Homework 1

Name: Mukul Sati
GTID: 902917135

February 9, 2016

Contents

Homework 1

1.1 PCA:

For my implementation of PCA, I show the 10th eigenvector instead of the 20th, as the 20th is zero as the data is 13-dimensional. The most important eigenvector, the 10th eigenvector and the plot for the ratio of the sorted eigenvalues to the sum of the eigenvalues for the wine dataset for two runs, one with 5 and the other time with 50 random samples of each class selected for training are:

`Using 5 samples for training:`
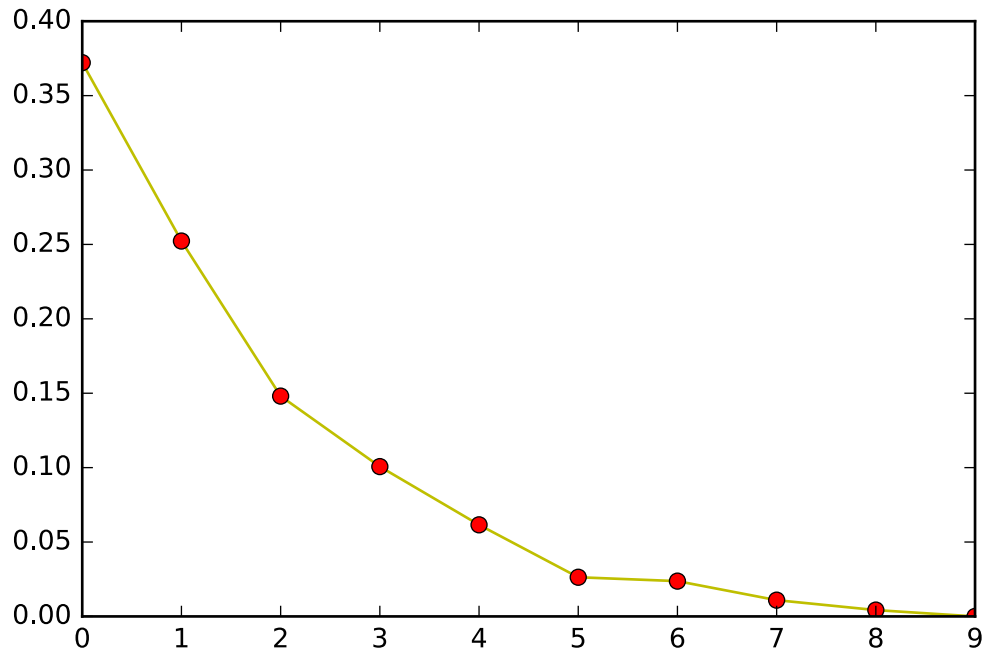

```
0th eigenvector:
[-0.34385994  0.13849928  0.11246626  0.36996116  0.19631731 -0.36142874
 -0.40236405 -0.20960933 -0.26041258 -0.32139478 -0.23303549 -0.15371937
 -0.29834503]
9th eigenvector:
[ 5.10793626e-01  -6.57173416e-03  -2.03910648e-01   5.32097708e-02
   3.51031169e-01   4.61235001e-02  -3.66222242e-01  -1.11559159e-01
   4.77567372e-01  -5.06606300e-05   6.83303340e-02   1.57324644e-01
  -4.06485843e-01]
```

