Assignment 1
07/06/2020

Ravi Raja raviraja@iisc.ac.in
SR-17054

1. Question 1: Valid Kernels

1.1 Q1 a.

A *Valid Kernel* is the kernel function whose gram matrix G is Symmetric and Positive Semi-Definite.

Given: *k(x,y)* = $\mathrm{x^T Ay}$

Gram Matrix G is defined as

$$G_i j = k(x_i, x_j)$$

, where

$$x_i, x_j \in R^d$$

For Gram matrix G to be Symmetric:

$$G = G^T$$

should hold. i.e.

$$x^T Ay = y^T A^T x$$

$$x^T Ay = \begin{bmatrix} x_1 & x_2 & ... & x_d \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & ... & a_{1d} \\ a_{21} & a_{22} & ... & a_{2d} \\ ... & ... & ... & ... \\ a_{d1} & a_{d2} & ... & a_{dd} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_d \end{bmatrix} = \sum_{i=1}^{d} \sum_{j=1}^{d} x_i y_j a_i j \tag{1}$$

$$y^T A^T x = \begin{bmatrix} y_1 & y_2 & ... & y_d \end{bmatrix} \begin{bmatrix} a^t_{11} & a^t_{12} & ... & a^t_{1d} \\ a^t_{21} & a^t_{22} & ... & a^t_{2d} \\ ... & ... & ... & ... \\ a^t_{d1} & a^t_{d2} & ... & a^t_{dd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_d \end{bmatrix} = \sum_{i=1}^{d} \sum_{j=1}^{d} y_i x_j a^t_{ij} \tag{2}$$

By equating (1) and (2) we get,

$$\sum_{i=1}^{d} \sum_{j=1}^{d} x_i y_j a_i j = \sum_{i=1}^{d} \sum_{j=1}^{d} y_i x_j a^t_{ij}$$
$$\implies a_{ij} = a^t_{ij}$$
$$\implies A = A^T$$

For Gram Matrix G to be Positive Semi-Definite

$$v^T G v \geq 0$$

, where v $\in$R^d

As G is symmetric it can be written in the form of spectral decomposition

$$vTGv = v^T U diag(\lambda) U^T v = \sum_{i=1}^{d} \lambda_i (v^T U)^2$$

where U is d x d orthogonal matrix containing eigen vectors of G and $\lambda_i^{\prime s}$ are the d eigen values of G.

The above equation is non-negative only when $\lambda_i^{\prime s}$ are all non-negative.

Now let's find out conditions on A for G to be psd

$$\begin{aligned}
v^T G v &\geq 0 \\
\implies \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j G_i j &\geq 0 \\
\implies \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j k(x, y) &\geq 0 \\
\implies \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j (x^T A y) &\geq 0 \\
\implies \sum_{i=1}^{d} v_i^2 (x^T A y) &\geq 0 \\
\implies x^T A y &\geq 0
\end{aligned}$$

