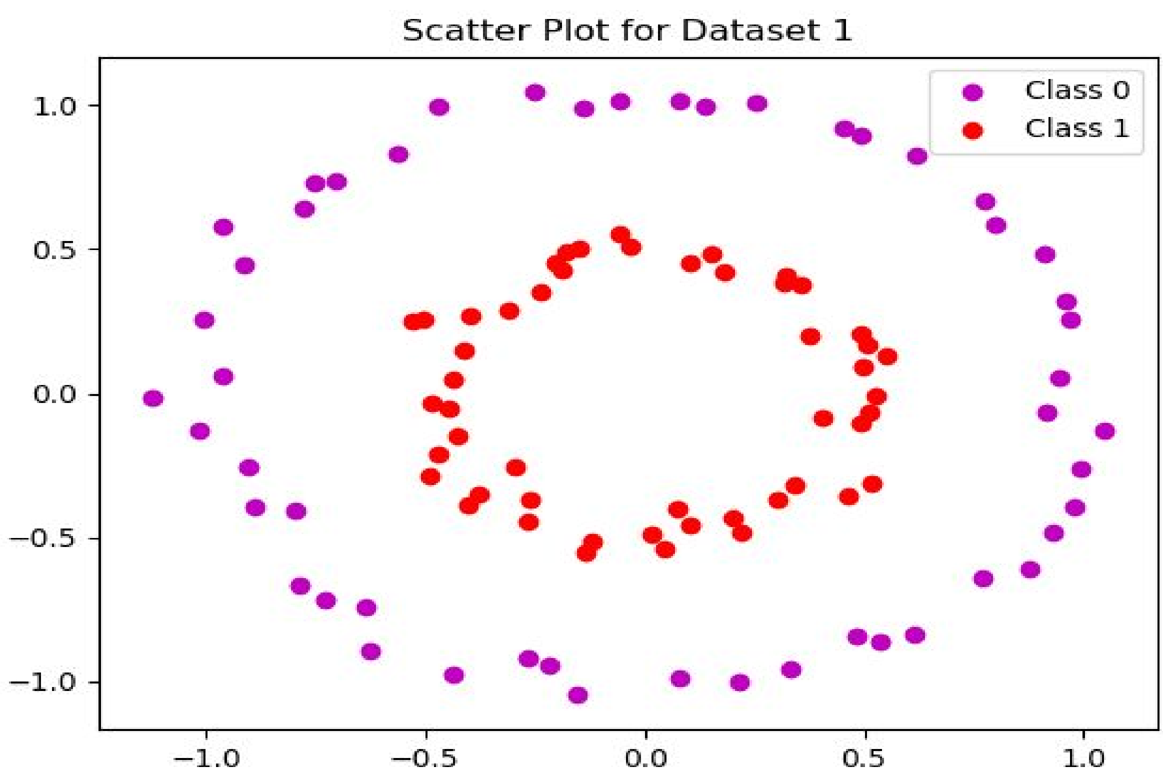
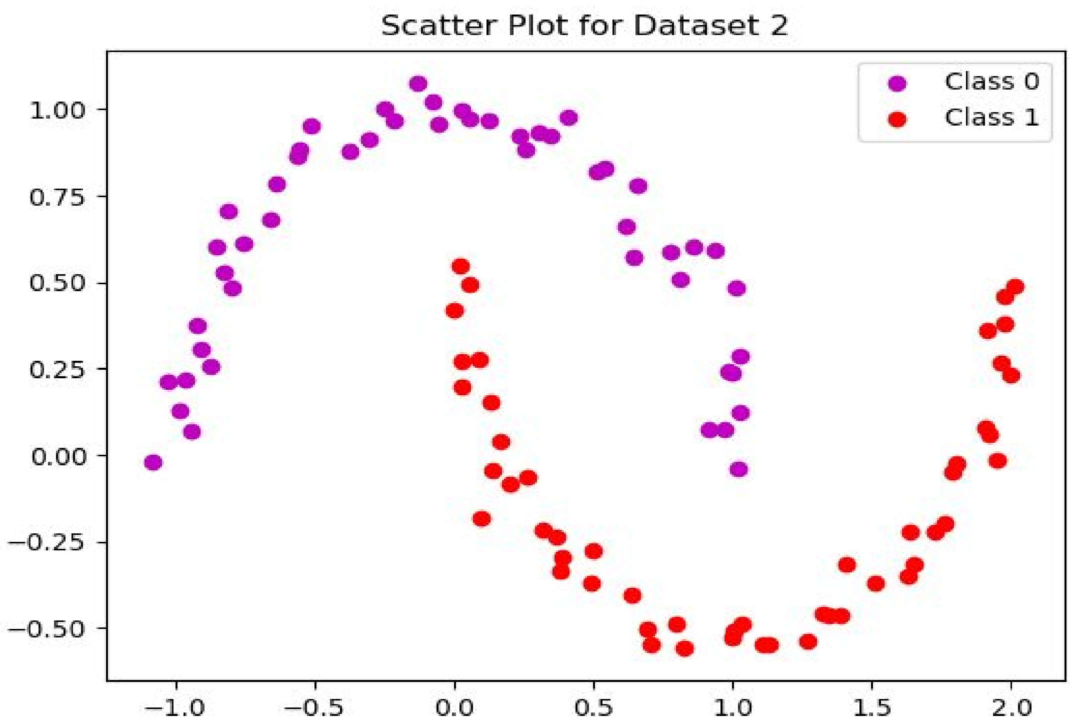
ML Assignment 2