`Using 50 samples for training:`
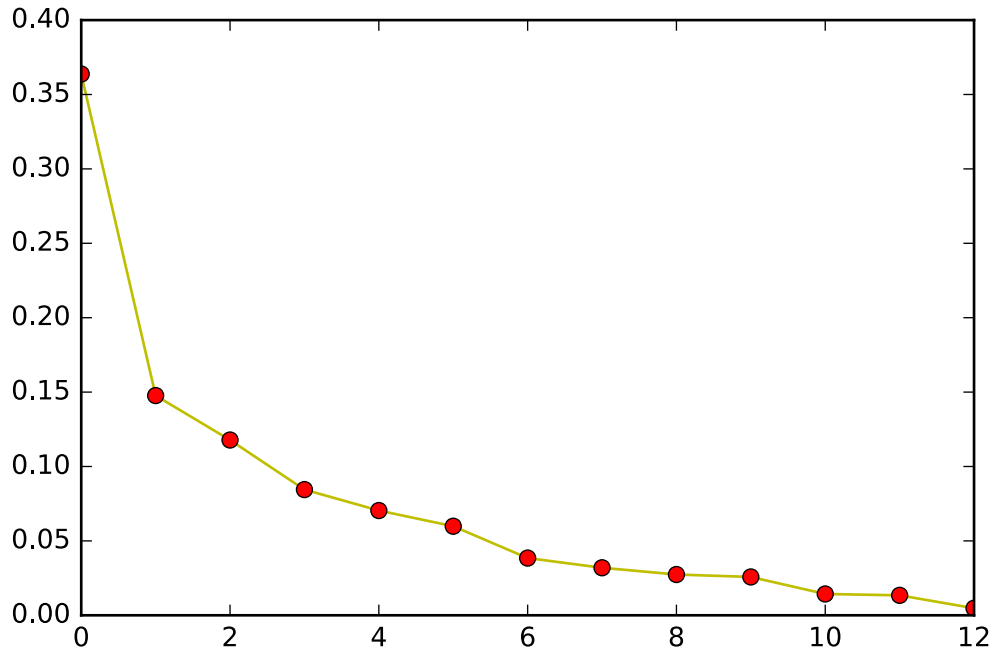
```
0th eigenvector:
[-0.36745926  0.02109651 -0.12377745  0.20876771 -0.18190286 -0.37929686
 -0.41195795  0.23876844 -0.23049816 -0.38430366 -0.0164171  -0.22820109
 -0.38572708]
9th eigenvector:
[-0.35338842  0.16218653  0.54318937 -0.34157764 -0.23783573 -0.03739143
  0.00716699 -0.4918175  -0.0624932  -0.04916445  0.05866241 -0.34014945
  0.1082448 ]
```

I also use a variable number of eigenvectors for the reconstruction, discarding those for which the ratio: (eigenvalue)/(sum of eigenvalues) falls below a threshhold (I use a threshhold of 0.03). Using this to reconstruct a test-example, I get the following reconstruction errors:
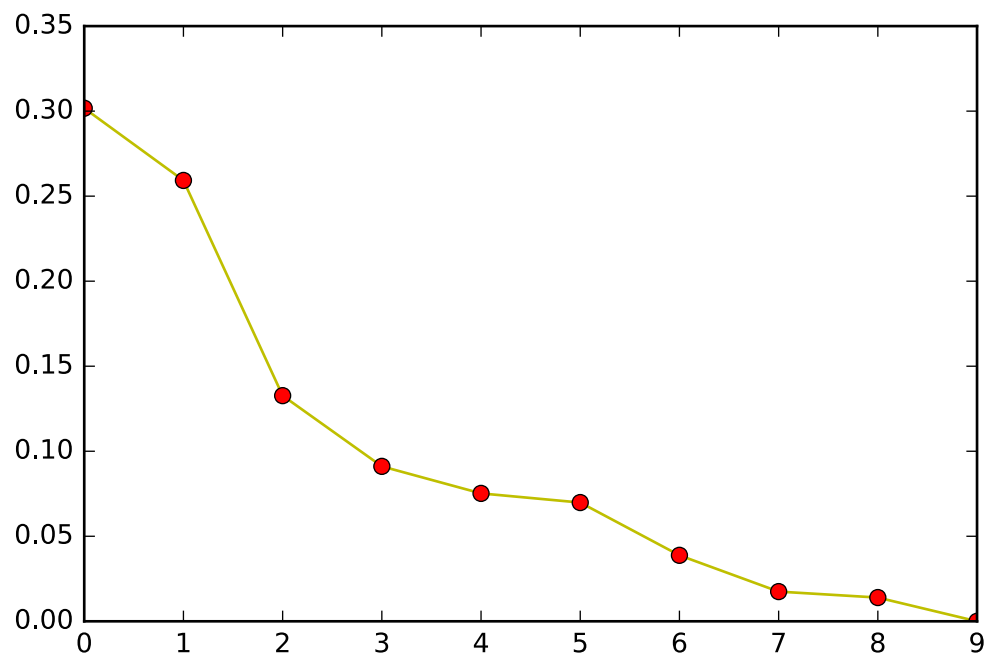
```
When using 5 samples for training:
Reconstruction error: 2.38816823748


When using 50 samples for training:
Reconstruction error: 0.418594431148
```