$$\implies \text{A should be positive semidefinite}$$

1.2 Q1. b

(i) Not Valid
(ii) Valid
(iii) Not Valid
(iv) Not Valid
(v) Not Valid

2. Question 2: Support Vector Machines

2.1 Q2. a

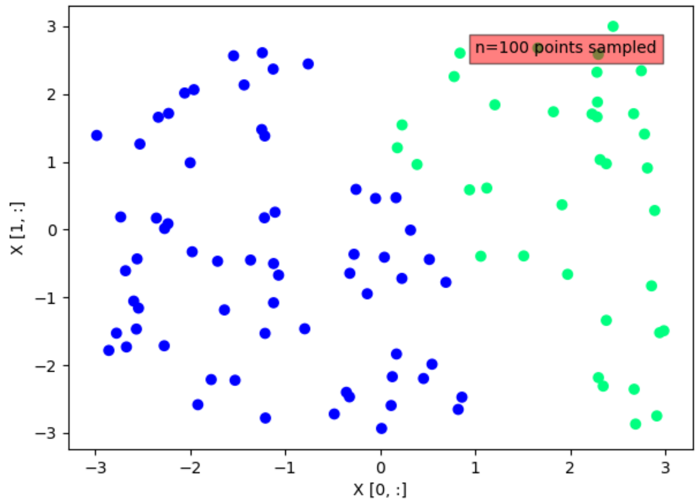

Figure 1: Plot for Question 2. a

2.2 Q2. b

Accuracy

$$TrainingAccuracy = 0.97$$

$$TestAccuracy = 0.9$$

Optimal Values of W b and C

$$W = \begin{bmatrix} 1.58e-04 \\ 6.17e-05 \end{bmatrix}$$

$$b = -0.00014097946258622763$$

$$C = 0.0001$$

2.3 Q2. c

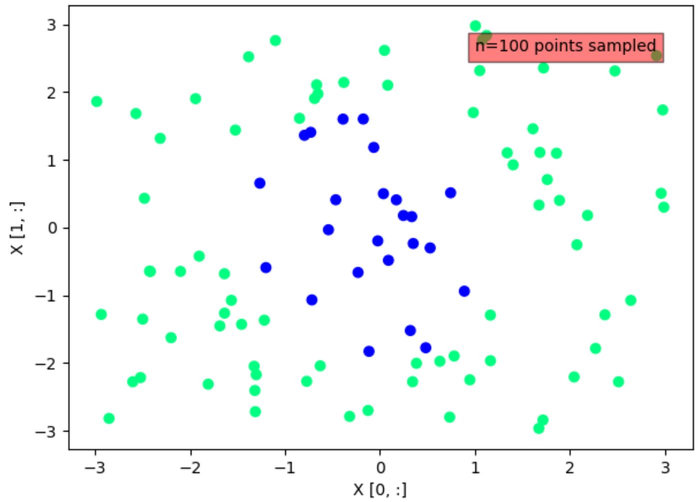

Figure 2: Plot for Question 2. c

Accuracy

$$TrainingAccuracy = 0.76$$

$$TestAccuracy = 0.64$$

Optimal Values of W b and C

$$W = \begin{bmatrix} 3.83e-10 \\ -3.09e-10 \end{bmatrix}$$

$$b = 2.2546004931838955e-08$$

$$C = 0.99$$

2.4 Q2. d

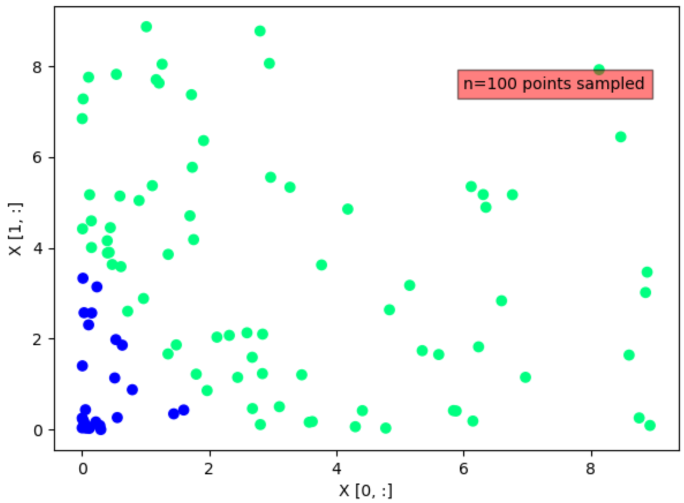

Figure 3: Plot for Question 2. d

Accuracy

$$TrainingAccuracy = 0.97$$

$$TestAccuracy = 0.92$$

Optimal Values of W b and C

$$W = \begin{bmatrix} 1.20e-04 \\ 6.86e-05 \end{bmatrix}$$

$$b = -0.00021143550955607615$$

$$C = 0.00015$$

Yes, the performance is better from what is reported in Q2. c.
This is because the data generated in Q2. c is not linearly separable. In Q2. d Non linearity is introduced in the data by using a mapping function ϕ

