Semester II, Academic Year 2022–2023

Machine Learning

[Course Code: COL341]

Submission: Assignment 2

Name	:	Mayand Dangi
Entry Number	:	2019PH10637

System Specifications

Processor	Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz
RAM	16 GB
Operating System	Windows 10

1 Binary Classification

In binary classification, we have tried to find a decision boundary seperating the two sets with some slack. We have implemented soft margin SVM which required solving following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w},b,\zeta} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{n=1}^{N}\zeta_n \\ \text{subject to} \quad & y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 - \zeta_n \text{ and } \zeta_n \geq 0 \quad (n = 1, 2, .., N) \end{aligned} \tag{1}$$

The langrangian form will be given as:

$$\begin{aligned} \min_{\boldsymbol{w},b,\zeta} \quad & \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m \boldsymbol{x}_n^T\boldsymbol{x}_m - \sum_{n=1}^{N}\alpha_n \\ \text{subject to} \quad & \alpha_n \geq 0 \text{ and } C \geq \alpha_n \geq 0 \quad (n = 1, 2, .., N) \end{aligned} \tag{2}$$

and the dual problem of soft SVM is given as:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha} Q_D^T \boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0 \text{ and } \mathbf{0} \leq \boldsymbol{\alpha} \leq C \cdot \mathbf{1} \end{aligned} \tag{3}$$

where,

$$Q_D = \begin{bmatrix} y_1 y_1 K_{11} & \dots & y_1 y_1 K_{1N} \\ y_2 y_1 K_{21} & \dots & y_1 y_1 K_{2N} \\ .. & \dots & .. \\ y_N y_1 K_{N1} & .. & y_1 y_1 K_{NN} \end{bmatrix}$$

$$A_D = \begin{bmatrix} \boldsymbol{y}^T \\ -\boldsymbol{y}^T \\ I_{N \times N} \end{bmatrix}$$

1.1 Analysis for Linear Kernel

- $C = 0.01$: Training accuracy: 99.63% and Validation accuracy: 87.18%
- $C = 0.1$: Training accuracy: 100% and Validation accuracy: 87.18%
- $C = 1.0$: Training accuracy: 100% and Validation accuracy: 87.18%
- $C = 10$: Training accuracy: 100% and Validation accuracy: 87.18%

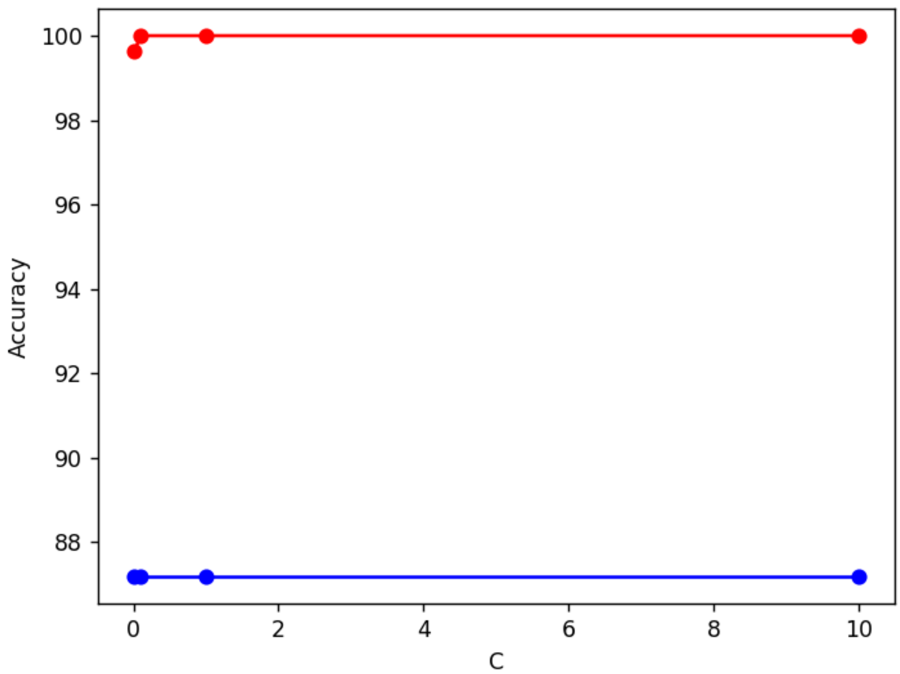
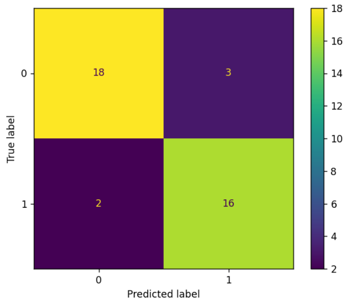
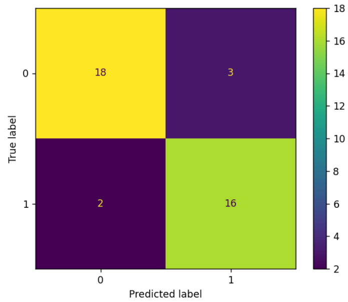


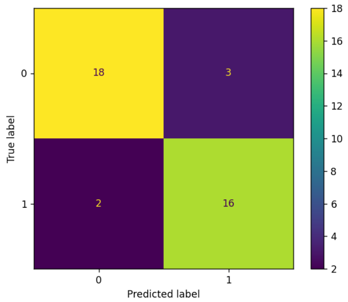
Figure 1: Blue curve represent accuracy of validation set and Red curve represent accuracy of training sets.



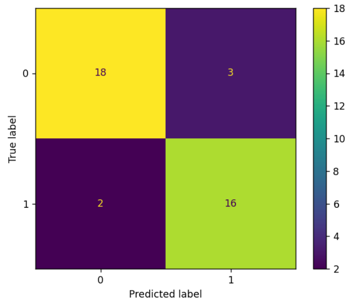
C = 0.01



C = 0.1



C = 1.0



C = 10.0

Figure 2: Confusion Matrix

1.2 Anylsis for RBF Kernel

We have tested accuracy on multiple hyper-parameters for RBF kernel, and results is shown below:

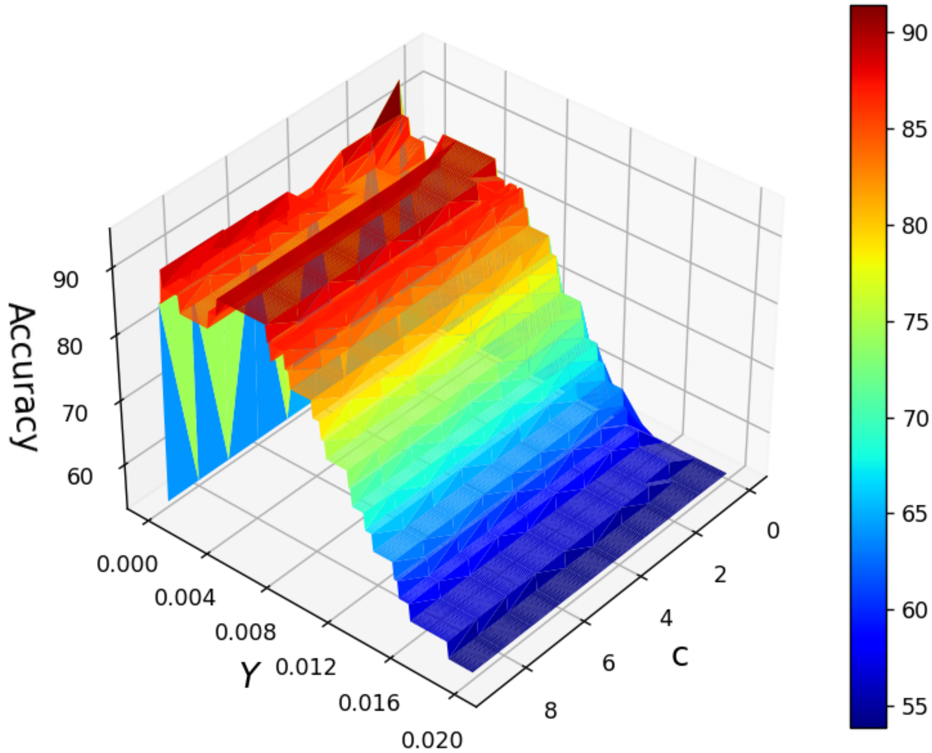


Figure 3: Plot of accuracy w.r.t C and γ

Following is the result of accuracy for different value of C and γ in sets $C = 0.01, 0.1, 1.0, 10$ and $\gamma = 0.1, 0.01, 0.001$.

