2. Principle component analysis using MATLAB.

a. Visualize images.

Below in Figure 1, the first line of three.txt is shown.

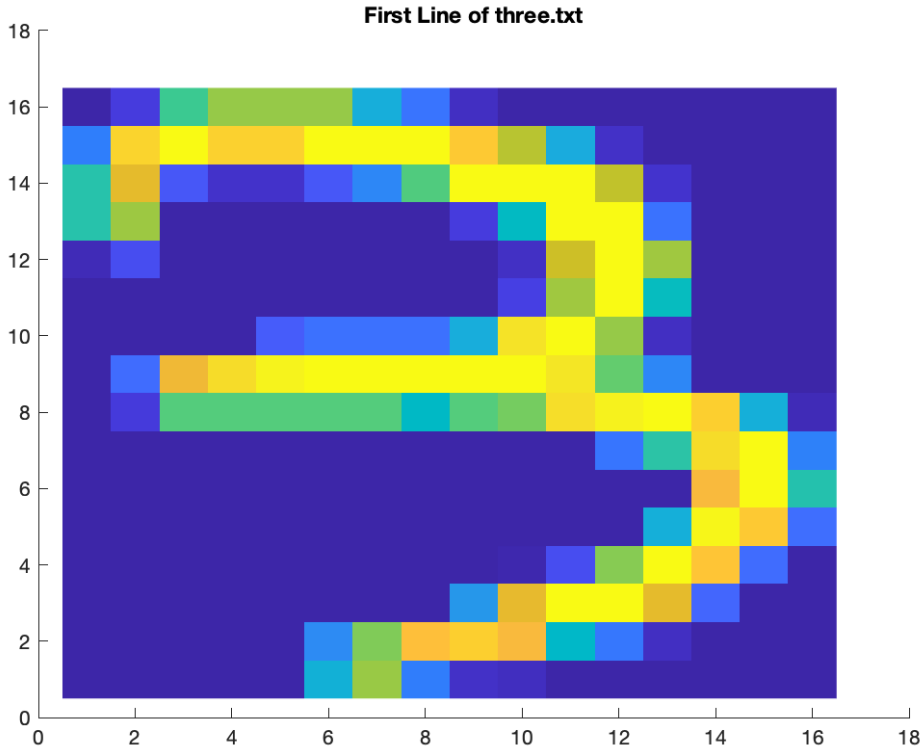

Figure 1: In MATLAB, visualization of the first line of three.txt is generated using the `imagesc()` function. The `imagesc()` function scales up the size of unconventional matrix slices that have to be displayed (such as this one line vector from three.txt) and maps the grayscale intensity values to colors based on a colormap for easier distinguishability. This produces the image as shown.

Below in Figure 2, the first line of eight.txt is shown.