For mnist, I show the 10th eigenvector when using just 5 samples from each class and the 20th otherwise. The most important eigenvector, the 10th eigenvector and the plot for the ratio of the sorted eigenvalues to the sum of the eigenvalues for the wine dataset for two runs, one with 5 and the other time with 50 random samples of each class selected for training are:
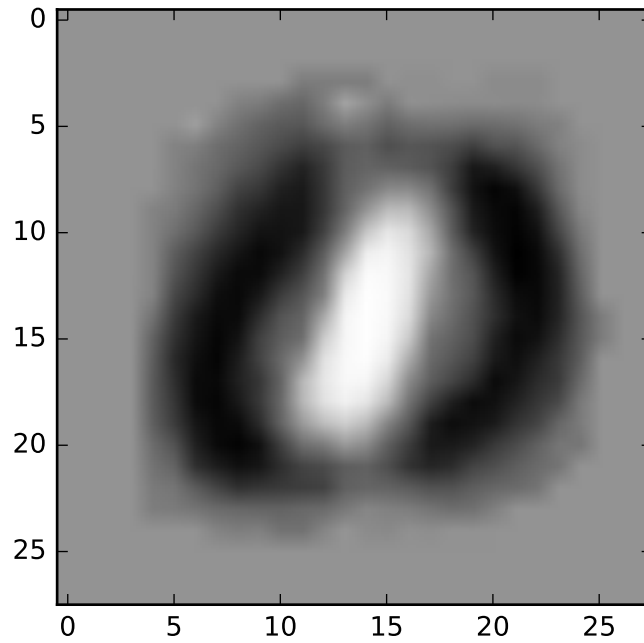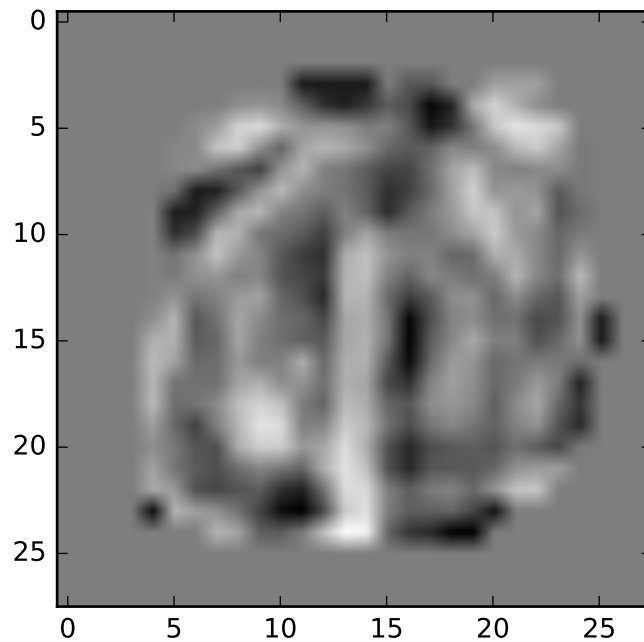
0th eigenvector:
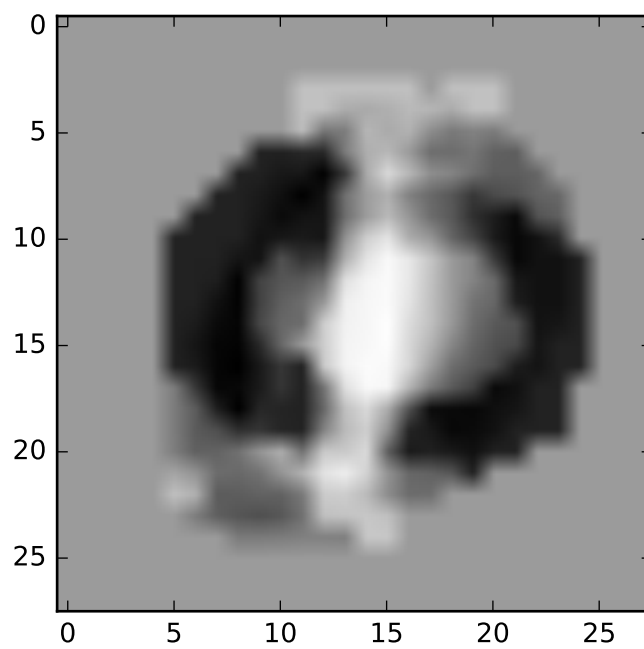
9th eigenvector:
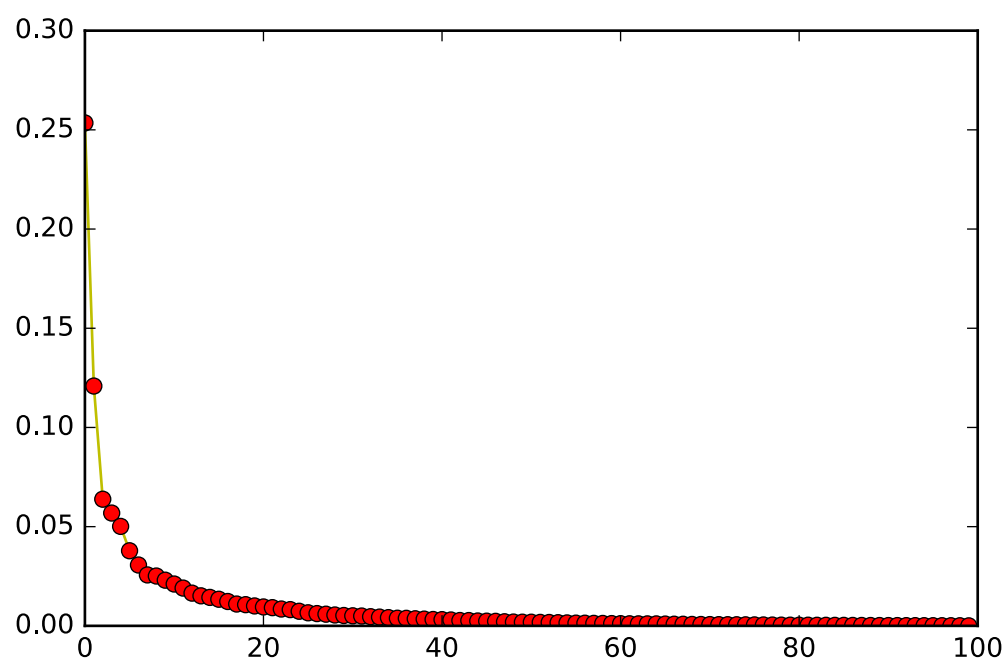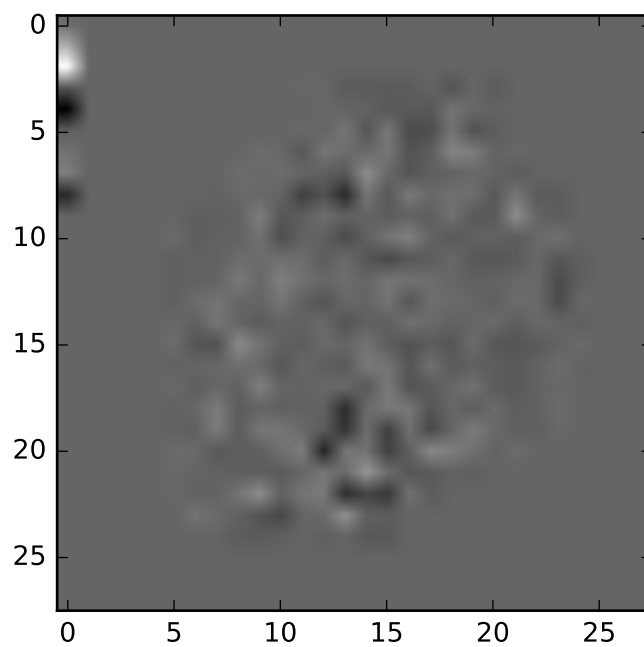
0th eigenvector:

20th eigenvector:

Similar to the wine data-set, I again use a variable number of eigenvectors for the reconstruction, discarding those for which the ratio: (eigenvalue)/(sum of eigenvalues) falls below a threshhold (I use a threshhold of 0.005 here). Using this to reconstruct a test-example, I get the following reconstruction and errors:

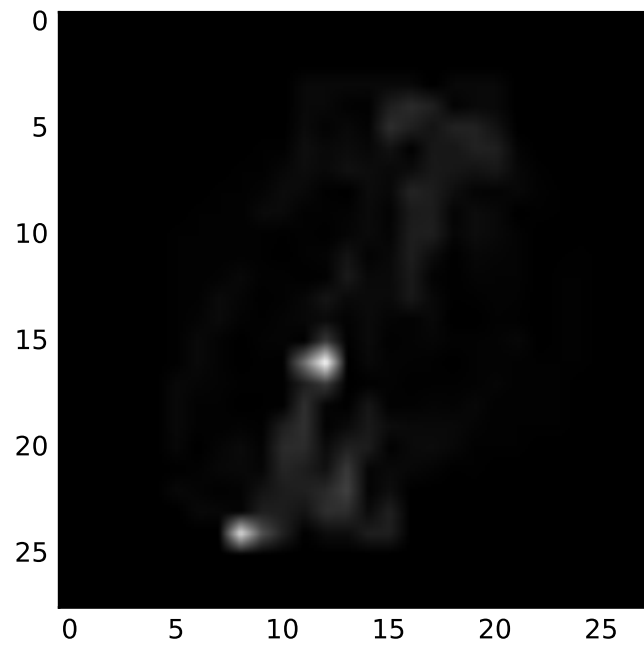When using 5 samples for training:
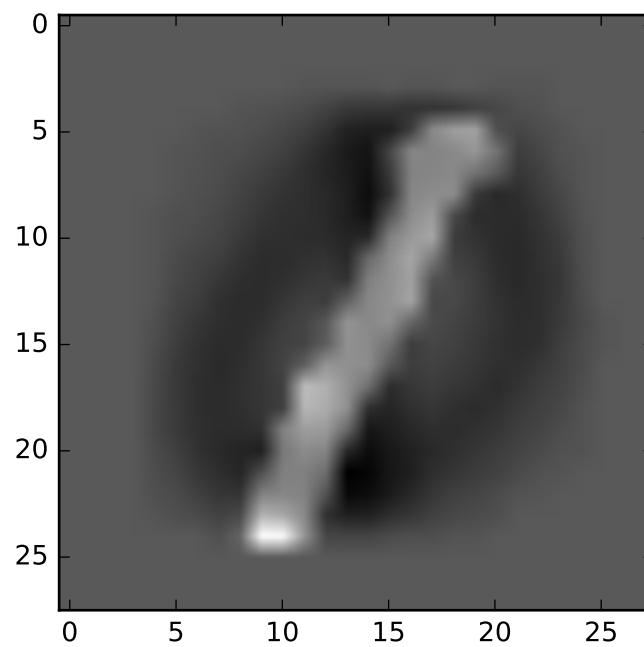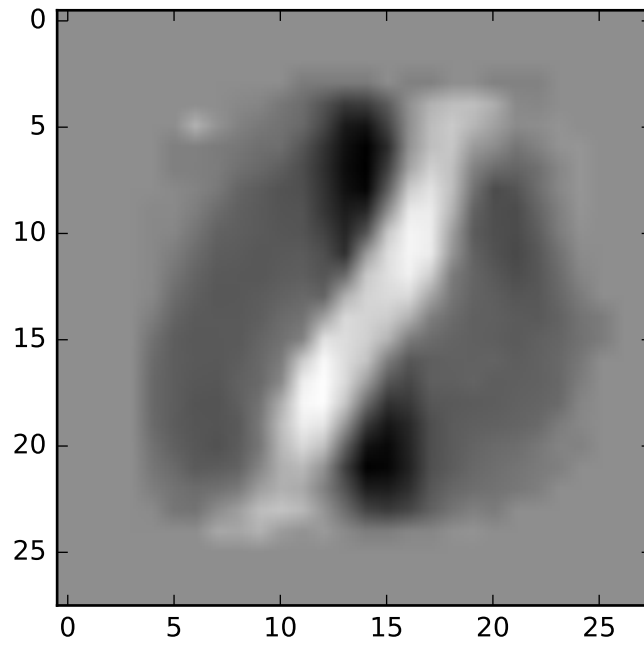Original sample
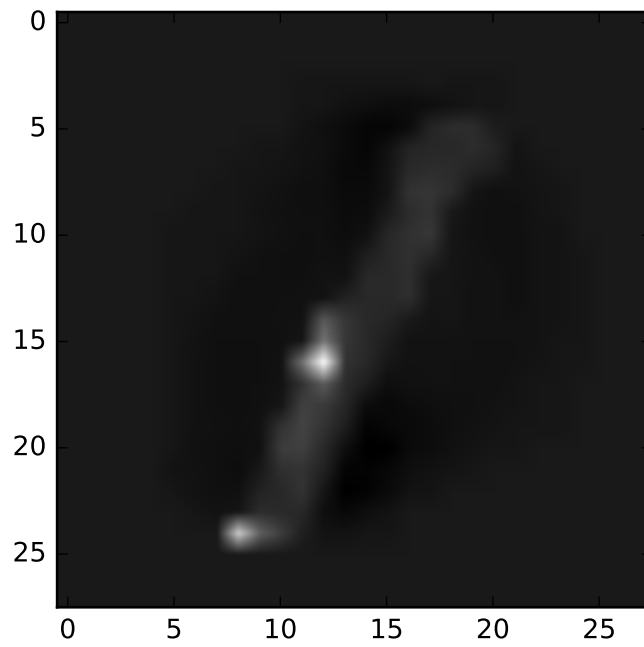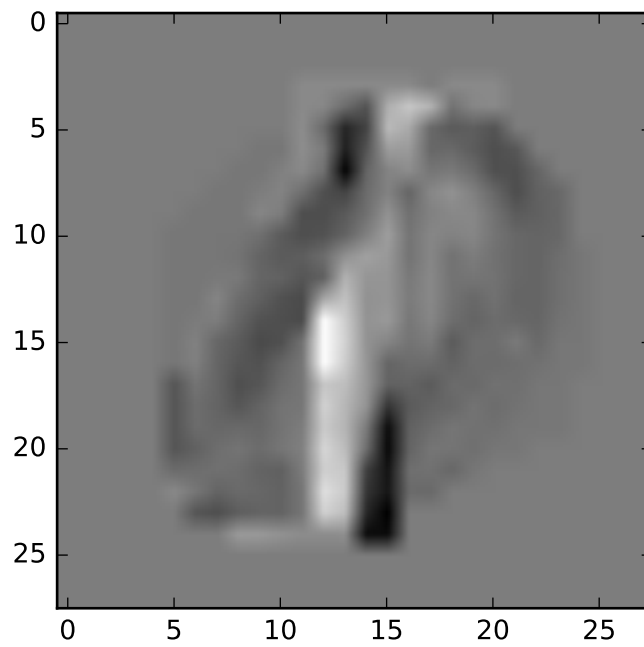
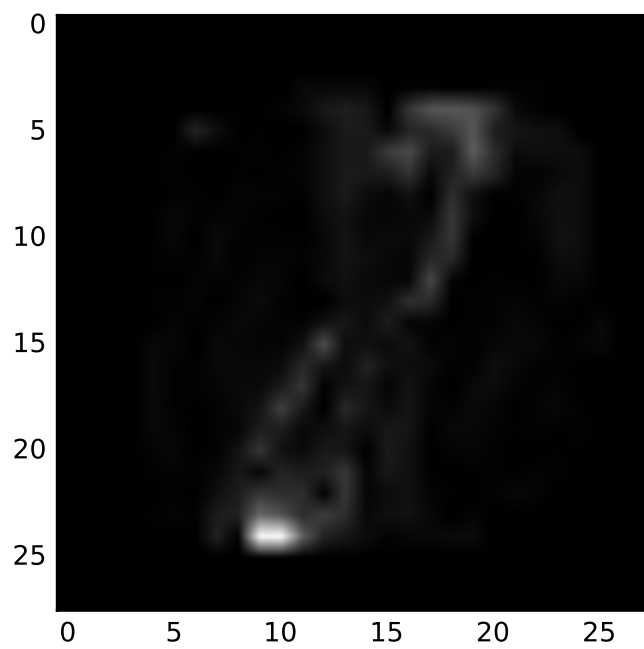Reconstruction