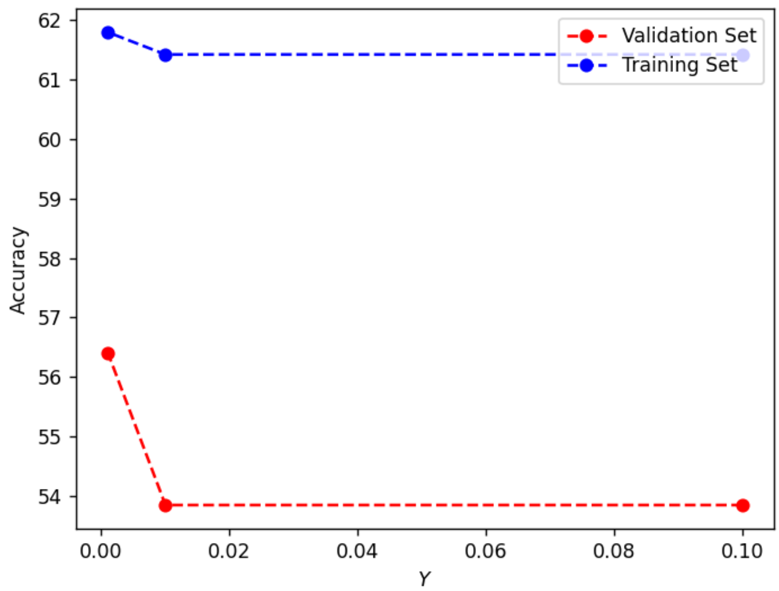


Figure 4: Plot of accuracy vs γ, for C = 0.01

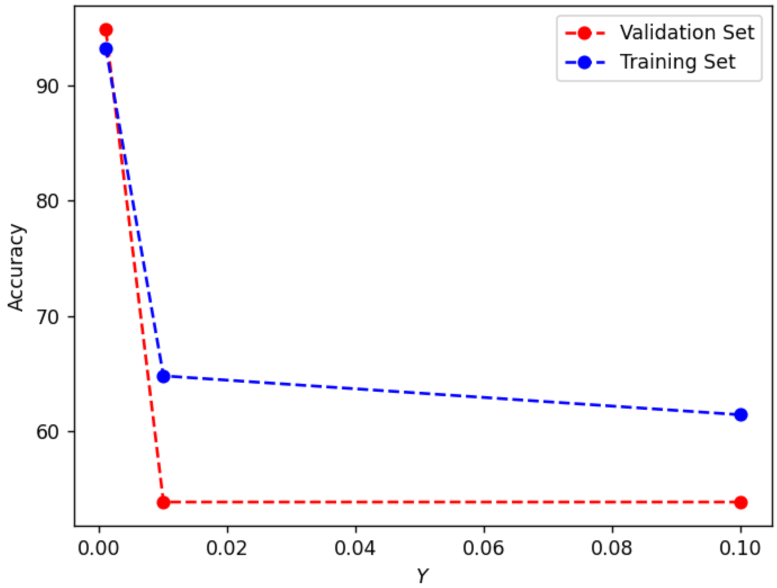


Figure 5: Plot of accuracy vs γ, for C = 0.1

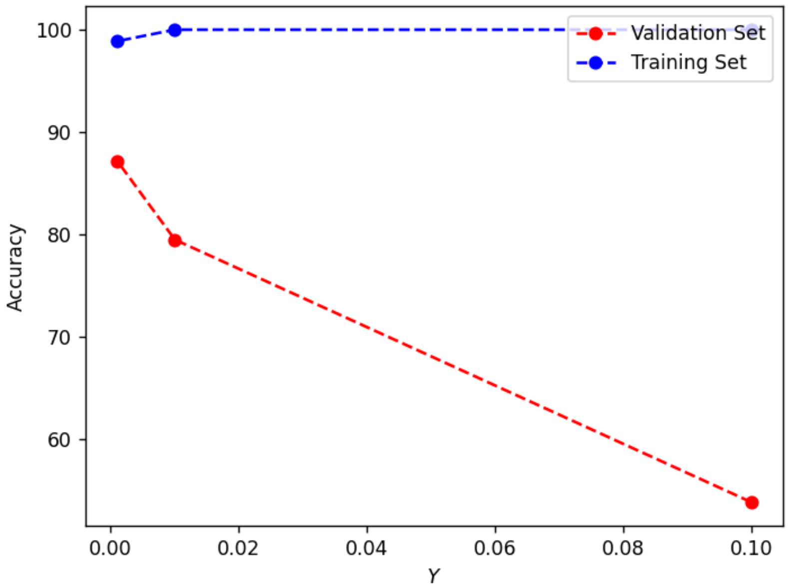


Figure 6: Plot of accuracy vs γ, for C = 1.0

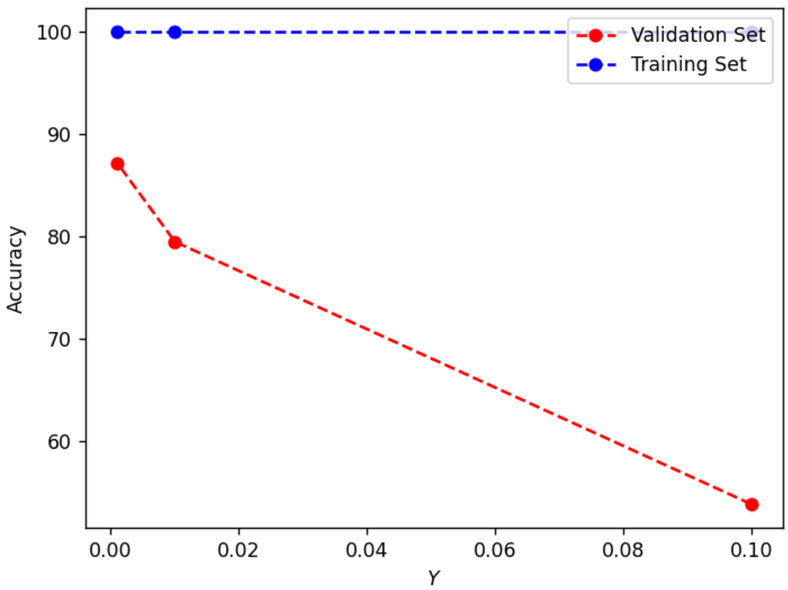


Figure 7: Plot of accuracy vs γ, for C = 10

Accuracy is evaluated on more data points of gamma, which produced following lists.