2.5 Q2. e

Expression for kernel

$$k(x, y) = \exp(-(||x - y||^2/(2 * \sigma^2)))$$

Accuracy

$$TrainingAccuracy = 1.0$$

$$TestAccuracy = 0.6$$

Hyperparameters

$$\sigma = 5.0$$

$$C = 0.1$$

3. Question 3: Kernelized-Regression

3.1 Q3. a

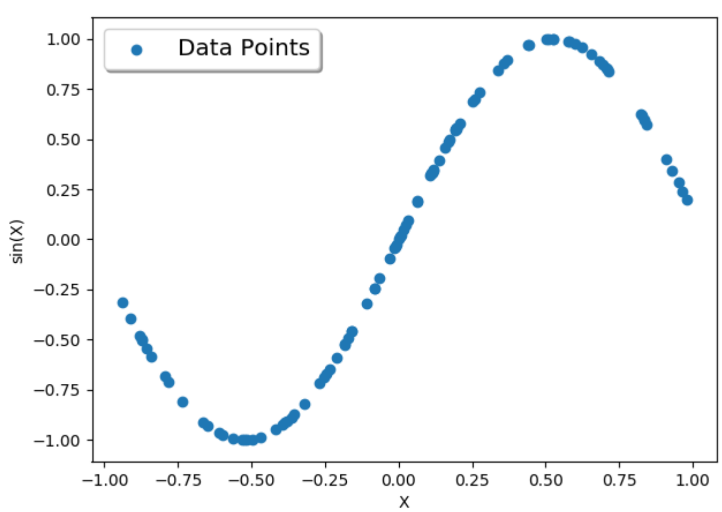

Figure 4: Plot for Question 3. a

3.2 Q3. b

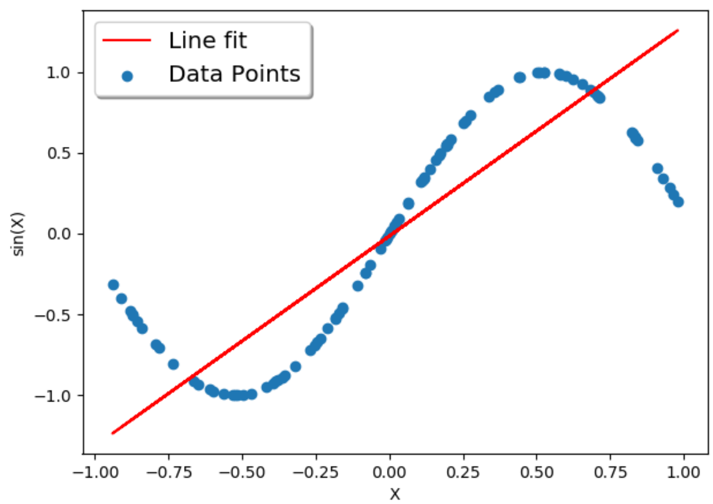

Figure 5: Plot for Question 3. b

Mean Squared Error = 14.898374467598645

3.3 Q3. c

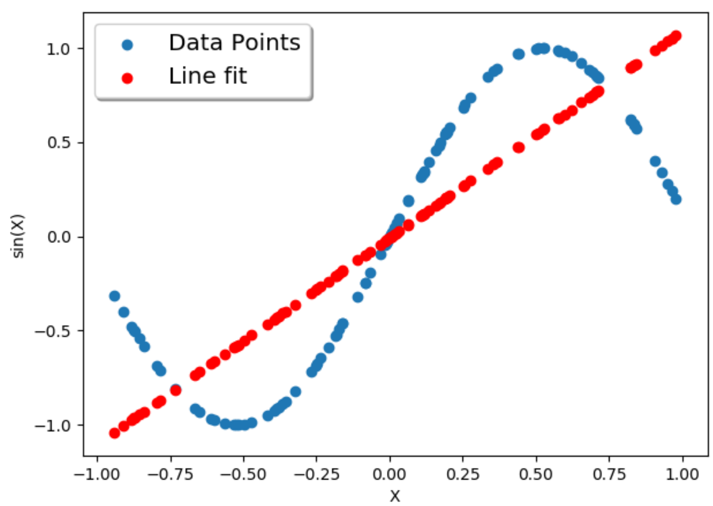

