Kernel Methods Assignment

Srishty Suman
srishtysuman@iisc.ac.in
16592

June 7th, 2020

Answer 1: Valid Kernels

(a) For $x^T Ay$ to be a valid kernel, the kernel matrix generated for x and y should be symmetric and positive semi-definite. i.e.;

$$x^T Ay \geq 0 \ \ \forall x, y \in \mathcal{R}^d \tag{1}$$

and if the above equation holds, then matrix A is positive semi-definite (from the property of positive semi-definite matrices).
Also, for k to be valid kernel, it has to be symmetric, i.e.;

$$x^T Ay = y^T Ax \ \ \forall x, y \in \mathcal{R}^d \tag{2}$$

Let's take d=2, x=[x_1 x_2], y=[y_1 y_2], A=[[a b],[c d]] be 2-by-2 matrix, which is positive semi-definite then,

$$x_1 a y_1 \ + \ x_2 c y_1 \ + \ + x_1 b y_2 \ + \ x_2 d y_2 = x_1 a y_1 \ + \ x_1 c \, y_2 \ + \ x_2 b y_1 \ + \ x_2 d y_2 \tag{3}$$

We can clearly see that for equation (3) to hold true, b==c, which means that matrix A is symmetric as well. (This can be generalized to higher dimension. For simplicity, I took smaller dimension).
Hence, for A to be a kernel, the necessary and sufficient condition on A is to be it a symmetric and positive semi-definite. i.e.; If A is PSD and symmetric, k is a kernel and if k is a kernel, then A is symmetric and PSD.

(b) **(a)** No. It is symmetric but not positive semi-definite.

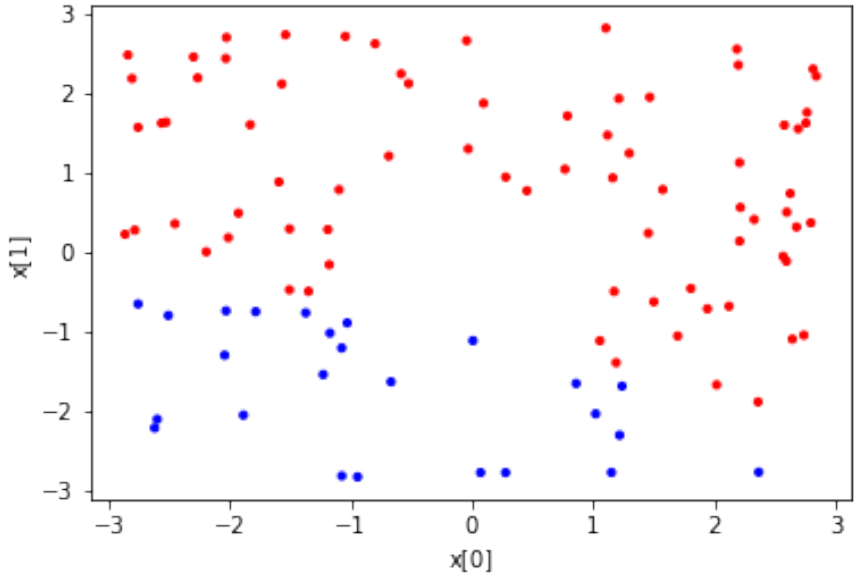
(b) Yes. It is likely to be a valid kernel, as empirically, it's smallest eigenvalues is getting calculated as -3.094884111823319e-07+0j, which can be treated as almost 0. If the smallest eigenvalue is almost 0, rest will be greater than or equal to 0 only. This implies that all eigenvalues are positive only, and hence the matrix is positive semi-definite. It is symmetric as well.

(c) No. It is neither symmetric nor positive semi-definite.

(d) No. This is symmetric but not positive semi-definite.

