ELİF GAMZE GÜLİTER

21802870

HOMEWORK 2

REPORT

QUESTION 1)

1.1

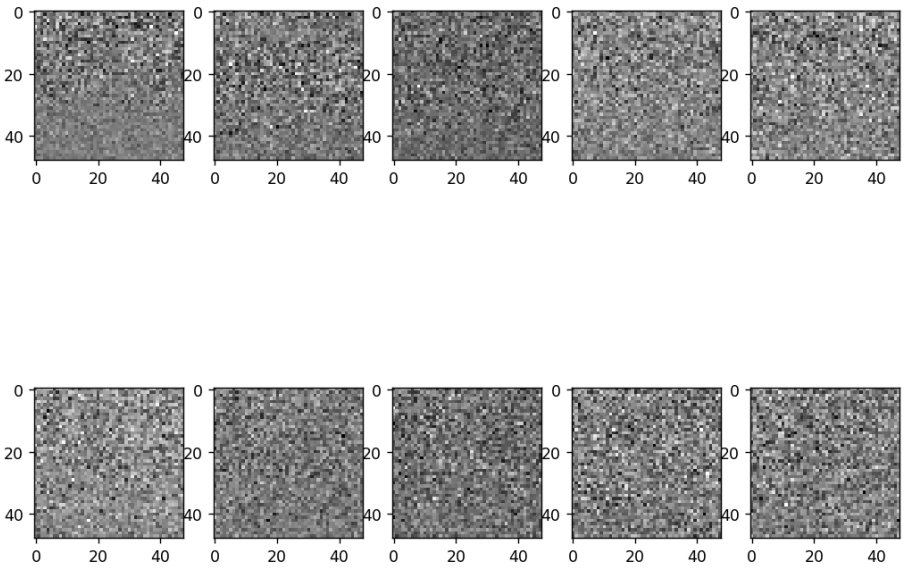

Figure 1: 48 x 48 eigen vectors

Above is the snapshoot of the reshaped eigenvectors with the size of 48 x 48. They are the first 10 eigenvectors in unsorted manner. Figure 2 is the eigen vectors of **top 10** eigenvalues.

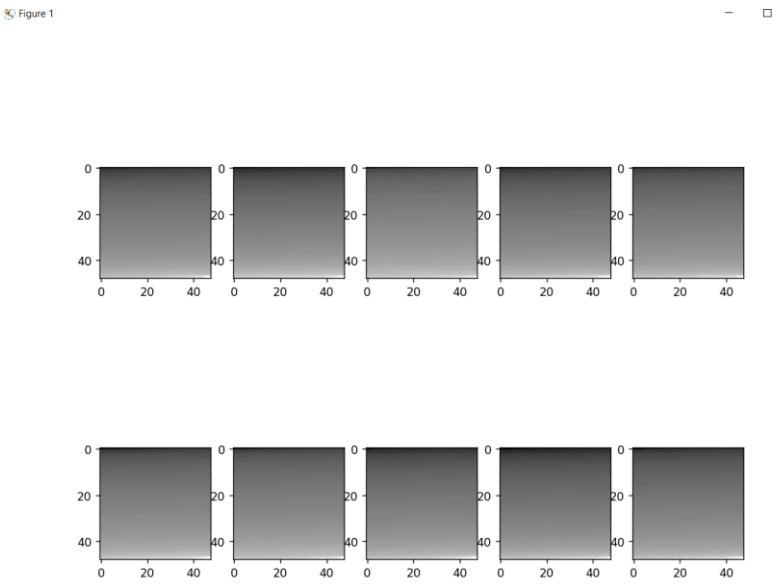

Figure 2: Top 10 eigenvectors

1.2

I found PVE for k = 10 value. In figure 3, each bar represents one PCA of 10 PCAs in total. The y-coordinate represents the percentage of PVE's of respective PCA