Reconstruction error: 30.7126937352

When using 50 samples for training:
Original sample

Reconstruction

Reconstruction error: 8.23648312092

1.2 LDA

I perform LDA for class discrimination maximizing dimensionality reduction after performing PCA for class agnostic dimensionality reduction. The most important eigenvector, the 10th eigenvector and the plot for the ratio of the sorted eigenvalues to the sum of the eigenvalues for the wine dataset for two runs, one with 5 and the other time with 50 random samples of each class selected for training are:

```
Using 5 samples for training:
```
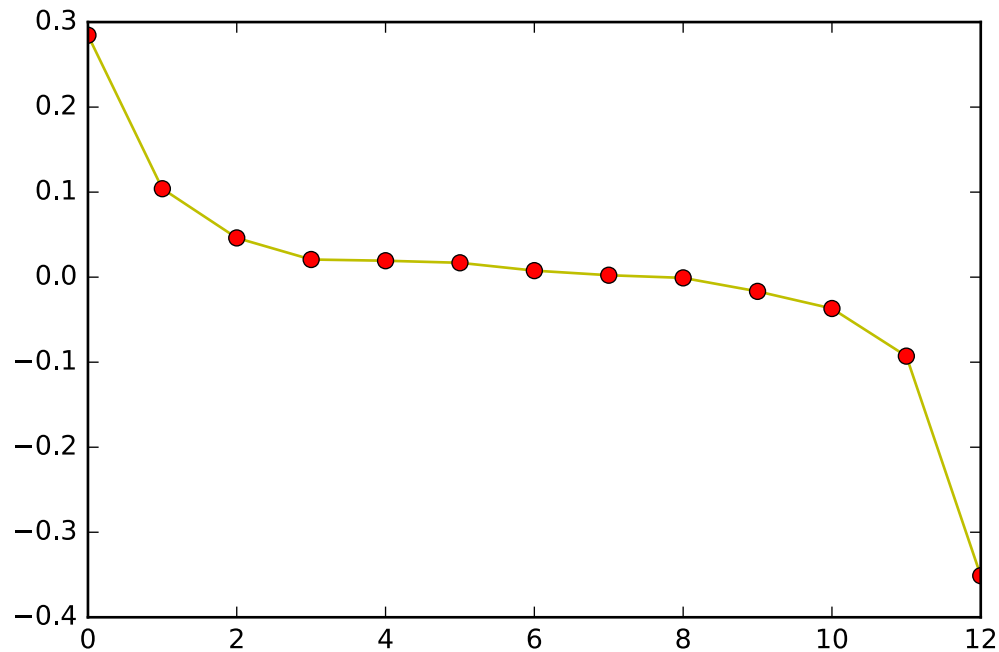


```
0th eigenvector:
[ 0.02512073 -0.05211056  0.26906833 -0.01806921 -0.22226442 -0.01071707
  0.19807376  0.3804282  -0.25404062 -0.37790185  0.39388645 -0.13488114
 -0.55656816]
9th eigenvector:
[ -3.44856908e-03   6.43474109e-04  -9.68625809e-03  -4.59535582e-02
  -1.54747729e-01   3.53055196e-01  -4.96445535e-01  -6.62494559e-02
  -1.43097520e-01   4.50400511e-02   6.78542002e-01   1.34447339e-01
   3.11992479e-01]

Using 50 samples for training:
```
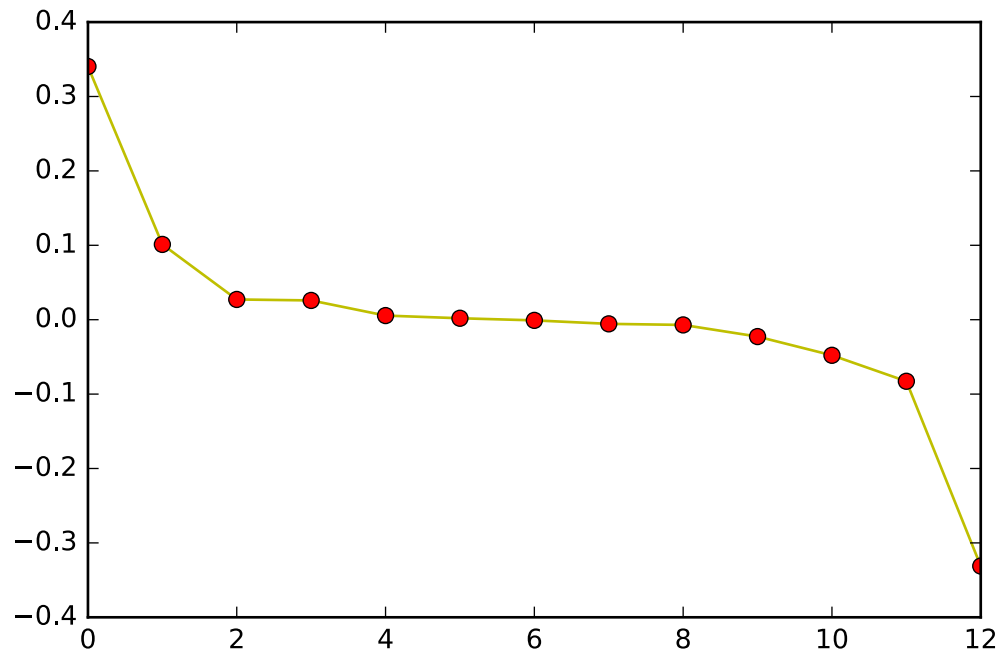
```
0th eigenvector:
[-0.06033156 -0.50261258 -0.15600335  0.11749617  0.18861993 -0.52206099
 -0.12781251 -0.15298264  0.05550418  0.16412112  0.09396104 -0.12852986
  0.54988032]
9th eigenvector:
[ 0.03461021 -0.0168285  -0.01562007 -0.05812101 -0.08590428  0.31766907
 -0.07870385 -0.81511878  0.17298554  0.40527802 -0.14196547 -0.01882478
 -0.03616571]
```

```
        File "<ipython-input-1693-c802a46ecf78>", line 2
      trainingMeans = [trainingDataUnlabeled[0].mean(), trainingDataUnlabeled[1].mean()]
        ^
    IndentationError: unexpected indent
```