# ML Assignment 2

Simran Kucheria (sk11645@nyu.edu)

## Problem 1

Q1
A

a) $K(x,\bar{x}) = \alpha K_1(x,\bar{x}) + \beta K_2(x,\bar{x}) \quad \alpha, \beta \geq 0$

$K_1, K_2$ are positive semi definite Kernels.

$K_1(x,\bar{x}) = \phi_1(x)^T \phi_1(\bar{x}) \cdot \quad K_2(x,\bar{x}) = \phi_2(x)^T \phi_2(\bar{x})$

$\alpha K_1(x,\bar{x}) = \alpha \phi_1(x)^T \phi_1(\bar{x})$ ① 

$\beta K_2(x,\bar{x}) = \beta \phi_2(x)^T \phi_2(\bar{x})$ ②

Since $\alpha, \beta \geq 0$ both these values are also positive semi definite.

① + ②

$= \alpha \phi_1(x)^T \phi_1(\bar{x}) + \beta \phi_2(x)^T \beta \phi_2 x$

The sum is also positive semi definite.

$\alpha K_1(x,\bar{x}) + \beta K_2(x,\bar{x})$ is a valid Mercer Kernel

b. $K(x,\bar{x}) = K_1(x,\bar{x}) \times K_2(x,\bar{x})$.

$= \phi_1(x)^T \phi_1(\bar{x}) \cdot \phi_2(x)^T \phi_2(\bar{x})$

$\sum_i \phi_1(x)_i \phi_1(\bar{x})_i \sum_j \phi_2(x)_j \phi_2(\bar{x})_j$

$= \sum_{ij} \phi_1(x)_i \phi_2 x_{(j)} \phi_1(\bar{x})_i \phi_2(\bar{x})_j = \phi(x)^T \phi(\bar{x})$

wher $\phi(x) = \phi_1(x) \otimes \phi_2(\hat{x})$

Which is the repn of a valid mercer kernel.

c) $K(x,\bar{x}) = f(K_1(x,\bar{x}))$

f is a polynomial fn.

$$f(x) = \sum_{i=0} \alpha_i x^i$$

$$f(K(x,\bar{x})) = \sum_{i=0}^{N} \alpha_i K(x,\bar{x})^i$$

As we have proved.

$K(x,\bar{x}) \cdot K(x,\bar{x})$ is a valid Mercer kernel.

$K(x,\bar{x})^i$ is a valid Mercer kernel.

Also if $(K(x,\bar{x}))$ is a valid Mercer kernel
we have proved that product of
Mercer kernel is also a valid Mercer kernel

$\alpha K(x,\bar{x})^i$ is a valid Mercer kernel.

d) $K(x,\bar{x}) = \exp(K_1(x,\bar{x}))$

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

$$e^{K(x,\bar{x})} = 1 + K(x,\bar{x}) + \frac{K(x,\bar{x})^2}{2!} + \frac{K(x,\bar{x})^n}{n!}$$

$K(x,\bar{x})^i$ is a valid Mercer kernel proven by (b).

$\alpha(K(x,\bar{x}) + \beta(K(x,\bar{x}))$ is also valid proved by (a).

$\therefore e^{(K(x,\bar{x}))}$ can be expressed using sum of multiplications and is also valid.

B

$$K(x,y) = e^{-1/2 \|x-y\|^2}$$

$$e^{-1/2\|x-y\|^2} = e^{-x^2/2} \cdot e^{-y^2/2} \cdot e^{xy}.$$

Using taylor expansion for the $e^{xy}$ term.

$$e^{xy} = \sum_{j=0}^{\infty} \frac{(xy)^j}{j!}$$

$$K(x,y) = \phi(x)\cdot\phi(y)$$

where $\phi(x) = e^{-x^2/2} \cdot \left(\sum_{j=0}^{\infty} \frac{x^j}{j!}\right)$

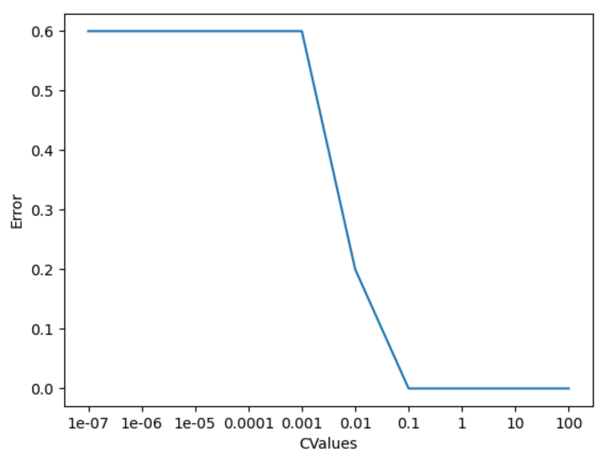
# Problem 2

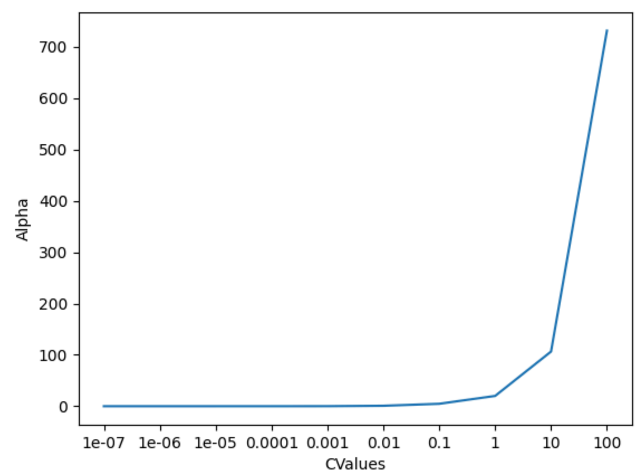
Used cross validation to split the data into 2 parts.
Experimented with linear, polynomial and rbf kernels. Along with varying the c, polynomial order and sigma values.