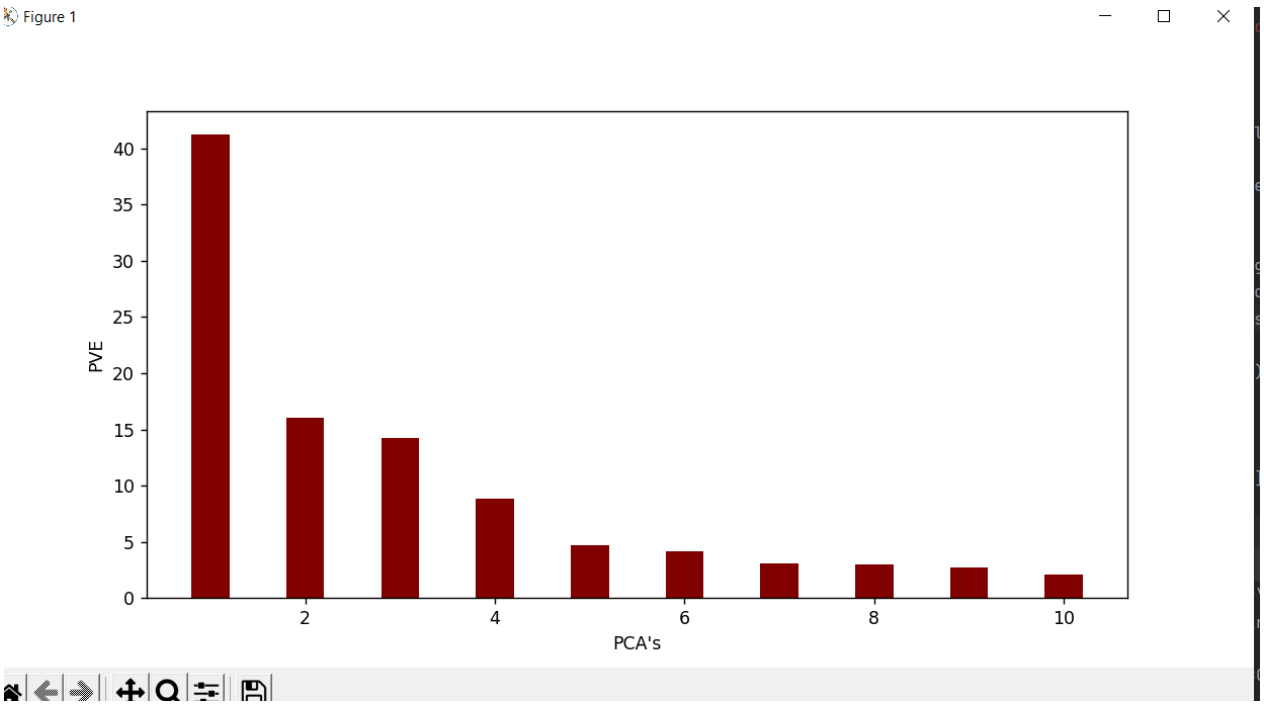

Figure 3: PVE- PCA plot of top-10 PCA 'S

1.3)

Figure 4 is the reconstructed eigenface with different k values (k = [1,10,50,100,500,2000] respectively). In the last picture k value is 2000.