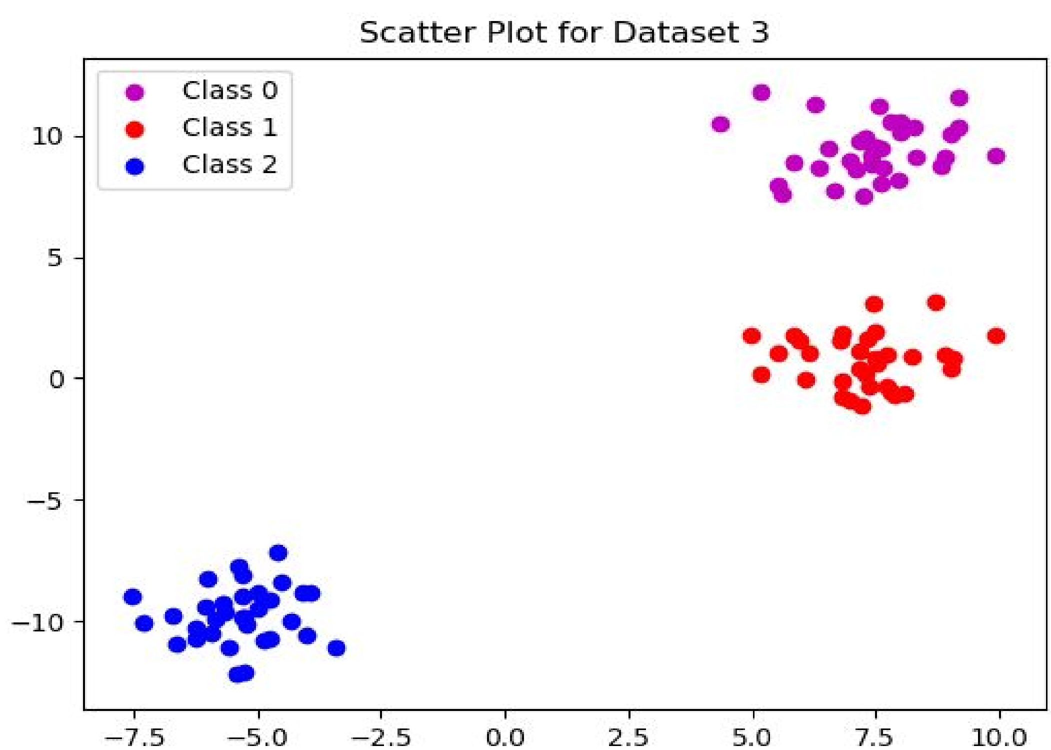
<u>Question.1</u>

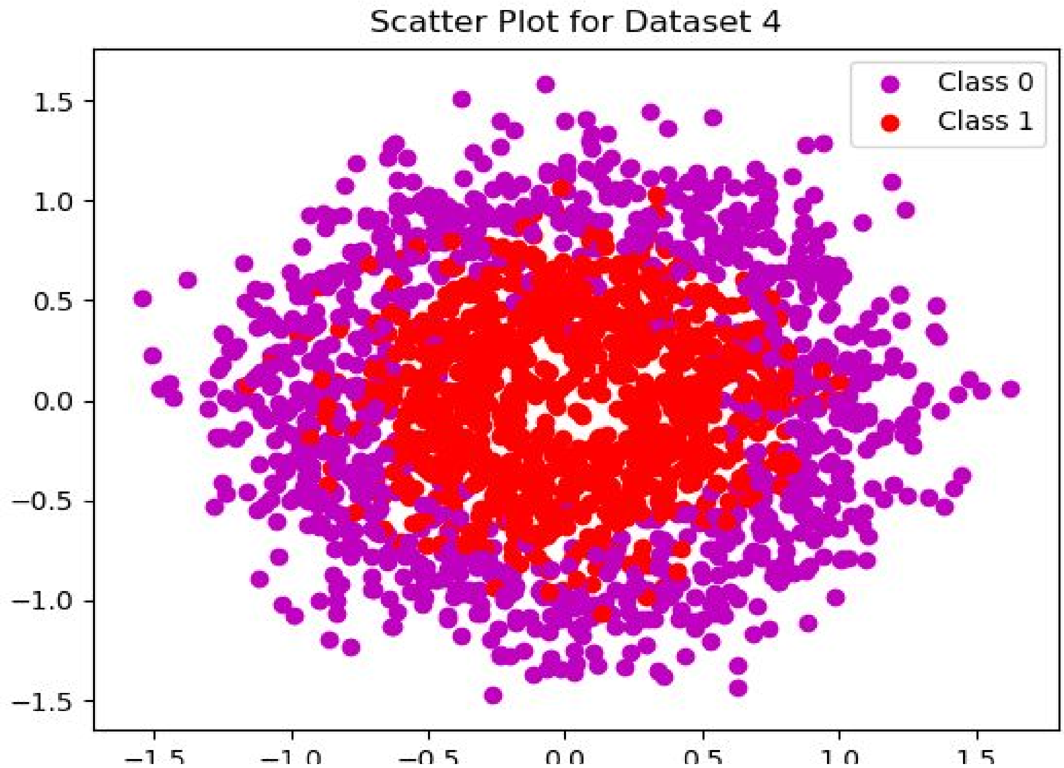
1.1

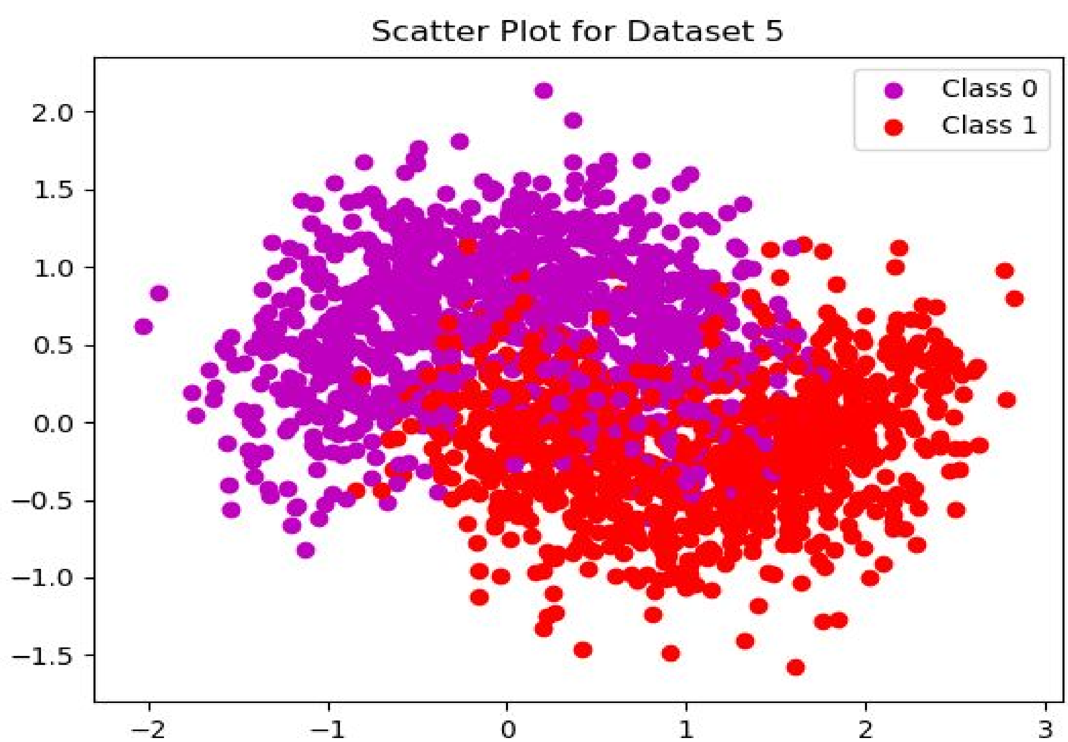
Dataset	Samples in each class	Number of classes	Number of samples	Outliers in each class
data_1.h5	50,50	2	100	0,0
data_2.h5	50,50	2	100	0,1
data_3.h5	34,33,33	3	100	3,4,6
data_4.h5	1000,1000	2	2000	7,49
data_5.h5	1000,1000	2	2000	61,54



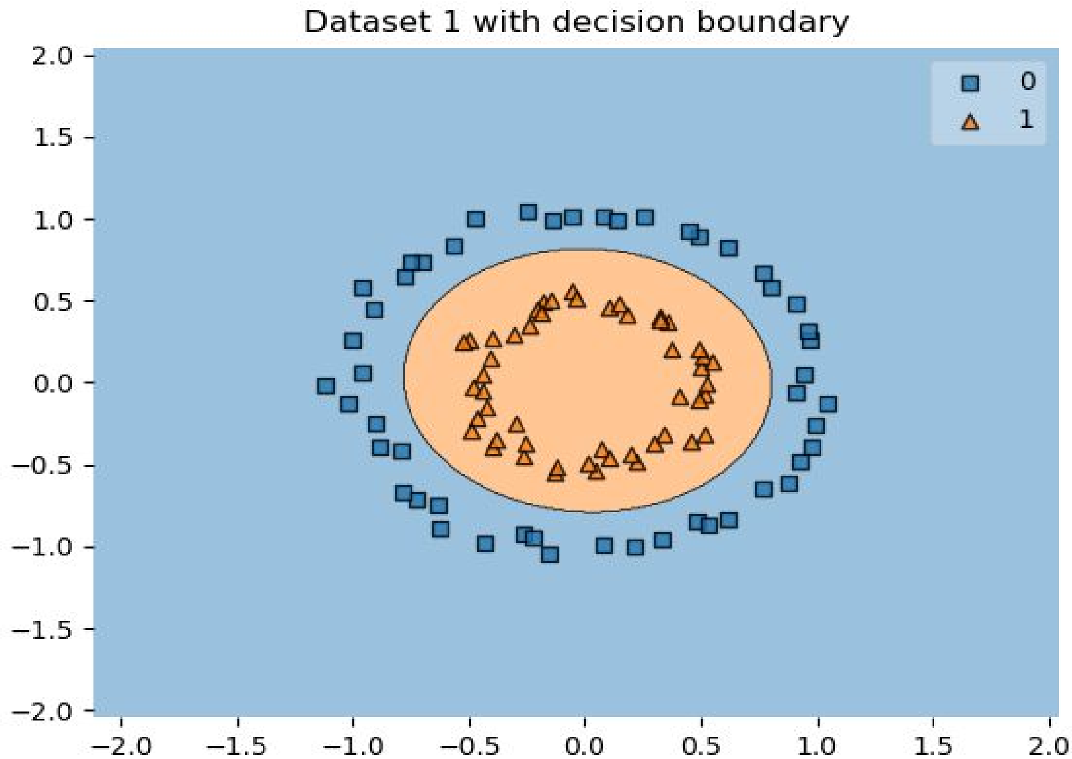

Scatter Plot for Dataset 2
Class 0
Class 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Scatter Plot for Dataset 3
Class 0
Class 1
Class 2
10
5
0
−5
−10
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

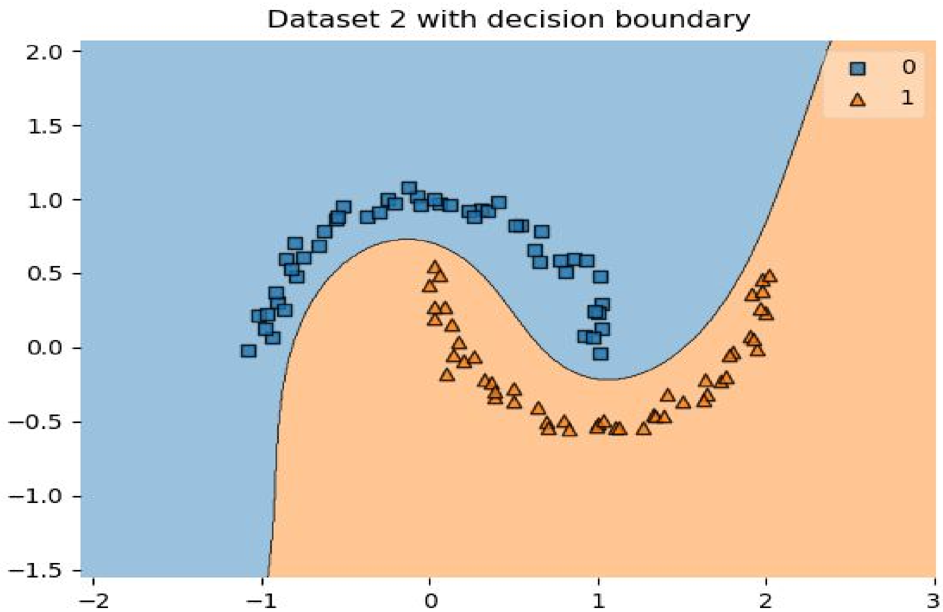

Scatter Plot for Dataset 4
Class 0
Class 1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


Scatter Plot for Dataset 5
Class 0
Class 1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2
−1
0
1
2
3

