1.1.

Given that x = $(x_1, x_2)^T$

$\emptyset(\mathrm{x}) \;=\; (x_1^2 , \sqrt{2}\, x_1 x_2 , x_2^2)^T$

We know that K(x,z) = $\emptyset(\mathrm{x}).\emptyset(\mathrm{z})$

K(x,z) = $(x_1^2 , \sqrt{2}\, x_1 x_2 , x_2^2)^T . (z_1^2 , \sqrt{2}\, z_1 z_2 , z_2^2)^T$

= $(x_1^2\, z_1^2 + \sqrt{2}\, x_1 x_2 \sqrt{2}\, z_1 z_2 + x_2^2\, z_2^2)^T$

K(x,z) = $(x_1 z_1 \;+\; x_2 z_2)^2$

K(x,z) = $(x.z)^2$

1.2

A. If we map the input vector to the feature space and then do the dot product on the mapped features. We have

$\emptyset(\mathrm{x}).\emptyset(\mathrm{z})$ = $(x_1^2 , \sqrt{2}\, x_1 x_2 , x_2^2)^T . (z_1^2 , \sqrt{2}\, z_1 z_2 , z_2^2)^T$

= $x_1^2\, z_1^2 + \sqrt{2}\, x_1 x_2 \sqrt{2}\, z_1 z_2 + x_2^2\, z_2^2$

In the above equation we have 3 + 5 + 3 = 11 multiplications and 2 additions.

B. If we use the equation I got in question 1

K(x,z) = $(x_1 z_1 \;+\; x_2 z_2)^2$

If we consider m = $x_1 z_1 \;+\; x_2 z_2$ then we have K(x,z) = m.m and in m we have 2 multiplications and 1 addition, so for K(x,z) we will have 3 multiplications and 1 addition, where 2 multiplications are needed for m and one multiplication needed for squaring m.

Actually both A and B refer to the same equation A is expansion of equation B so we actually should get the same no of multiplications and additions for both.

2.1.

K(x,x') can be written as floor($\frac{x.x'}{||x||*||x'||}$) which is 1 when x=x' else 0. $\emptyset(\mathrm{x})$ should be of the form [1 0 0……] it should be an m dimensional vector there only one element is 1 and rest all are zero, only then we can have $\emptyset(\mathrm{x}).\emptyset(\mathrm{x'})$ as 1 when x=x' else 0. The closest function I can get for this scenario is $\emptyset(\mathrm{x})$ = $\frac{x}{||x||}$, since dimension of X is not given I am assuming it to be m-dimension.

2.2

The given kernel is a linear kernel since it is represented as K(x,z) = x.z and since this is a valid kernel we will have a linear separator which can classify the given data. We know that w.x can be written as w.x = $a_1 K(x_1, x) \;+\; a_2 K(x_2, x) + \ldots\ldots\ldots$ if K(x,x') is x.x' then w.x will be a linear equation and hence we get a linear separator.

2.3

This will be a bad idea when m is very large or dimension of X is smaller than m. It will be computationally expensive and also will be hard to find a function which maps accordingly.

3.