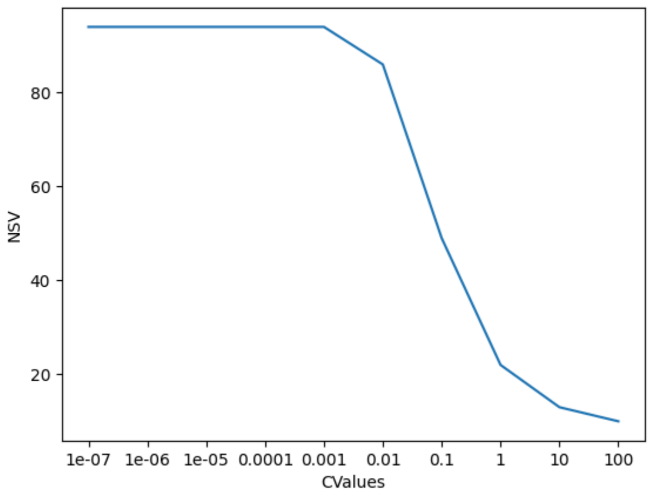
## For Linear kernel

Experimented with the following C Values:
`[0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100]`

The error reached 0 and stayed 0 for C > 0.01






On comparing with the Sum(Alpha) values graph and the NSV graph, the alphas remain constant until C>1 and the number of support vectors also reduces after C>1. So 0.01>C>1 are good values of C

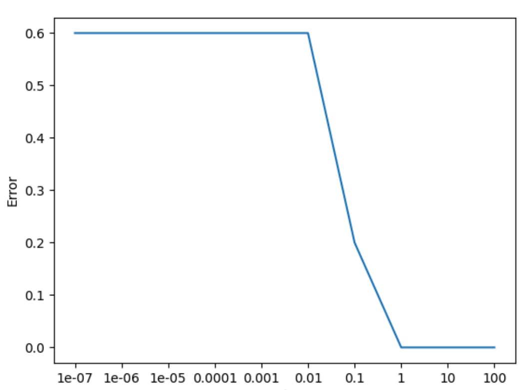
## For Polynomial kernel

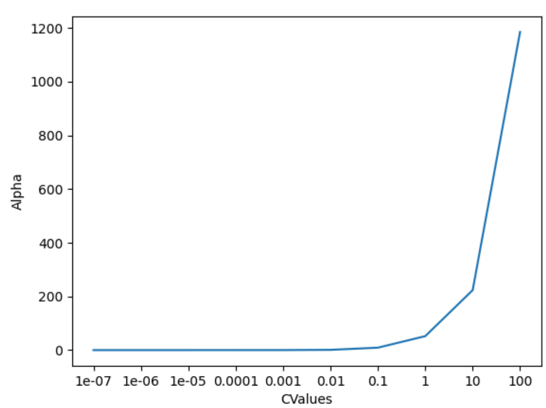
Experimented with the following
C Values:
`[0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100]`

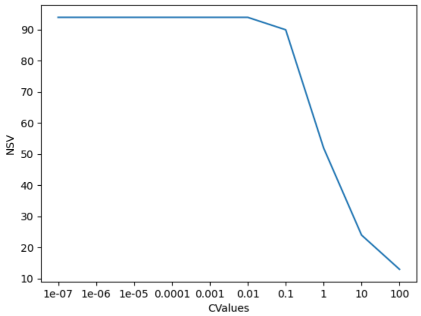
Degree Values:
[1,2,3,4,5]
We can vary C as well as the degree of the polynomial