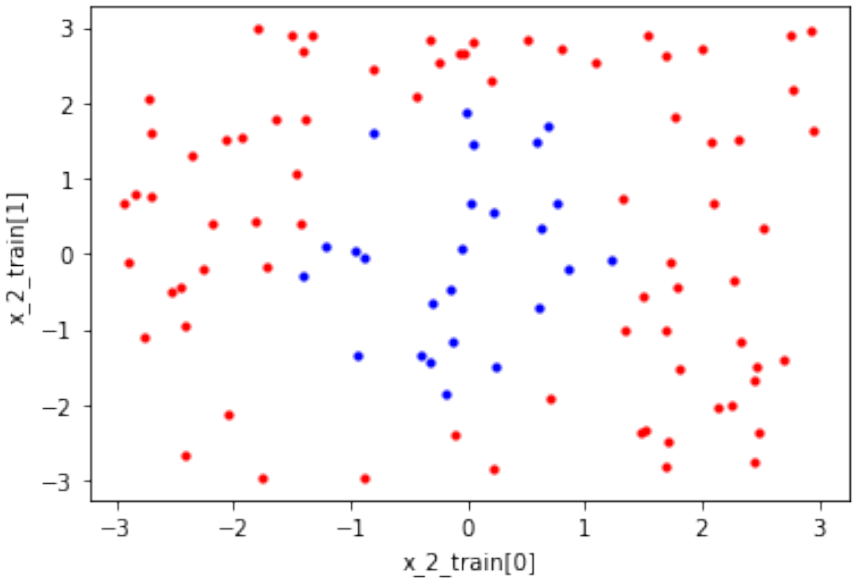
(e) No. This is symmetric but not positive semi-definite.

Answer 2: Support Vector Machines



(a)
Scatter Plot for data generated as per question 2a.

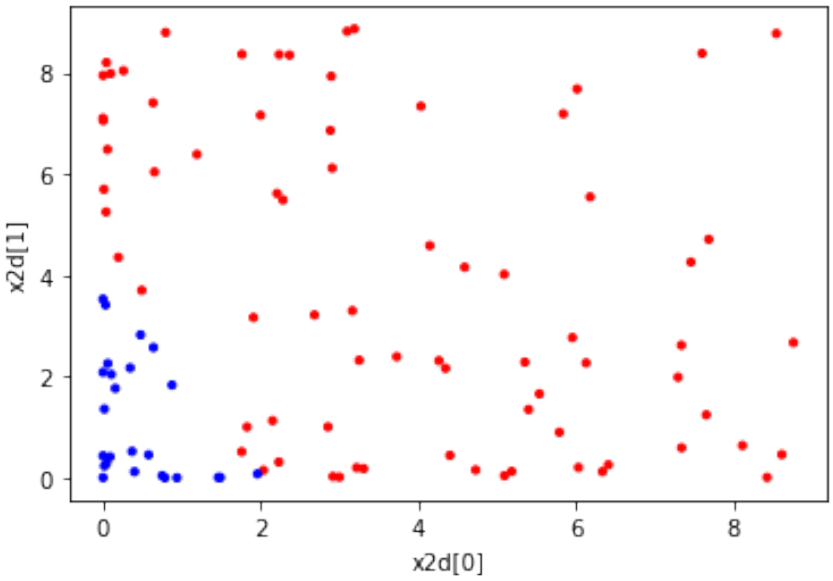
(b) Accuracy on training set is: [0.99]
Optimal weight vector w is: [-0.49246572 -1.44775596]
Optimal intercept b is: [-1.69477202]
Optimal value of C is: [0.1001]
Accuracy on test set is: [1.0]



(c) **(i)**

(ii) Accuracy on training set is: [0.75]
Optimal weight vector w is: [-4.09849796e-14 -8.06076385e-14]
Optimal intercept b is: [-1.]

Optimal C is: [5.9001]
test Accuracy is: 0.62



(d) **(i)**

(ii) Training accuracy is: [1.]
Optimal weight vector w is: [-3.23001388 -1.54467185]
Optimal intercept b is: [6.45432954]
Optimal C is: [1.4001]
Accuracy on test set is 0.86

