CSCI 5525 Machine Learning

Assignment 2

Ritiz Tambi

tambi004

Using 1 Grace Day

Problem 2 partner - Krishna Chaitanya Bandi (bandi014)

# Problem 1

### A)

Lagrange Dual for a non-separable soft-margin case:

To fit the above equation for CVXOPT

max(1T𝛼 – ½ 𝛼T𝐾𝛼) => min(½ 𝛼T𝐾𝛼 - 1T𝛼 )

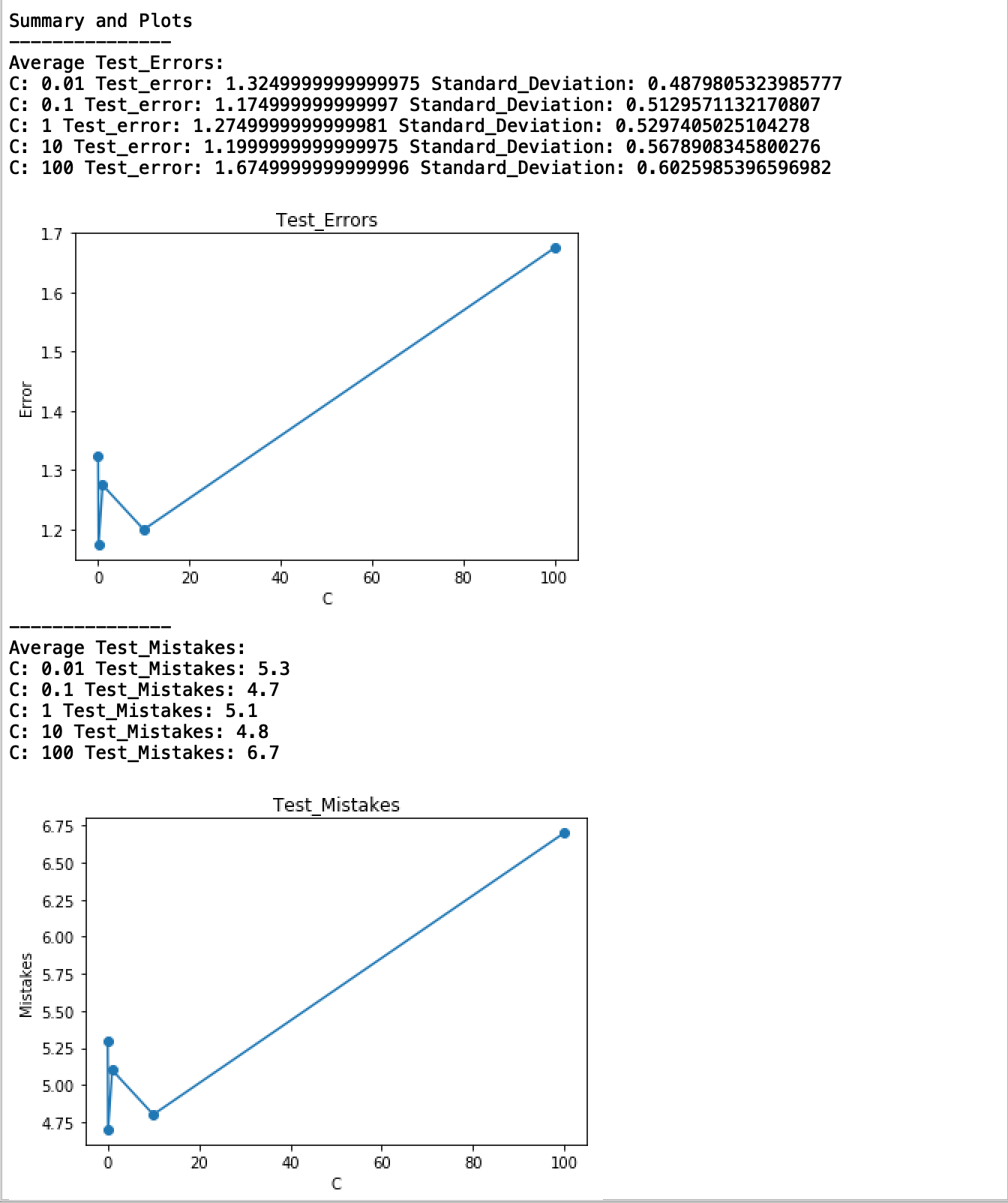
Constraints:

𝛼 ≥ 0, -𝛼 ≥ -C, = 0

### B)

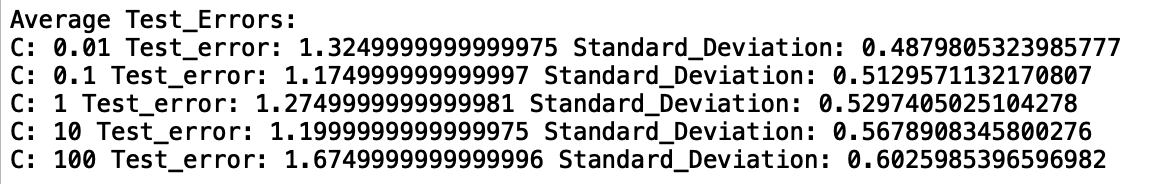
#### (i)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **C** | **Run\_1** | **Run\_2** | **Run\_3** | **Run\_4** | **Run\_5** | **Run\_6** | **Run\_7** | **Run\_8** | **Run\_9** | **Run\_10** |
| **0.01** | 1.5 | 0.5 | 0.75 | 1.25 | 1.75 | 1.75 | 1 | 2.25 | 1.25 | 1.25 |
| **0.1** | 1.75 | 0.75 | 0.5 | 2.0 | 0.75 | 0.75 | 1.25 | 1.75 | 1.5 | 0.75 |
| **1** | 1.25 | 2 | 1.75 | 0.5 | 0.75 | 1.25 | 1.5 | 2 | 1.25 | 0.5 |
| **10** | 1.25 | 0.5 | 0.75 | 1.5 | 1.25 | 2 | 0.75 | 0.75 | 2.25 | 1.25 |
| **100** | 2.25 | 0.75 | 1.25 | 2 | 1 | 1.5 | 1.5 | 3 | 1.5 | 1.75 |



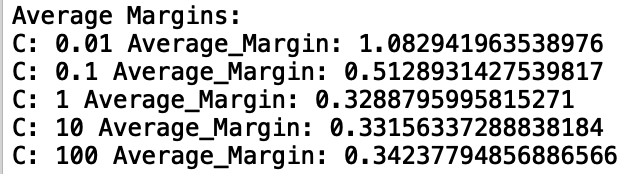
### 

#### (ii)



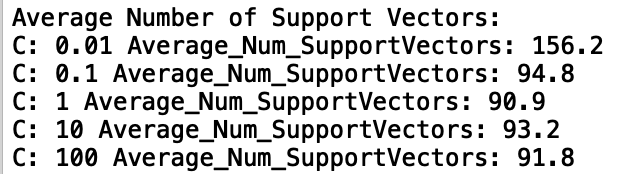
Initially as C increases test-error decreases. With further increase in C values, test-error also increases as depicted by the test-error values and plots above.

#### (iii)



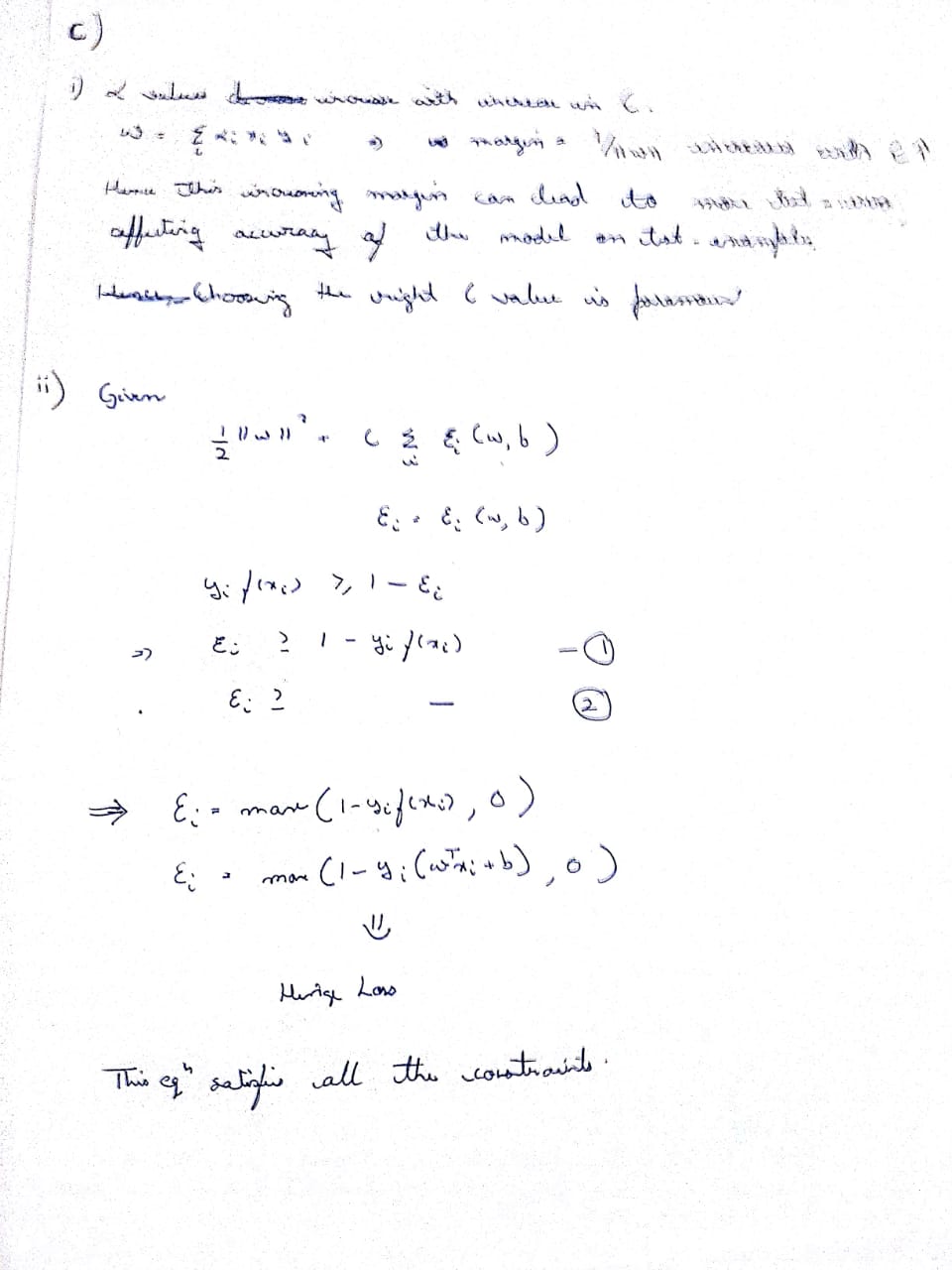
As depicted above by the plots and average geometric-margin values per C; with increase in C values geometric margins decrease