Varying C while keeping the degree fixed yields the following graphs

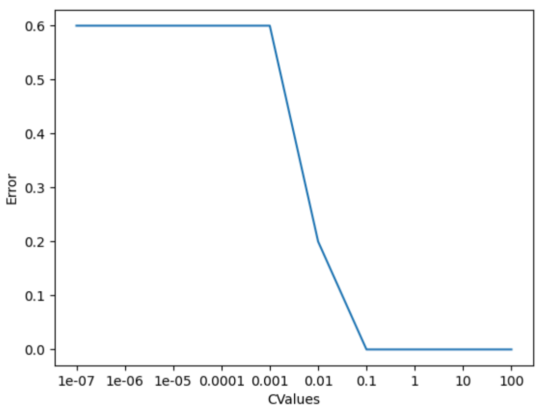
For degree = 1

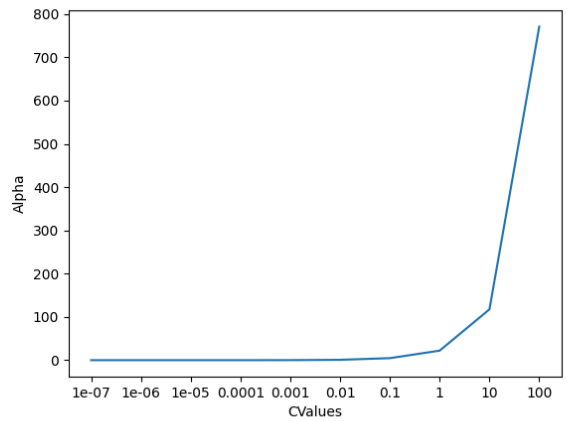


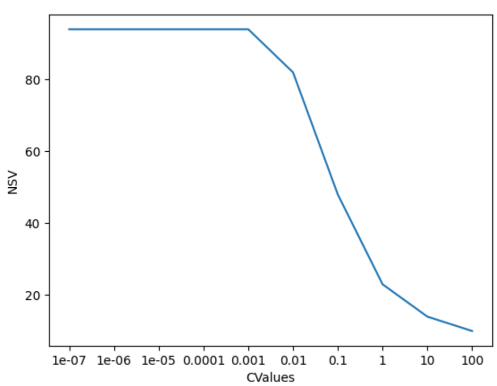





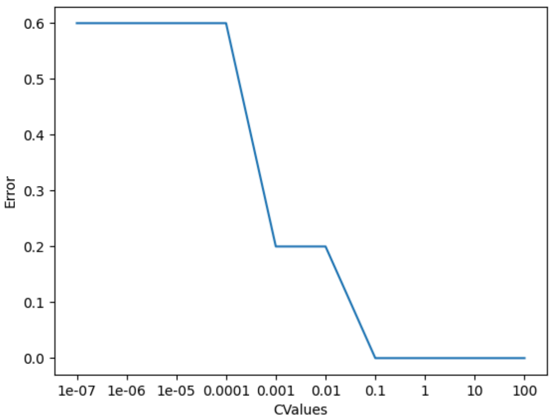

For degree = 2

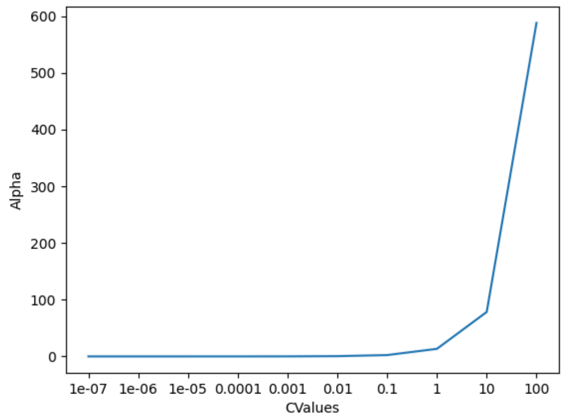


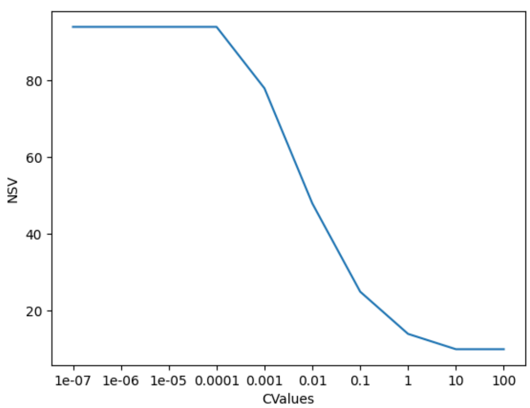



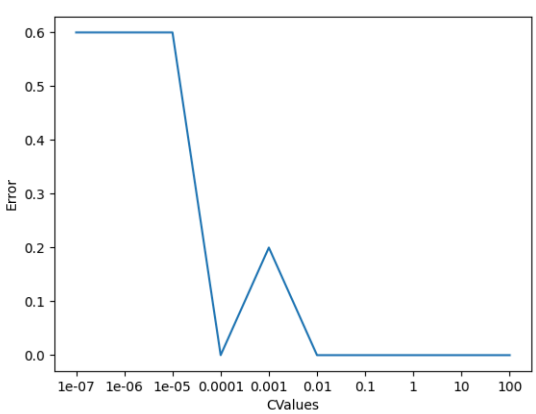
For degree = 3

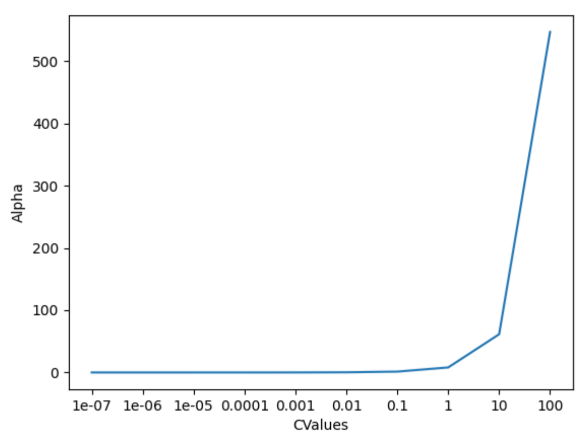


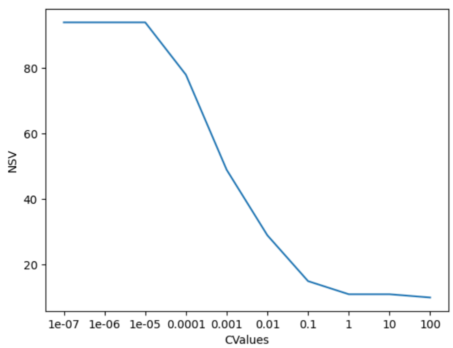



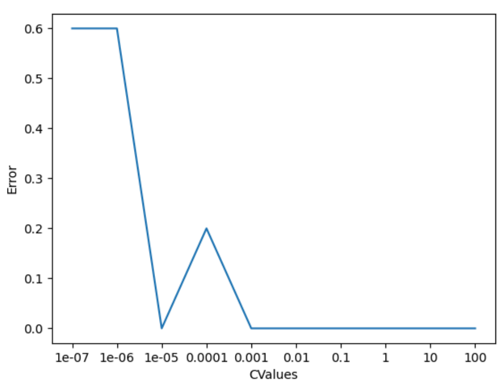

For degree = 4