Figure 6: Plot for Question 3. c with k =1

Mean Squared Error = 13.819606456519109

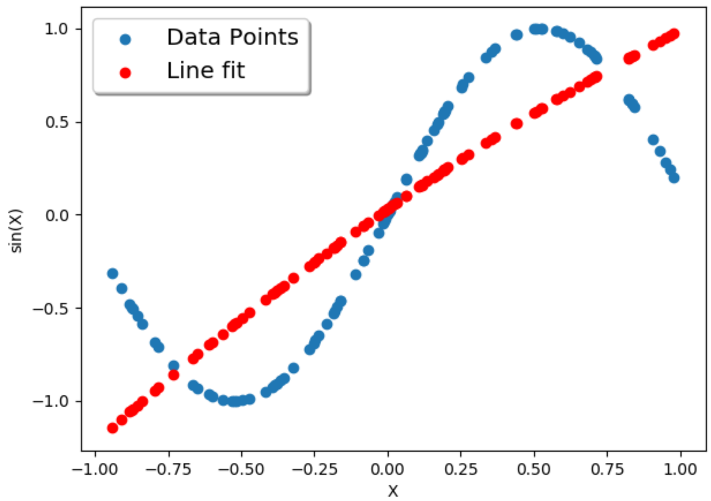

Figure 7: Plot for Question 3. c with k = 2

Mean Squared Error = 13.636266256819166

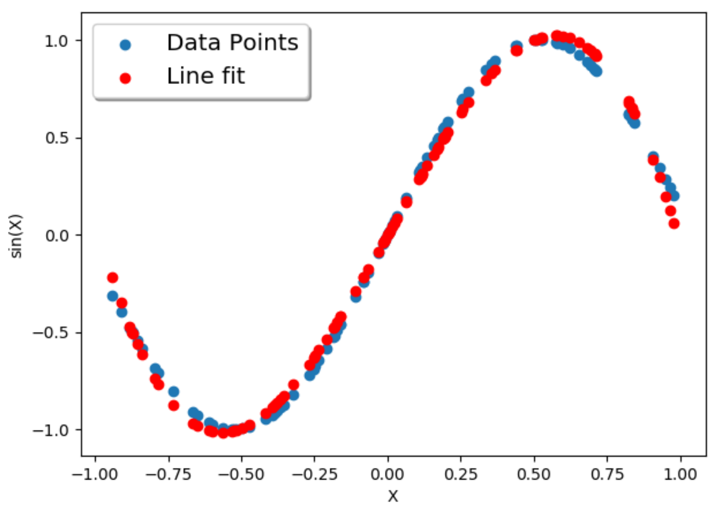

Figure 8: Plot for Question 3. c with k = 3

Mean Squared Error = 0.22155907938819272

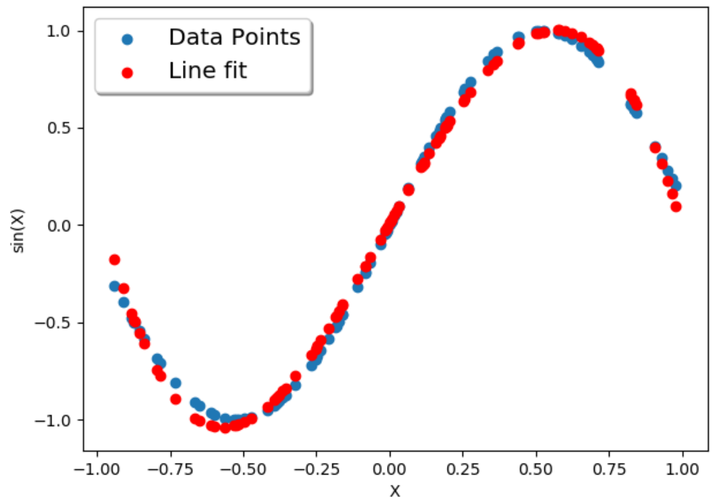

Figure 9: Plot for Question 3. c with k = 4

Mean Squared Error = 0.19669492545897022