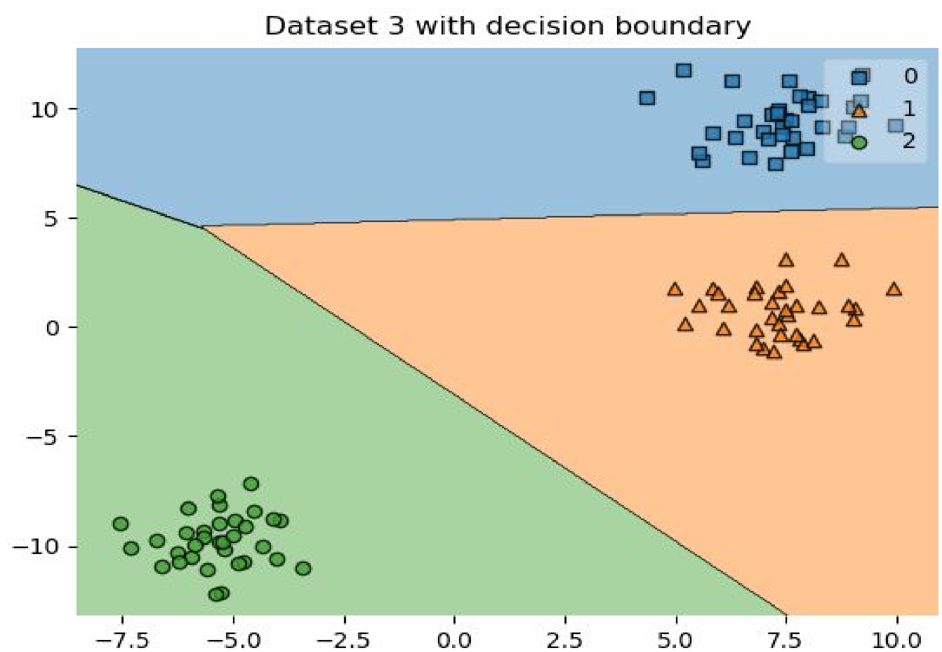
1.2



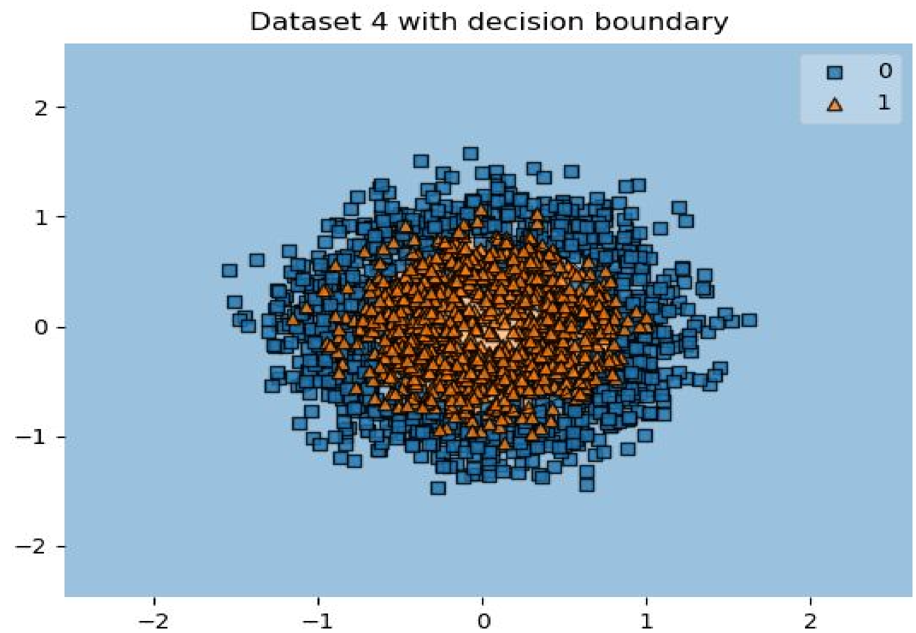
Reason: As the scatter-plot is in the form of a circle. Therefore a polynomial kernel of degree 2 can be used to separate the two classes. RBF kernel can also be used.



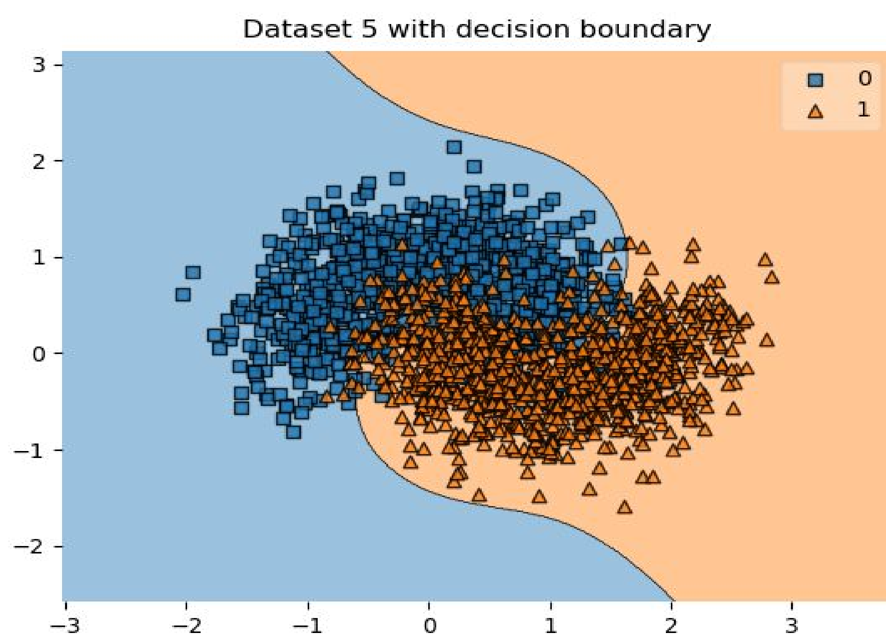
Reason: As the scatter-plot is in the form of a cubic function. Therefore a polynomial kernel of degree 3 can be used to separate the two classes.



Reason: It is intuitive that 3 classes can be separated by 3 lines. Therefore a linear kernel can be used to separate the three classes. Polynomial kernel of degree 2 can also be used.

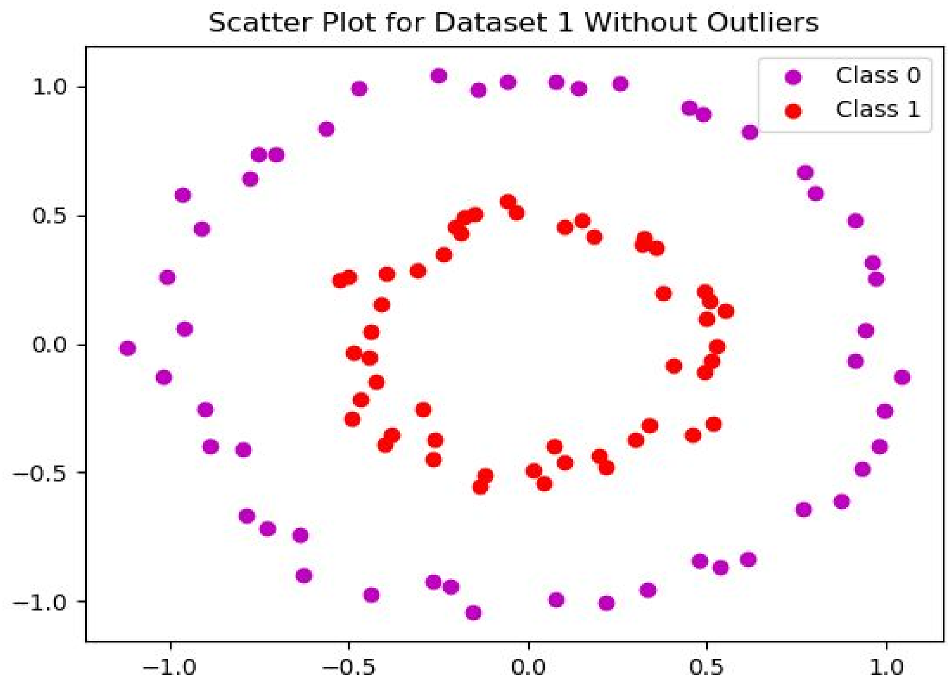


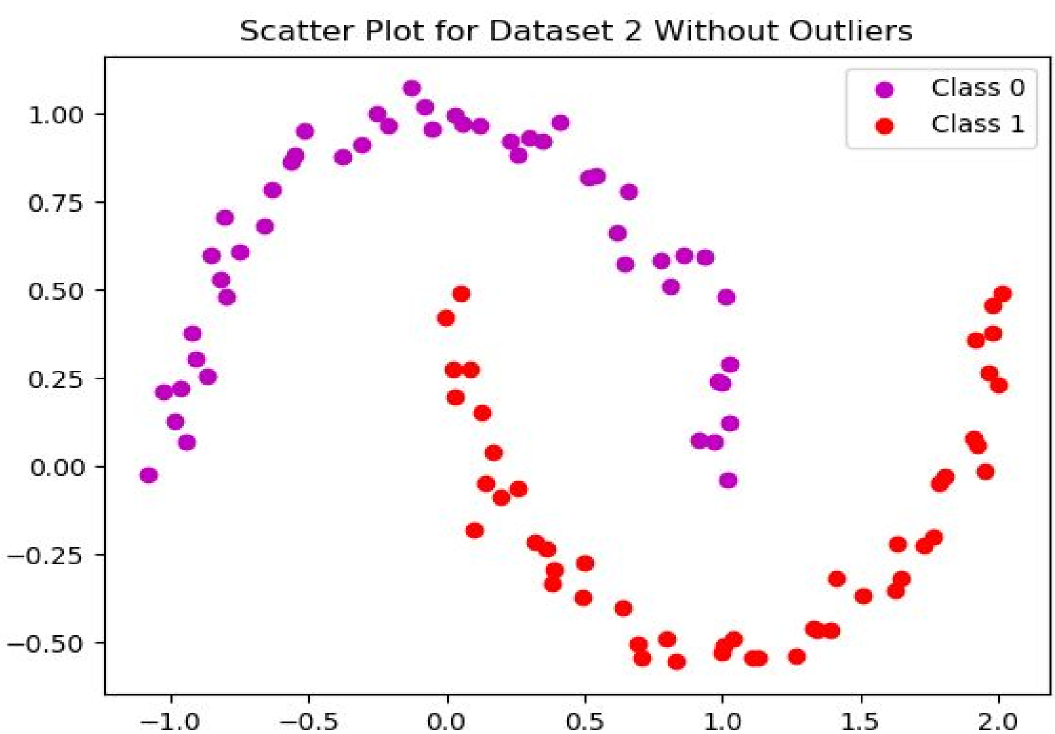
Reason: As the scatter-plot is in the form of a circle. Therefore a polynomial kernel of degree 2 can be used to separate the two classes. RBF kernel can also be used.

