1. Weighted NN

1b.)

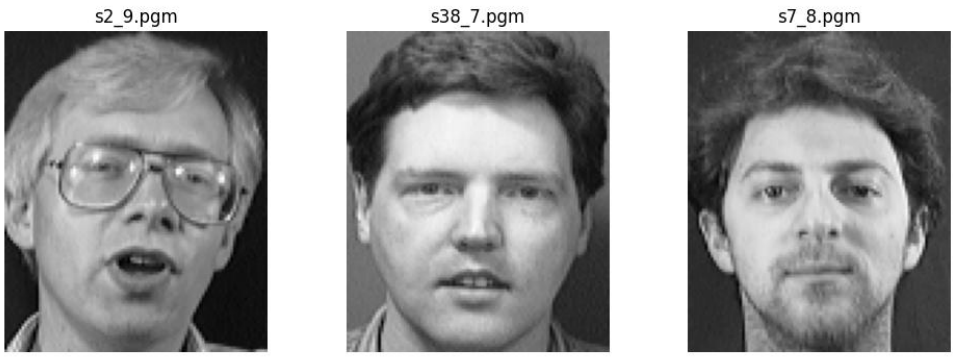
Sigma:	.01	.07	.15	1.5	2	4
Accuracy:	0.68	0.92	0.92	0.8	0.72	0.76

It seems that the best sigma values are around .07-.15. Lower sigmas undervalue the weights, therefore underfitting the data. However, higher sigmas cause the model to overfit to the testing data, causing lower accuracy in the prediction.