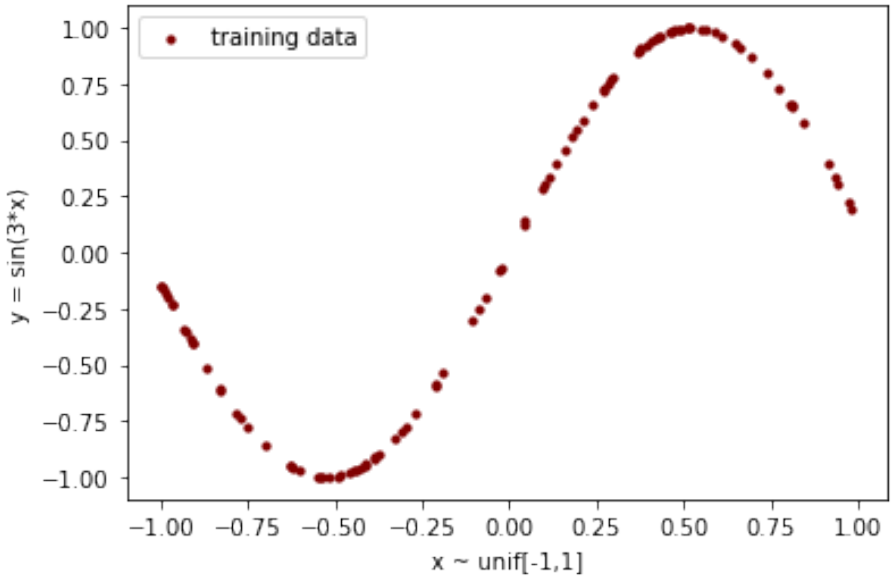
(iii) Yes, the performance is better as compared to that in 2(c). This is because, the data was not linearly seperable in the previous dimensions. So, we did feature transform to project the data to higher dimension, and there it got linearly seperable.

(e) I used RBF kernel. Expression for the same is as follows:

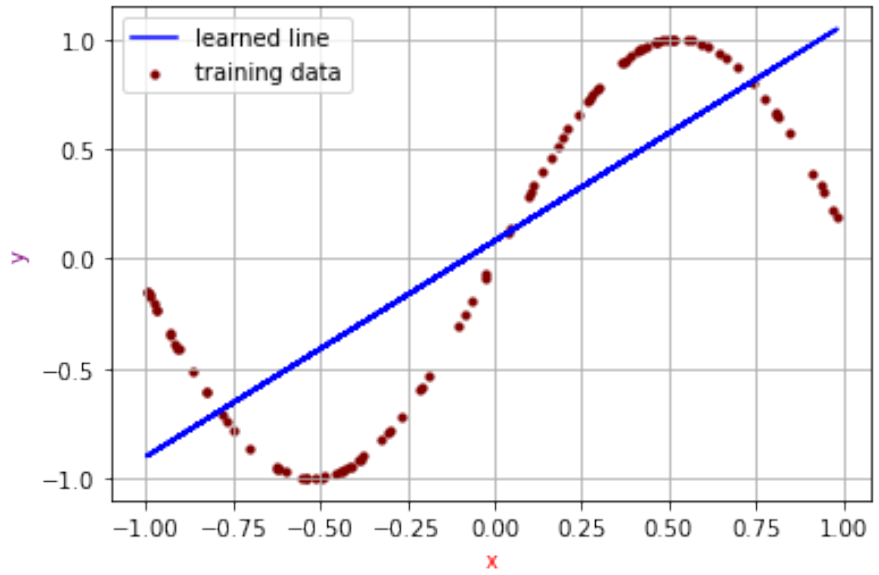
$$K[x, y] = np.exp(-gamma * np.linalg.norm(x - y) * *2) \tag{4}$$

where gamma is hyperparameter to be tuned to get optimal performance.
Accuracy on training set is: [0.96]
Optimal weight vector w is: [-0.11844115 -0.09463962]
Optimal b is: [0.33433062]
Optimal C is: [0.001396]
gamma: 0.5
Accuracy on test set is: 0.94

Answer 3: Kernelised-Regression



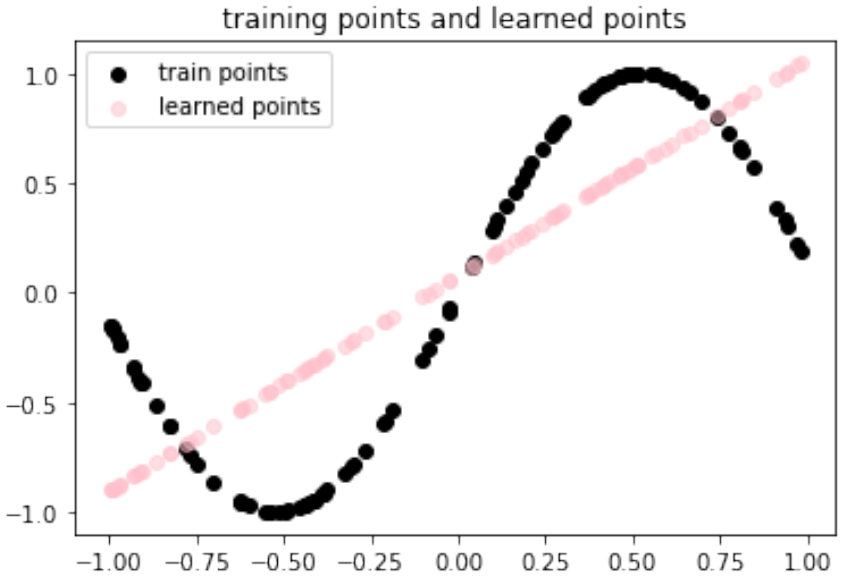
(a)