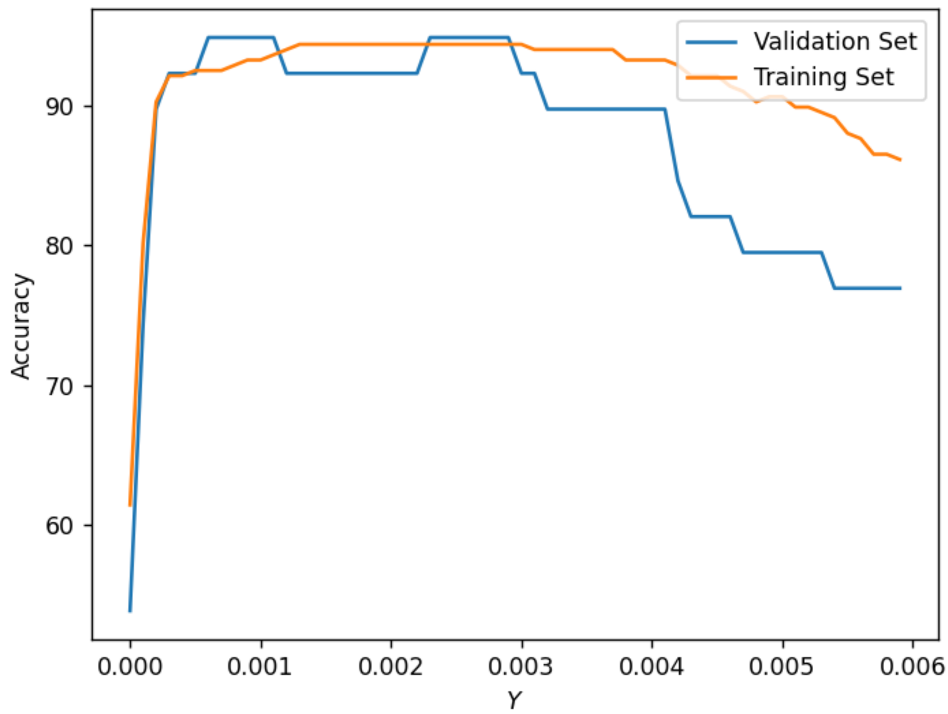


Figure 8: Plot of accuracy vs γ, for C = 0.1

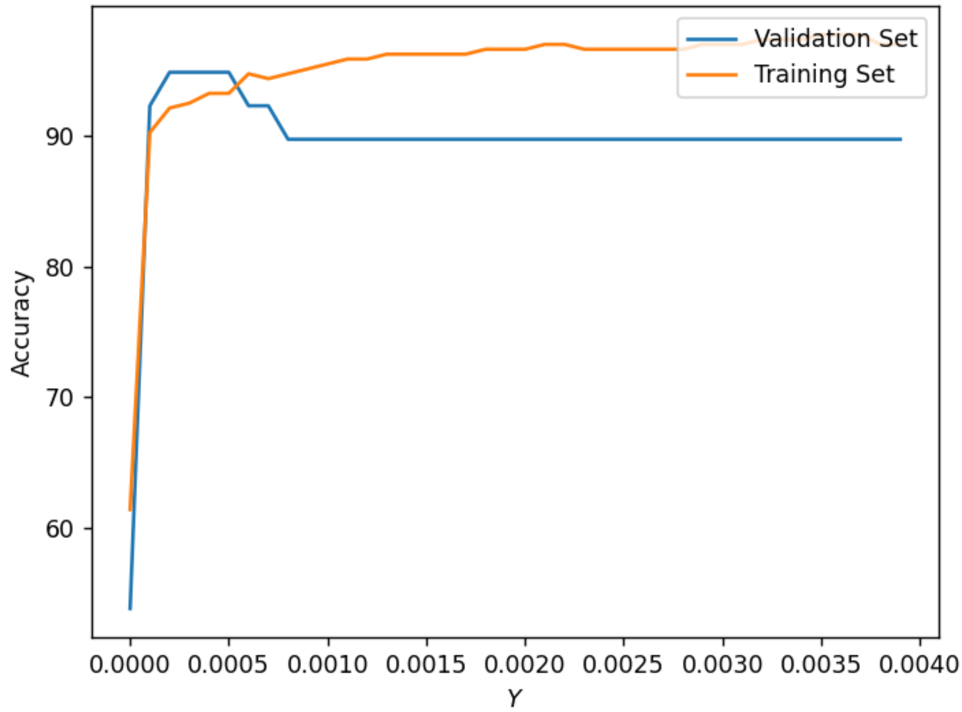


Figure 9: Plot of accuracy vs γ, for C = 0.2

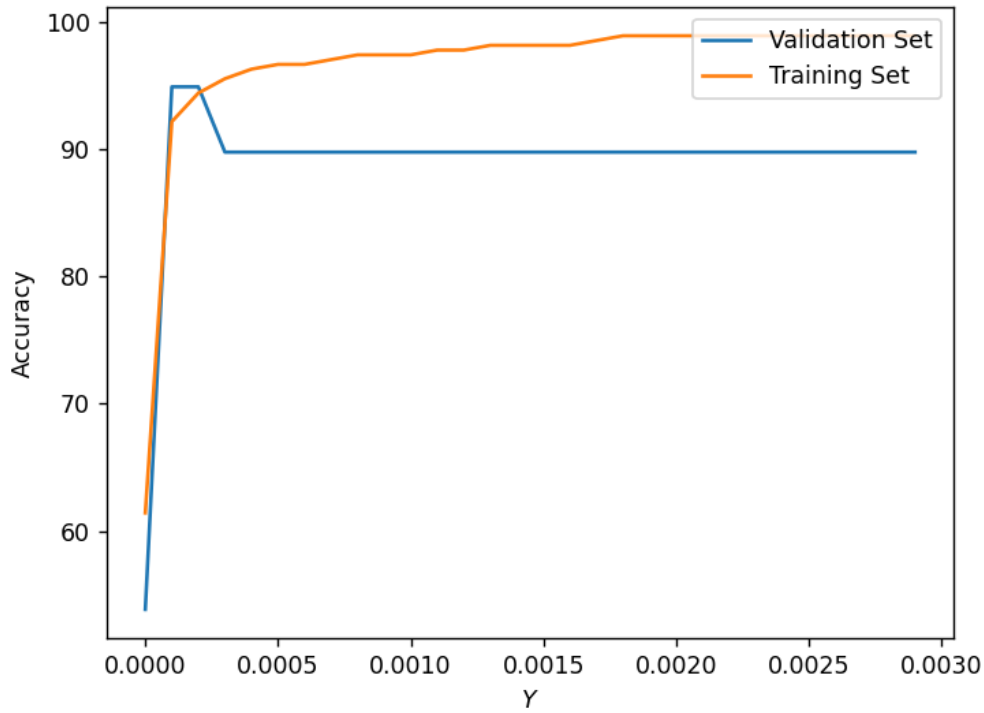
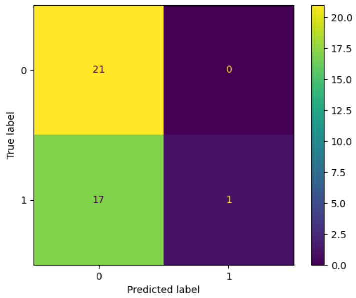
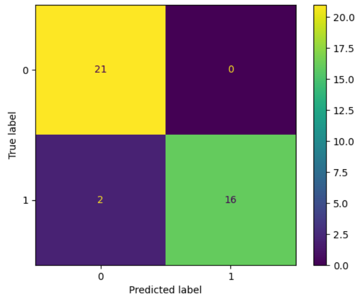


Figure 10: Plot of accuracy vs γ, for C = 0.5

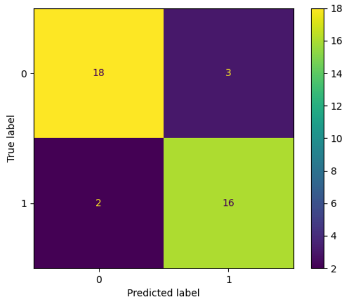
The best pair of hyper-parameters found is for $C = 0.5$ and $\gamma = 0.00025$, giving training accuracy of 95.13% and validation accuracy of 94.87%. The best pair of hyper-parameters found is for $C = 0.2$ and $\gamma = 0.00025$, giving training accuracy of 94.38% and validation accuracy of 92.31%. The best pair of hyper-parameters found is for $C = 0.1$ and $\gamma = 0.00025$, giving training accuracy of 94.37% and validation accuracy of 92.30%.



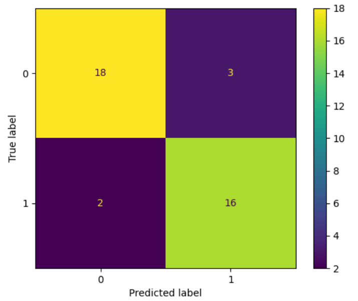
C = 0.01



C = 0.1

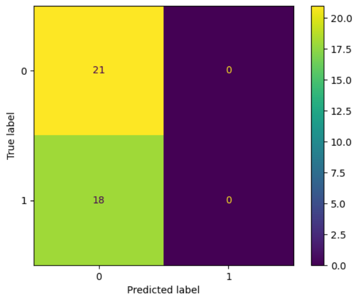


C = 1.0

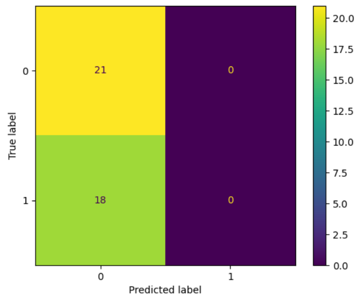


C = 10.0

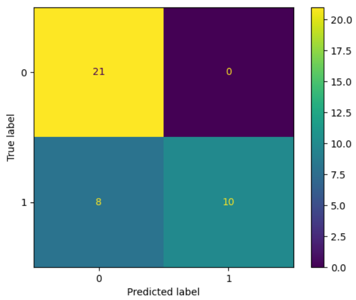
Figure 11: Confusion Matrix for $\gamma = 0.001$



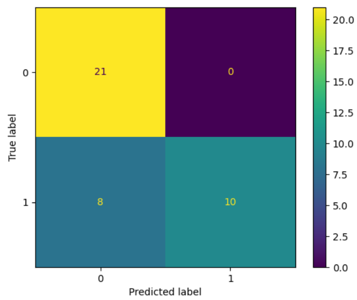
C = 0.01



C = 0.1

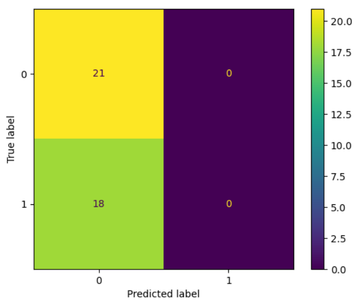


C = 1.0

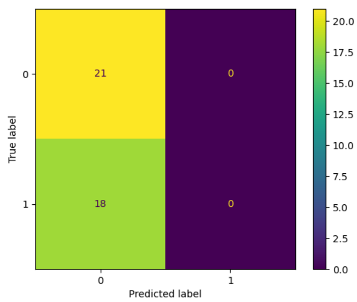


C = 10.0

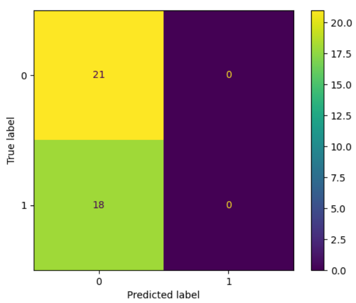
Figure 12: Confusion Matrix for $\gamma = 0.01$



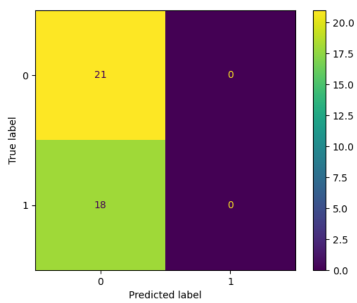
C = 0.01



C = 0.1



C = 1.0



C = 10.0

Figure 13: Confusion Matrix for $\gamma = 0.1$

1.3 Analysis for Polynomial Kernel

We have tested accuracy on multiple hyper-parameters for the polynomial kernel, and results is shown below:

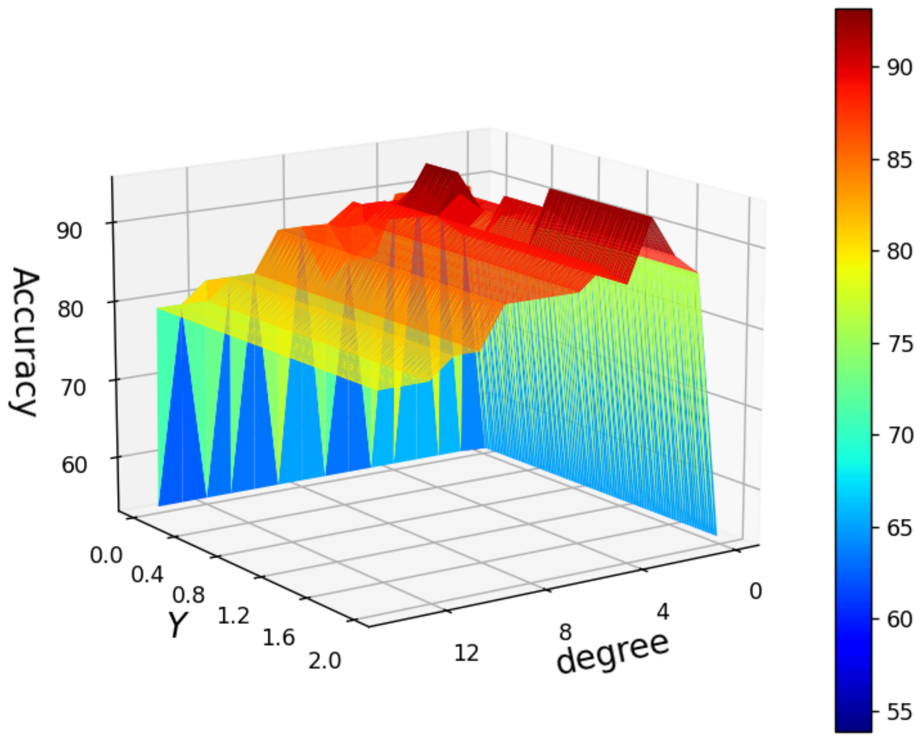


Figure 14: Plot of accuracy w.r.t C and γ

Following is the result of accuracy for different value of C and d in sets $C = 0.01, 0.1, 1.0, 10$ and $d = 3, 5$.