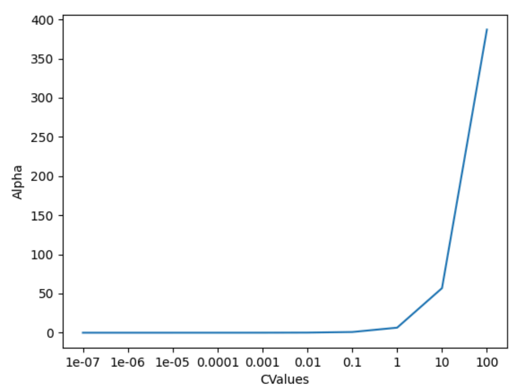


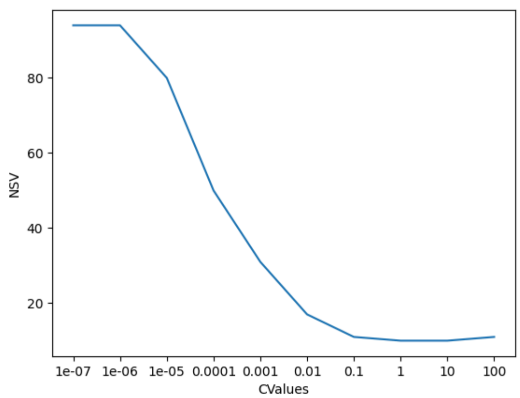






For degree = 5



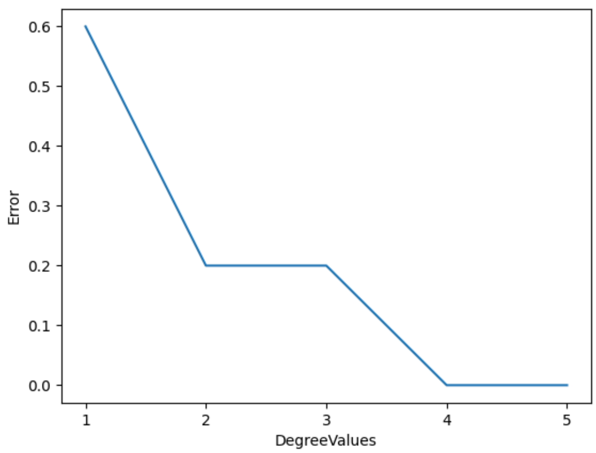

Looking at the Error graphs for the different degrees, the error becomes 0 for all degrees when C > 0.01

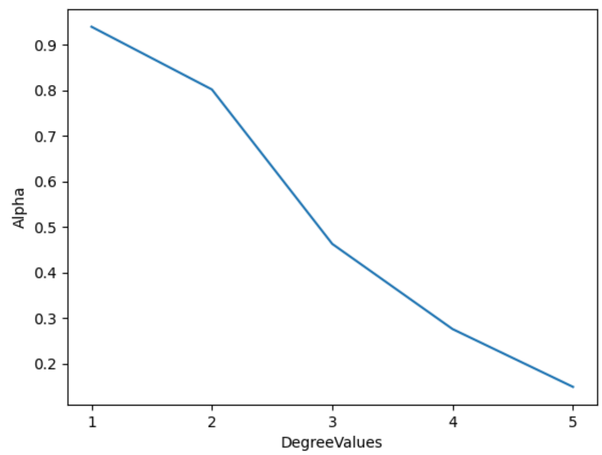
The number of support Vector graphs across different degrees shows that the number of support vectors reduces after C >1 which means it doesnt fit the data as well

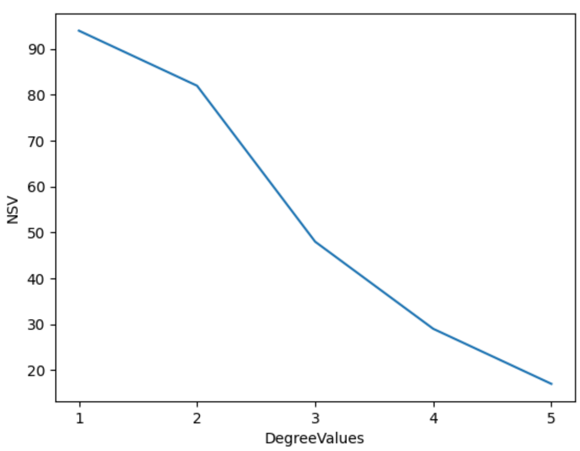
**So Optimum Value of C for the polynomial kernel is between 0.01 and 1**