Optimal Hyper Parameters are

Learning rate = 0.0001

No of epochs = 50

Regularization = 0.85

Training Accuracy = 92.531

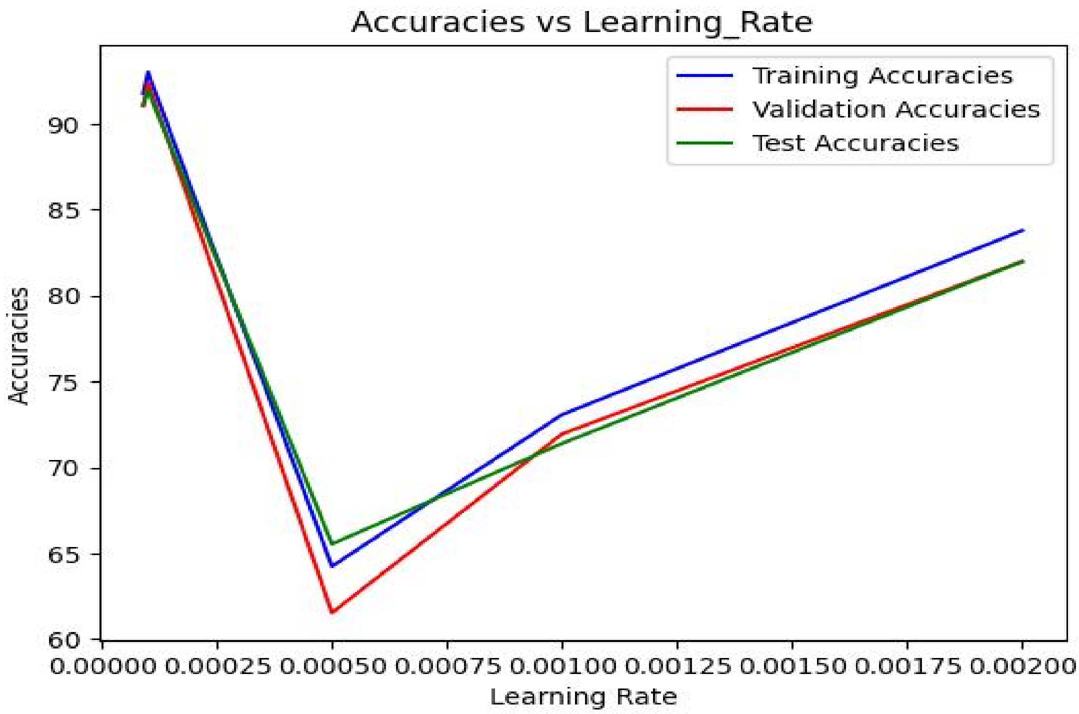
Validation Accuracy = 91.630

Testing Accuracy = 92.061

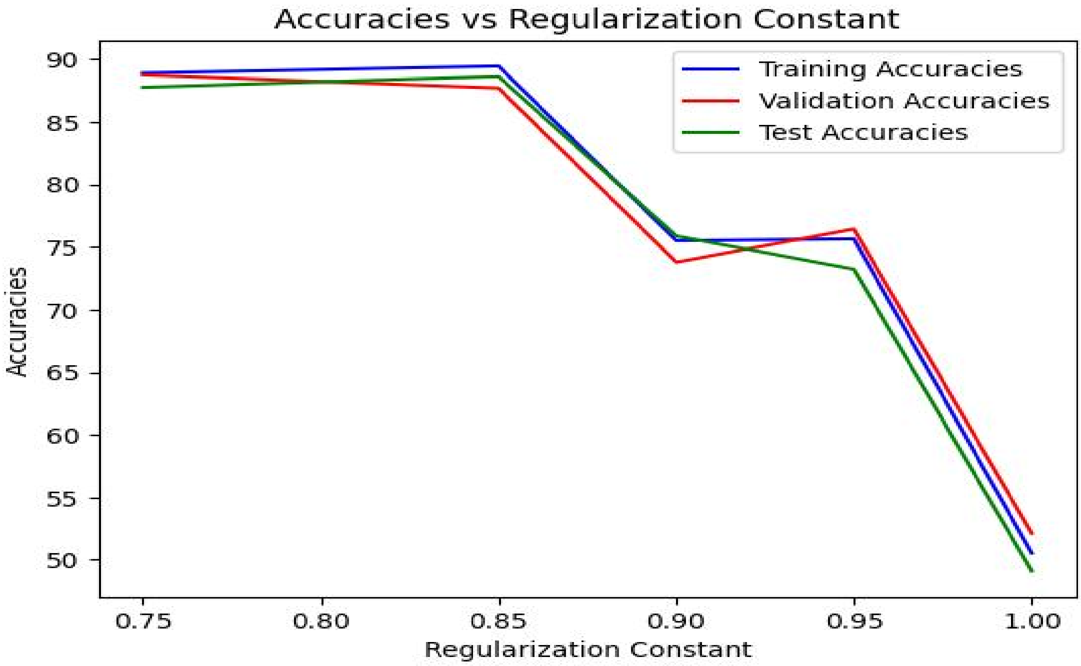
The code for the same is attached

The plots are as follows

Accuracies vs Learning rate



Accuracies vs Regularization Constant


Accuracies vs Regularization Constant
Training Accuracies
Validation Accuracies
Test Accuracies
Accuracies
90
85
80
75
70
65
60
55
50
0.75
0.80
0.85
0.90
0.95
1.00
Regularization Constant