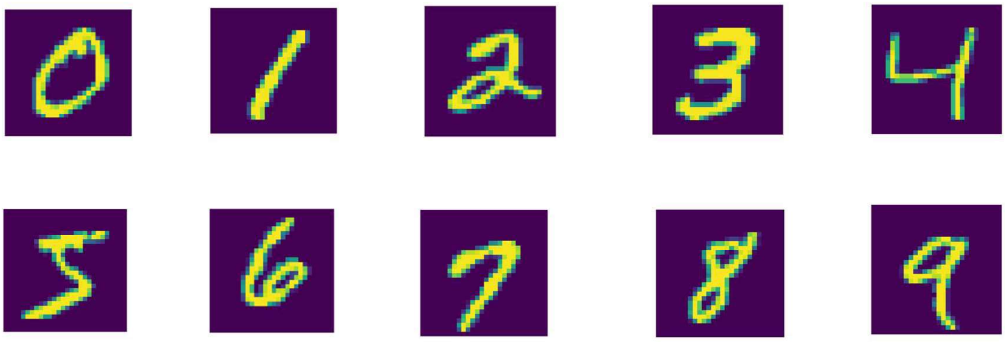
HW2 - ML

Submitters: Shahar Stahi 305237257, Gal Barak 204233688

Q1:

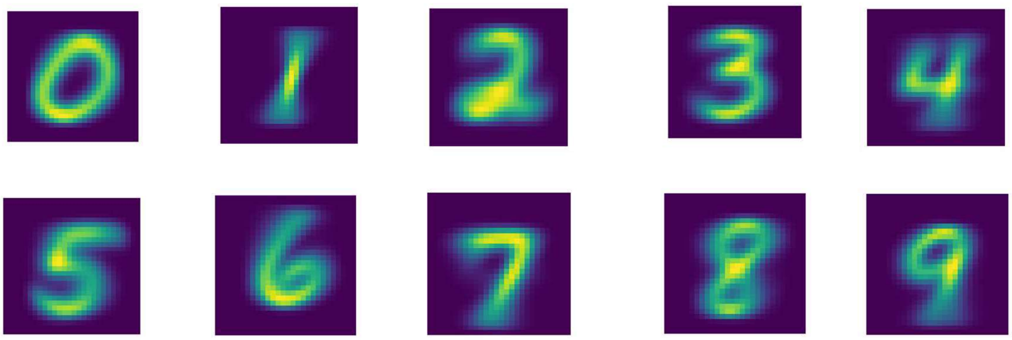
Step 1:

Classes plots:



Step 2:

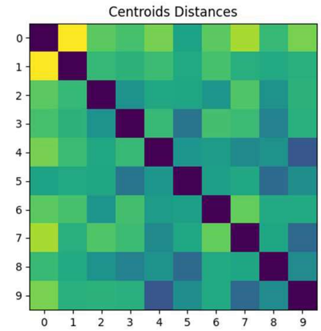
Classes centroids:



Step 3:

Centroids Distances Matrix and heat map:

```
Centroids Distances Matrix:
[[    0 31252 23342 22212 24802 18080 23353 27146 21016 25019]
 [31252     0 20933 20234 21475 19001 22110 19700 19094 19848]
 [23342 20933     0 16110 18585 18520 16379 22798 16034 20227]
 [22212 20234 16110     0 21223 12125 22046 21379 13893 19836]
 [24802 21475 18585 21223     0 16373 17294 14921 15958  8430]
 [18080 19001 18520 12125 16373     0 17646 18744 10886 15575]
 [23353 22110 16379 22046 17294 17646     0 23904 19132 19115]
 [27146 19700 22798 21379 14921 18744 23904     0 18448 10674]
 [21016 19094 16034 13893 15958 10886 19132 18448     0 15080]
 [25019 19848 20227 19836  8430 15575 19115 10674 15080     0]]
```



