ECE586 MPX Principal Component Analysis (PCA)

Libo Zhang (NetID: lz200)

**Exercise 1. (15 pts)**

***For each p = [0, 1, 2], print the resulting post-PCA data matrix with mean added back in and rounded to 2 decimal digits.*** The results are shown below.

Calendar

Description automatically generated

Note: I feel a bit **confused** because in the **MPX handout Exercise 1**, we **are only** asked to report the post-PCA data matrix. In **MPX Python notebook template**, however, we are asked to report both the post-PCA data matrix and the error in the Frobenius norm. Since I have showed the above results to the professor during Tuesday’s Office Hour and the professor confirmed the correctness of my results, I will follow the requirements in the pdf handout and choose not to report the error in the Frobenius norm. Thanks for your understanding.

Please continue to the next page for Exercise 2.

**Exercise 2. (25 pts)**

Plot with mean removal. (Note: The biggest number “2” first showing up is the mean.)

Table

Description automatically generated

Plot without mean removal. (Note: The biggest number “2” first showing up is the mean.)

Table

Description automatically generated

***What do you observe about these images? How many of these look a lot like digit 2?***

These greyscale images, compared with the original digit 2 images, seem to distort, or mingle the boundaries between the actual white color digit and the black color background, and have comprehensive grey color area as the background. As for the plots with mean removal, I think the first 24 images look a lot like digit 2. As for the plots without mean removal, I think the first image looks extremely like digit 2. However, all in all, only the first 18 images look a lot like digit 2 and the rest images are heavily distorted. It seems that with mean removal, although the first image may not look “extremely” like digit 2, the rest images are more “average” like digit 2 than images without mean removal.

**Exercise 3. (30 pts)**

The results are shown below.

Chart, line chart

Description automatically generated

Calendar

Description automatically generated

Chart, line chart

Description automatically generated

A picture containing text

Description automatically generated

Please continue to the next page for more results.

Chart, line chart

Description automatically generated

A picture containing text

Description automatically generated

Chart, line chart

Description automatically generated

A screen shot of a person's face

Description automatically generated with low confidence

Please continue to the next page for more results.

Chart, line chart

Description automatically generated

Calendar

Description automatically generated with low confidence

Chart, line chart

Description automatically generated

A picture containing calendar

Description automatically generated

Please continue to the next page for more results.

Chart, line chart

Description automatically generated

A picture containing calendar

Description automatically generated

Chart, line chart

Description automatically generated

A picture containing calendar

Description automatically generated

Please continue to the next page for more results.

Chart, histogram

Description automatically generated

Calendar

Description automatically generated with medium confidence

Chart, line chart

Description automatically generated

Calendar

Description automatically generated

Please continue to the next page for Exercise 4.

**Exercise 4. (30 pts)**

First, I show the results after using PCA to reduce dimension from 28 \* 28 = 784 dimensions to 500 dimensions. The results are shown below.

A picture containing calendar

Description automatically generated

Please continue to the next page for Exercise 4.

Second, I show the original results acquired from the Least Squares project. The original results are shown below.

Calendar

Description automatically generated

***How does the error rate compare to the results from the Least Squares project?***

From the above two figures, we know the PCA results have a training error of 13.83%, which is a little bit higher but very similar to the Least Squares project training error 13.56%. Similarly, the PCA results have a testing error of 15.51%, which is only a little bit higher but very similar to the Least Squares project testing error 15.28%.

To conclude, after reducing dimensions from 784 to 500 with PCA, we can still achieve very good or very competitive training and testing accuracies compared with the original Least Squares results.