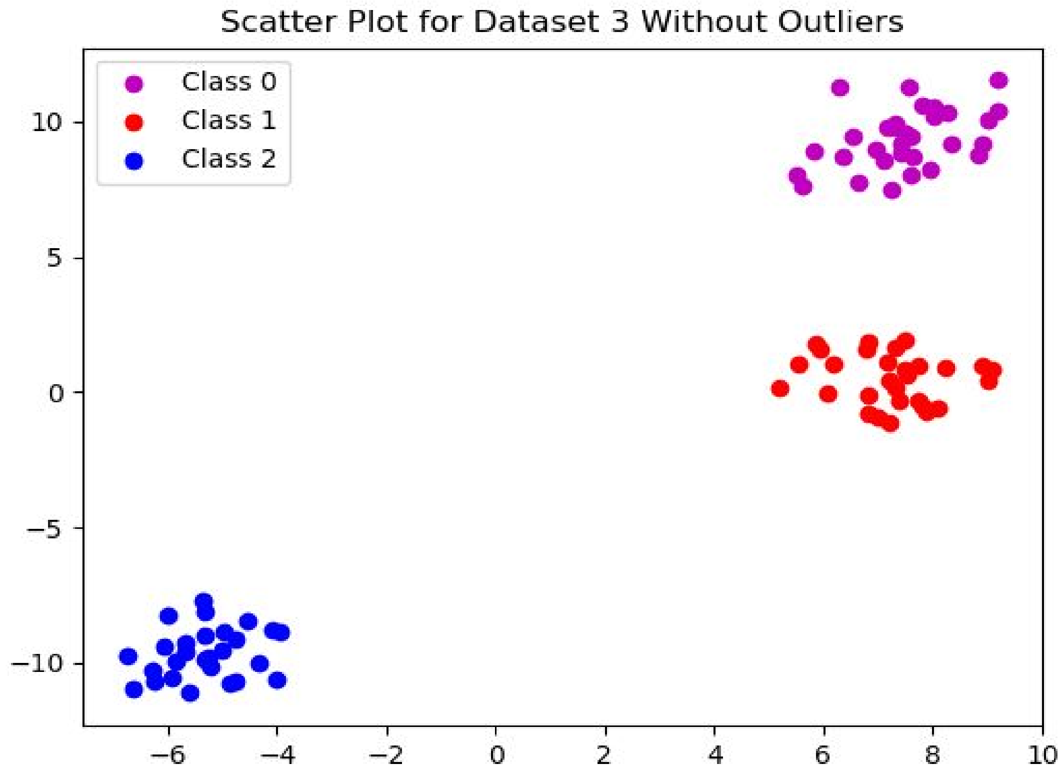


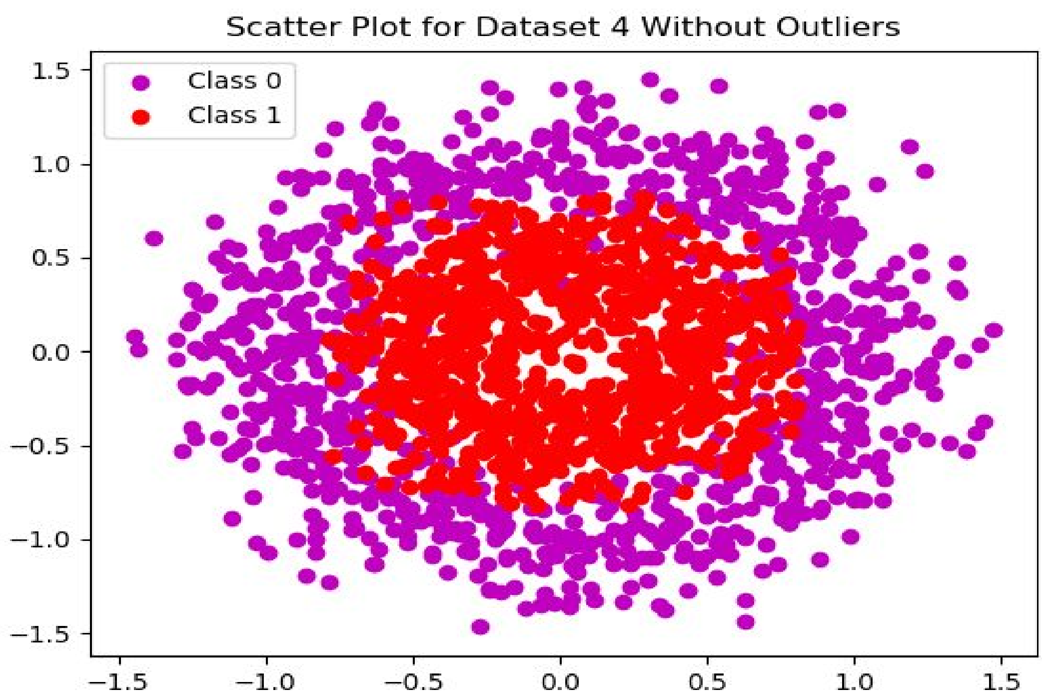
Reason: As the scatter-plot is in the form of a cubic function. Therefore a polynomial kernel of degree 3 can be used to separate the two classes.

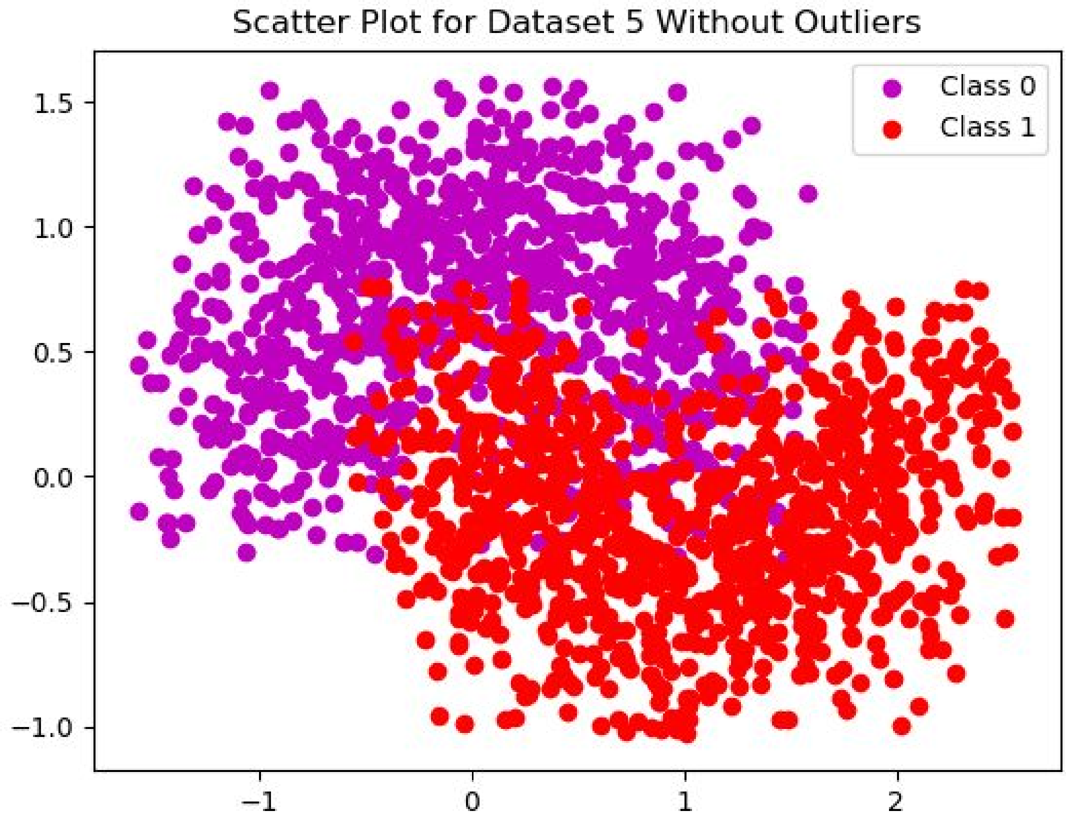
1.3


Scatter Plot for Dataset 1 Without Outliers
Class 0
Class 1
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0