2. PCA Face Recognition

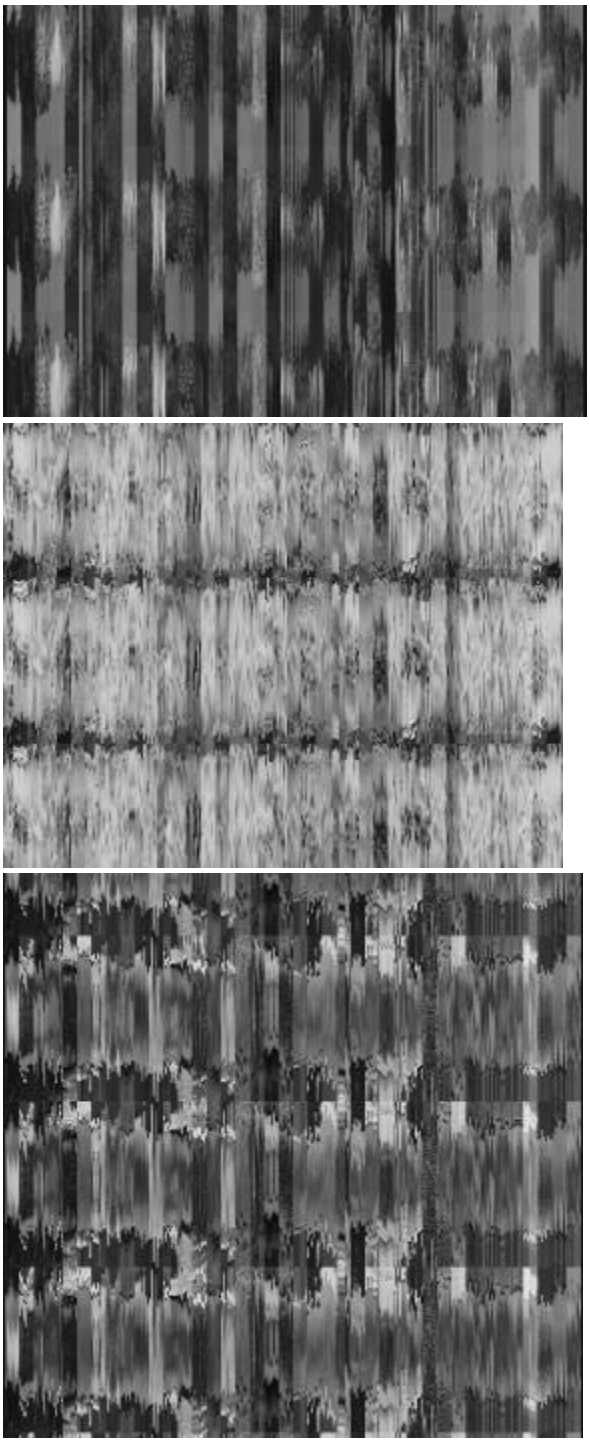
2.0 Data Pre-Processing

1 ps5-2-0.png



2.1 PCA Analysis

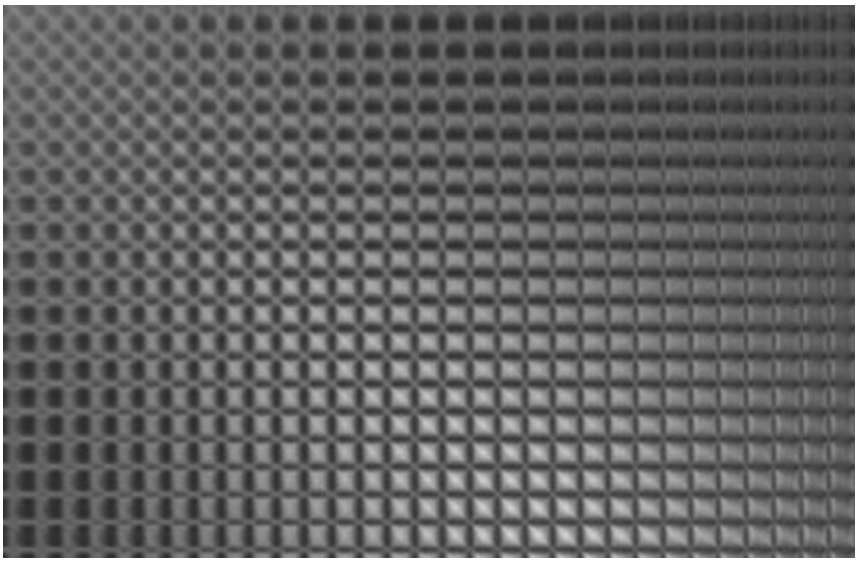
a. ps5-2-1-a.png

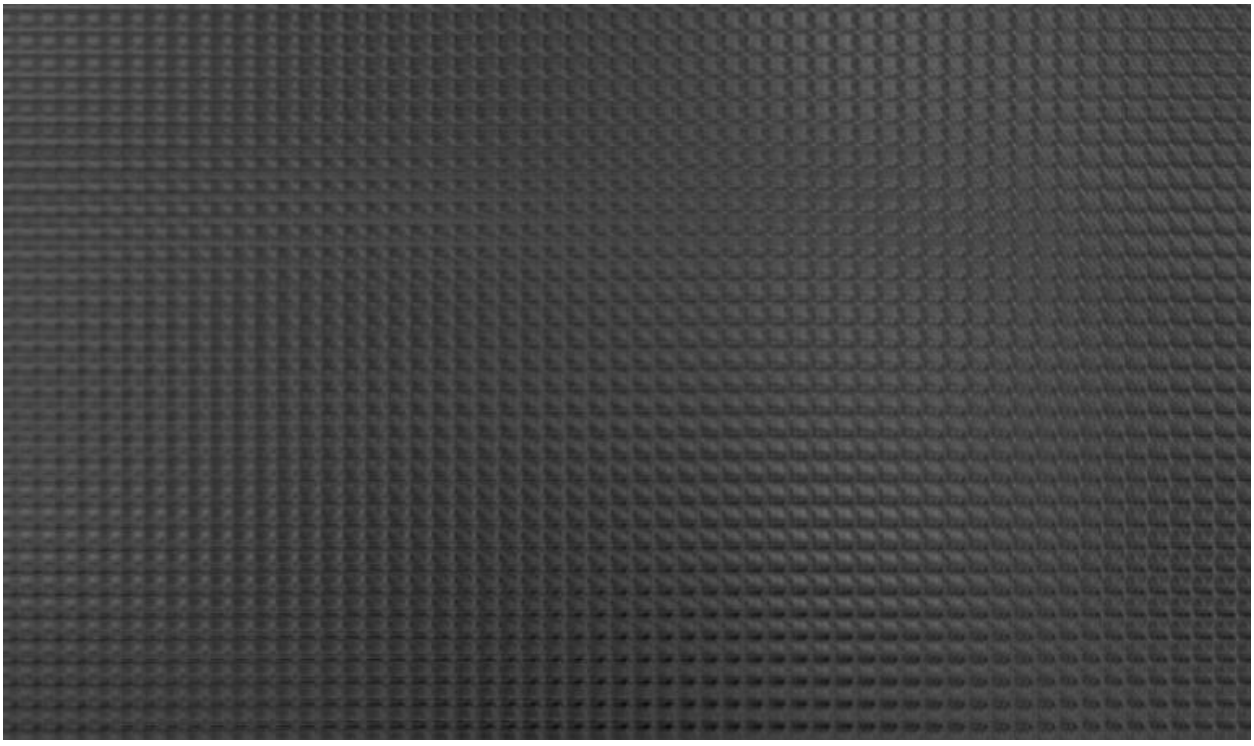


b. The mean face is very blurry with a incredibly receded hairline. All facial features are very muted and spread out.