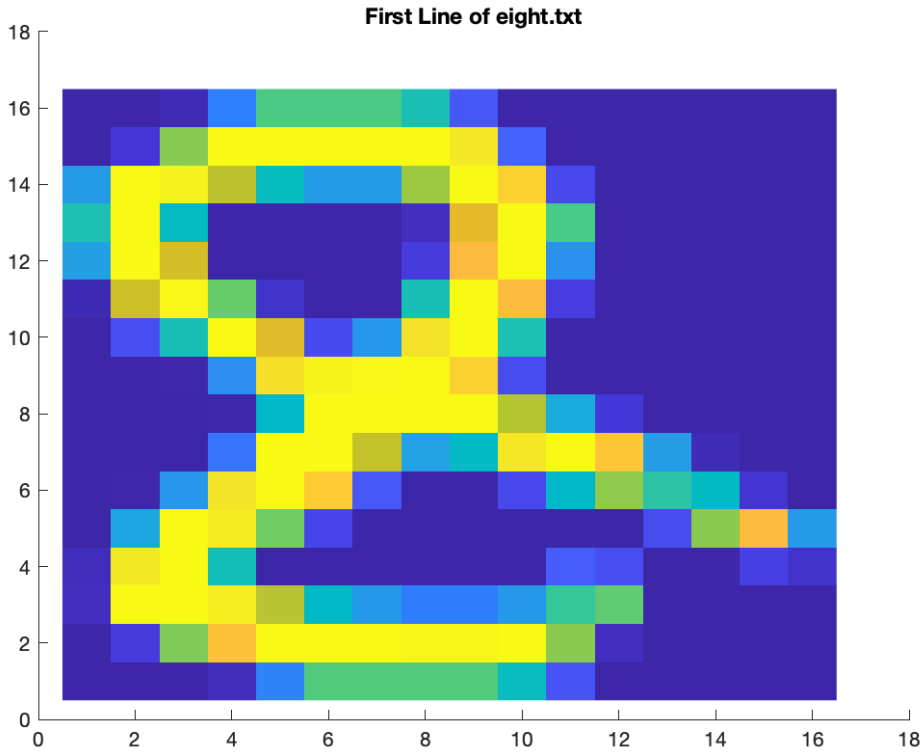

Figure 2: Again in MATLAB, visualization of the first line of eight.txt is generated using the `imagesc()` function.

b. Sample mean.

In Figure 3 below, $\bar{X}$ is displayed as a 16×16 grayscale image.

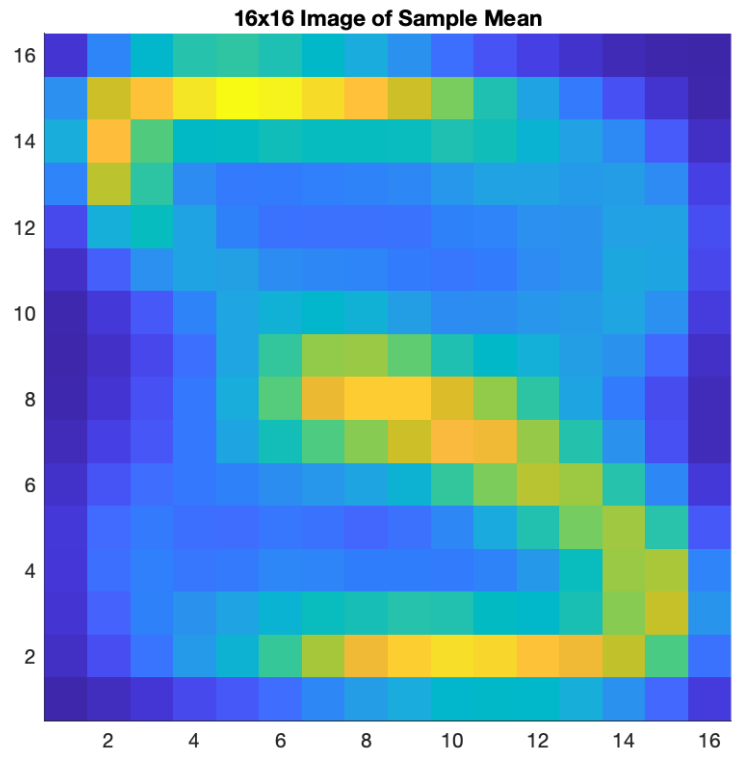

Figure 3: $\bar{X}$ was computed in MATLAB and displayed as a 16×16 grayscale image using `imagesc()`.

c. Covariance submatrix.

The sample covariance matrix is shown below as the output of the computation performed in MATLAB.

```
  59.167      142.15      28.682     -7.1786     -14.336
  142.15      878.94      374.14      24.128     -87.128
  28.682      374.14      1082.9      555.23      33.724
 -7.1786      24.128      555.23      1181.2      777.77
 -14.336     -87.128      33.724      777.77        1430
```

d. Compute eigenvalues.

The two largest eigenvalues computed in MATLAB are shown below.

```
237155.246290486
145188.352686825
```

In Figure 4 below, eigenvector v_1 is shown.