Scatter Plot for Dataset 2 Without Outliers
Class 0
Class 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


Scatter Plot for Dataset 3 Without Outliers
Class 0
Class 1
Class 2
10
5
0
−5
−10
−6
−4
−2
0
2
4
6
8
10


Scatter Plot for Dataset 4 Without Outliers
Class 0
Class 1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5



1.4

Data Set 4

Accuracy on training Data with implemented predict function (Linear SVM): 54.0

Accuracy on training Data with Sklearn predict function (Linear SVM): 54.0

Accuracy on testing Data with implemented predict function (Linear SVM): 52.75

Accuracy on testing Data with Sklearn predict function (Linear SVM): 52.75

Data Set 4

Accuracy on training Data with implemented predict function (RBF Kernel): 89.375

Accuracy on training Data with Sklearn predict function (RBF Kernel): 89.375

Accuracy on testing Data with implemented predict function (RBF Kernel): 86.25

Accuracy on testing Data with Sklearn predict function (RBF Kernel): 86.25

Data Set 5

Accuracy on training Data with implemented predict function (Linear SVM): 85.4375

Accuracy on training Data with Sklearn predict function (Linear SVM): 85.4375

Accuracy on testing Data with implemented predict function (Linear SVM): 80.75

Accuracy on testing Data with Sklearn predict function (Linear SVM): 80.75

Data Set 5

Accuracy on training Data with implemented predict function (RBF Kernel): 88.5625

Accuracy on training Data with Sklearn predict function (RBF Kernel): 88.5625

Accuracy on testing Data with implemented predict function (RBF Kernel): 85.25

Accuracy on testing Data with Sklearn predict function (RBF Kernel): 85.25

Question.2

FOLD: 1

Accuracy of Linear kernel for one-vs-one: 29.9

Accuracy of Linear kernel for one-vs-all: 29.9

Accuracy of RBF kernel for one-vs-one: 40.9

Accuracy of RBF kernel for one-vs-all: 40.9

Accuracy of Polynomial kernel for one-vs-one: 38.3

Accuracy of Polynomial kernel for one-vs-all: 38.3

FOLD: 2

Accuracy of Linear kernel for one-vs-one: 30.099999999999998

Accuracy of Linear kernel for one-vs-all: 30.099999999999998

Accuracy of RBF kernel for one-vs-one: 40.699999999999996

Accuracy of RBF kernel for one-vs-all: 40.699999999999996

Accuracy of Polynomial kernel for one-vs-one: 39.900000000000006

Accuracy of Polynomial kernel for one-vs-all: 39.900000000000006

FOLD: 3

Accuracy of Linear kernel for one-vs-one: 30.4

Accuracy of Linear kernel for one-vs-all: 30.4

Accuracy of RBF kernel for one-vs-one: 42.3

Accuracy of RBF kernel for one-vs-all: 42.3

Accuracy of Polynomial kernel for one-vs-one: 38.6

Accuracy of Polynomial kernel for one-vs-all: 38.6

FOLD: 4

Accuracy of Linear kernel for one-vs-one: 30.0

Accuracy of Linear kernel for one-vs-all: 30.0

Accuracy of RBF kernel for one-vs-one: 41.8

Accuracy of RBF kernel for one-vs-all: 41.8

Accuracy of Polynomial kernel for one-vs-one: 37.7

Accuracy of Polynomial kernel for one-vs-all: 37.7

FOLD: 5

Accuracy of Linear kernel for one-vs-one: 27.900000000000002

Accuracy of Linear kernel for one-vs-all: 27.900000000000002

Accuracy of RBF kernel for one-vs-one: 40.5

Accuracy of RBF kernel for one-vs-all: 40.5

Accuracy of Polynomial kernel for one-vs-one: 38.6

Accuracy of Polynomial kernel for one-vs-all: 38.6

Mean Accuracy

- Mean accuracy over 5 folds of Linear kernel: **29.658%**
- Mean accuracy over 5 folds of RBF kernel: **41.238%**
- Mean accuracy over 5 folds of the quadratic polynomial kernel: **38.62%**

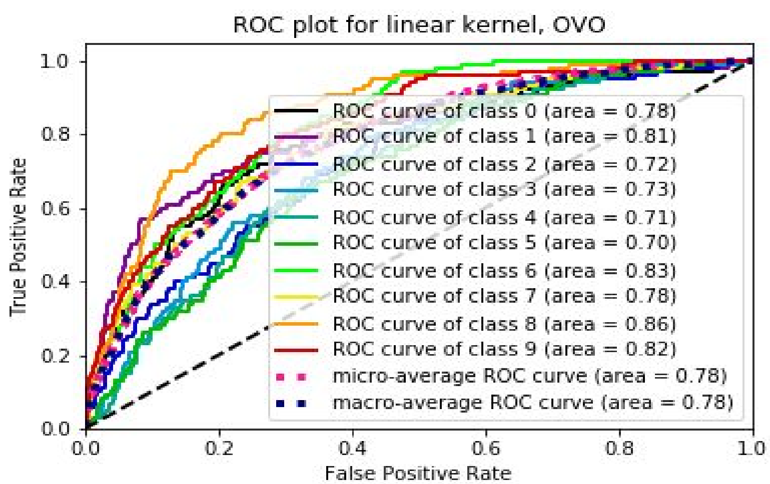
From the accuracy, we can see that the RBF kernel performed the best.

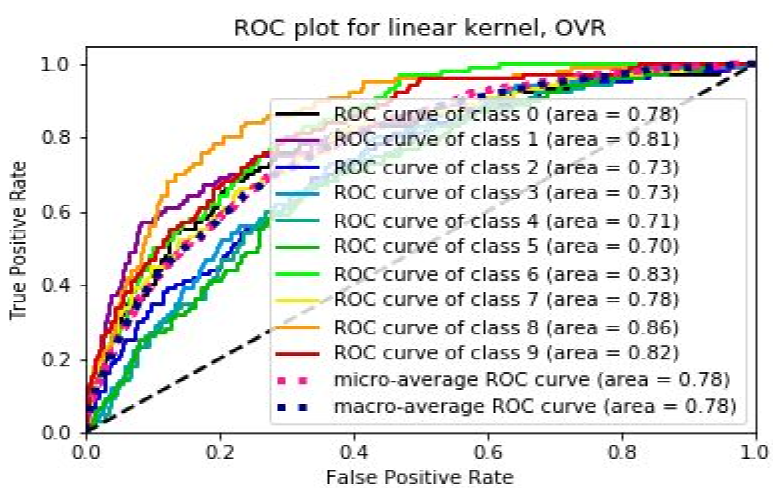
Mean F-Score

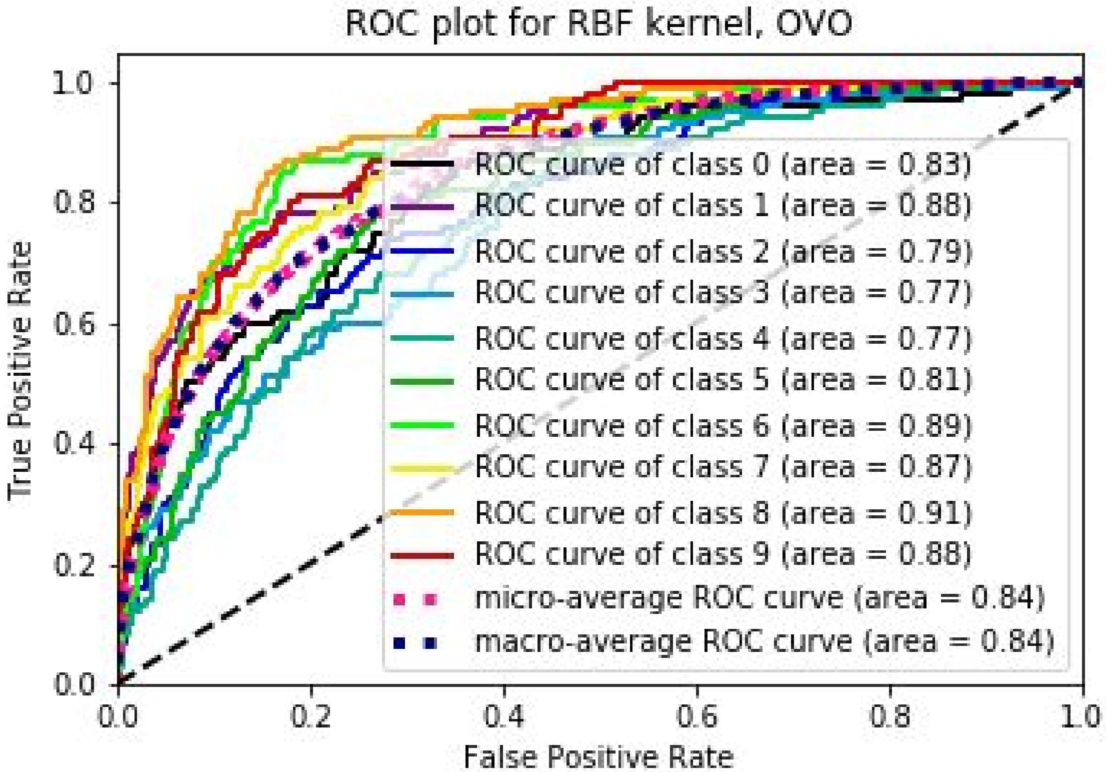
- Mean F-Score over 5 folds of Linear kernel: **29.85671673006336%**
- Mean F-Score over 5 folds of RBF kernel: **41.03430923324565%**
- Mean F-Score over 5 folds of the quadratic polynomial kernel: **38.26563098903506%**