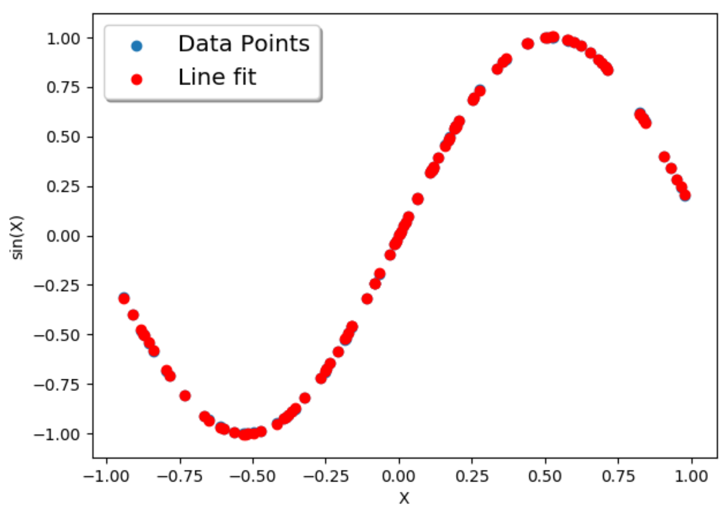

Figure 10: Plot for Question 3. c with k = 5

Mean Squared Error = 0.0007461556221027676

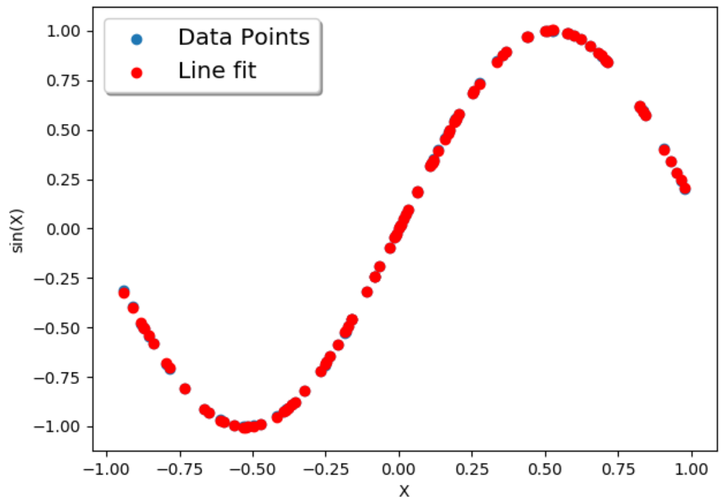

Figure 11: Plot for Question 3. c with k = 6

Mean Squared Error = 0.0006701207085641931

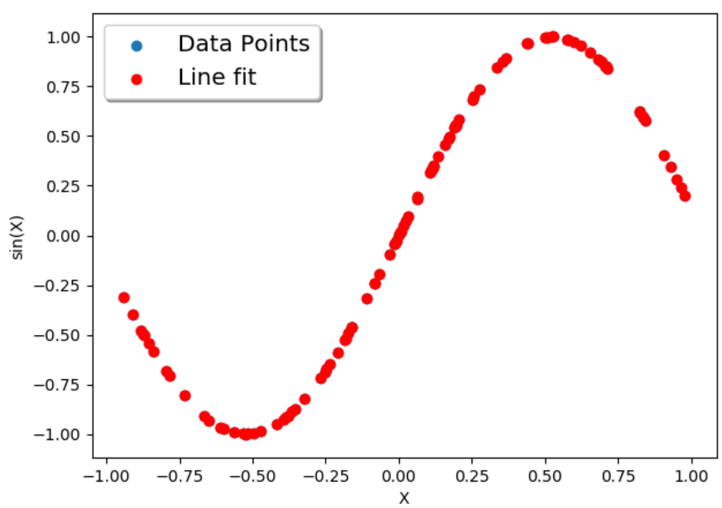

Figure 12: Plot for Question 3. c with k = 7

Mean Squared Error = 6.286393008119336e-07

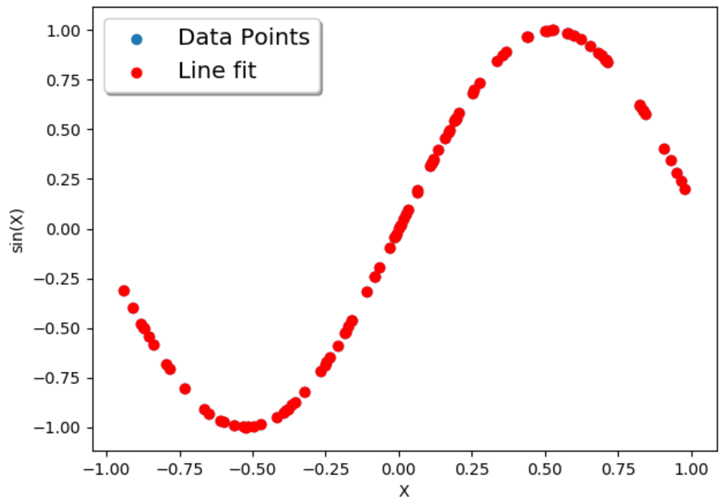

Figure 13: Plot for Question 3. c with k = 8

