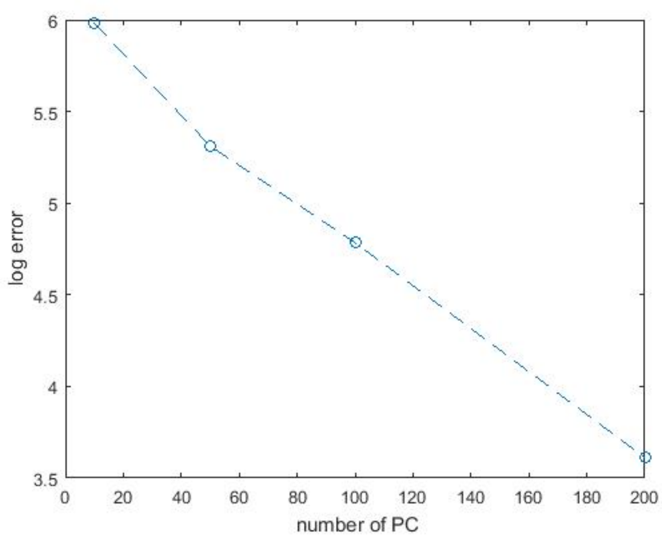

Figure 6: p = 200