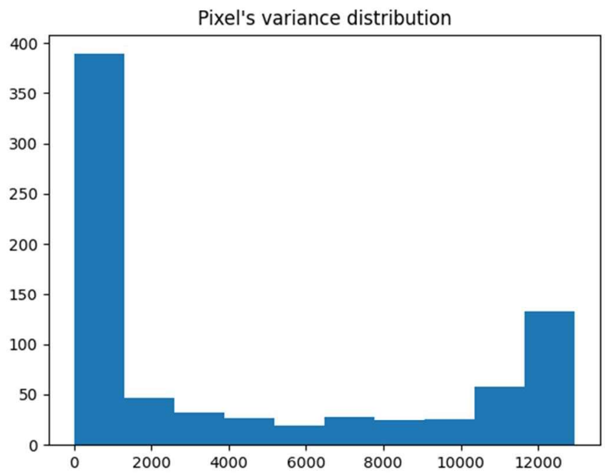
By observing the matrix and heat map we can see which digits are easy to classify and which are harder:

By the heat map, the "colder" the color is, the easier it's matching numbers to classify.

An Example for digits that are easy to distinguish: 1 and 0

An Example for digits that are hard to distinguish: 4 and 9

Step 4:



There are 784 pixels overall.

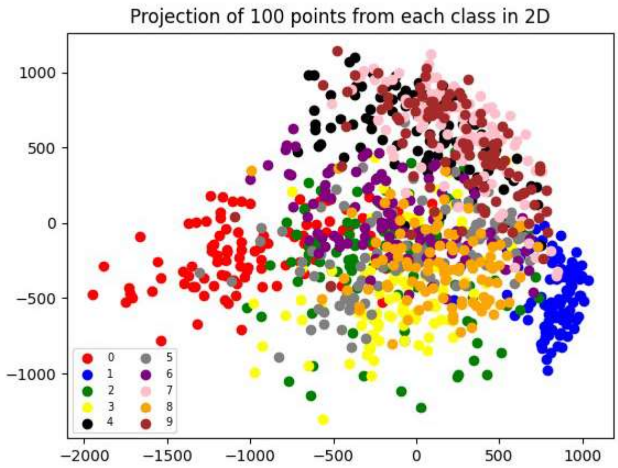
We can see that most of the pixels has low to 0 variance (Around 500, range: 0-4000) and they have low to no impact on the classification decision.

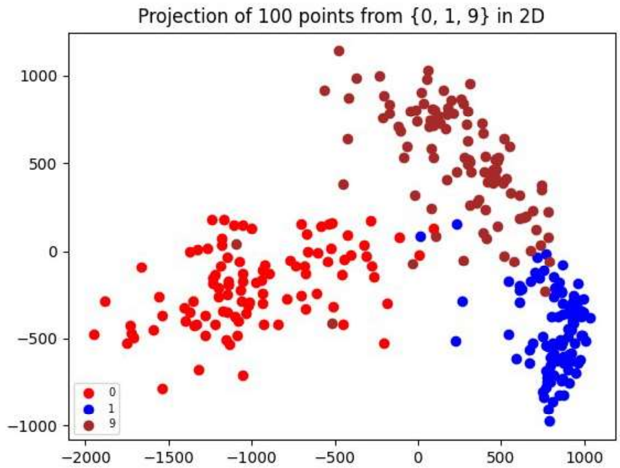
There are another 100 (approx.) pixels which have a medium impact (4000-8000)

And the rest (around 200) have a high impact (>10000).

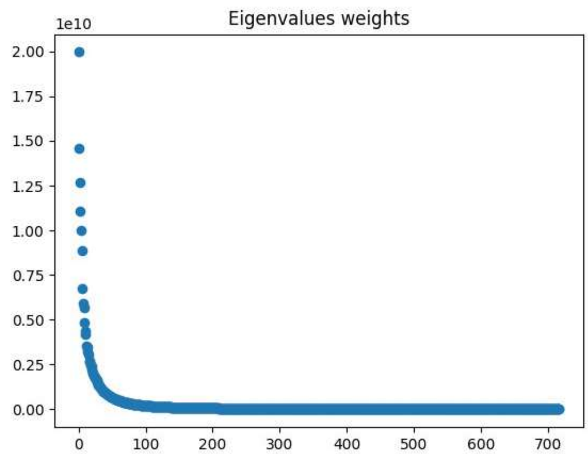
Later in our classification, we'll focus on this last part (actually on its higher end)