Varying the degrees while keeping C fixed yields the following graphs

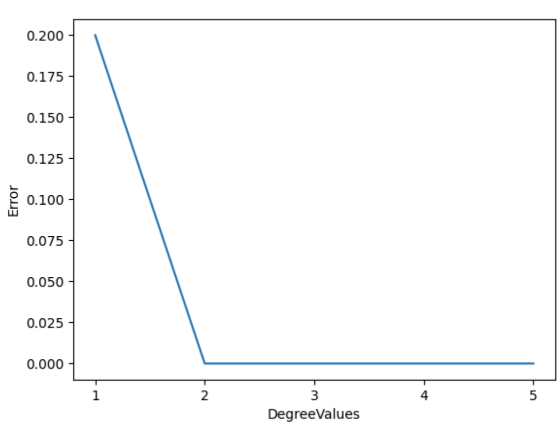
For C = 0.01

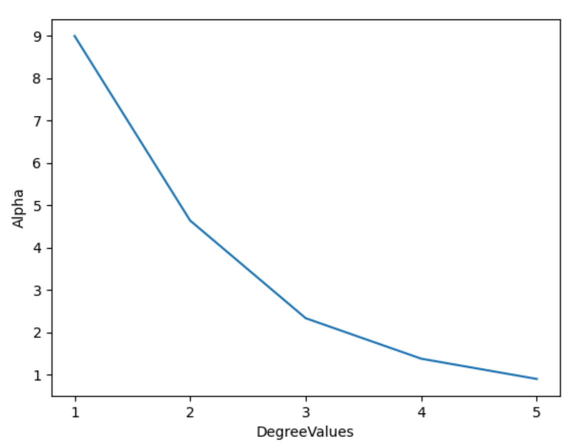


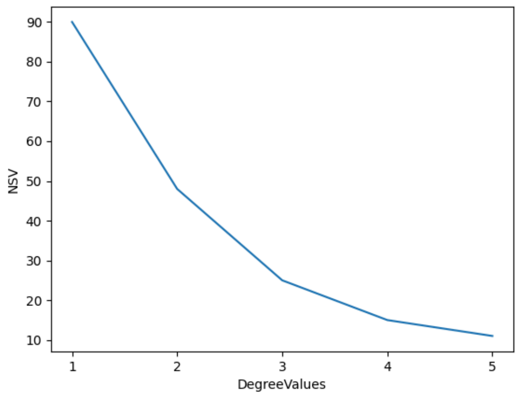



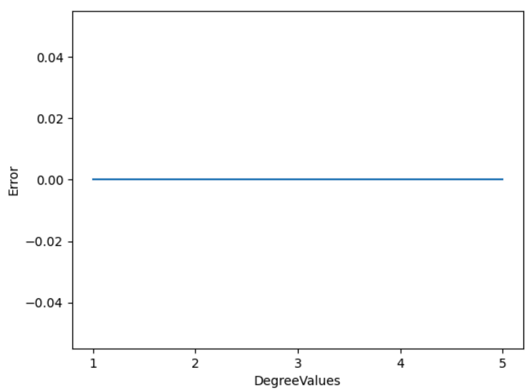

For C = 0.1

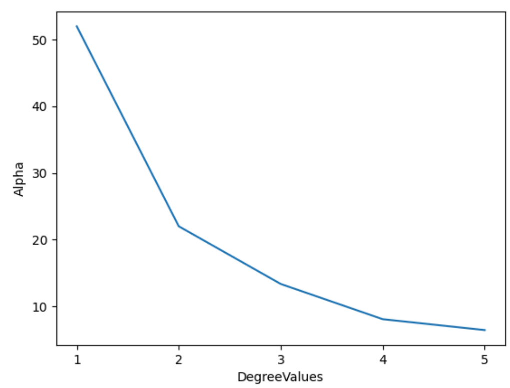


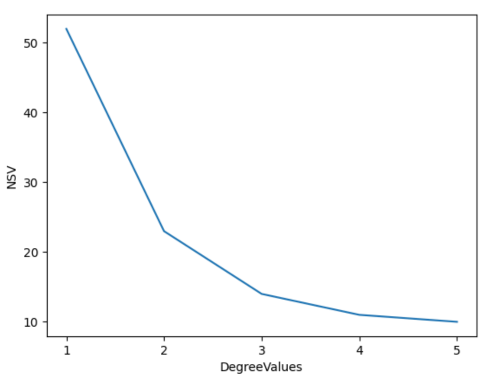






For C = 1




Across all values of C looking at the error graphs with respect to the degree, the error is 0 for **degrees 4 and 5 which makes it the optimum.**

## For RBF kernel

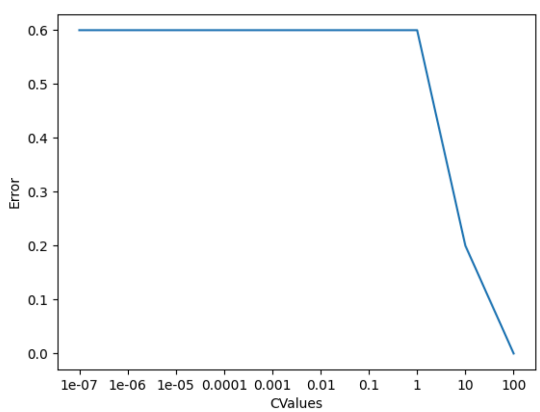
Experimented with the following
C Values:
`[0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100]`

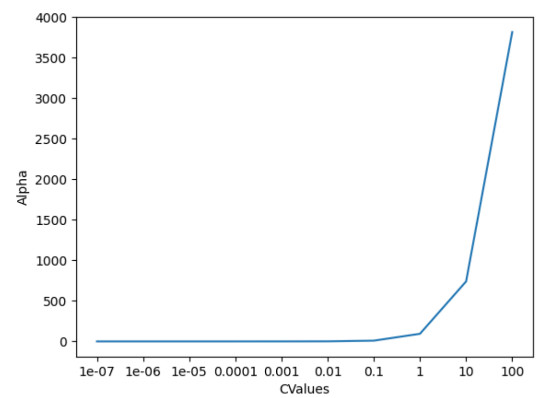
Sigma Values:
[0.001,0.01,0.1,1,10]