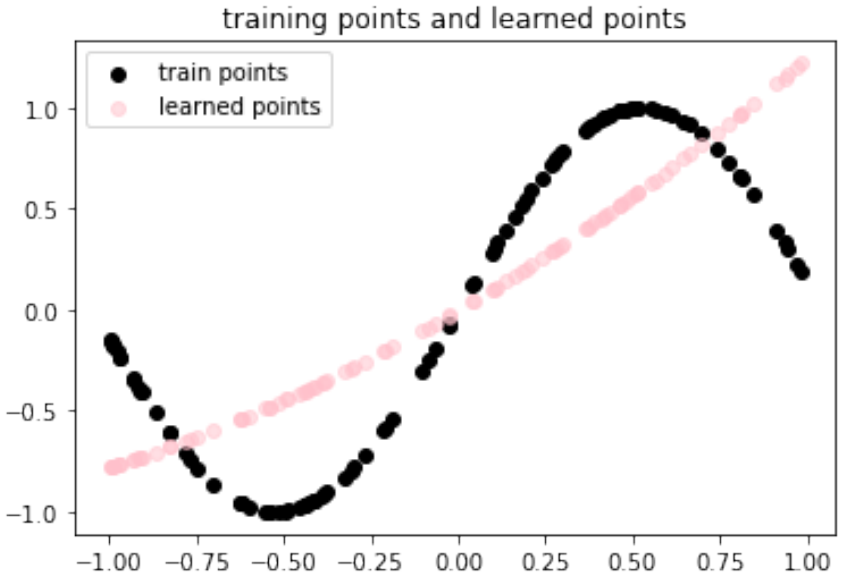
(b)

mean squared error is: [0.15882603]

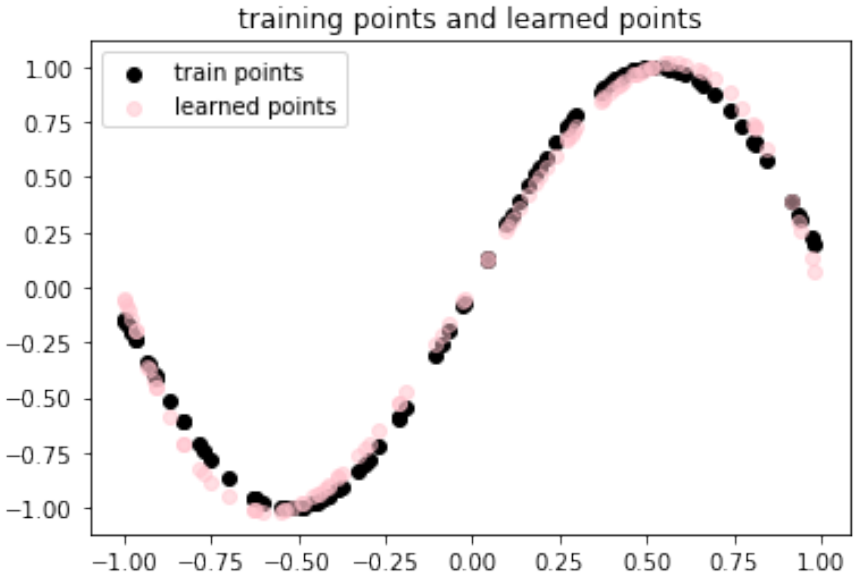
(c) for iteration 1, the plot obtained is as follows:



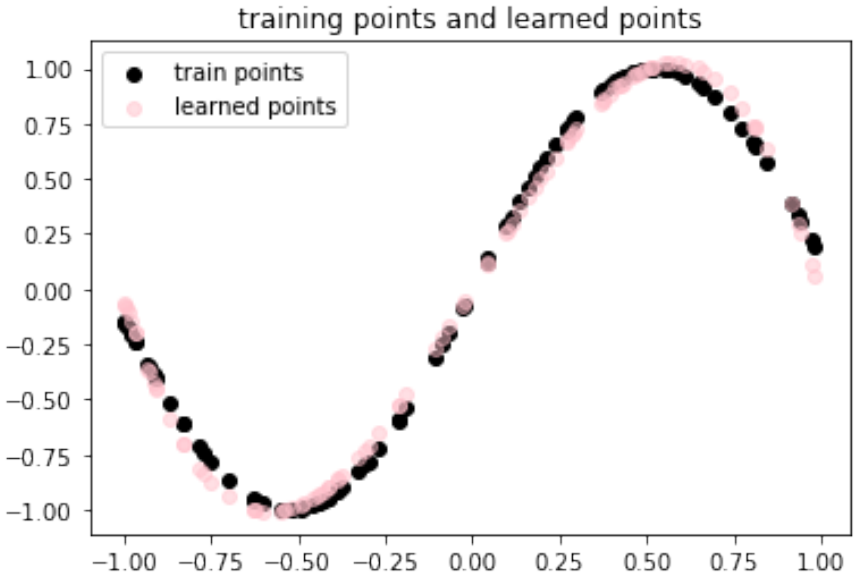
for iteration 2, the plot obtained is as follows:



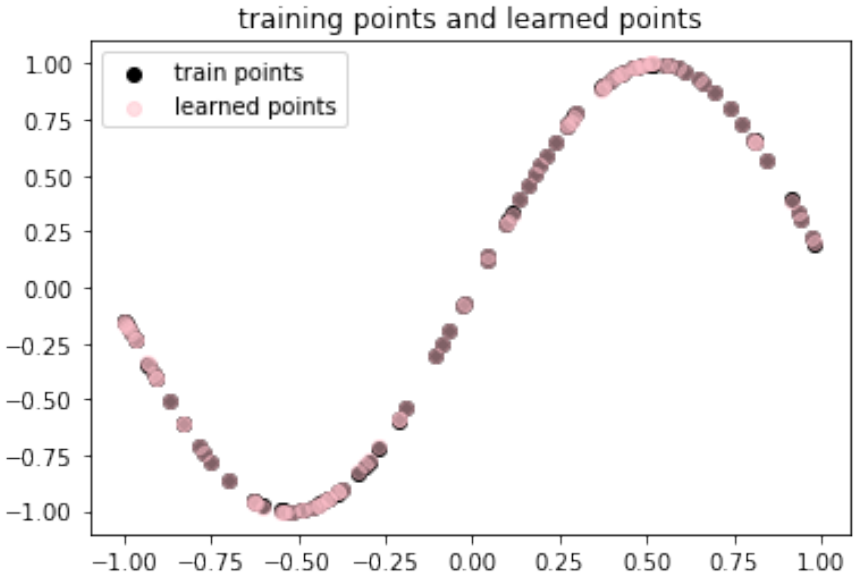
for iteration 3, the plot obtained is as follows:



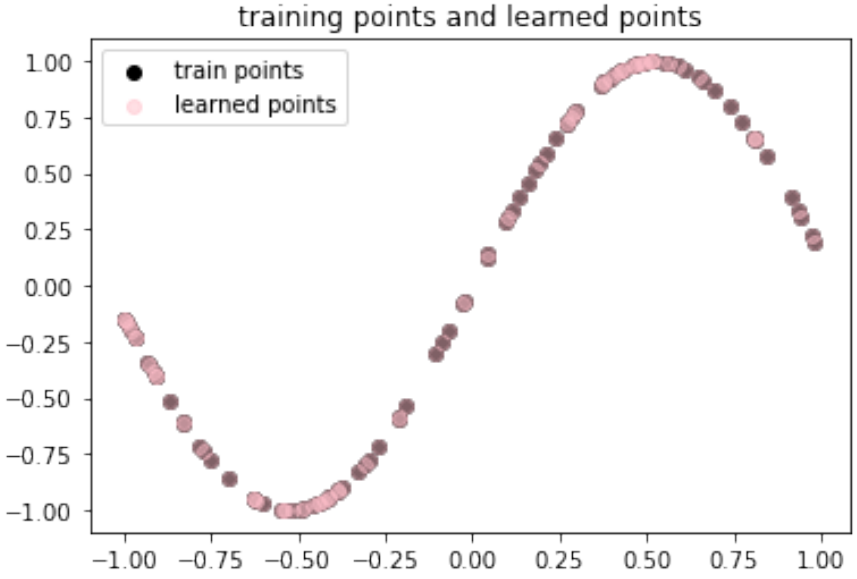
For iteration 4, the plot is as:



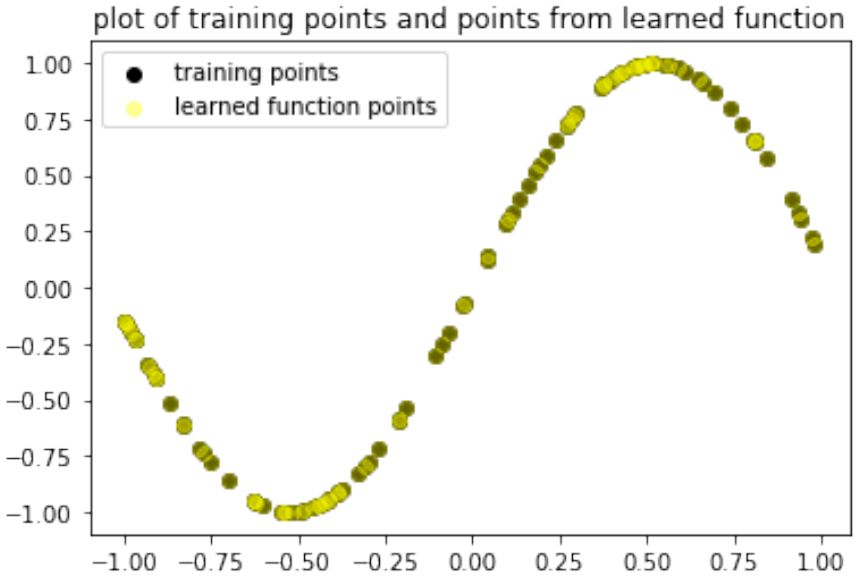
plot for iteration 5:



There was not much changes hereafter, hence not adding plots for iteration 7,8,9. for iteration 10, the plot obtained is as follows:



(d) Kernel used by me: RBF kernel.
Training set mean squared error: 1.2619395213471602e-07 for gamma = 50000002



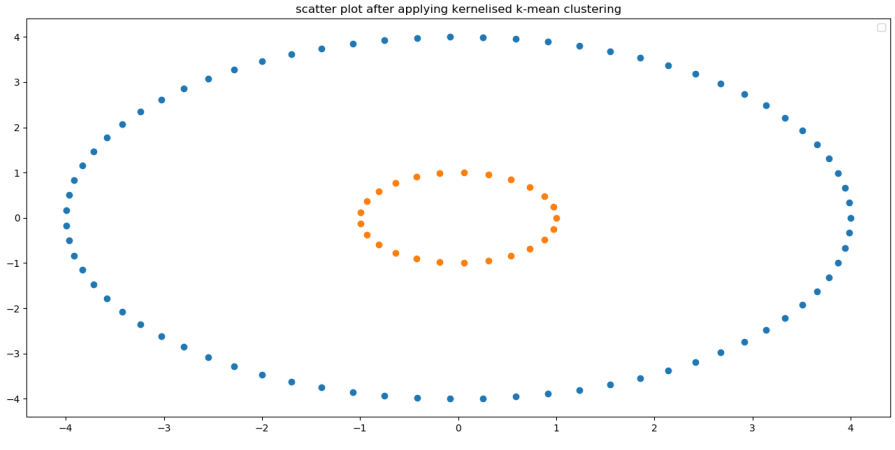
test set mean squared error: 0.010106092824380162