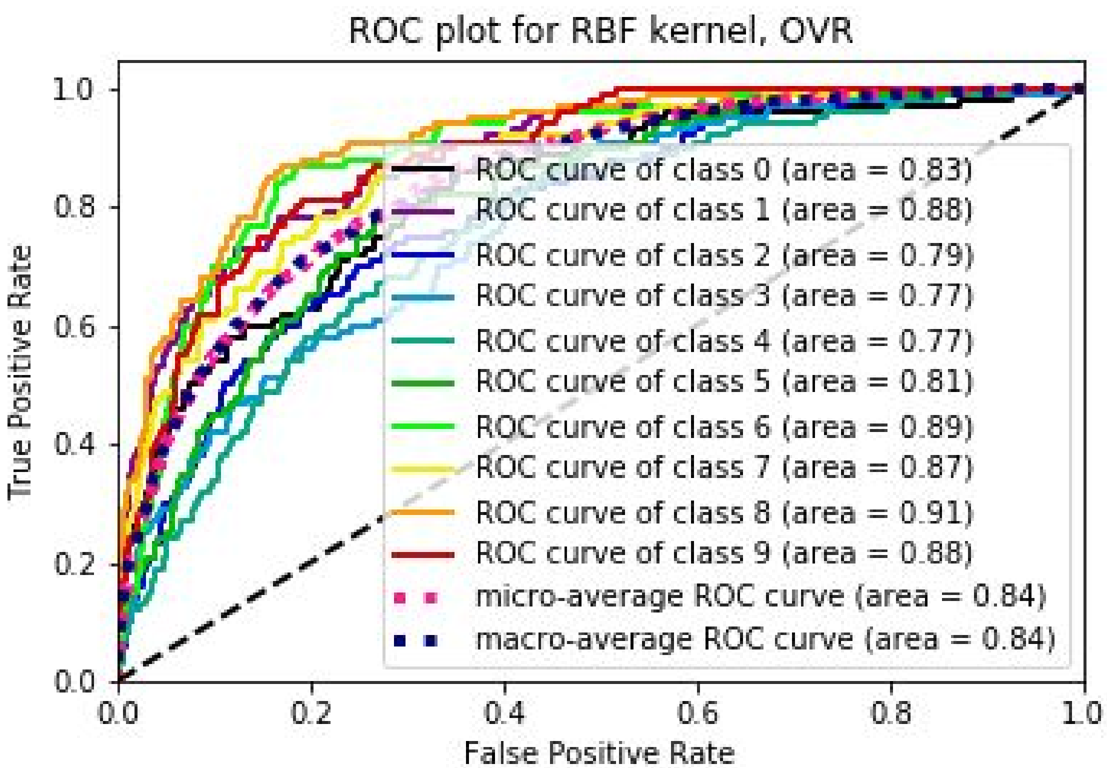
From the F-Score, we can see that the RBF kernel performed the best.

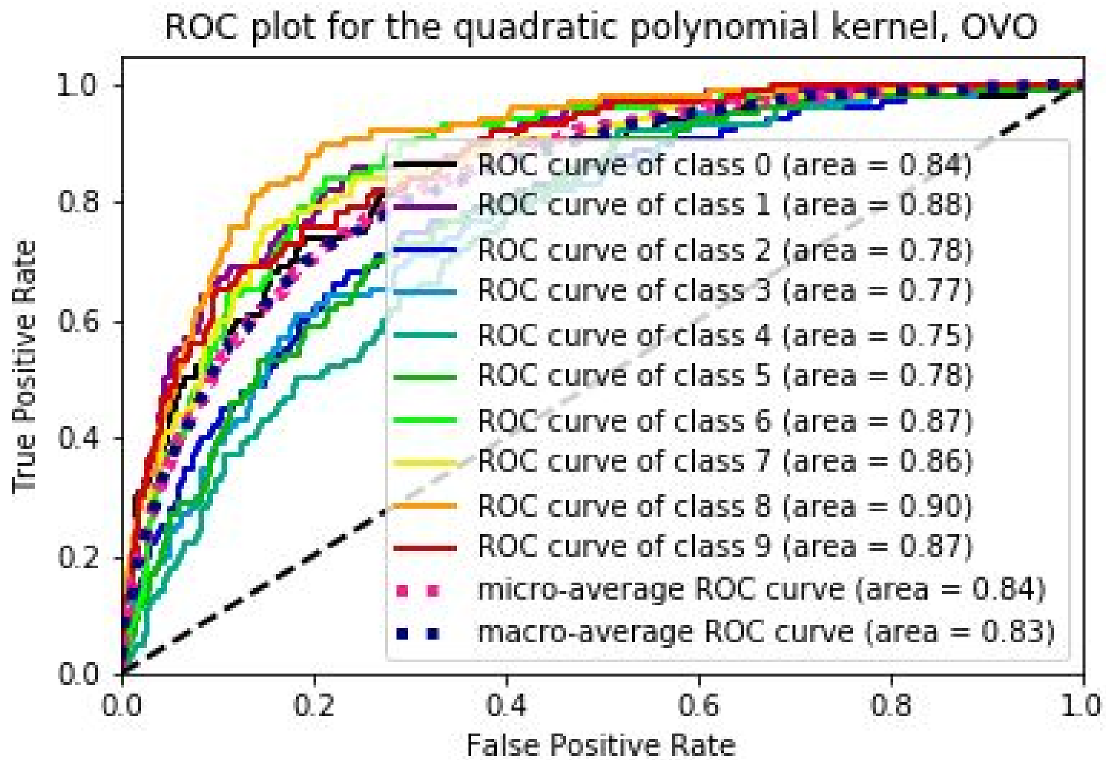
ROC Curve

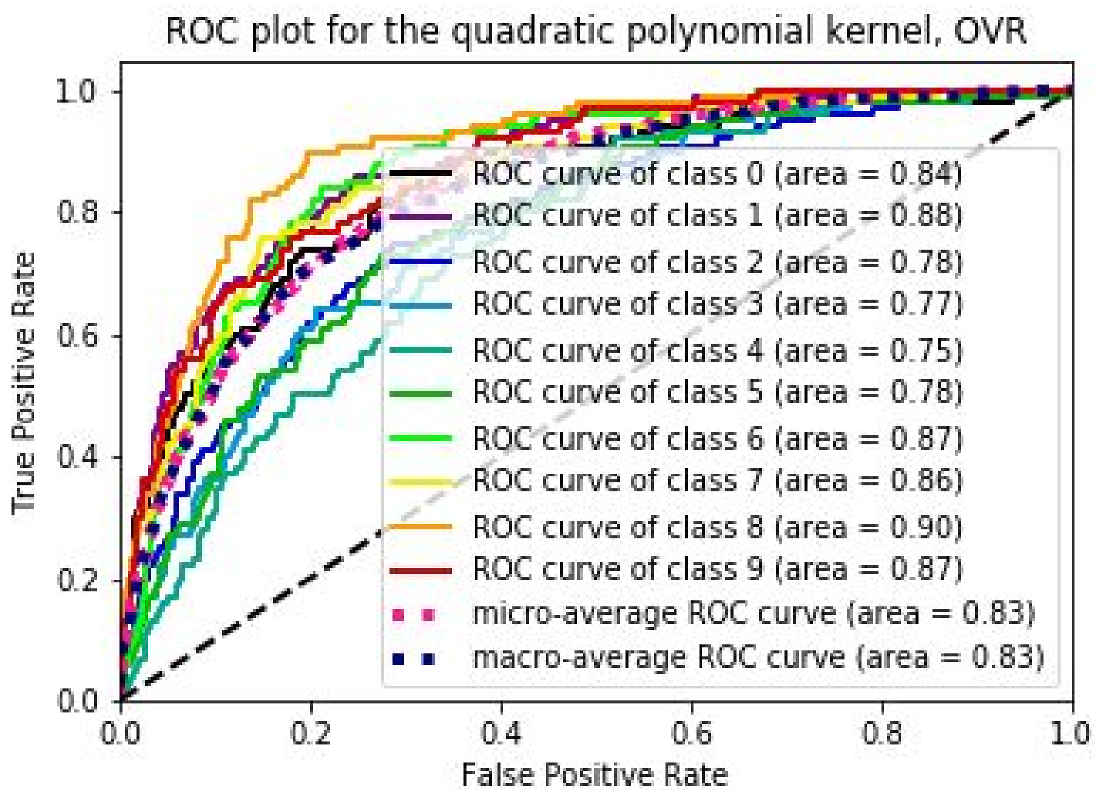





ROC plot for RBF kernel, OVO
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve of class 0 (area = 0.83)
ROC curve of class 1 (area = 0.88)
ROC curve of class 2 (area = 0.79)
ROC curve of class 3 (area = 0.77)
ROC curve of class 4 (area = 0.77)
ROC curve of class 5 (area = 0.81)
ROC curve of class 6 (area = 0.89)
ROC curve of class 7 (area = 0.87)
ROC curve of class 8 (area = 0.91)
ROC curve of class 9 (area = 0.88)
micro-average ROC curve (area = 0.84)
macro-average ROC curve (area = 0.84)