Mean Squared Error = 5.726874175515086e-07

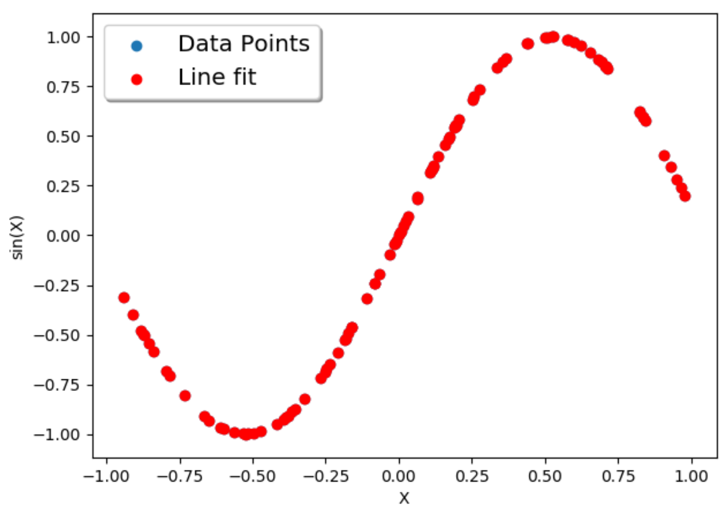

Figure 14: Plot for Question 3. c with k = 9

Mean Squared Error = 2.597737764449787e-10

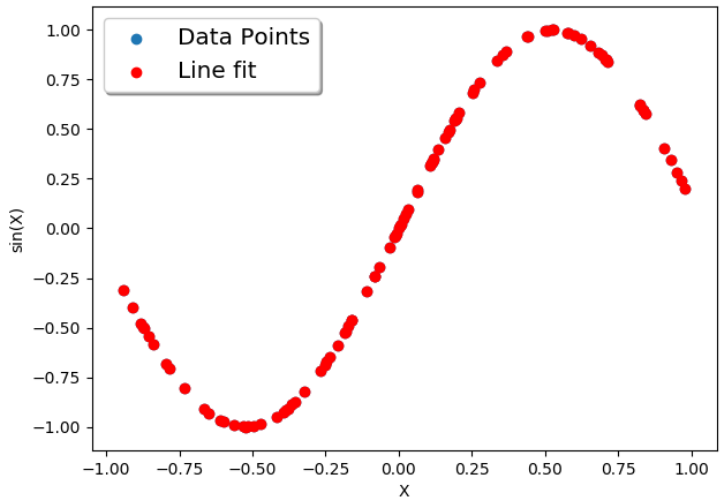

Figure 15: Plot for Question 3. c with k = 10

Mean Squared Error = 2.2614191795620724e-10

3.4 Q3. d

Expression for kernel

$$k(x,y) = \exp(-(||x-y||^2/(2*\sigma^2)))$$

Training set Mean Squared Error = 2.9779655984849057e-22
Test set Mean Squared Error = 92.0351748212746