Answer 4: Kernel K-Means

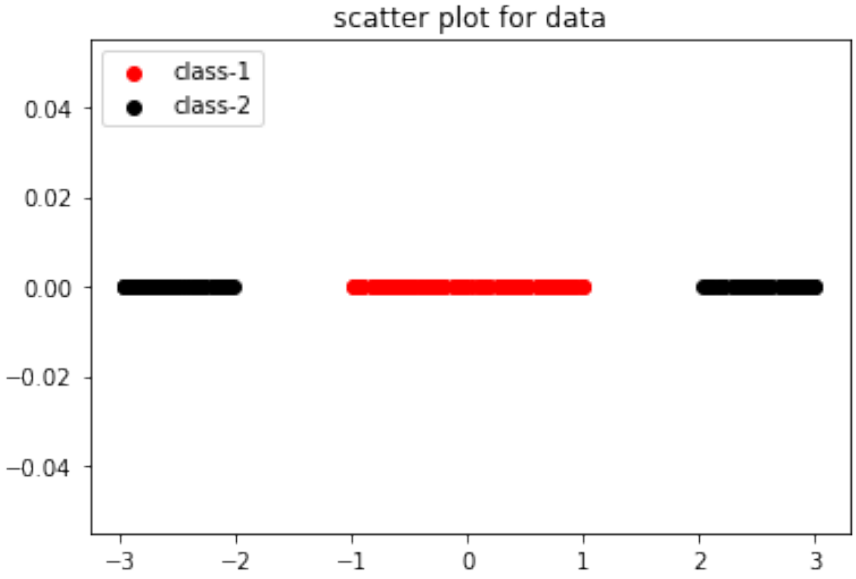
(a) Sum of all entries of D for d=10 is 0.0 .

(b) Here also, sum is 0.0 only.

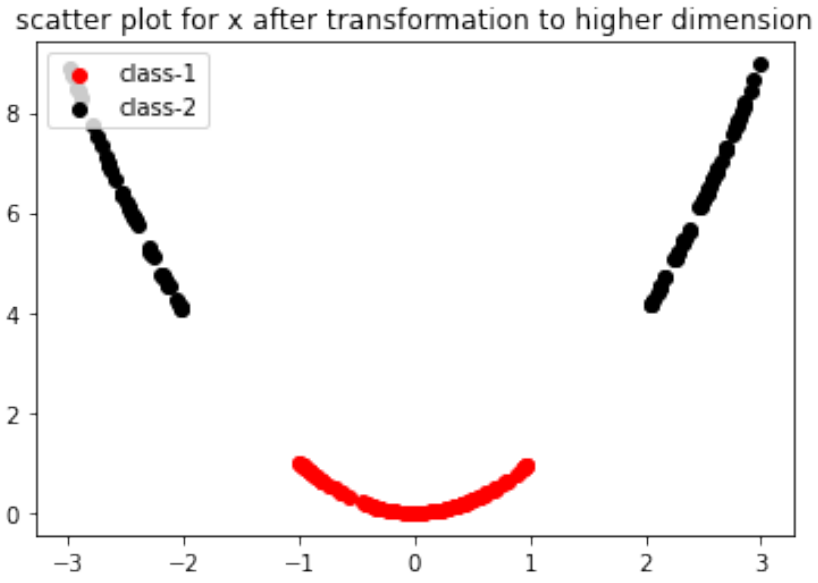
(c)



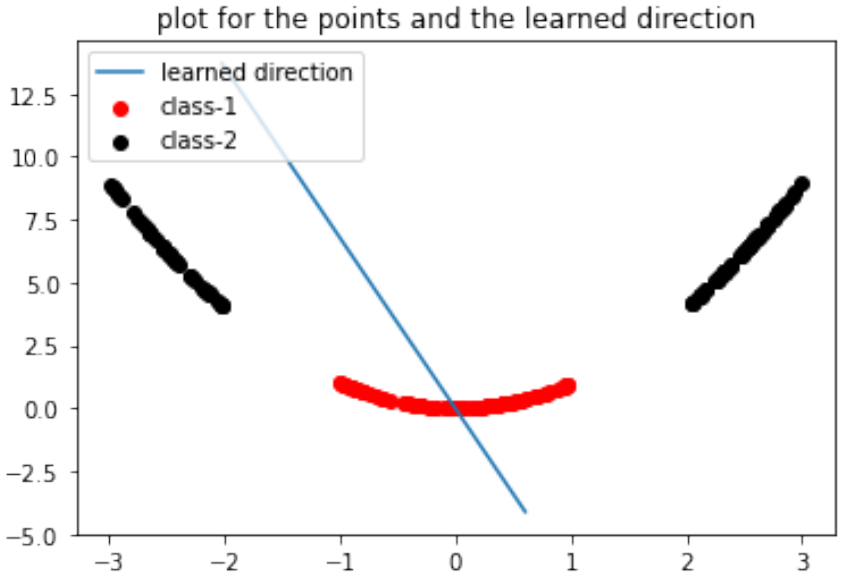
Answer 5: Kernel Fischer's Discriminant Analysis