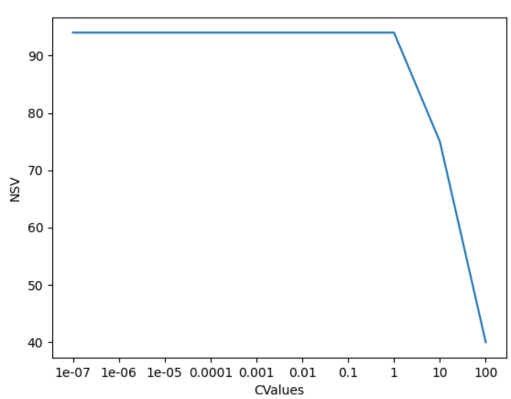
We can vary C as well as the sigma

Varying C while keeping the sigma fixed yields the following graphs

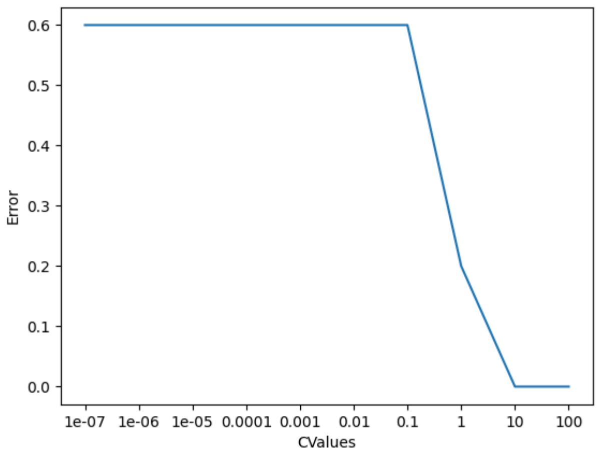
For Sigma = 0.001

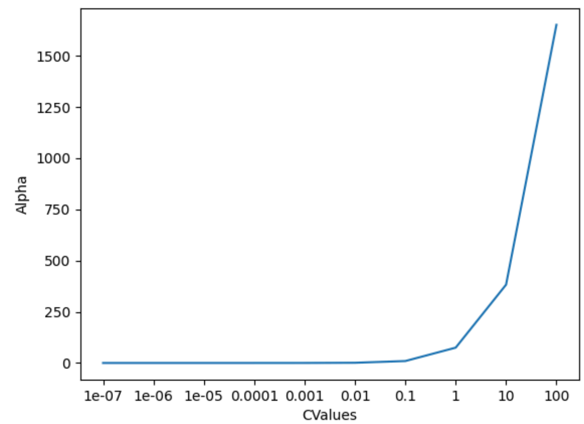


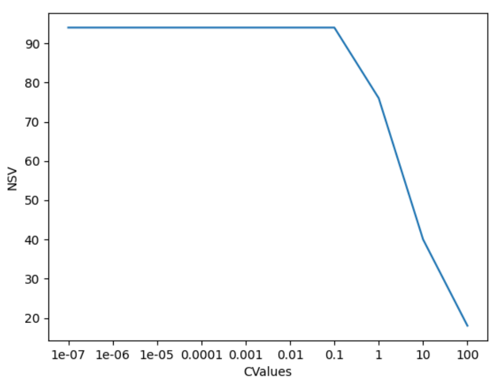



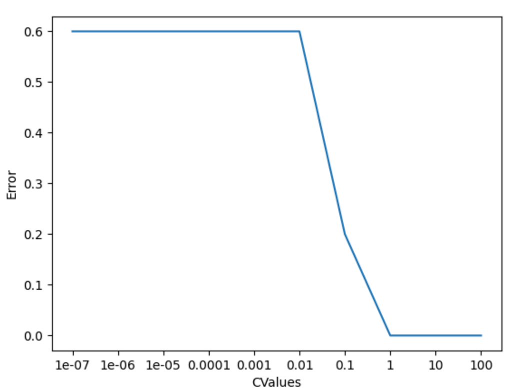
For Sigma = 0.01

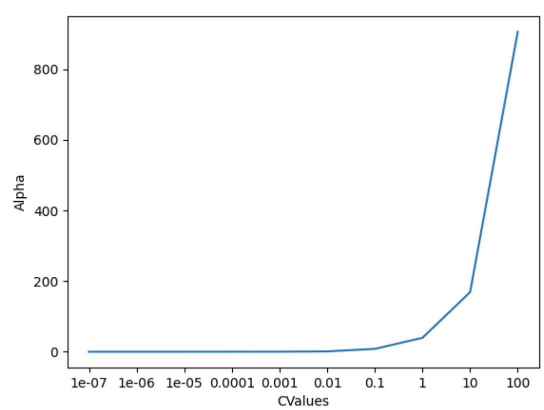


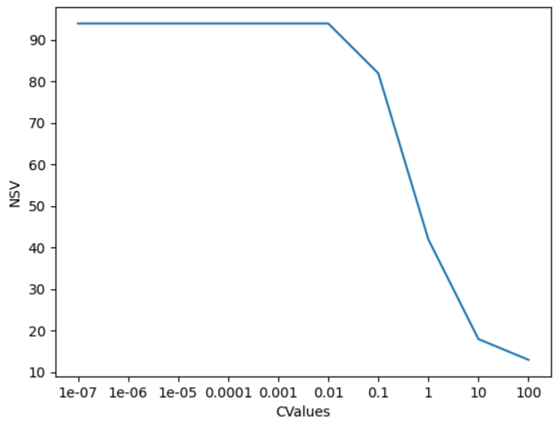



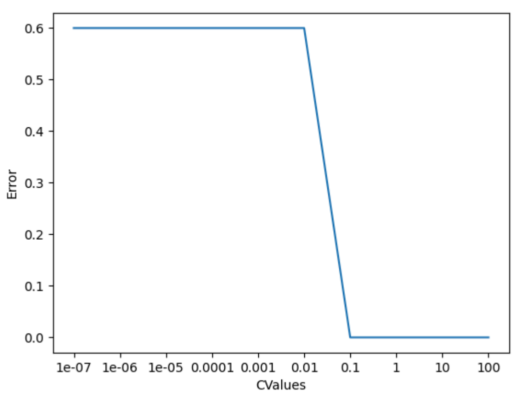

For Sigma = 0.1

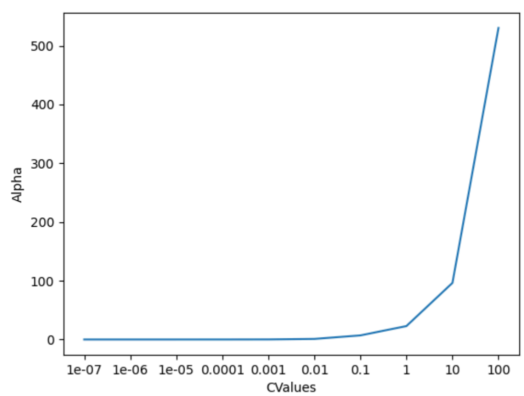


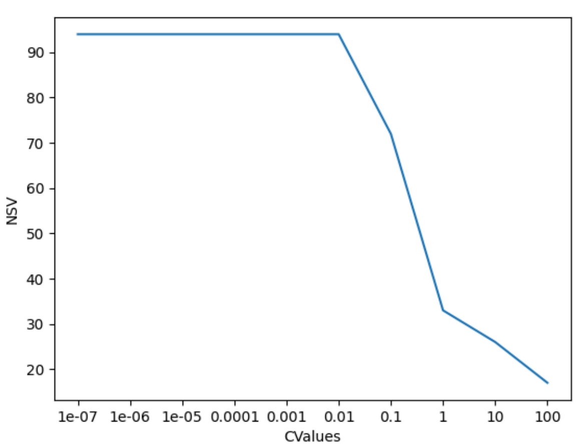




NSV
90
80
70
60
50
40
30
20
10
1e-07
1e-06
1e-05
0.0001
0.001
0.01
0.1
1
10
100
CValues