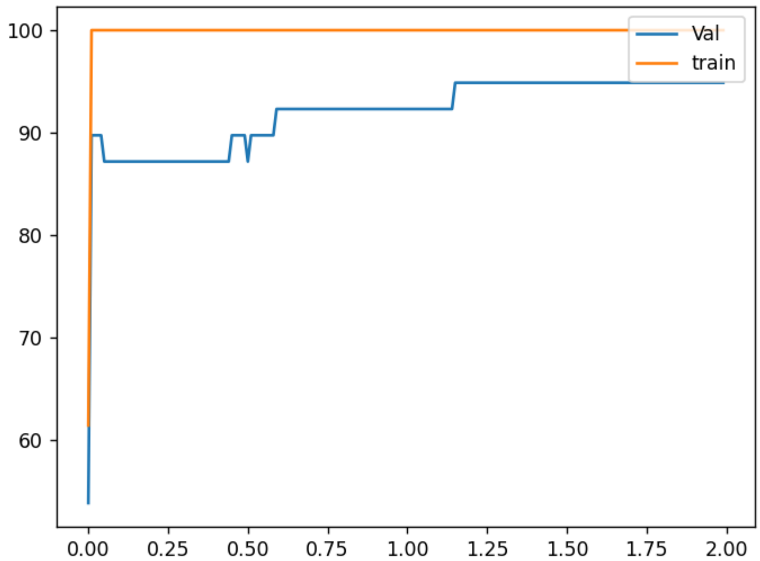


Figure 15: Plot of accuracy vs γ, for C = 0.01 and degree = 3, best hyper-parameters deduce is around $\gamma = 1.5$

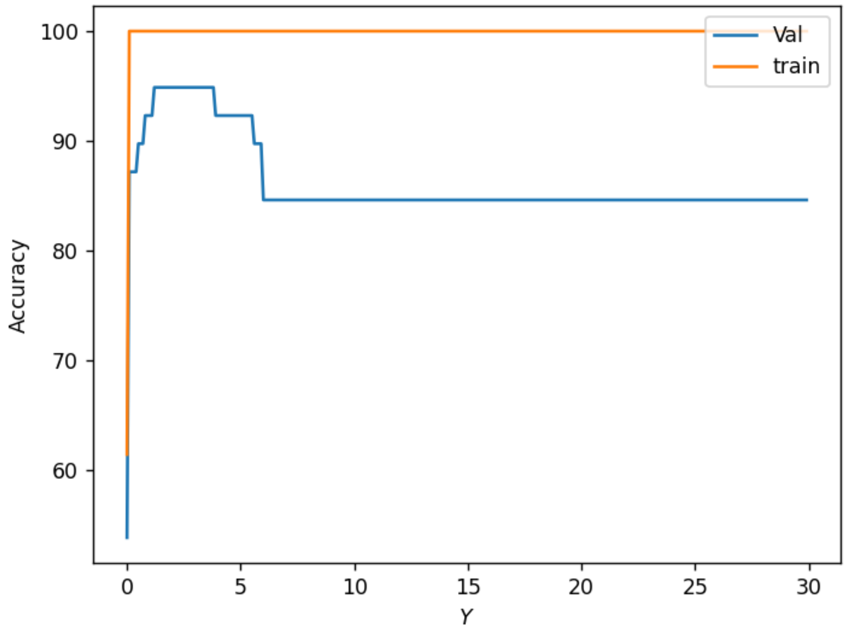


Figure 16: Plot of accuracy vs γ, for C = 0.1 and degree = 3, best hyper-parameters deduce is around $\gamma = 2.0$

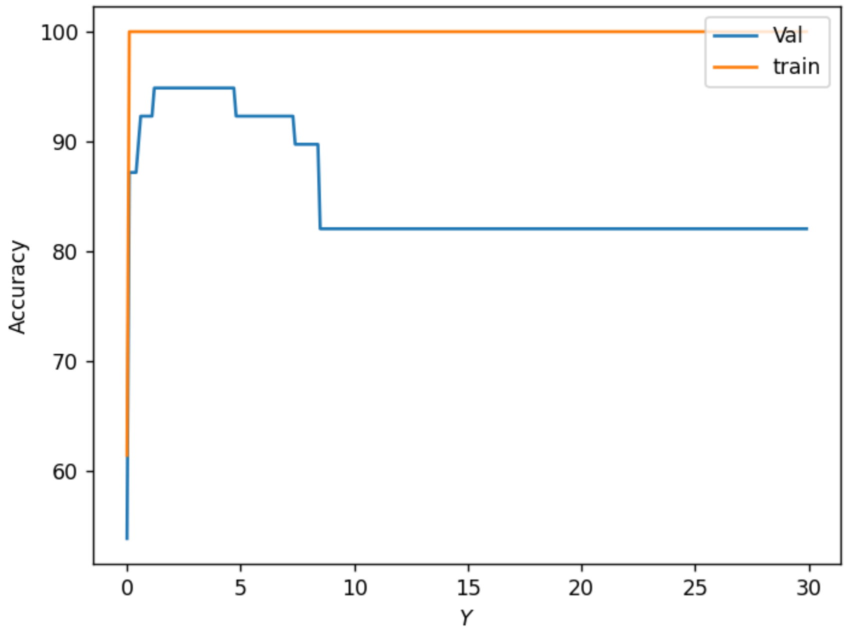


Figure 17: Plot of accuracy vs γ, for C = 1 and degree = 3, best hyper-parameters deduce is around $\gamma = 2.0$

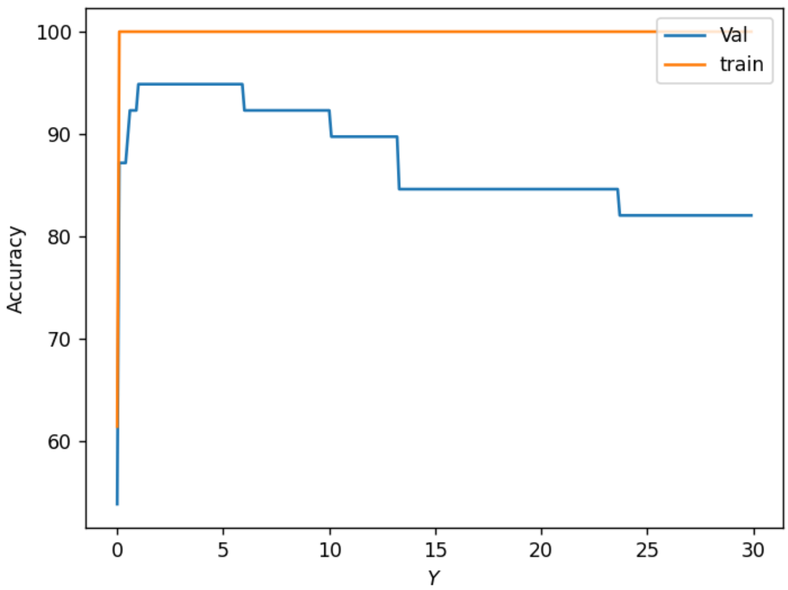


Figure 18: Plot of accuracy vs γ, for C = 10 and degree = 3, best hyper-parameters deduce is around $\gamma = 2.0$

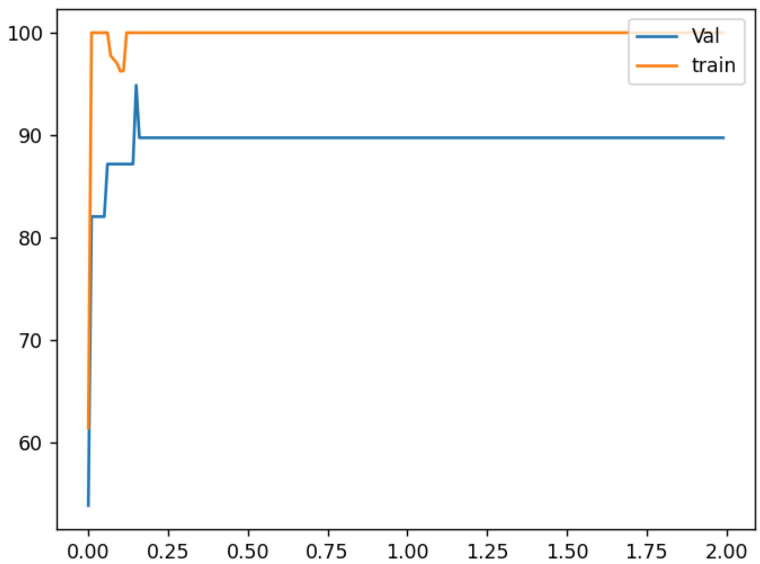


Figure 19: Plot of accuracy vs γ, for C = 0.01 and degree = 5, best hyper-parameters deduce is around $\gamma = 0.15$

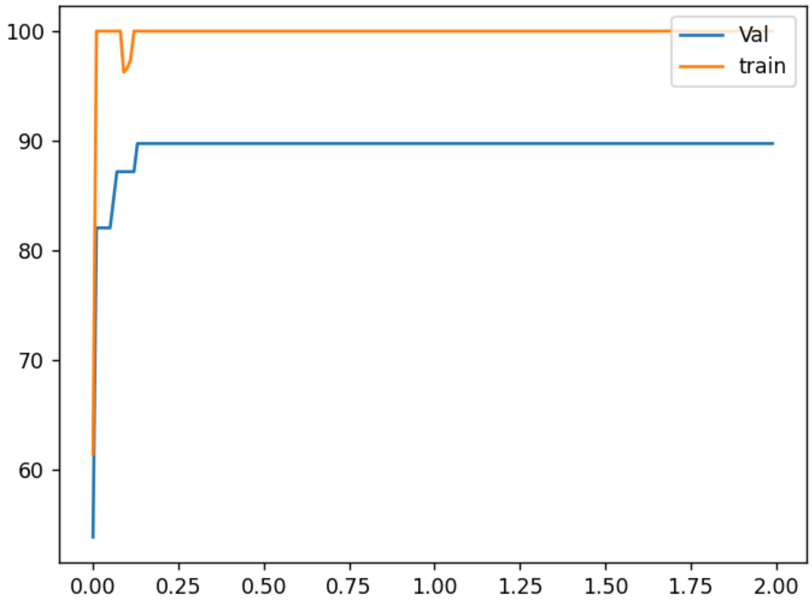


Figure 20: Plot of accuracy vs γ, for C = 0.1 and degree = 5, best hyper-parameters deduce is around $\gamma = 1.0$

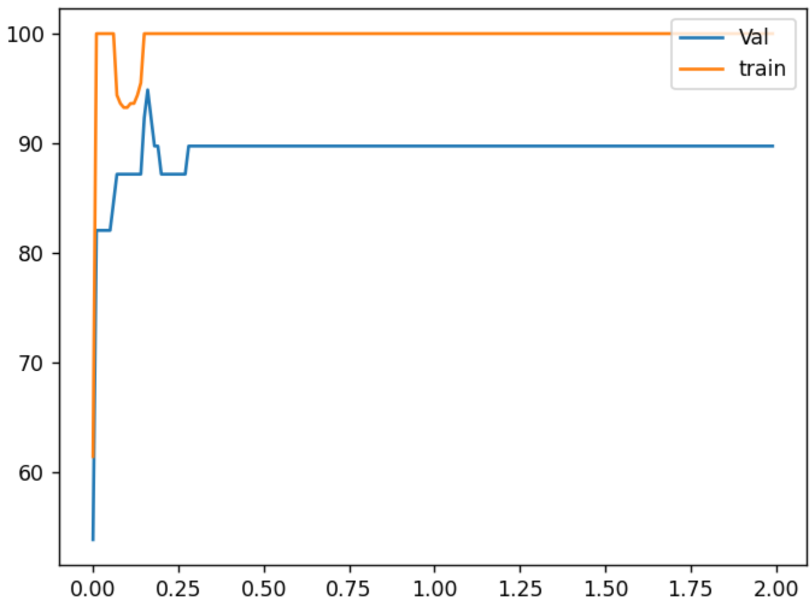


Figure 21: Plot of accuracy vs γ, for C = 1 and degree = 5, best hyper-parameters deduce is around $\gamma = 0.16$