ROC plot for RBF kernel, OVR
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve of class 0 (area = 0.83)
ROC curve of class 1 (area = 0.88)
ROC curve of class 2 (area = 0.79)
ROC curve of class 3 (area = 0.77)
ROC curve of class 4 (area = 0.77)
ROC curve of class 5 (area = 0.81)
ROC curve of class 6 (area = 0.89)
ROC curve of class 7 (area = 0.87)
ROC curve of class 8 (area = 0.91)
ROC curve of class 9 (area = 0.88)
micro-average ROC curve (area = 0.84)
macro-average ROC curve (area = 0.84)


ROC plot for the quadratic polynomial kernel, OVO
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve of class 0 (area = 0.84)
ROC curve of class 1 (area = 0.88)
ROC curve of class 2 (area = 0.78)
ROC curve of class 3 (area = 0.77)
ROC curve of class 4 (area = 0.75)
ROC curve of class 5 (area = 0.78)
ROC curve of class 6 (area = 0.87)
ROC curve of class 7 (area = 0.86)
ROC curve of class 8 (area = 0.90)
ROC curve of class 9 (area = 0.87)
micro-average ROC curve (area = 0.84)
macro-average ROC curve (area = 0.83)


ROC plot for the quadratic polynomial kernel, OVR
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve of class 0 (area = 0.84)
ROC curve of class 1 (area = 0.88)
ROC curve of class 2 (area = 0.78)
ROC curve of class 3 (area = 0.77)
ROC curve of class 4 (area = 0.75)
ROC curve of class 5 (area = 0.78)
ROC curve of class 6 (area = 0.87)
ROC curve of class 7 (area = 0.86)
ROC curve of class 8 (area = 0.90)
ROC curve of class 9 (area = 0.87)
micro-average ROC curve (area = 0.83)
macro-average ROC curve (area = 0.83)

<u>Mean Confusion Matrix</u>

- Mean Confusion Matrix over 5 folds of Linear kernel:

```
[ [37.4  2.4  9.8  5.2  3.4  1.2  3.4  6.8 21.8  8.6]
  [ 7.  42.8  7.   6.2  4.   5.   3.2  5.4  8.6 10.8]
  [14.8  4.8 30.2 11.2 12.   7.8  7.   4.6  6.   1.6]
  [ 4.2  7.6 10.  22.4 14.6 16.2 10.2  5.6  5.4  3.8]
  [ 9.   4.  23.  10.2 22.2  6.6  8.8 10.2  3.4  2.6]
  [ 9.   6.4 13.4 22.8 15.8 13.6  8.8  5.4  3.8  1. ]
  [ 3.8  3.8  9.2 21.  14.2 10.  29.4  3.6  3.8  1.2]
  [ 9.2  4.8 11.2  7.2 11.6 10.8  4.8 32.   3.8  4.6]
  [22.2  9.2  2.4  3.2  4.6  5.2  1.2  2.  43.4  6.6]
  [ 7.8 24.6  3.8  4.2  3.6  3.   2.2  9.8 15.4 25.6] ]
```

- Mean Confusion Matrix over 5 folds of RBF kernel:

```
[ [43.4  1.6  3.   3.8  5.   4.   3.2  2.8 22.4 10.8]
  [ 4.6 49.2  3.   5.4  2.2  1.4  6.4  2.8  7.8 17.2]
  [14.6  2.4 31.8  8.4 16.8  6.  10.4  4.2  3.4  2. ]
  [ 8.   5.4  7.8 24.4  9.8 18.  14.8  6.2  0.8  4.8]
  [ 6.   4.4 23.   3.4 25.4  7.4 15.2  7.   3.4  4.8]
  [ 3.4  3.8 11.2 21.2  9.4 31.8 10.2  4.2  2.6  2.2]
  [ 2.6  2.  10.   5.6 19.   3.8 48.8  3.4  1.   3.8]
  [ 4.6  3.4  6.8  8.4 12.6  7.2  7.8 41.4  1.4  6.4]
 [ 8.8  2.2  1.8  2.   3.   6.   0.   1.8 64.  10.4]
 [ 4.6 10.6  1.2  3.4  0.4  3.4  3.8  7.2 12.4 53. ] ]
```

- Mean Confusion Matrix over 5 folds of the quadratic polynomial kernel:

```
[ [45.6  1.6  7.2  3.2  5.4  2.2  5.   4.4 18.   7.4]
  [ 4.6 49.2  7.4  3.2  4.   3.   3.4  4.  10.2 11. ]
  [12.8  2.  35.8  9.2 14.6  4.4 10.8  5.6  4.2  0.6]
  [ 2.8  5.2 10.6 29.   9.  16.4 16.6  3.6  2.8  4. ]
  [ 6.2  3.6 22.6  5.8 28.6  4.8 14.2  9.   3.   2.2]
  [ 3.   3.6 13.2 23.  13.2 20.  11.8  8.   3.   1.2]
  [ 1.8  1.6  9.8 12.8 17.   1.8 50.8  2.8  1.   0.6]
  [ 4.6  0.8  9.4  7.4 17.4  6.   6.4 42.2  1.6  4.2]
  [20.   5.   1.2  3.6  2.8  3.8  1.2  1.6 53.4  7.4]
  [ 6.6 19.2  3.2  4.8  4.2  2.   4.   5.4 12.  38.6] ]
```

From the mean accuracy, F-Score, ROC Curves and Confusion Matrix, we can see that the RBF kernel performed the best as compared to linear and quadratic polynomial kernel.

Question 5

Name - Harshit Rai
Roll No. - 2017152

1

classmate
Date
Page

Ans-5) $X \rightarrow$ has 2 features, x_1 and x_2

$\therefore$ we will get 2 weights $\Rightarrow w_1$ and w_2

As, $W = \sum_{i=1}^{N} \alpha_i \cdot d_i \cdot x_i$

$$W_1 = \sum_{i=1}^{10} \alpha_i \cdot d_i \cdot x_{1i}$$

and, $W_2 = \sum_{i=1}^{10} \alpha_i \cdot d_i \cdot x_{2i}$

So, $W_1 = (4 \times 0.414 \times 1) + 0 - 0 - (2.5 \times 1.18 \times 1) + 0 - 0 + (3.5 \times 1.18)$
$- 0 - (2 \times 0.414) + 0$

$\boxed{W_1 = 2.008}$ — (1)

Similarly, $W_2 = (2.9 \times 0.414 \times 1) + 0 - 0 - (1 \times 1.18 \times 1) + 0 - 0 + (4 \times 1.18)$
$- 0 - (2.1 \times 0.414) + 0$

$\boxed{W_2 = 3.8712}$ — (2), $w = [2.008 \quad 3.8712]^T$

Now, we know that our decision ~~boundary~~ function is given by $\Rightarrow$

$1 = w^T X + b$

P.T.O.

2

At $(4, 2.9) \Rightarrow +1 = (2.008 \times 4) + (2.9 \times 3.8712) + b_1$

$b_1 = 1 - 19.25848 = \boxed{-18.25848}$

At $(2.5, 1) \Rightarrow +1 = (2.008 \times 2.5) + (1 \times 3.8712) + b_2$

$b_2 = \boxed{-7.8912}$

At $(3.5, 4) \Rightarrow +1 = (2.008 \times 3.5) + (4 \times 3.8712) + b_3$

$b_3 = \boxed{-21.5128}$

At $(2, 2.1) \Rightarrow +1 = (2.008 \times 2) + (2.1 \times 3.8712) + b_4$

$b_4 = \boxed{-11.14552}$

$$b = \frac{b_1 + b_2 + b_3 + b_4}{4} = \boxed{-14.702}$$

(A) $b = \boxed{-14.702}$

(B) x_1 $(4, 2.9)$
x_4 $(2.5, 1)$
x_2 $(3.5, 4)$
x_3 $(2, 2.1)$

$\Rightarrow$ These are the support vectors.
Because their coefficients or "α" $\neq 0$.

Question 6

(3)

Date

Page

(c) $W = \begin{bmatrix} 2.008 \\ 3.8712 \end{bmatrix}$

Now, $(2.008)x_1 + (3.8712)x_2 - 14.702 = Y$

Equation of optimal Hyper Plane or O.H.P.

for, $X = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \Rightarrow Y = 2.9356 > 0$

$\therefore X(3,3) \Rightarrow Y = 1$

and, $\alpha = 0 \Rightarrow$ As it is not a support vector.

Ans. 6)

We have RBF kernel given by,

$$K(u, v) = (e)^{-\gamma \|u - v\|^2}$$

It is given that "u" and "v" have only 1 feature i.e.

u and v are real Numbers.

P.T.O.