(a)

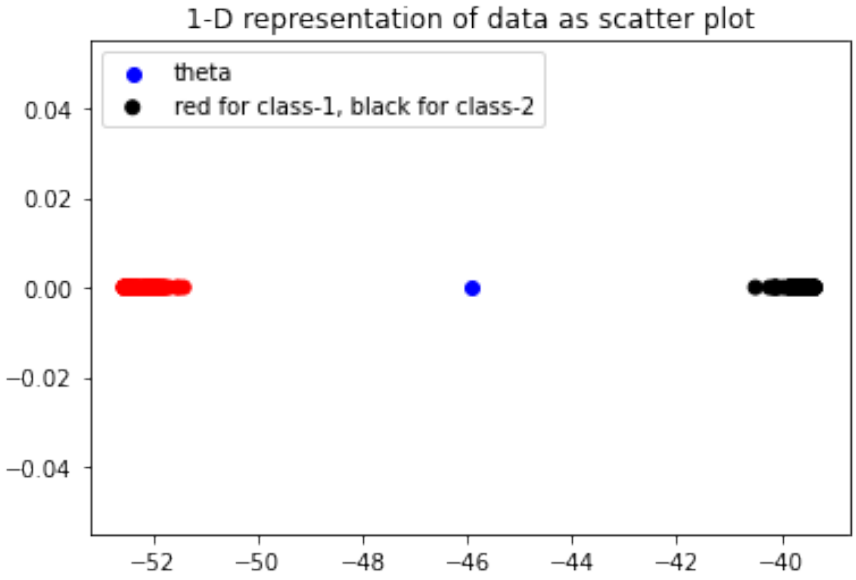


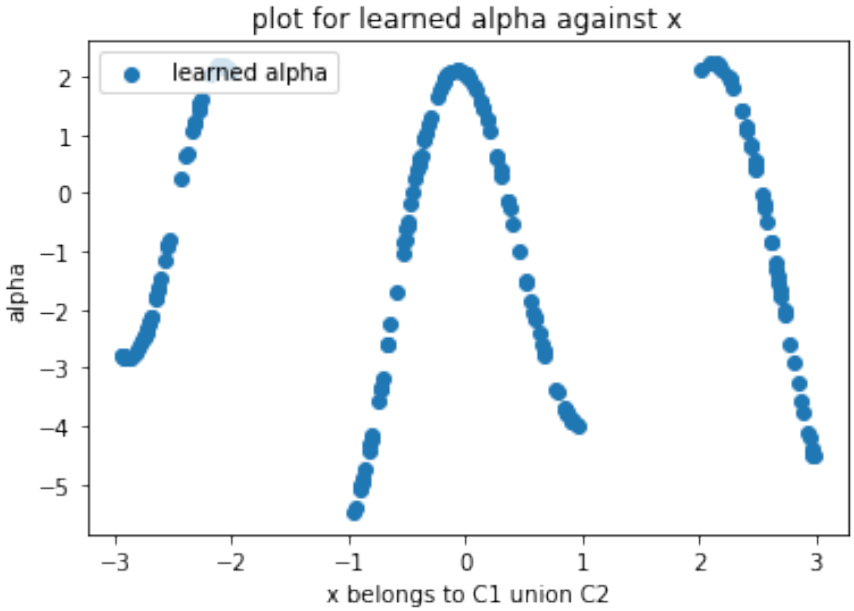
(b)



(c)

(d) I used RBF kernel. Classification accuracy is 1.0. Plot for 1d representation of data is as follows:




plot for learned alpha against x
learned alpha
alpha
x belongs to C1 union C2
2
1
0
−1
−2
−3
−4
−5
−3
−2
−1
0
1
2
3