

# Homework 5 Report

Peide Li

April 18, 2016

1. The plot of the five points is shown below:

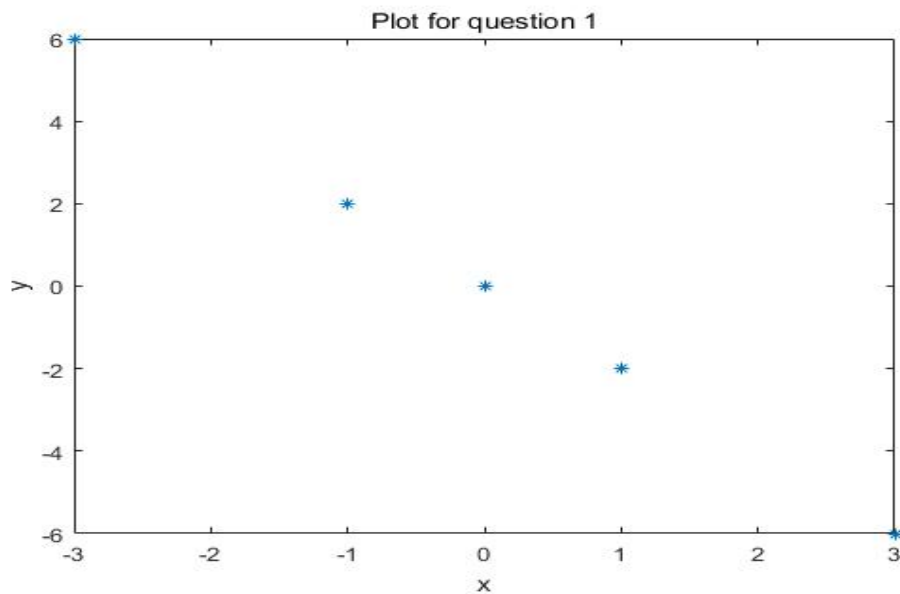


Figure 1: Figure 1

The plot shows that:

- 1 The first principle component is  $y = 0$ . Because all the data are symmetrically distributed between the line  $y = 0$  and the absolute value of the slope is larger than 1. So it looks more like a horizontal line than a vertical line.
- 2 The second principle component is  $x = 0$ . Because all the data are also symmetrically distributed between the line  $x = 0$ .

2. I applied Principle component analysis to the usps data set by doing eign-decomposition to the correlation coefficient matrix of the data set. In my code, I transpose the data set A so that each column would be a data of a imagine. I use  $p = 10, 50, 100, 200$  components to construct the feature spaces. And I reconstruct the first two imagines of the data set. The original first two imagines are shown here:



(a) The first imagine



(b) The second imagine

Figure 2: Original first two imagines

And the reconstructed imagines are shown below:



(a) 10 components



(b) 50 components



(c) 100 components



(d) 200 components

Figure 3: Reconstruction of the first imagine

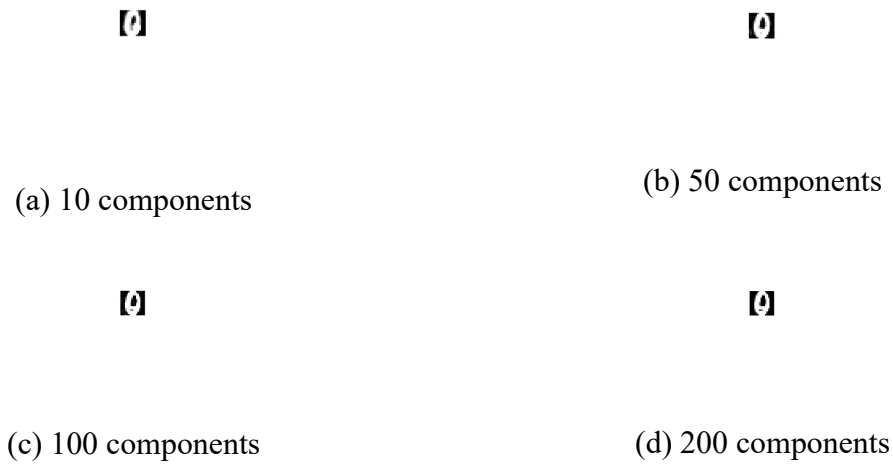


Figure 4: Reconstruction of the second image

And the total error for the reconstruction are:

p	10	50	100	200
Error	$2.0920 \times 10^4$	$1.0692 \times 10^4$	$6.2755 \times 10^3$	$1.7588 \times 10^3$

My code and all outcome are uploaded in my Github account: [https://github.com/Kira233767/CSE847-Homework\\_5.git](https://github.com/Kira233767/CSE847-Homework_5.git)