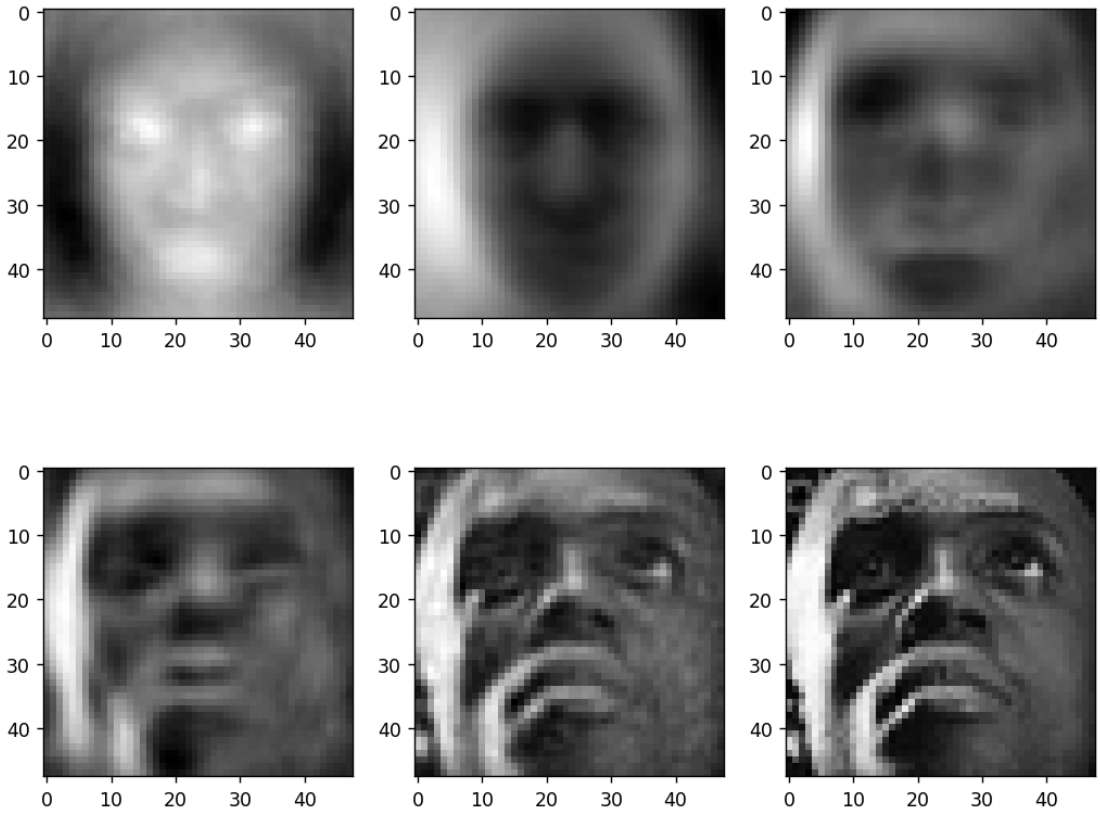

Figure 4: Reconstructed eigenfaces with different k values

QUESTION 2)

2.1

Derived formula is $\beta^{RSS} = XX^T\beta$

2.2

Using the function **np.linalg.matirx_rank(features.fot(features.T))** we found that rank is 13.

2.3

Figure 5 is the plot of linear regression model. Figure 6 is the MSE and coefficients of linear regression model.

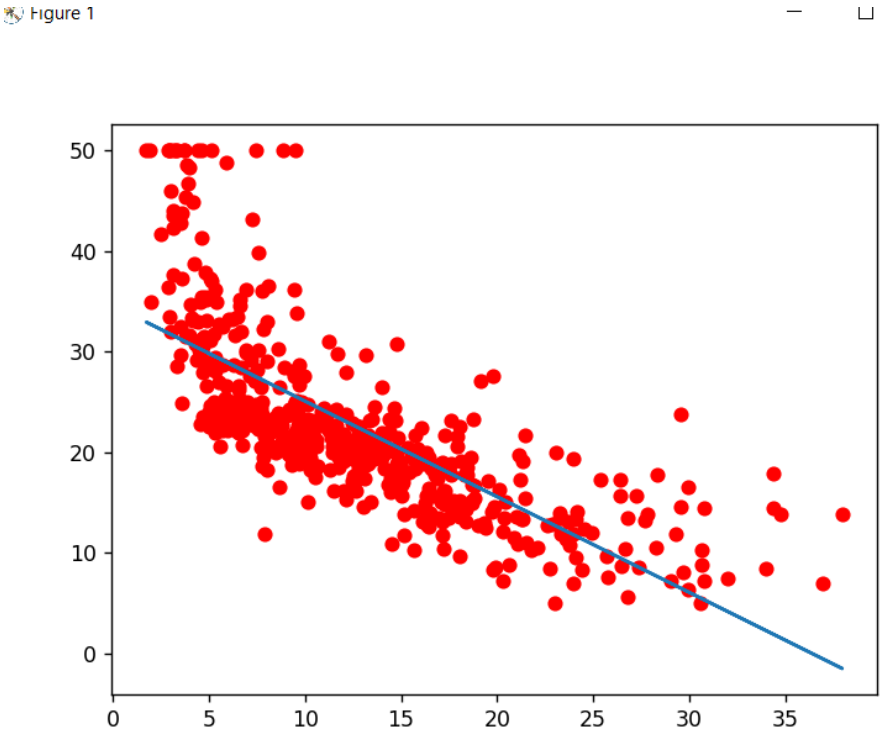

Figure 5: Linear regression model

```
weight 0 :  [34.55384088] weight 1 :  [-0.95004935]
MSE FOR 2.3 : [38.48296723]
```

Figure 6: Coefficients and MSE of linear regression model

2.4

Figure 7 is the plot of polynomial regression model. Figure 8 is the MSE and coefficients of polynomial regression model.