Step 5:





Step 6:



We investigated the data and will try to save 80%-85% of the energy (In dimensions: 44-59)

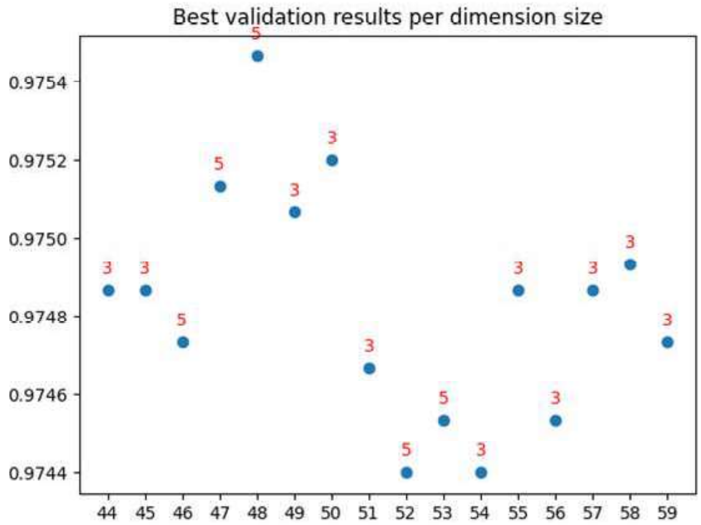
<u>Step 7:</u>

In this step we learnt the best k and dimension (if two combinations gave the same results we preferred the lower dimension option for faster calculations).

```
-----Starting to determine the best parameters-----
Results for 44 dimensions with k = 1 are: 97.38666666666667%
Results for 44 dimensions with k = 2 are: 97.14666666666668%
Results for 44 dimensions with k = 3 are: 97.48666666666666%
Results for 44 dimensions with k = 4 are: 97.34666666666666%
Results for 44 dimensions with k = 5 are: 97.46666666666667%
Results for 44 dimensions with k = 6 are: 97.27333333333334%
Results for 44 dimensions with k = 7 are: 97.2%
Results for 44 dimensions with k = 8 are: 97.13333333333334%
Results for 44 dimensions with k = 9 are: 97.07333333333334%
---The best k for 44 dimensions is 3 with 97.48666666666666% success.---
Results for 45 dimensions with k = 1 are: 97.35333333333334%
Results for 45 dimensions with k = 2 are: 97.06666666666666%
```

By the validation set, we got the best combination is: **Dimension = 48, K = 5** (In red-K)

```
The best overall results (97.54667%) are in 48 dimensions with k = 5
```



Then, we applied our classification on the test set:

```
---Starting the test process---
Done. success rate: 97.48000%
```

Q2:

Classifier for the data in the first column

As we can see from the plot of the training data, the Red class is found on the left side and the Blue class is found on the right side of the plot. However, there are some points that mixed in the border between the classes so we will use a small (But not tiny) C-Value.

For that reason, we used the linear classifier with the parameters C_val = 0.5 & mrg = 1.

This classifier gave us 95% success rate.

Classifier for the data in the second column

In this data set, the Blue class is above the Red class, so we used an RBF classifier.

This time, the value C and mrg will not affect us as much as gamma will.

For that reason, we set the parameters of the RBF classifier to C_val = 1, mrg = 1 & gamma=4

This classifier gave us 97.5% success rate.

Classifier for the data in the third column

In this data set, the Red class is surrounded by the Blue class. That is why we chose the RBF classifier.

To get the best success rate, we chose a big C-Value and a small gamma value.

So as explained, the parameters of the RBF classifier are C_val = 10 & mrg = 1 & gamma=0.1

This classifier gave us 95% success rate.