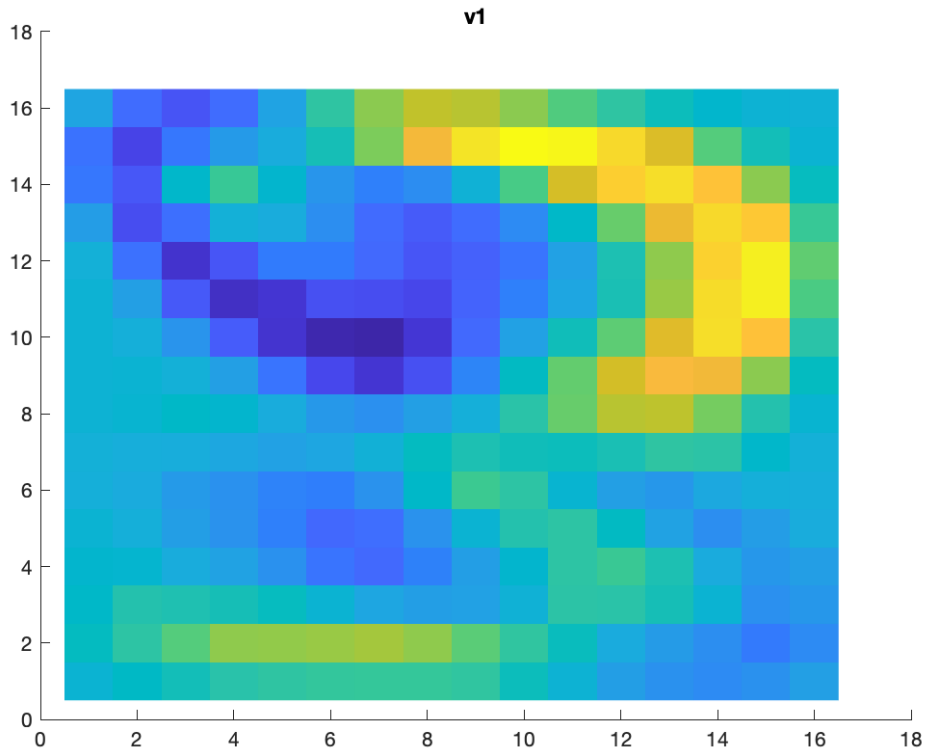

Figure 4: In MATLAB, eigenvector v_1 is displayed using `imagesc()`.

In Figure 5 below, eigenvector v_2 is shown.

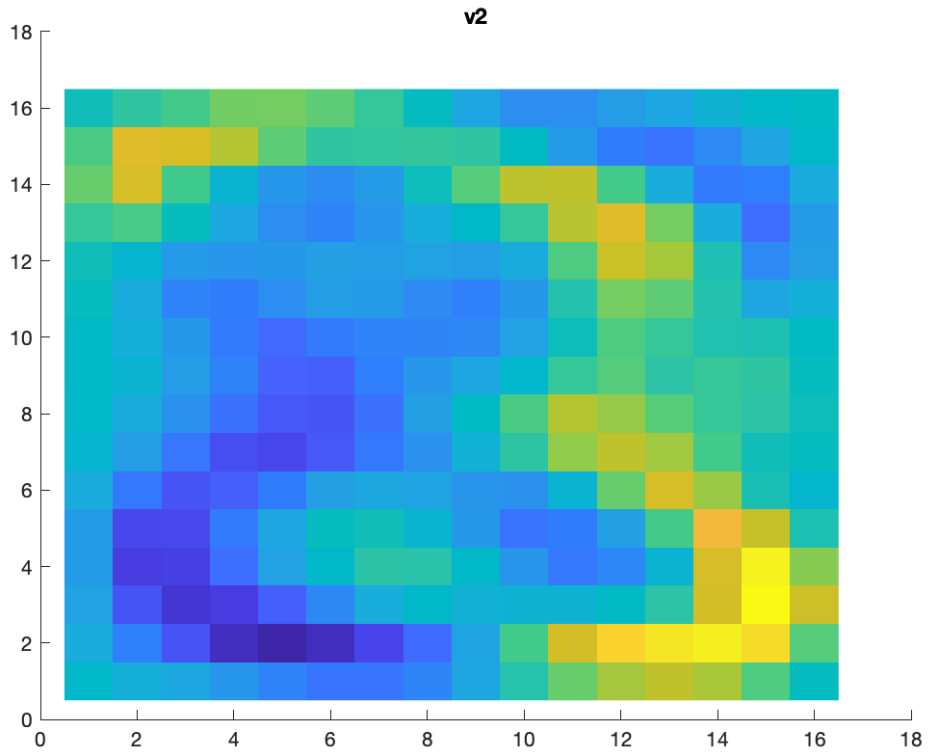

Figure 5: In MATLAB, eigenvector v_2 is displayed using `imagesc()`.

e. Matrix projection.