④

Date ______
Page ______

for simplicity, $\gamma = 1$

Now, $K(u, v) = e^{-(u-v)^2}$ $\left[\|u-v\|^2 = (u-v)^2, \text{ as because now "u" and "v" are real No.} \right]$

$K(u, v) = e^{-u^2 - v^2 + 2uv}$

$K(u, v) = e^{-u^2 - v^2} \cdot e^{2uv}$

$K(u, v) = e^{-(u^2+v^2)} \cdot e^{2uv}$ — ①

Now, using Taylor series expansion of

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} - - \cdots$$

$$\therefore e^{2uv} = 1 + \frac{2uv}{1!} + \frac{(2uv)^2}{2!} + \frac{(2uv)^3}{3!} + - - - - - \frac{(2uv)^n}{n!}$$

Put this in ①

$$K(u, w) = (e)^{-(u^2+w^2)} \cdot \left[1 + \frac{2uw}{1!} + \frac{(2uw)^2}{2!} + \frac{(2uw)^3}{3!} \cdots + \frac{(2uw)^n}{n!}\right]$$

$$K(u, w) = (e)^{-[u^2+w^2]} \cdot \left[(1 \cdot 1) + \left(u\sqrt{\frac{2}{1!}} \cdot u\sqrt{\frac{2}{1!}}\right) + \left(u^2\sqrt{\frac{2^2}{2!}} \cdot w^2\sqrt{\frac{2}{2!}}\right) \cdots \cdots + \left(u^n\sqrt{\frac{2^n}{n!}} \cdot w^n\sqrt{\frac{2^n}{n!}}\right)\right]$$

$\cancel{K(u, w) = [(e)^{-2u^2} \ldots]}$

This can be written in the form of $\Rightarrow$

$$K(u, w) = \phi^T(u) \cdot \phi(w) = \{\phi(u), \phi(w)\}$$

Dot Product.

$$\text{where, } \boxed{\phi(u) = (e)^{-u^2} \cdot \left[1, u\sqrt{\frac{2}{1!}}, u^2\sqrt{\frac{2^2}{2!}}, u^3\sqrt{\frac{2^2}{3!}} \cdots \cdots u^n\sqrt{\frac{2^n}{n!}}\right]^T}$$

Ans //

P.T.O.

(6)

Now, coefficient of "μ" $= (e)^{-\mu^2} \cdot \sqrt{\frac{(2)^n}{n!}}$

when, $n \to \infty$

coeff. of "μ" $= \lim_{n \to \infty} (e)^{-\mu^2} \sqrt{\frac{(2)^n}{n!}}$ — (2)

we have to check that whether (2) converges or diverges as $n \to \infty$

$\Rightarrow T_n = (e)^{-\mu^2} \sqrt{\frac{(2)^n}{n!}}$

$T_{n+1} = (e)^{-\mu^2} \sqrt{\frac{(2)^{n+1}}{(n+1)!}}$

$\frac{T_{n+1}}{T_n} = \sqrt{\frac{2}{n+1}} = R$ (radius of convergence).

as $n \to \infty$

$R = \sqrt{\frac{2}{\infty}} = 0$ $\left[\text{as } \frac{1}{\infty} = 0\right]$

Question 4

7

Date
Page

As $R = 0 \Rightarrow$ it converges.

or we can say that at higher values say (1000000 = n).

$$(e)^{\mu^2} \cdot \sqrt{\frac{2^n}{n!}} = 0$$

Therefore, The coefficients of higher order terms of "U" present in $\phi(4)$ $\boxed{\text{becomes } 0 \text{ or } = 0}$

Ans //

Ans.4) $K(X, X') = [1 + X^T X']^2$ — ①

Polynomial Kernel of degree 2.

given, $X = [x_1, x_2]^T$

$X' = [x'_1, x'_2]^T$

Puting X and x' in the Kernel ①.

P.T.O.

(8)

$$X^T X = \underset{(1\times2)}{[x_1 \;\; x_2]} \underset{(2\times1)}{\begin{bmatrix} x_1' \\ x_2' \end{bmatrix}} = \underset{(1\times1)}{[x_1 x_1' + x_2 x_2']}.$$

Now, $K(X, X') = [1 + X^T X]^2$

$$K(X, X') = [1 + x_1 x_1' + x_2 x_2']^2$$

$$K(X, X') = 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2 x_1 x_1' + 2 x_2 x_2' + 2 x_1 x_2 x_1' x_2'$$

$$K(X, X') = (1 \cdot 1) + (x_1^2 \cdot x_1'^2) + (x_2^2 \cdot x_2'^2) + (\sqrt{2} x_1 \cdot \sqrt{2} x_1') + (\sqrt{2} x_2 \cdot \sqrt{2} x_2')$$
$$+ (\sqrt{2} x_1 x_2 \cdot \sqrt{2} x_1' x_2')$$

$$K(X, X') = \phi^T(X) \cdot \phi(X') = \{\phi(X), \phi(X')\}$$

Dot Product.

where, $\boxed{\phi(X) = [1, x_1^2, x_2^2, \sqrt{2} x_1, \sqrt{2} x_2, \sqrt{2} x_1 x_2]^T}$

Ans //

"$\phi(X)$" contains "6" features. (1 bias and 5 actual features).

Question 3

(9)

Date

Page

Ans.3) Yes we can model the XOR operation using an S.V.M.

Input vector, x	Desired response, d
(-1, -1)	-1
(-1, +1)	+1
(+1, -1)	+1
(+1, +1)	-1

We will be choosing a Polynomial kernel of degree 2 for this promblem, which is given by $\Rightarrow$

$$K(x, x_i) = [1 + x^T x_i]^2 \quad (1)$$

our "x" has two features, $x = [x_1, x_2]^T$

and, $x_i = [x_{i1}, x_{i2}]^T$

Put these in (1)

$$K(x, x_i) = [1 + x_1 x_{i1} + x_2 x_{i2}]^2$$

$$K(x, x_i) = 1 + x_1^2 x_{i1}^2 + x_2^2 x_{i2}^2 + 2 x_1 x_{i1} + 2 x_2 x_{i2} + 2 x_1 x_{i1} x_2 x_{i2}$$

P.T.O.

Now, the image of the input vector "x" induced in the feature space is therefore deduced to be =>

$$\phi(x) = \left[1, x_1^2, \sqrt{2}\, x_1 x_2, x_2^2, \sqrt{2}\, x_1, \sqrt{2}\, x_2\right]^T$$

Similarly, $\phi(x_i) = \left[1, x_{i1}^2, \sqrt{2}\, x_{i1} x_{i2}, x_{i2}^2, \sqrt{2}\, x_{i1}, \sqrt{2}\, x_{i2}\right]^T$

$, i = 1, 2, 3 \ldots$

we define a Kernel-gram matrix "K"

$\Downarrow$

$$K = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix} = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & x_1 \cdot x_{i1} & x_1 \cdot x_{i2} \\ x_2 \cdot x_1 & x_2 \cdot x_2 & x_2 \cdot x_{i1} & x_2 \cdot x_{i2} \\ x_{i1} \cdot x_1 & x_{i1} \cdot x_2 & x_{i1} \cdot x_{i1} & x_{i1} \cdot x_{i2} \\ x_{i2} \cdot x_1 & x_{i2} \cdot x_2 & x_{i2} \cdot x_{i1} & x_{i2} \cdot x_{i2} \end{bmatrix}$$

(A)

The objective function for the dual form is given by =>

$$\theta(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i, x_j)$$

$$\theta(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\Big[9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4$$
$$+ 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^3 - 2\alpha_3\alpha_4 + 9\alpha_4^2\Big]$$

(B)

Now, optimizing $\theta(\alpha)$ with respect to the Lagrange multiplier

$$\frac{d\theta}{d\alpha_1} = 0 \Rightarrow 9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$
$$\frac{d\theta}{d\alpha_2} = 0 \Rightarrow -\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1$$
$$\frac{d\theta}{d\alpha_3} = 0 \Rightarrow -\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$
$$\frac{d\theta}{d\alpha_4} = 0 \Rightarrow \alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

Solving these 4 equation

we get,

$$\alpha_{0,1} = \alpha_{0,2} = \alpha_{0,3} = \alpha_{0,4} = \frac{1}{8}$$

The optimum value of, $\theta_0(\alpha) = \frac{1}{4}$

P.T.O.

We may write $\frac{\|w_o\|^2}{2} = Q(\alpha) = \frac{1}{4}$

$\|w_o\| = \frac{1}{\sqrt{2}}$, $w_o \rightarrow$ optimal weight vector.

Now, we know that the optimal weight vector is given by $\Rightarrow$

$$w_o = \sum_{i=1}^{N} \alpha_i d_i \phi(x_i)$$

$$w_o = \frac{1}{8}\left[(-1)\phi(x_1) + (1)\phi(x_2) + (1)\phi(x_3) + (-1)\phi(x_4)\right]$$

$$w_o = \left(\frac{1}{8}\right)\left\{ -\begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ -\sqrt{2} \\ \sqrt{2} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} \right\}$$

$$w_o = \left(\frac{1}{8}\right)\begin{bmatrix} 0 \\ 0 \\ -4\sqrt{2} \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The first element ~~[illegible]~~ of w_0 denotes the bias "b" which is "0" in this case.

Therefore equation of optimal hyperplane is given by $\Rightarrow W_0^T \phi(x) = 0$

$$\text{or,} \quad \begin{bmatrix} 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix}_{(1\times 6)} \begin{bmatrix} 1 \\ x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ x_2^2 \\ \sqrt{2}\, x_1 \\ \sqrt{2}\, x_2 \end{bmatrix}_{(6\times 1)} = 0$$

$$\left(\frac{-1}{\sqrt{2}}\right)\left(\sqrt{2}\, x_1 x_2\right) = 0$$

$$\boxed{-x_1 x_2 = 0} \quad \text{ⓒ}$$

Ans

Therefore the XOR problem can be modelled using S.V.M. with the polynomial function ⓒ given above.