Classifier for the data in the last column

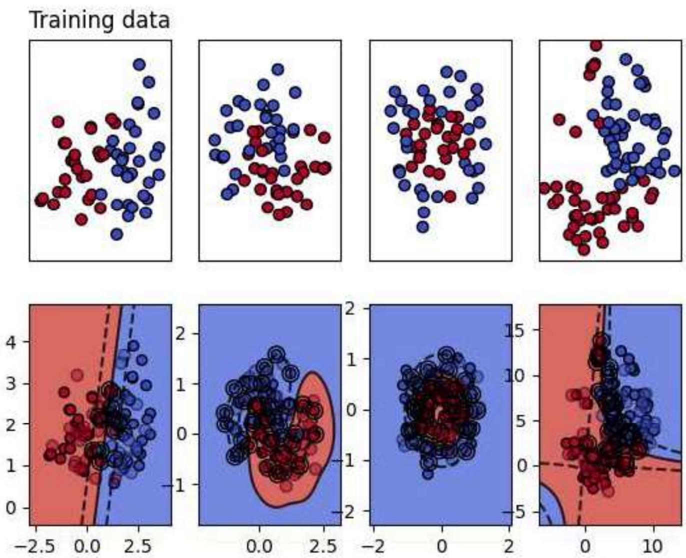
In this data set, the Blue class is in the upper right side of the plot. That is why we chose the poly classifier.

To get the best success rate, we set mrg = 4 to deal with the boundary between the classes

So as explained, the parameters of the poly classifier are C_val = 0.1 & mrg = 4

This classifier gave us 94.8% success rate.

```
linear
0.95
rbf
0.975
rbf
0.95
poly
0.9482758620689655
```



Q3:

Given the kernel normalized linear kernel

$$K(x, x') = \frac{x^T \times x'}{||x^T|| \times ||x'||}$$

According to the input space in Figure1.a, we can see that we have six training points from two classes. Four points from blue class and the other two from red class.

3.1

The feature vector $\phi(x)$ corresponding to this kernel function is

$$K(x, x') = \frac{x^T x'}{||x^T||||x'||} = \frac{x^T}{||x^T||} \times \frac{x'}{||x'||} = \phi(x^T) \times \phi(x')$$

And that is why

$$\phi(x) = \frac{x}{||x||} \to \phi\binom{x_1}{x_2} = \left(\frac{x_1}{||x||}, \frac{x_2}{||x||}\right)$$

3.2

We will use the feature vector above to map the points from input space to their new position

$$\phi(0.2 \quad 0.4) = \left(\frac{0.2}{\sqrt{0.2^2 + 0.4^2}} \quad \frac{0.4}{\sqrt{0.2^2 + 0.4^2}}\right) = (0.447 \quad 0.894)$$

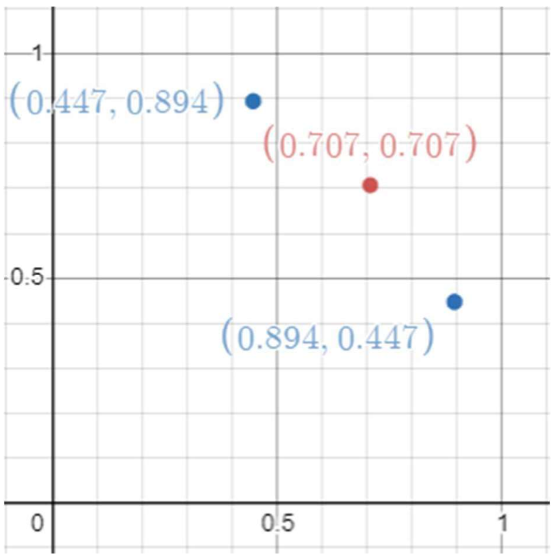
$$\phi(0.4 \quad 0.2) = \left(\frac{0.4}{\sqrt{0.4^2 + 0.2^2}} \quad \frac{0.2}{\sqrt{0.4^2 + 0.2^2}}\right) = (0.894 \quad 0.447)$$

$$\phi(0.4 \quad 0.8) = \left(\frac{0.4}{\sqrt{0.4^2 + 0.8^2}} \quad \frac{0.8}{\sqrt{0.4^2 + 0.8^2}}\right) = (0.447 \quad 0.894)$$

$$\phi(0.8 \quad 0.4) = \left(\frac{0.8}{\sqrt{0.8^2 + 0.4^2}} \quad \frac{0.4}{\sqrt{0.8^2 + 0.4^2}}\right) = (0.894 \quad 0.447)$$