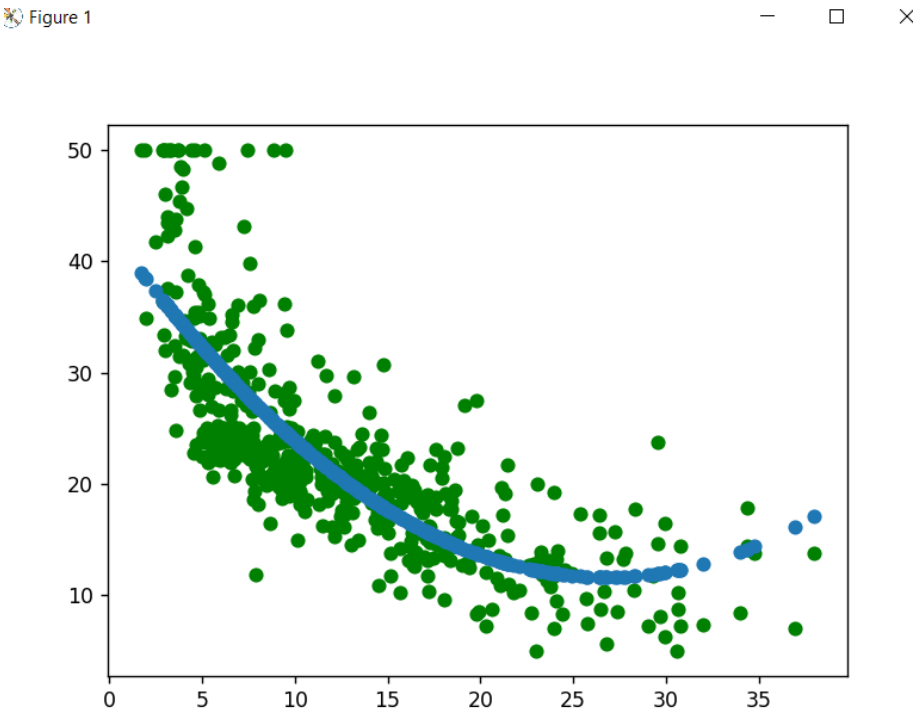

Figure 7: Plot of polynomial regression model

```
weight 0 :  [42.86200733] weight 1 :  [-2.3328211] weight 2 :  [0.04354689]
MSE FOR 2.4 : [30.33052008]
```

Figure 8: Coefficients and MSE of polynomial regression model

QUESTION 3)

3.1

Figure 9 is the metrics of full-batch gradient descent with learning rate = 1/10000. I believe that there is a mistake in mathematical form of my logistic regression model. However, I had not time to optimize it.

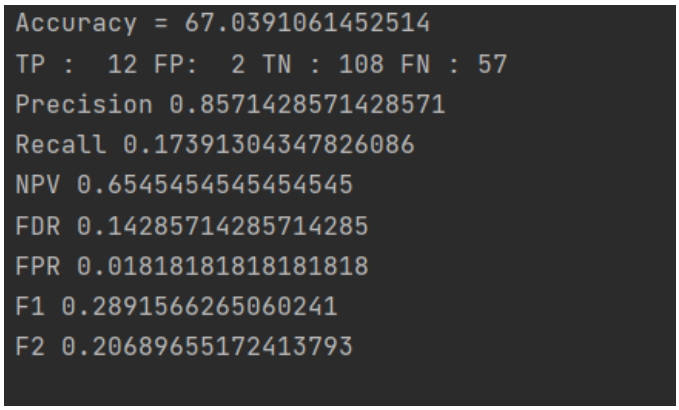

Figure 9: metrics of full-batch gradient descent with learning rate = 1/10000

3.2

I implemented the code of the LR with SGD. However, the code runs forever, and I had no time to fix it. You can see the code in between lines 139-169 in my code.

3.3

I think the simple metrics such as recall, and precision might not be enough because sometimes the data is imbalanced and we can even have zero values in confusion matrix. Therefore, it is good idea to also use other metrics.