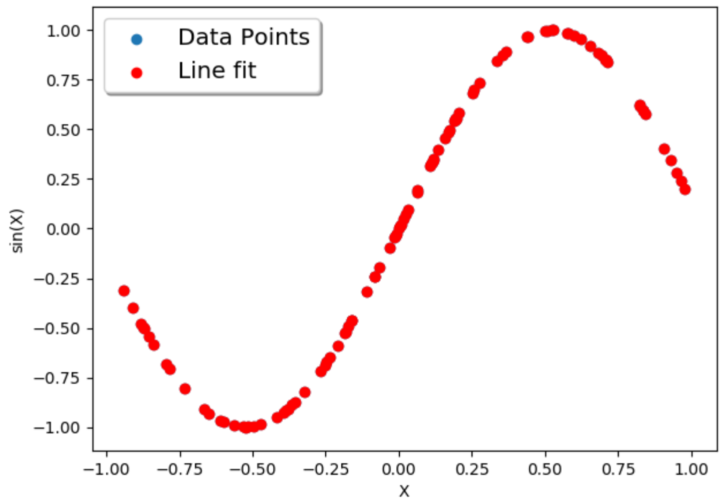

Figure 16: Plot for Question 3. d with Gaussian kernel

4. Question 4: Kernel K-Means

4.1 Q.4 a

Sum of Distances = 0.0

4.2 Q.4 b

For d = 10

$$\sum_{i=1}^{d} d_i = 0.0$$

4.3 Q.4 c

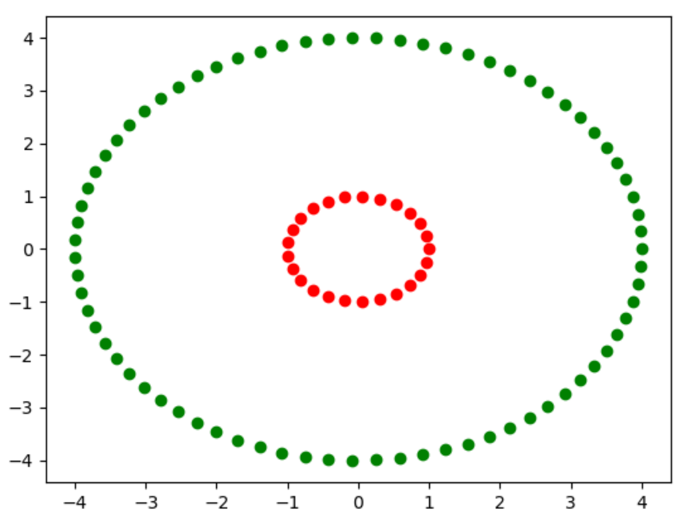

Figure 17: Plot for Question 4. c

5. Question 5: Kernel Fisher's Discriminant Analysis

5.1 Q.5 a

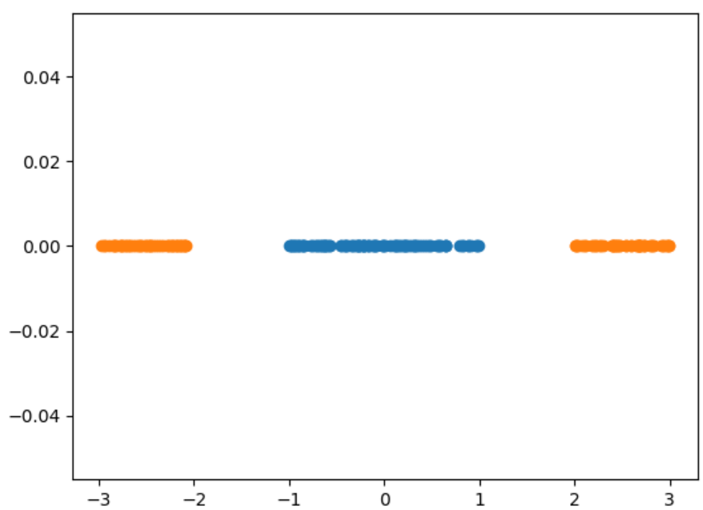

Figure 18: Plot for Question 5. a

5.2 Q.5 b