#### (iv)



As depicted above by the plots and average number of Support Vector values per C; with increase in C values number of support vectors decrease. This is due to decreasing number of points within the margins as margins value decrease with C.

### C)



# Problem 2

a)

Implementing the **Pegasos algorithm** stated in the Research Paper, we obtained the following mean and standard deviation of runtime results. We observe that the loss function value decreases rapidly with increasing number of iterations.

Note: Our graphs depict for epochs till 400, but loss decreases further till n\_epochs = 2000

Batch Size (k) = 1

Number of runs = 5

Mean Runtime = 0.0540613174 seconds

Standard Deviation Runtime = 0.0054093021 seconds

A screenshot of a cell phone

Description generated with very high confidence

Batch Size (k) = 20

Number of runs = 5

Mean Runtime = 2.5996386051 seconds

Standard Deviation Runtime = 1.4536558864 seconds

A screenshot of a cell phone

Description generated with high confidence

Batch Size (k) = 200

Number of runs = 5

Mean Runtime = 3.8312240600 seconds

Standard Deviation Runtime = 1.8121650722 seconds

A screenshot of a cell phone

Description generated with very high confidence

Batch Size (k) = 1000

Number of runs = 5

Mean Runtime = 6.1666266918 seconds

Standard Deviation Runtime = 0.7572492165 seconds

A screenshot of a cell phone

Description generated with very high confidence

Batch Size (k) = 2000

Number of runs = 5

Mean Runtime = 1.6600903987 seconds

Standard Deviation Runtime = 0.0935010046 seconds

A screenshot of a cell phone

Description generated with very high confidence

### b)

Derivation of the Gradient of the **Softplus** function is as follows,

A close up of text on a white background

Description generated with high confidence

The mean and standard deviation of runtimes we obtained using the **Softplus algorithm** are as follows for different k values.

Batch Size (k) = 1

Number of runs = 5

Mean Runtime = 43.4661227703 seconds

Standard Deviation Runtime = 30.3375075746 seconds

A pencil and paper

Description generated with high confidence

Batch Size (k) = 20

Number of runs = 5

Mean Runtime = 12.4473050117 seconds

Standard Deviation Runtime = 2.3757711437 seconds

A close up of a device

Description generated with high confidence

Batch Size (k) = 200

Number of runs = 5

Mean Runtime = 15.3314484596 seconds

Standard Deviation Runtime = 1.4831238377 seconds

A close up of a device

Description generated with high confidence

Batch Size (k) = 1000

Number of runs = 5

Mean Runtime = 22.4613066673 seconds

Standard Deviation Runtime = 0.7229130157 seconds

A picture containing screenshot

Description generated with high confidence

Batch Size (k) = 2000

Number of runs = 5

Mean Runtime = 14.7364604949 seconds

Standard Deviation Runtime = 4.7252624717 seconds

A picture containing screenshot

Description generated with high confidence

The **Pegasos** algorithm performs stochastic gradient descent on a non-smooth function (hinge loss)  
whereas the other version uses a soft function variant of hinge loss called **Softplus**, we   
observed that the Cost function has a steeper curve or the cost value decreased faster (better convergence) in **Softplus** variant than in **Hinge loss** variant (Pegasos) as we know **stochastic gradient descent** reaches the optimum value much faster in smoother functions compared to non-smoother functions.