

Intro to ML  
PS5 Report

Name: Timothy Horrell  
ID: 4475990

Questions:

1) .

- a. .
- b. The closer to zero that sigma gets, the prediction becomes like a 1NN. It makes sense the lowest sigma value would have lower accuracy because of this. In addition, the larger the sigma parameter, the closer the prediction becomes like the dataset average. Sigma 0.05 and 0.2 seem to be the strongest values of sigma that balance each of these extremes.

Sigma	Accuracy (%)
0.01	68
0.05	92
0.2	92
1.5	80
3.2	72
5	72

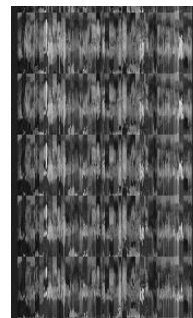
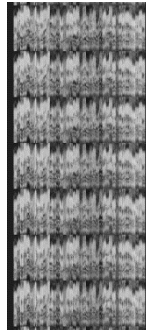
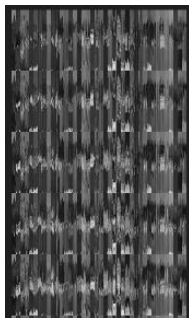
2) Image output below



ps5-2-0.png

2.1)

a) left edge, middle, right edge of images respectively



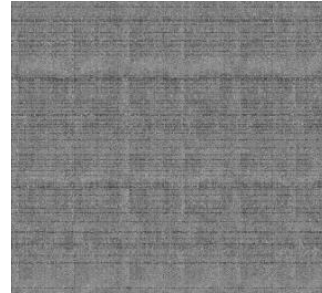
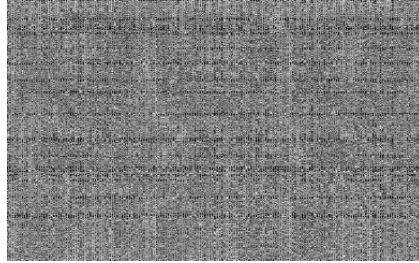
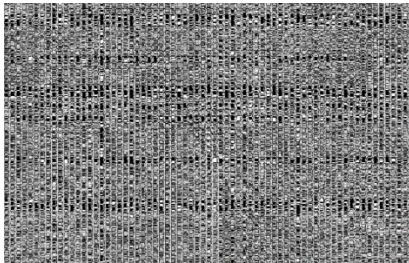
ps5-1-a.png

b) The results are the average face across each of the testing images. It has humanoid features and a general shape but lacks sharpness.



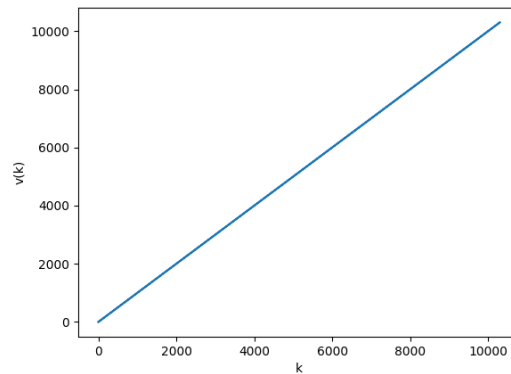
ps5-2-1-b.png

c) close zoom, medium zoom, least zoom respectively



ps5-2-1-c.png

d) 160 eigenvectors capture 95% of the variance in the data



ps5-2-1-d.png

e) matrix U dimensions: (10304, 160)

The results below of the eigen faces are supposed to represent highlight areas where the faces have high levels of variance. This image is the first 9 eigenvectors stacked vertically on top of one another. Only the first image has a face-like appearance. The rest have certain areas where the image is darker and lighter. I think this could be attributed to how the image is rendered. I am using PIL to load the image, which requires values to be integers. Given my eigenfaces don't appear like they do in the class notes, I think this may be the problem which I didn't solve in the time working on this.



2.2).

a).

b)  $W_{\text{train}}$  dimensions: (320, 160)

$W_{\text{test}}$  dimensions: (80, 160)

2.3).

a) As  $K$  grows larger, the accuracy of the prediction goes down. This is the opposite of other KNN implementations we've done, where there seems to be a middle range of  $K$  that provides high accuracy without trending towards the average of the dataset. I believe that in this instance, the higher KNN values predict too close to the mean of the dataset and therefore begin to lose accuracy. 1NN is the best classifier.

KNN	Accuracy (%)
1	81
3	68
5	58
7	54
9	54
11	51

b) The performance difference between each SVM classification paradigm is negligible in each metric of training time and testing accuracy. The Linear kernel function performs the best in accuracy yet takes significantly longer to train. The polynomial and RBF kernel functions train much quicker yet perform at around half the accuracy of the linear model. This could be due to model overfitting, where a kernel function fits too closely to the training data. All in all, the linear one-vs-one appears to be the strongest fit due to its marginally quicker training time and accuracy over other paradigms.

The linear SVM model also outperforms the best KNN classifying model which is 1NN at 81% vs. the 86% of the SVM. Given the inaccuracy of higher KNN values, the linear one-vs-one SVM is the strongest classifier of faces.

Training Time per Model (ms)		
	One-vs-one	One-vs-all
Linear	155	158
Polynomial	14	13
RBF	7	7

Testing Time per Model (ms)		
	One-vs-one	One-vs-all
Linear	0.99	1
Polynomial	1	1
RBF	3	2

Testing Accuracy per Model (%)		
	One-vs-one	One-vs-all
Linear	86	86
Polynomial	44	44
RBF	36	36

3) I would choose the following features for the following reasons:

**\*\*Note that these are all in relation to 'stretches' of highway and attributes of the area they are close to**

Distance from existing chargers: considering this avoids stacking chargers on top of each other, rather than spreading them out.

Population density in the surrounding area: this is important to understand how many cars may be in the area by 2030

Rate of population change: this helps to build an idea of whether people will be moving into an area to service it with chargers accordingly.

Average Income: income can indicate how likely somebody may be to buy an electric car

Number of electric car households: will give insight into how many electric cars are in the area around the highway

Highway throughput: This gives an idea as to how much traffic may be passing through from other areas