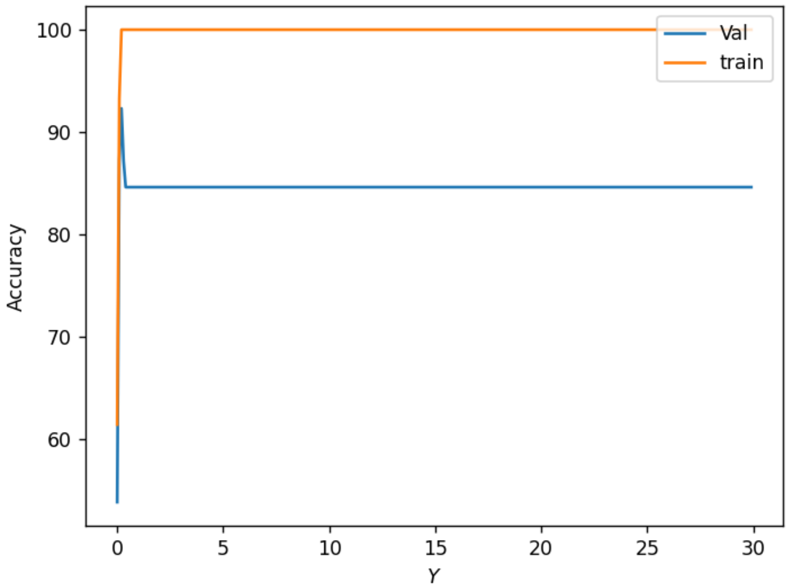
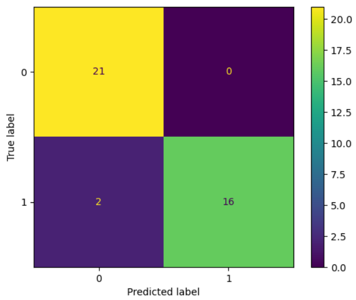
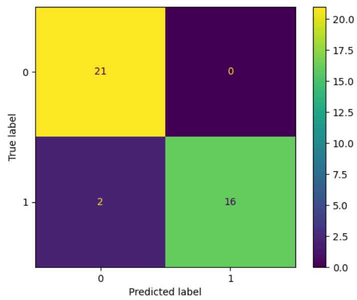


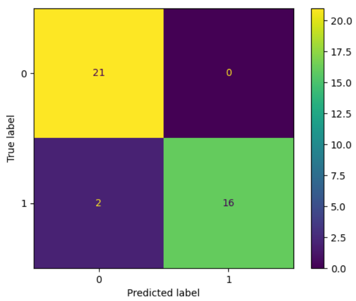
Figure 22: Plot of accuracy vs γ, for C = 10 and degree = 5, best hyper-parameters deduce is around $\gamma = 0.2$



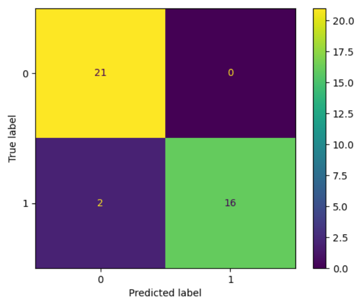
Confusion Matrix for $C = 0.01$, $\gamma = 2.0$ and degree = 3



Confusion Matrix for $C = 0.1$, $\gamma = 2.0$ and degree = 3



Confusion Matrix for $C = 1$, $\gamma = 2.0$ and degree = 3



Confusion Matrix for $C = 10$, $\gamma = 2.0$ and degree = 3

Figure 23: Confusion Matrix for $\gamma = 2$

1.4 Analysis for Sigmoid Kernel

We have tested accuracy on multiple hyper-parameters for the sigmoid kernel, and results is shown below:

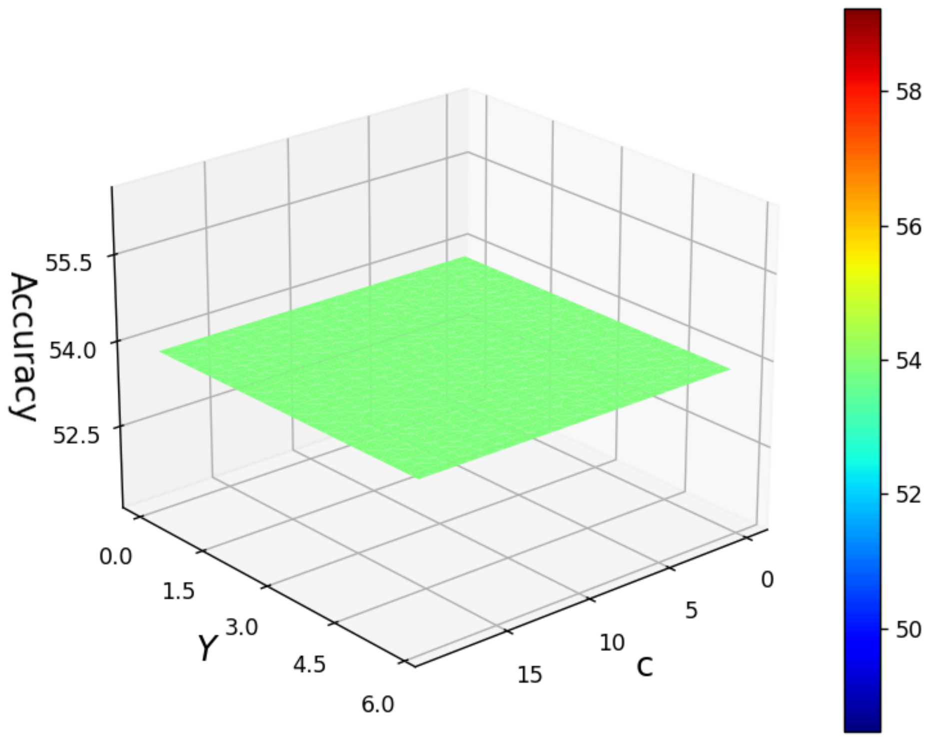
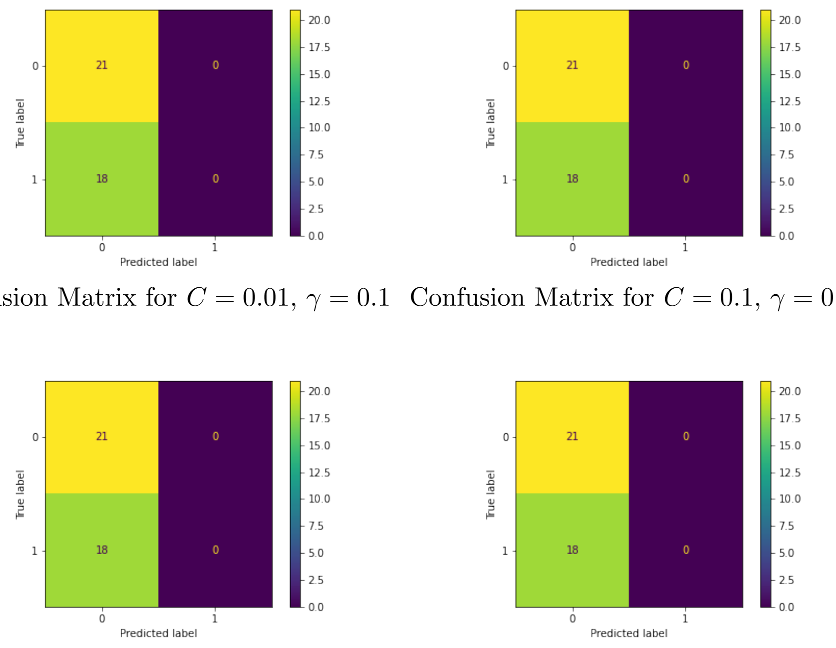


Figure 24: Plot of accuracy w.r.t C and γ



Confusion Matrix for $C = 0.01$, $\gamma = 0.1$ Confusion Matrix for $C = 0.1$, $\gamma = 0.1$

Confusion Matrix for $C = 1$, $\gamma = 0.1$ Confusion Matrix for $C = 10$, $\gamma = 0.1$

Figure 25: Confusion Matrix for $\gamma = 0.1$ and offset $= 0$

1.5 Analysis for Laplacian Kernel

The accuracy of classifier for the Laplacian kernel w.r.t to C and γ and results is shown below:

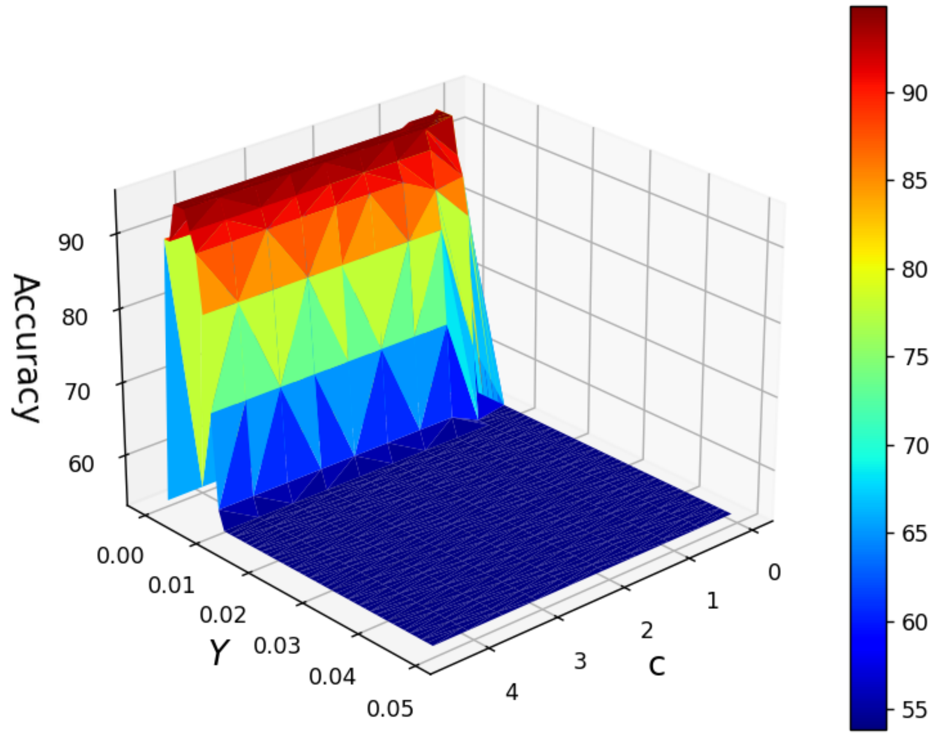


Figure 26: Plot of accuracy w.r.t C and γ

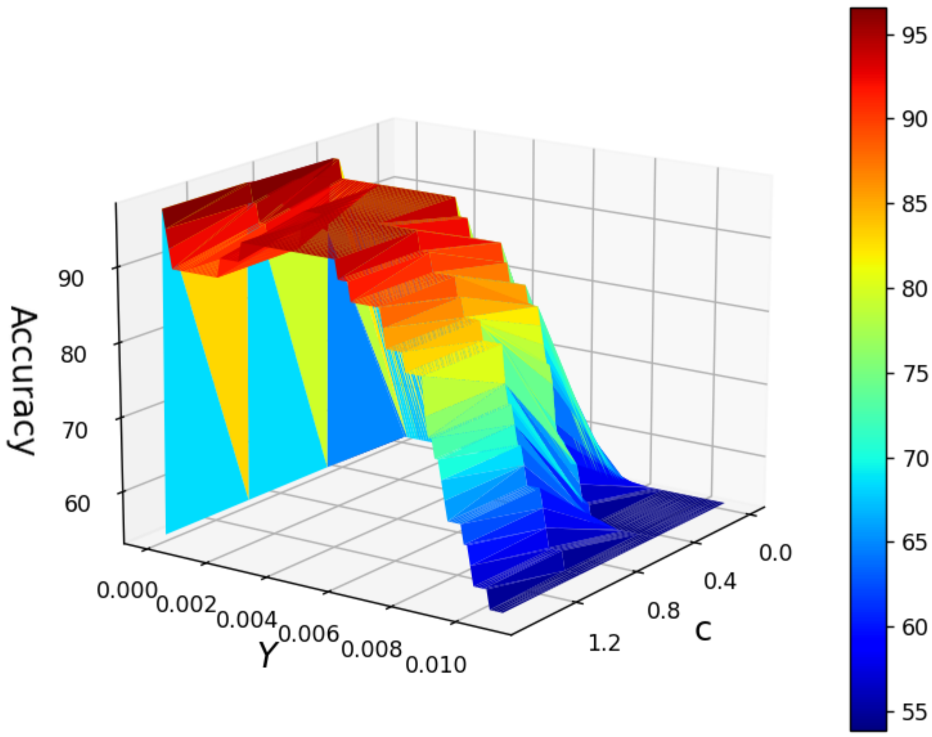
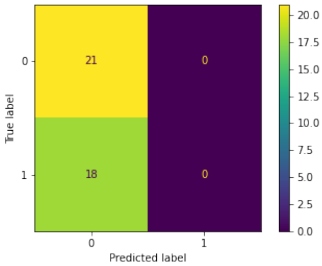
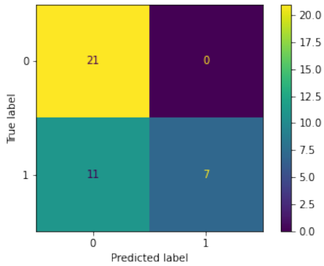


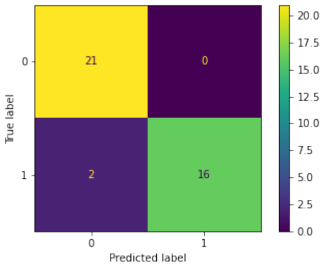
Figure 27: Plot of accuracy w.r.t C and γ. From this graph the optimal value of hyper-parameters obtains are: $C = 0.5$ and $\gamma = 0.0003$ giving accuracy on training set of 100% and on validation set of 97.44%



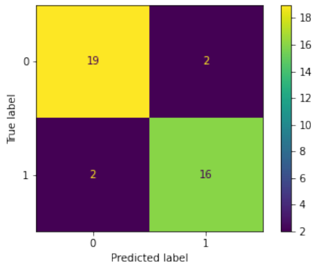
Confusion Matrix for $C = 0.01$, $\gamma = 0.0003$



Confusion Matrix for $C = 0.1$, $\gamma = 0.0003$



Confusion Matrix for $C = 1$, $\gamma = 0.0003$



Confusion Matrix for $C = 10$, $\gamma = 0.0003$

Figure 28: Confusion Matrix for $\gamma = 0.0003$

The best hyper-parameter found is $C = 0.5$ and $\gamma = 0.0003$ giving accuracy on training set of 100% and on validation set of 97.44%.

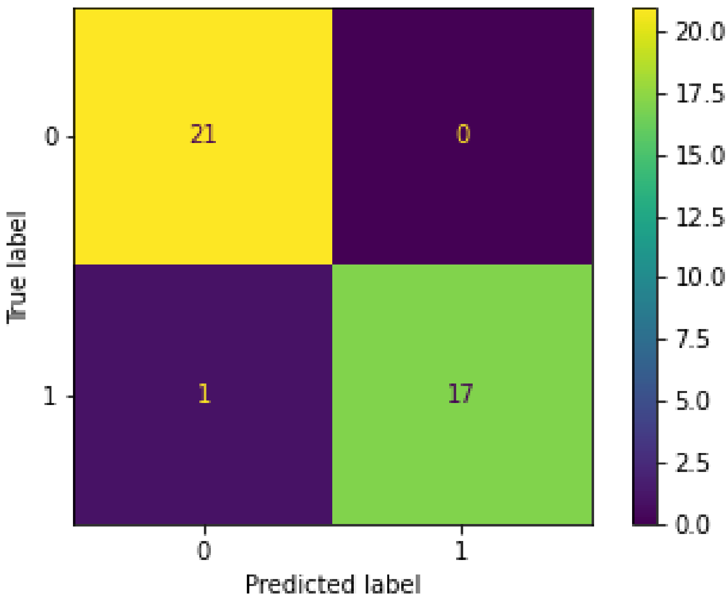


Figure 29: Confusion matrix for $C = 0.5$ and $\gamma = 0.0003$

There one thing to observe that as we increases value of C, Validation error drops off as the slackness in model also decreases.

2 Multi-class classification

Multi-class classification can be done using two approach one is named as One-vs-One and other as One-vs-All.

2.1 One-vs-All

In this scenario we determine a hyperplane which separate between a class and all others at once for each class.

2.1.1 RBF Kernel

The accuracy of classifier for the RBF kernel w.r.t to C and γ and results is shown below:

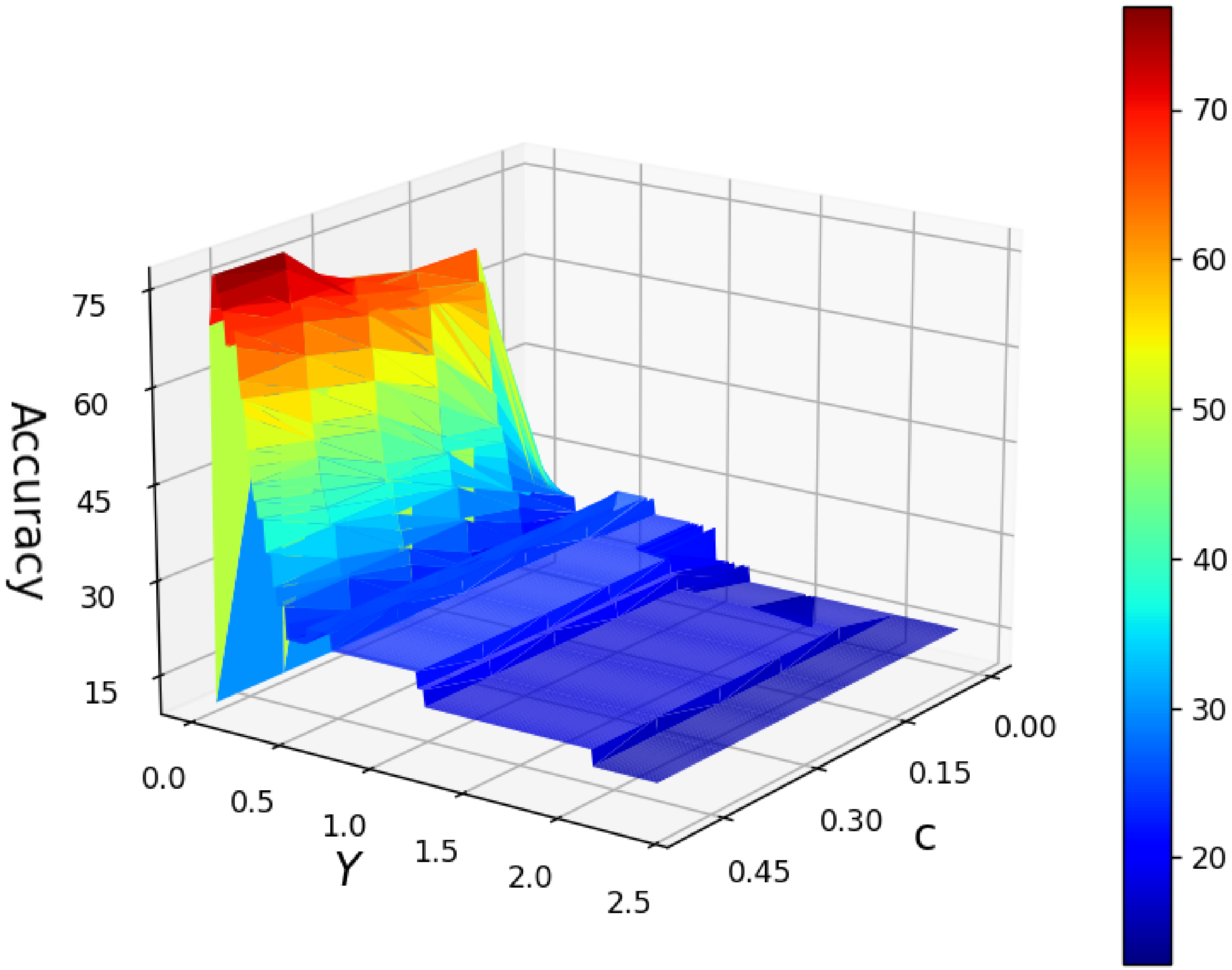


Figure 30: Plot of accuracy w.r.t C and γ

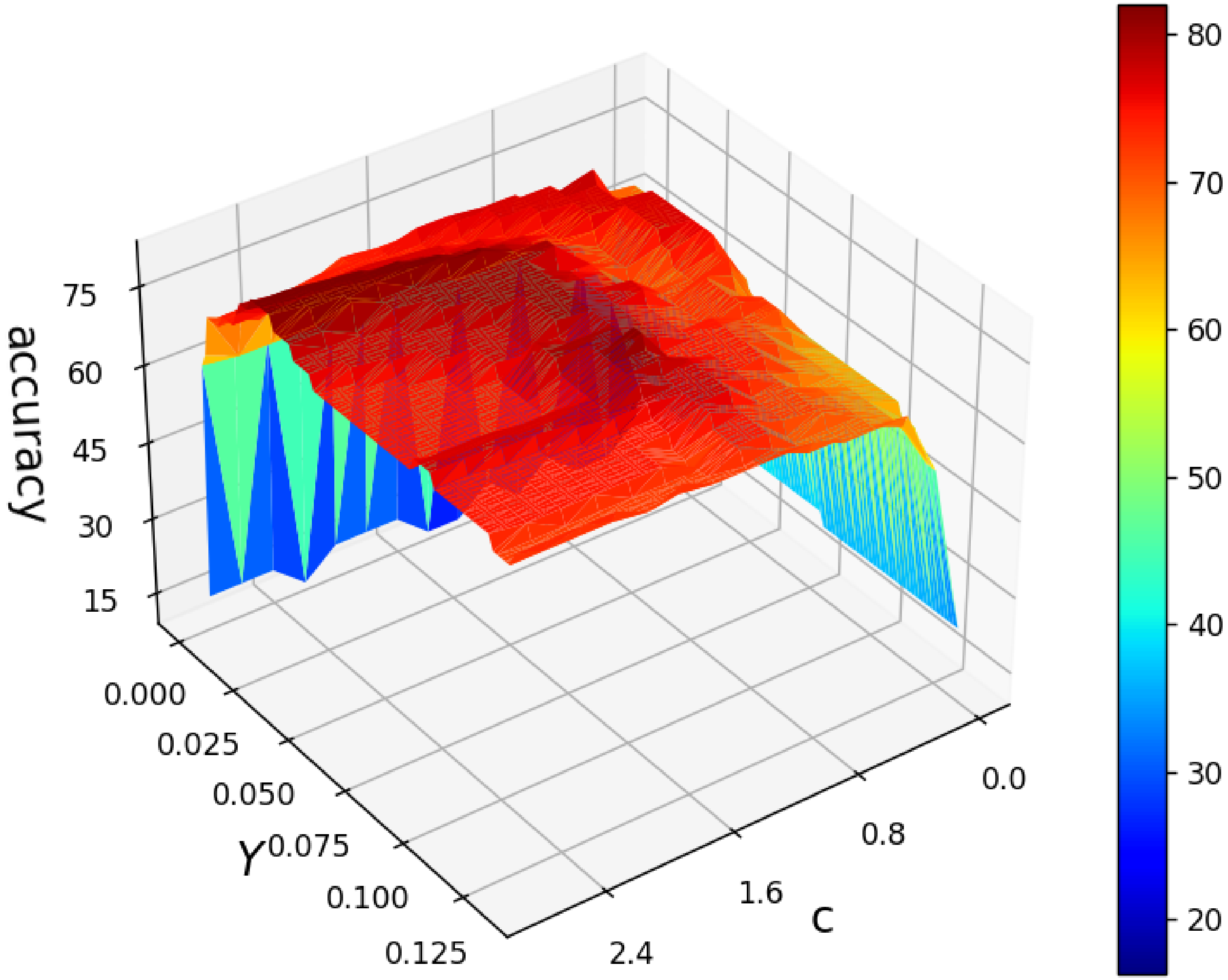


Figure 31: Plot of accuracy w.r.t C and γ. From this graph the optimal value of hyper-parameters obtains are: $C = 1.4$ and $\gamma = 0.088$ giving accuracy on training set of 99.25% and on validation set of 82.05%

Following is the result of accuracy for different value of C and d in sets $C = 0.1, 1.0$ and $\gamma = 0.1$.

- $\gamma = 0.1$, $C = 0.1$: Accuracy on training set of 76.40% and on validation set of 53.84%
- $\gamma = 0.1$, $C = 1.0$: Accuracy on training set of 98.87% and on validation set of 74.35%

The best set found is for $C = 1.4$ and $\gamma = 0.088$ giving accuracy on training set of 99.25% and on validation set of 82.05%.

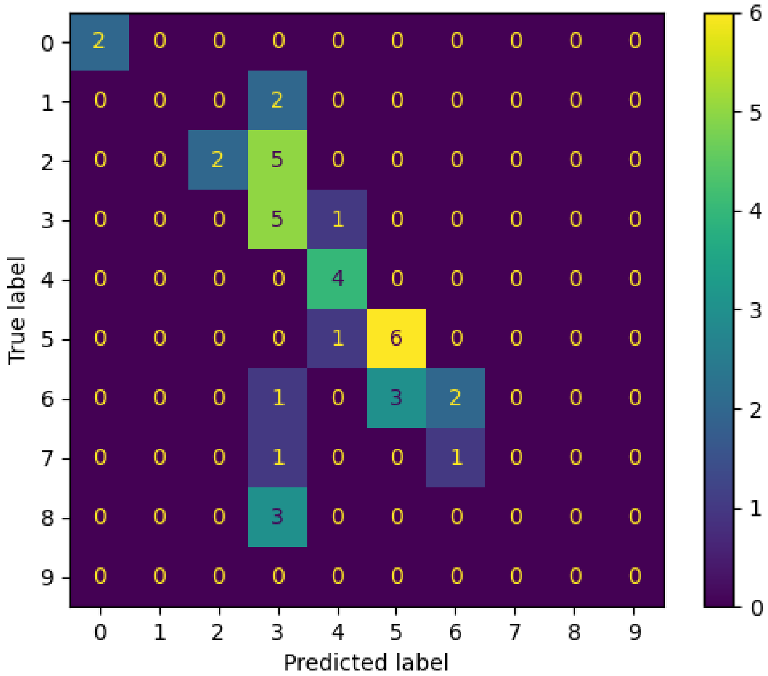


Figure 32: Confusion Matrix for $\gamma = 0.1$, $C = 0.1$

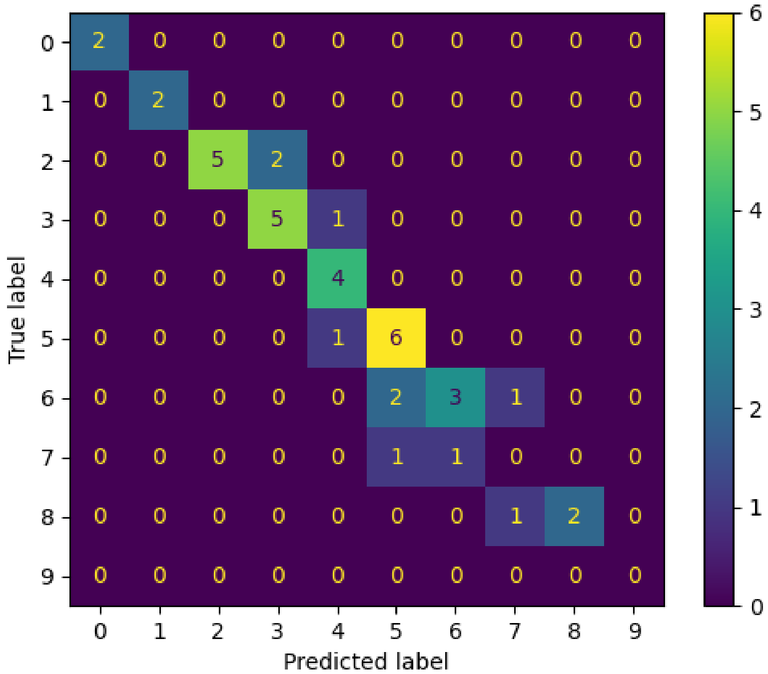


Figure 33: Confusion Matrix for $\gamma = 0.1$, $C = 1$

2.1.2 Polynomial Kernel

The accuracy of classifier for the RBF kernel w.r.t to C and γ keeping degree $= 5$ and results is shown below:

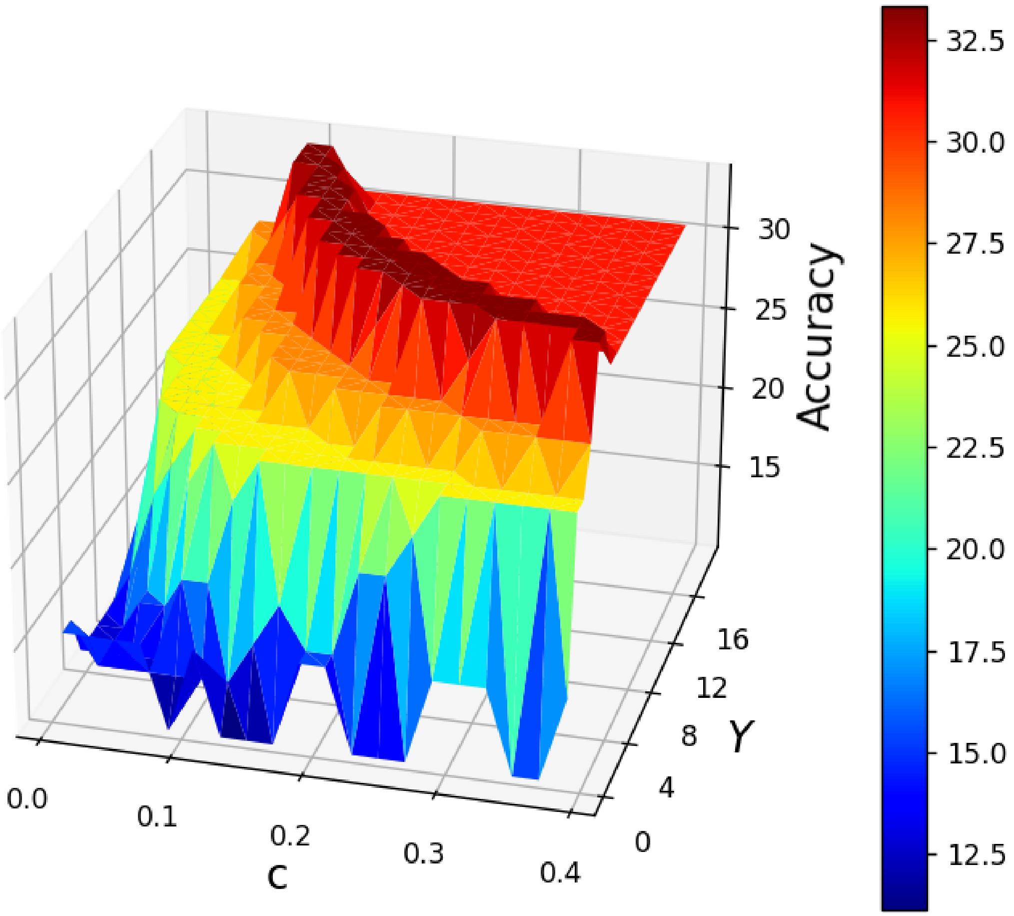


Figure 34: Plot of accuracy w.r.t C and γ. From this graph the optimal value of hyper-parameters obtains are: $C = 0.1$ and $\gamma = 17$ giving accuracy on training set of 65.99% and on validation set of 33.33%

2.2 One-vs-One

In this type of classifier, we determine a hyperplane which separate between one class and another class neglecting the all others class. In total we obtain of $N(N-1)/2$ hyperplanes.