Ps5-2-1-b.png

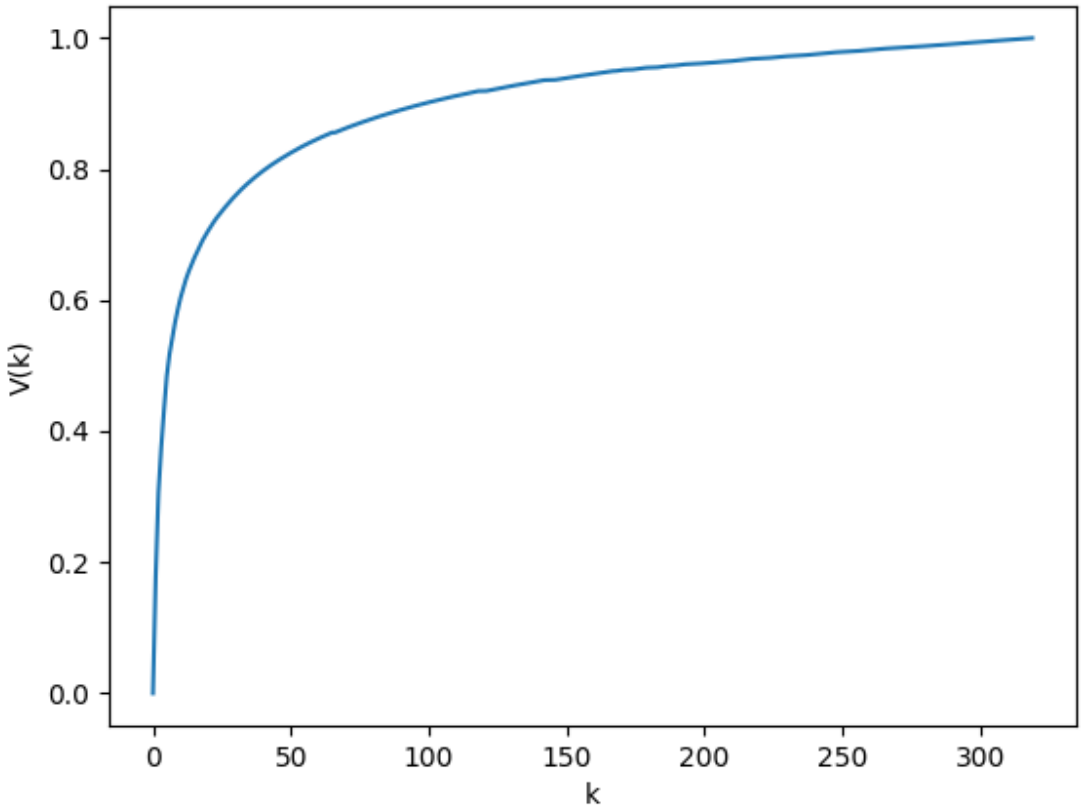
c. ps5-2-1-c.png





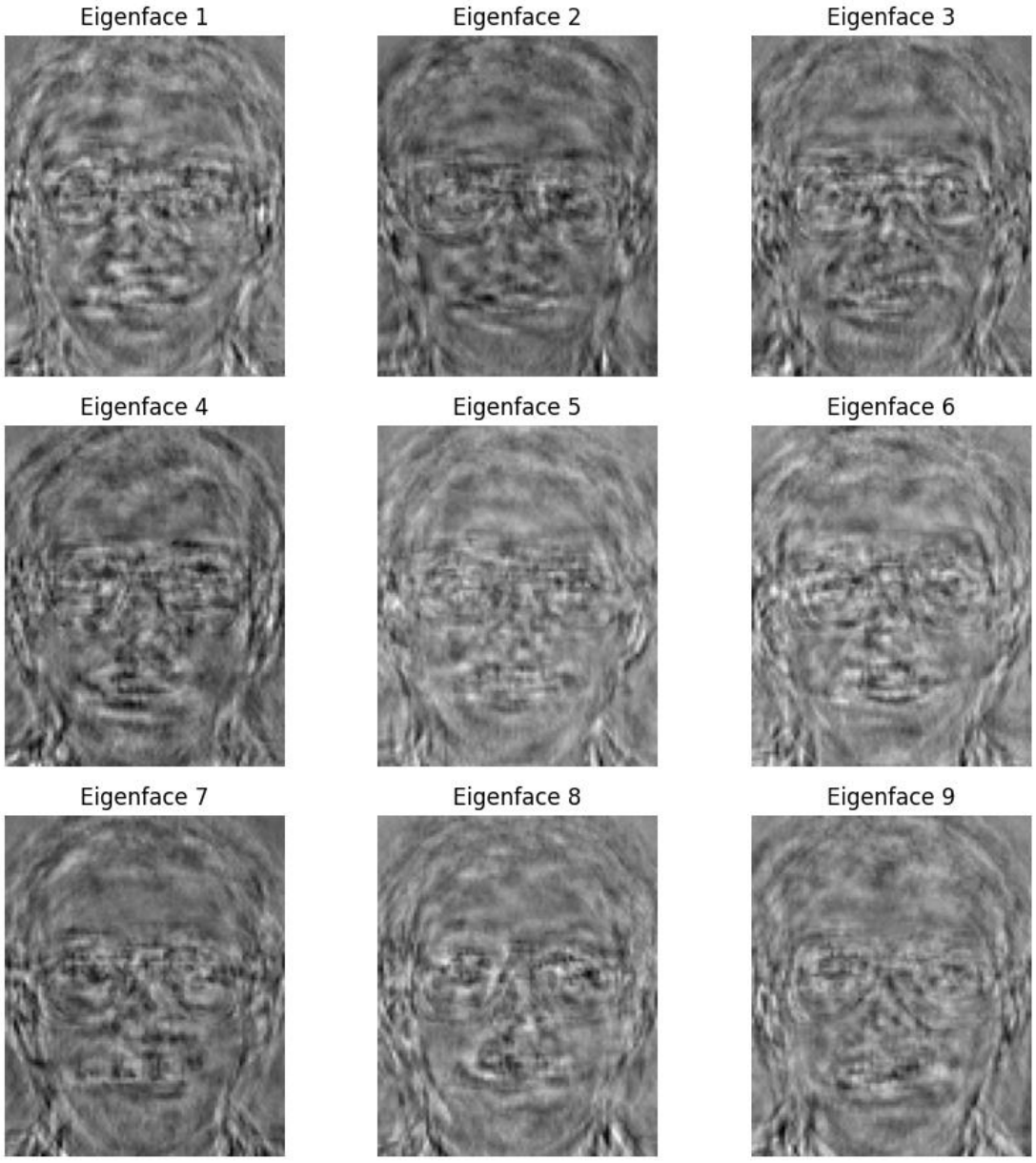
d. 172 Eigenvectors capture 95% of the training data

ps5-2-1-d.png



e. U Dimensions: (10304, 172), the eigenfaces are very rippled and a face is barely identifiable in each of them. This makes sense as they show the vectors of with the most variance in the faces.

Ps5-2-1-e.png



2.2 Feature extraction for face recognition

b. W_training dimensions (320, 172)

W_testing dimensions: (80, 172)

2.3 Face Recognition

a.

K	1	3	5	7	9	11
Accuracy	0.9625	0.95	0.9125	0.925	0.8625	0.775

K = 1 has the most accuracy, this means that the data is heavily clustered with low overlap.

b.

Training Time (s)	One v. All	One v. One
Linear	0.0157	0.5531
Polynomial	0.0117	0.5650
RBF	0.0167	0.7460

Accuracy	One v. All	One v. One
Linear	0.9625	0.9625
Polynomial	0.75	0.85
RBF	0.9875	0.9875

The RBF kernel seemed to have the most accuracy, however it also had the highest training time. This was followed by linear, then by polynomial for accuracy. One vs One had higher training time across the board, but only affected accuracy for Polynomial, with .85 compared to .75 for One v all.

Overall, the SVM classifier with the RBF Kernel had the most accuracy, and KNN with 1 nearest neighbor and Linear SVM are tied for second place.

3. Case Study

This scenario is a regression problem, as the probability that a borrower defaults is a continuous variable. Some features that can be used are income, expenses, and net worth.