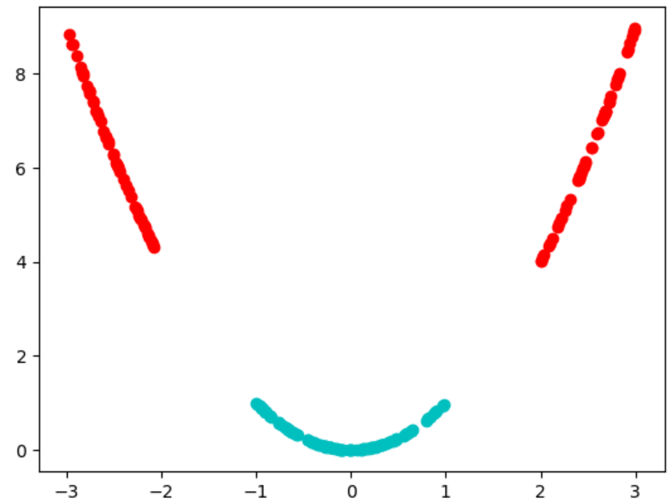

Figure 19: Plot for Question 5. b

5.3 Q.5 c

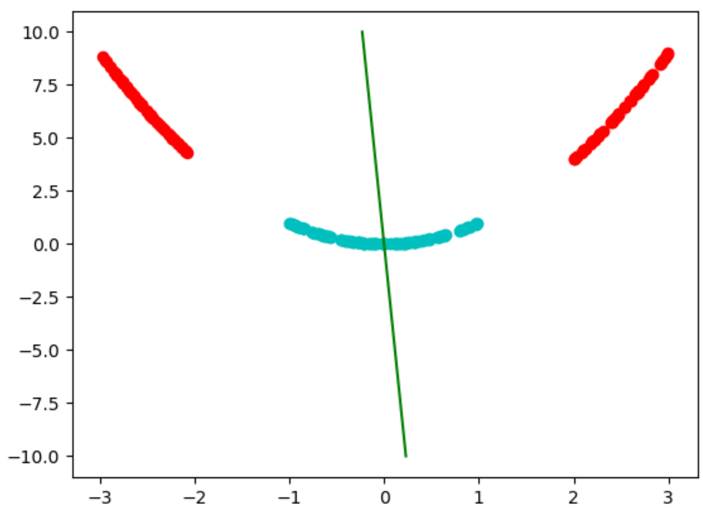

Figure 20: Plot for Question 5. c

5.4 Q.5 d

Expression for kernel

$$k(x, y) = \exp(-(||x - y||^2/(2 * \sigma^2)))$$

$\sigma = 3.8$

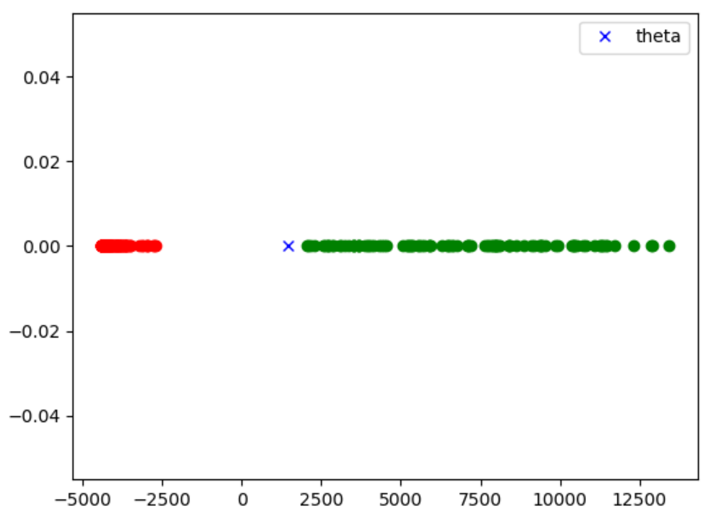

Figure 21: Plot for Question 5. d one-dimensional representation of data

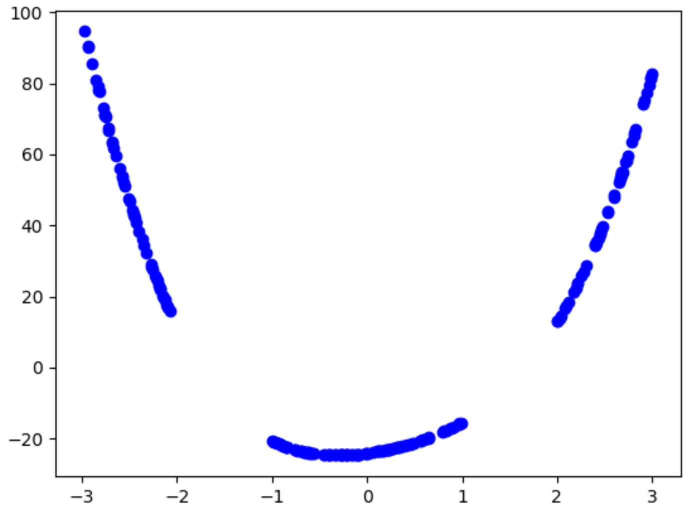

Figure 22: Plot for Question 5. d α vs x