2.2.1 RBF Kernel

The accuracy of classifier for the RBF kernel w.r.t to C and γ and results is shown below:

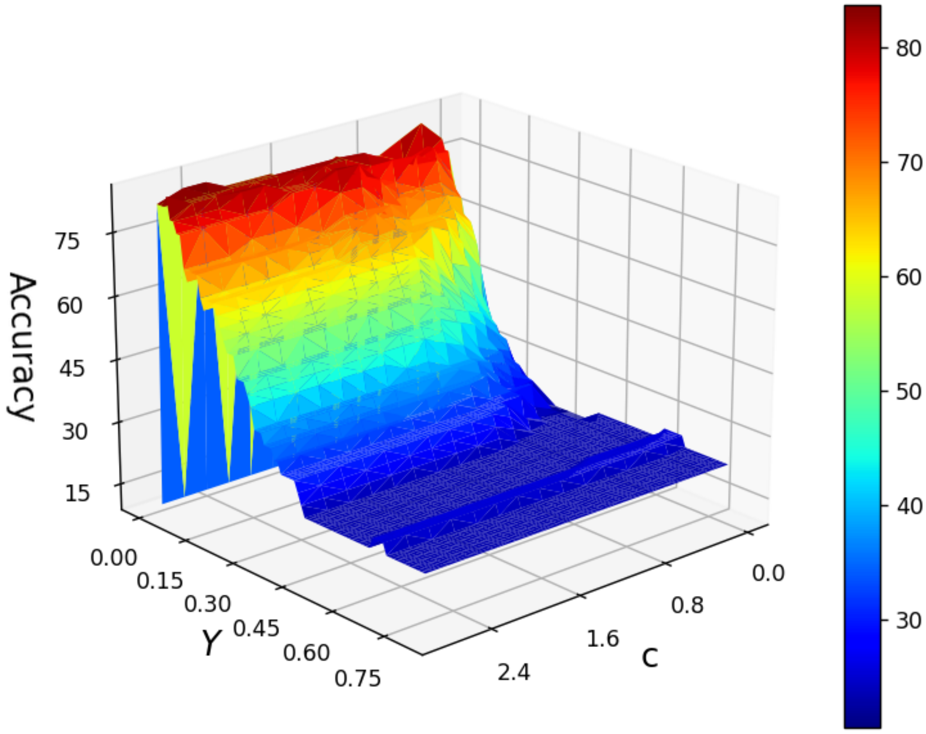


Figure 35: Plot of accuracy w.r.t C and γ

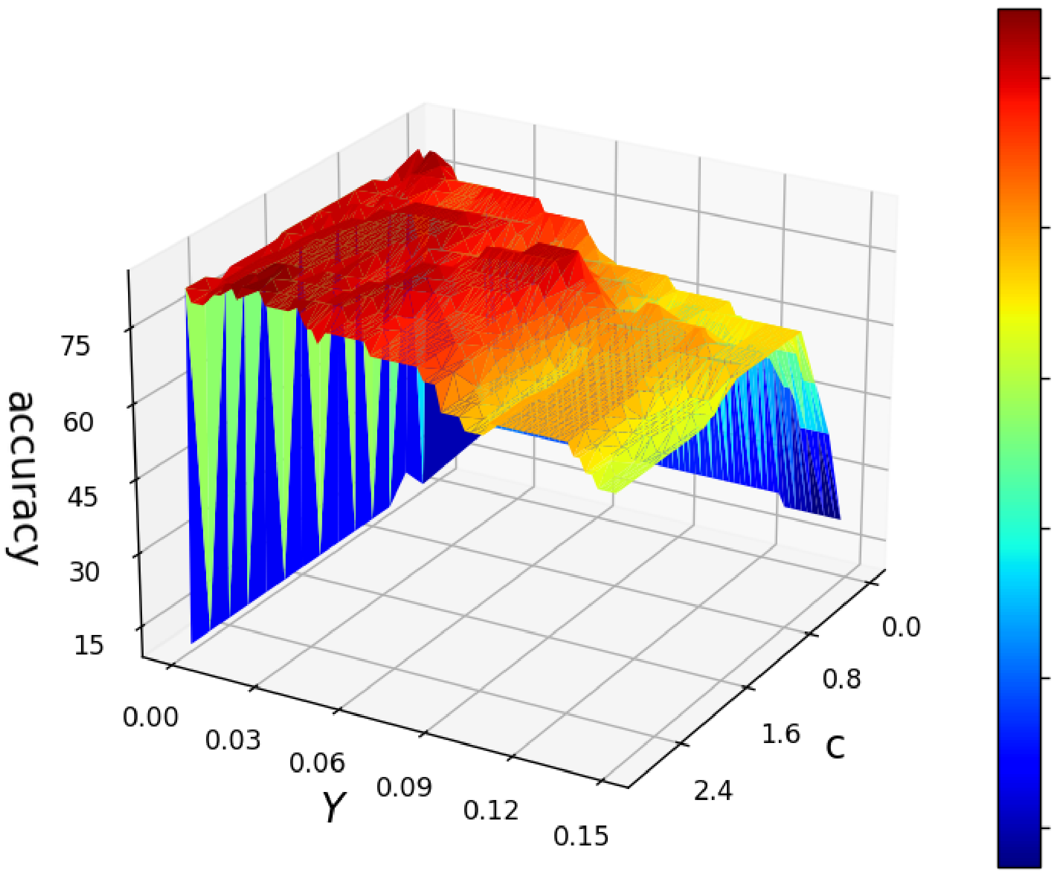


Figure 36: Plot of accuracy w.r.t C and γ. From this graph the optimal value of hyper-parameters obtains are: $C = 3$ and $\gamma = 1.7$ giving accuracy on training set of 93.4% and on validation set of 85%

Following is the result of accuracy for different value of C and d in sets $C = 0.1, 1.0$ and $\gamma = 0.1$.

- $\gamma = 0.1$, $C = 0.1$: Accuracy on training set of 76.40% and on validation set of 53.84%
- $\gamma = 0.1$, $C = 1.0$: Accuracy on training set of 98.87% and on validation set of 74.35%

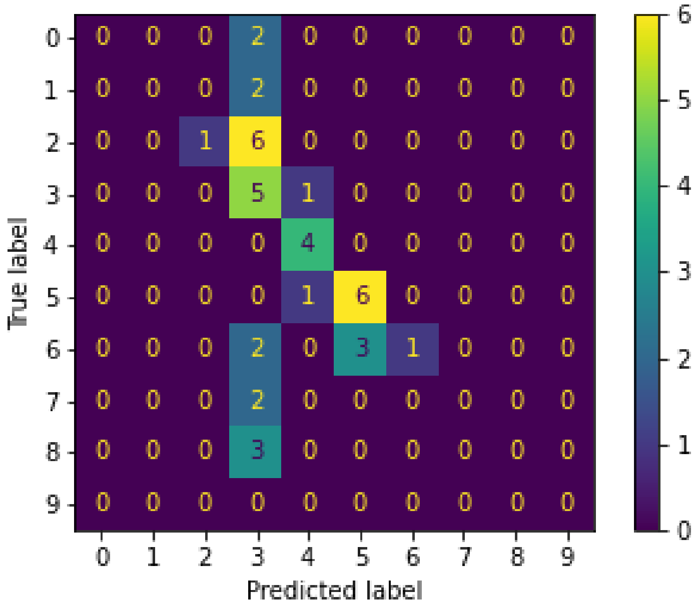


Figure 37: Confusion Matrix for $\gamma = 0.1$, $C = 0.1$

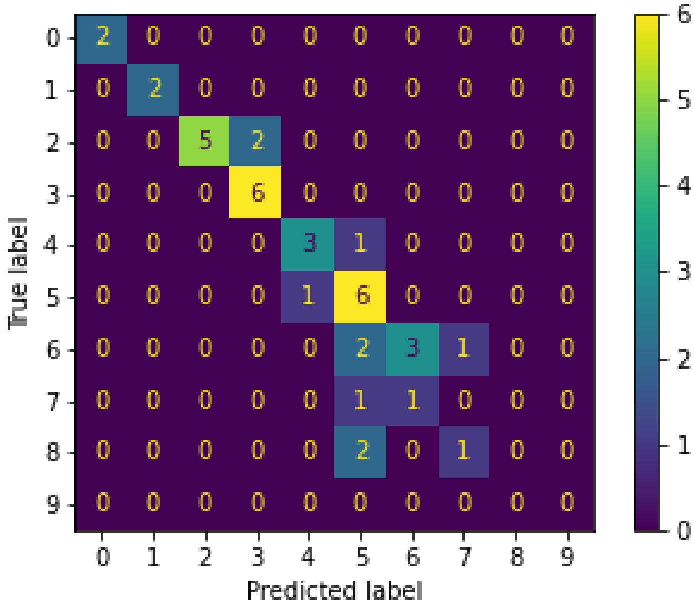


Figure 38: Confusion Matrix for $\gamma = 0.1$, $C = 1$