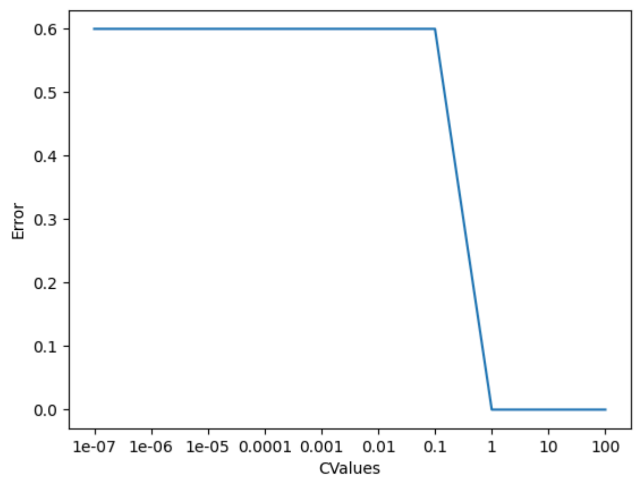
For sigma = 1

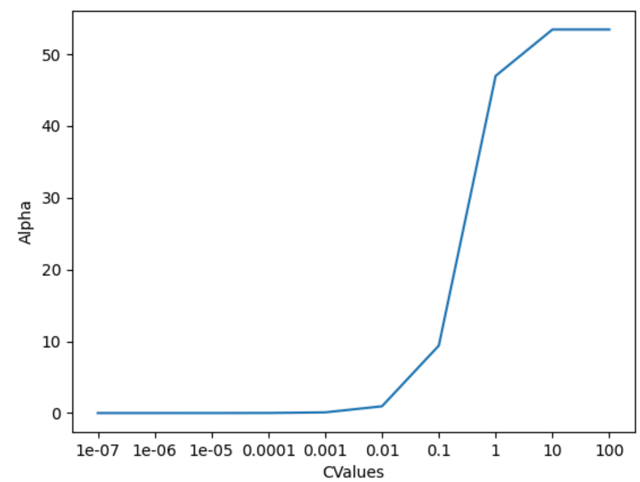

Error
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1e-07
1e-06
1e-05
0.0001
0.001
0.01
0.1
1
10
100
CValues

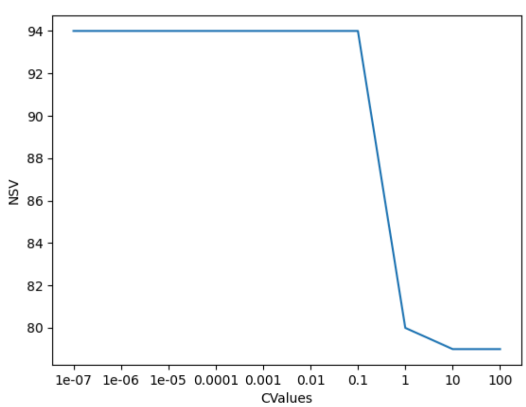

Alpha
500
400
300
200
100
0
1e-07
1e-06
1e-05
0.0001
0.001
0.01
0.1
1
10
100
CValues


NSV
90
80
70
60
50
40
30
20
1e-07
1e-06
1e-05
0.0001
0.001
0.01
0.1
1
10
100
CValues

For Sigma = 10



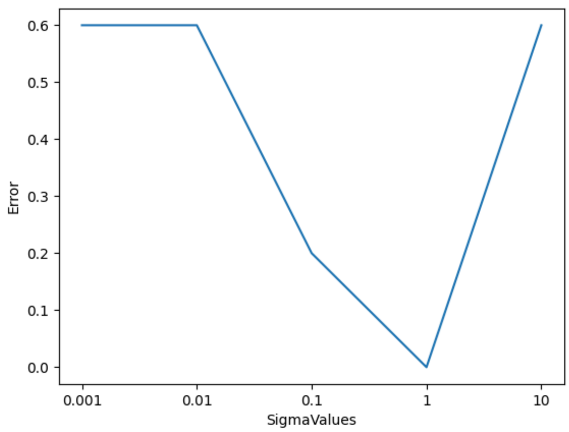



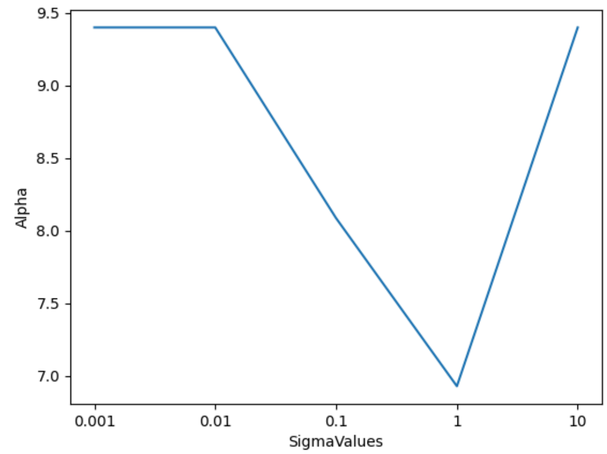

Looking at the Error graphs for the different Sigmas, the error becomes 0 for different degrees when C > 0.1

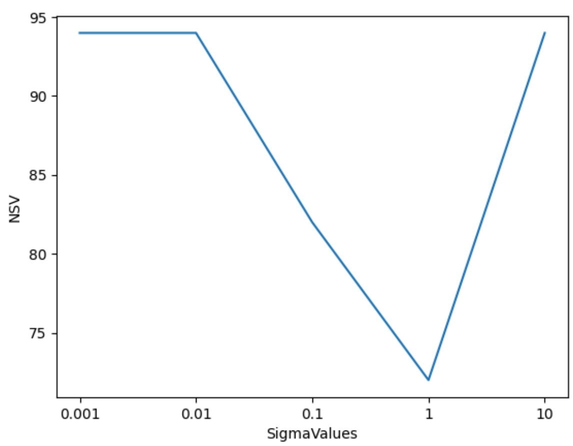
**So Optimum Value of C for the RBF kernel is C>0.1**