Matrix multiplication was performed in MATLAB to get the projected coordinates of the first line of three.txt and eight.txt respectively as reported below.

```
p3 =

240.9        297.49

p8 =

-208       -594.72
```

f. Average reconstruction error.

The average reconstruction error was computed in MATLAB as shown below. The code is pasted to show my work, and the result is pasted as the last line.

```
A = V*V';
err = [];
for i = 1:size(X,1)
    err = [err; X(i,:)*A - X(i,:)];
end
size(err);
s = sum(err.^2,2);
size(s);
format long g;
disp(sum(s)/size(X,1));

1405766.85128888
```

g. Plot 2D point cloud.

In Figure 6 below, the plot of the 2D point cloud is shown.

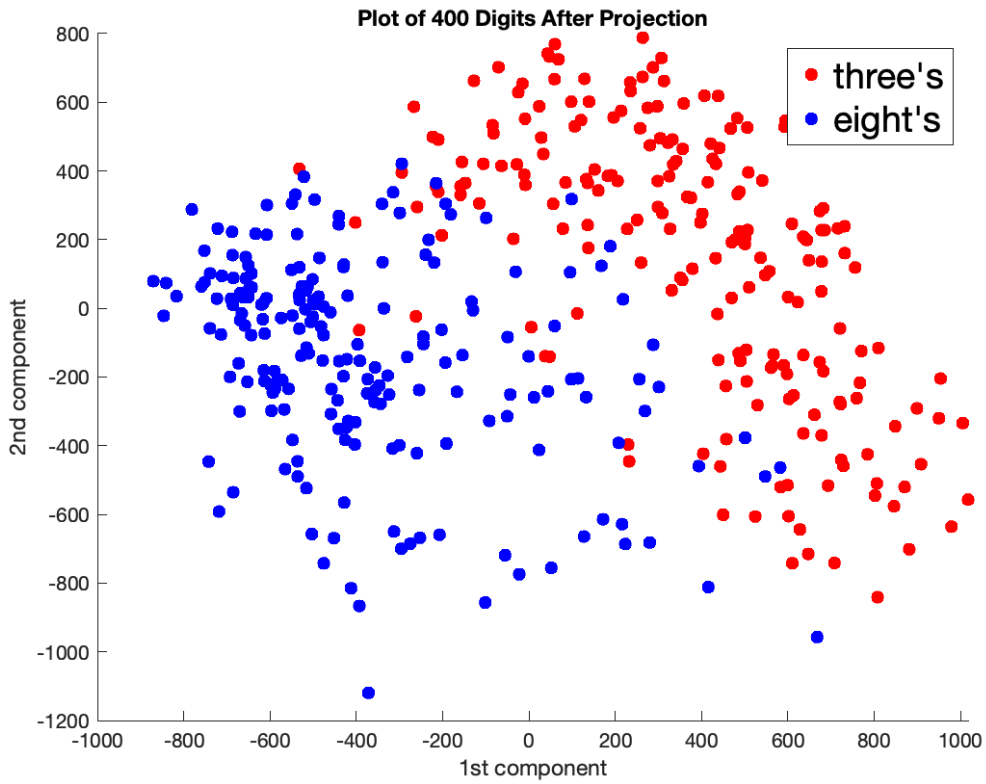

Figure 6: In MATLAB, the 400 digits after projection by PCA is shown with red dots indicating three's and blue dots indicating eight's. The axes are the first and second principle components from projection.

3. Naive Bayes

a. Estimating π probabilities.

b. Class conditional distribution for English.

c. Make bag-of-words.

d. Computing $\hat{p}(x|y)$.

e. Posterior $\hat{p}(x|y)$.

f. Evaluating performance of classifier.

g. Limiting training sample.