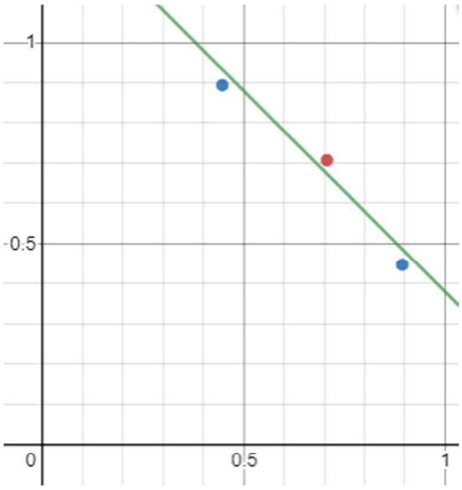
$$\phi(0.4 \quad 0.4) = \left(\frac{0.4}{\sqrt{0.4^2 + 0.4^2}} \quad \frac{0.4}{\sqrt{0.4^2 + 0.4^2}}\right) = (0.707 \quad 0.707)$$

$$\phi(0.8 \quad 0.8) = \left(\frac{0.8}{\sqrt{0.8^2 + 0.8^2}} \quad \frac{0.8}{\sqrt{0.8^2 + 0.8^2}}\right) = (0.707 \quad 0.707)$$



3.3

The maximum margin decision boundary in the feature space will be something like that:



For example: the function $y = -x + 1.38$ (The green line) is a good separator between those classes according to the given data.

3.4

The answer is NO

Explanation: assume we scale each vector by positive number S

$$\phi(Sx) = \left(\frac{Sx_1}{||Sx||}, \frac{Sx_2}{||Sx||}\right) = \left(\frac{Sx_1}{S||x||}, \frac{Sx_2}{S||x||}\right) = \left(\frac{x_1}{||x||}, \frac{x_2}{||x||}\right) = \phi(x)$$

So, as we can see, even after the scaling we will get the same point coordinates.

3.5

As we saw, an example for the decision boundary might be the function

$$y = -x + 1.38$$

Now, if we compare this function to the function $x^2 + y^2 = 1$ we will get 2 points:

$(0.8446, 0.5354)$ and $(0.5354, 0.8446\,)$

These two points are laid on the decision boundary on the feature space and they are already normalized so they will keep their position at the input space and at the feature space.

Now, to find the decision boundary in the input space, we need to find all the points in the input space that are mapped to these two intersection points after we use the feature vector as we explained above.

Let p' be the point in the feature space so we can say that $p' = \frac{p}{||p||}$

Now consider $x' = \frac{x}{\sqrt{x^2+y^2}}$ and $y' = \frac{y}{\sqrt{x^2+y^2}}$

So, we will get the line $y = \frac{y\prime}{x\prime}x$

Let's calculate the decision boundary represented by the point $(0.8446, 0.5354)$

$$x' = \frac{x}{\sqrt{x^2+y^2}} = \frac{0.8446}{\sqrt{0.8446^2 + 0.5354^2}} = 0.8445$$

$$y' = \frac{y}{\sqrt{x^2+y^2}} = \frac{0.5354}{\sqrt{0.8446^2 + 0.5354^2}} = 0.5353$$

And the decision boundary is $y = \frac{y\prime}{x\prime}x = \frac{0.5353}{0.8445}x = 0.6338x$

If we do the same process for the second point, we will get the line $y = \frac{y\prime}{x\prime}x = \frac{0.8445}{0.5353}x = 1.5776x$

That is why the decision boundary in the original input space is