Varying the Sigma while keeping C fixed yields the following graphs

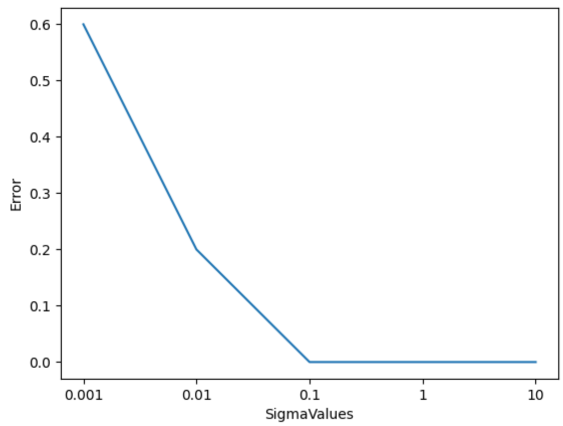
For C = 0.1

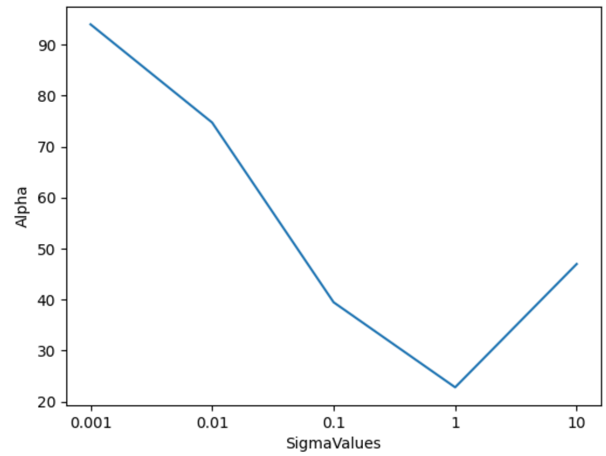


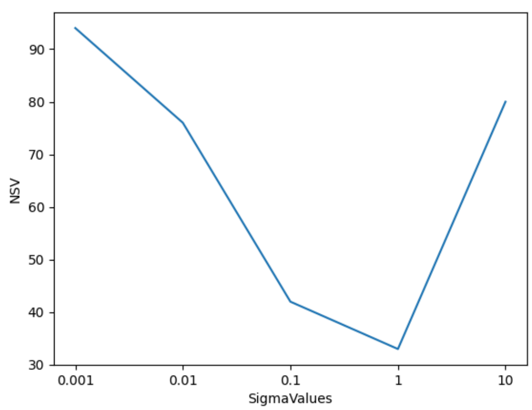



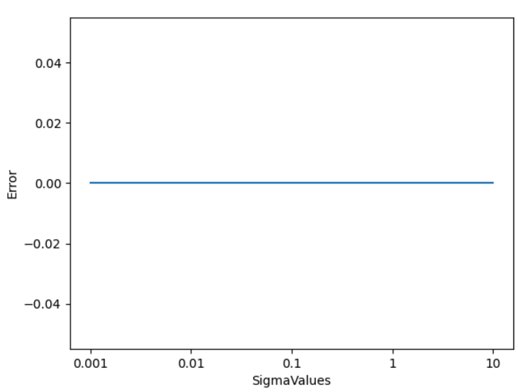
For C = 1

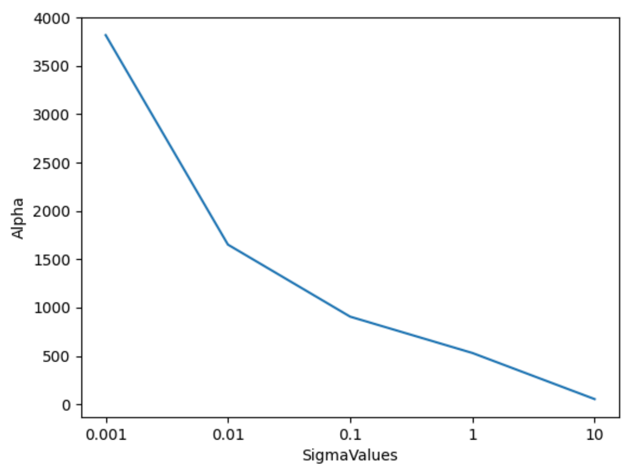


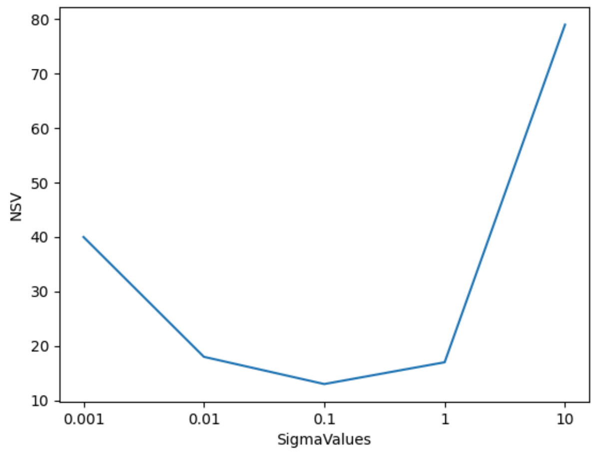




For C = 100






**Looking at the error and alpha graphs sigma values of 1 seems optimum**

# Problem 3

$\cancel{f(x|\alpha)} =$

$$f(x|\alpha) = \alpha^{-2}\, x\, e^{-x/\alpha}$$

$$f(x_i|\alpha) = \prod_{i=1}^{n} \frac{1}{\alpha^2} e^{-x_i/\alpha} = \left(\frac{1}{\alpha^2}\right)^n e^{\frac{-\sum_{i=1}^{n} x_i}{\alpha}}$$

$$= \alpha^{-2n}\, e^{-\sum_{i=1}^{n} x_i/\alpha}$$

Taking log likelihood.

$$\log f(x_i|\alpha) = -2n \log \alpha - \sum_{i=1}^{n} \frac{x_i}{\alpha}$$

Taking derivative to max loglikelihood.

$$\frac{-2n}{\alpha} + \sum_{i=1}^{n} x_i / \alpha^2 = 0$$

$$-2\alpha n + \sum_{i=1}^{n} x_i = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{n} = 2\alpha$$

$$\bar{x} = 2\alpha$$

$$\alpha = \frac{\bar{x}}{2}$$

$x_1 = 0.25$, $x_2 = 0.75$, $x_3 = 1.5$, $x_4 = 2.5$, $x_5 = 2$.

$\bar{x} = \frac{0.25 + 0.75 + 1.5 + 2.5 + 2}{5} = 7/5 \quad \alpha = 7/10 = 0.7$