**Project1**

[Gongjie.cs@gmail.com](mailto:Gongjie.cs@gmail.com)

**Problem1**

1. The three data points is 。we can calculate the covariance matrix is . The eigenvector matrix is , and the corresponding eigenvalue are 0 and 2. So the principal component is .
2. We choose vector V = , thus the projected data is P = S \* V= . The mean value of this projected data is 0 , so the variance of the projected data is 2.
3. If we reconstruct the original data we’ll get (−1,−1),(0,0),(1,1) respectively, So the reconstruction error is 0.

**Problem2**

**PART1**

I write this program in matlab, the code of MyPCA , PCAProjection，PCAReconstruction can be found in attachments

**PART2**

**Task1,**

(1)I choose the number 3;



**Comments：**

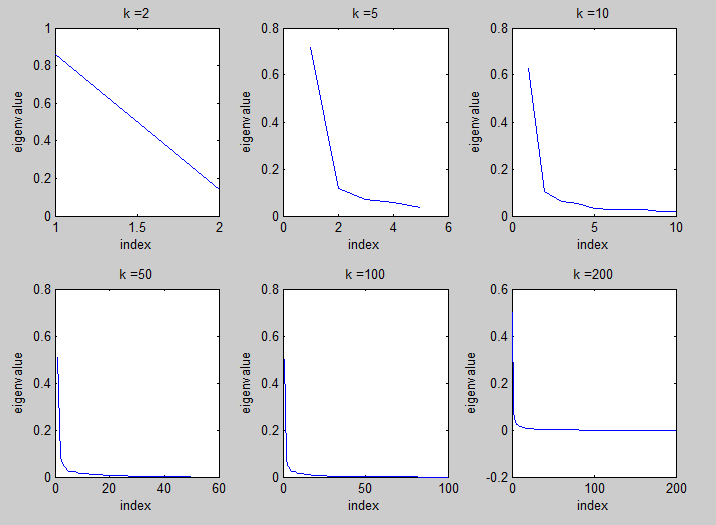
From this picture, we can see clearly that nearly every image that is showed here has the shape of 3. This to some extent shows that the eigenvectors that corresponds to the larger eigenvalues records the main shape of number ‘3’ while the latter ones only captures the marginal information.

**Task1.2,**

1.2.1 This is the eigenvalue and corresponding percentage in decreasing order:



**1.2.2 , 1.2.3** For k is 2,5,10,50,100,200 the corresponding figure is: Their percentages the others are very close to 0,and can be ignored. The sum of the k-largest eigenvalues is shown below.



**Task 1.3**

1. The reconstruction error is showed in this table bellows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| k | 2 | 5 | 10 | 50 | 100 |
| error | 71.4578 | 53.4893 | 37.4219 | 6.3125 | 0.9000 |

From this table, we can infer that the value of reconstruction error is growing slower and slower, until we choose all the eigenvectors. However, the reconstruction error varies not large enough and we could not get enough information from the 5 numbers. But when k = 100 we can see that the error is not so high and image is very clear , they can cover the main features of the digit, soI`ll choose 100 eigenvectors.

**Task 1.4**

K = 100

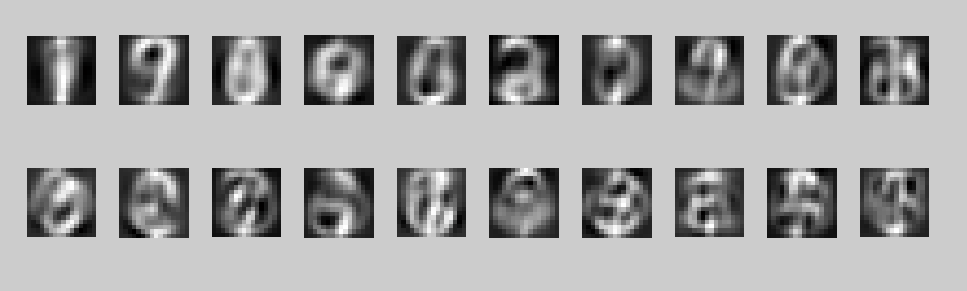


K = 50

From these pictures, we can see that the result is very good, which shows that PCA analysis is very useful for the case of one number class. We can recognize the exact number just by using only a few eigenvectors.

**Task 2.1**

1. The result is as follows, here I also list the top 20:

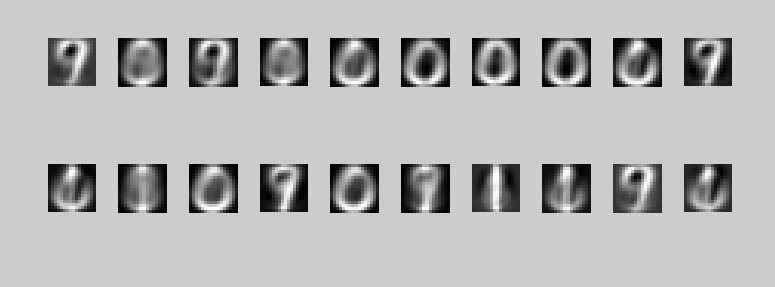


Comment:

These numbers are not as clear as the ones in case number 7, we could recognize the shape of several numbers reluctantly.

This is partly because the images of different numbers are so much different. Thus when we apply PCA, the eigenvectors are by no means capture the main features of each number. As a result, the images listed here are very vague and very difficult to recognize.

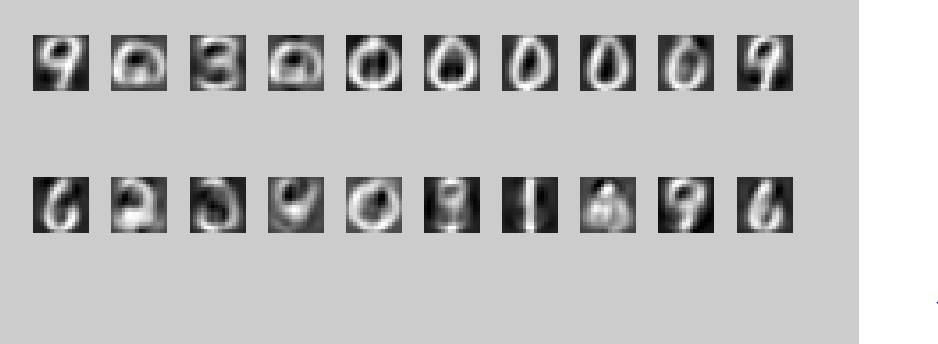
**Task 2.2**



K=2



K = 5



K = 10



K = 50



K = 100



K = 200 Error = 0.4304

Comment:

With k’s increasing, the images reconstructed better and better, and the error getting lower and lower. Compared with k=100 and k=200 , We will see a highly increasing of image quality , so I think k=200 is much better.