

Predicting handwritten digits using Support vector machine

Ishank Pahwa

Binary Classification Analysis (Using all features)

Using SVM from Sklearn which implements libsvm
Binary classification for labels 0 and 1 initially

Linear Kernel:

Hyperparameters is

1. C (Regularization parameter)

Using k-fold cross validation with 6 folds and ranging 50 values of c from 0.01 to 0.00001

Max Cross Validation Score : 100%

Optimal Value of C : 0.00408755

Now using train test split with 75% training set and 25% test set

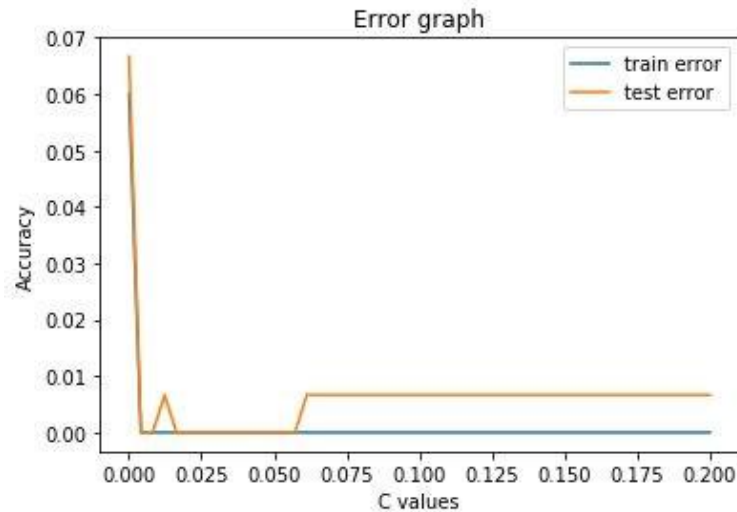
Train Accuracy using optimal values : 100%

Test Accuracy Using optimal values : 100%

Visualization of variation of hyperparameters

Range of C : 0.0001 to 0.2

Train-Test Errors Plot with variation in C (75% split) :



RBF Kernel

Hyperparameters is

1. C (Regularization parameter)
2. Gamma

Grid search was performed to find the optimal values with k-fold cross validation k-fold=6

50 values of C between 0.01 to 0.00001

20 values of Gamma between 0.1 to 0.001

Max Cross Validation Score : 0.996667

Optimal Value of C : 0.00857286

Optimal Value of Gamma : 0.0166316

Train-Test Split : 75% Train, 25% Test

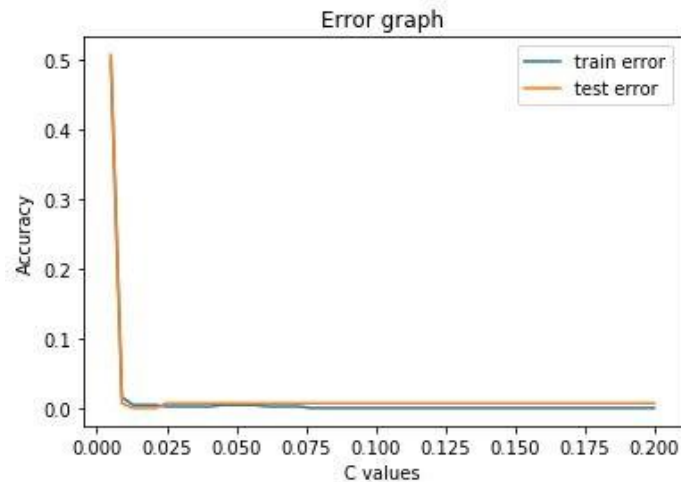
Train Accuracy : 0.9777

Test Accuracy : 0.9733

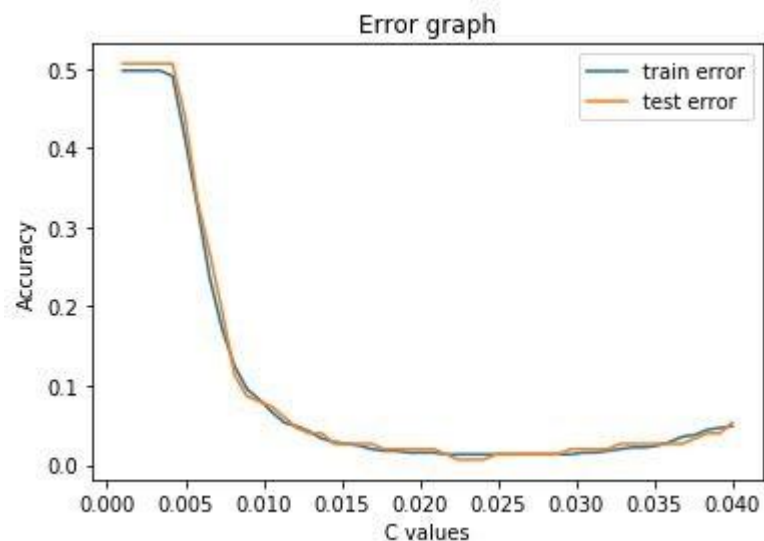
Visualization of variation of hyperparameters 1.

Using optimal Gamma and Varying C :

Gamma : 0.0166316



2. Using optimal C and Varying Gamma: C
: 0.00857286



Polynomial Kernel

Hyperparameters is

1. C (Regularization parameter)
2. Gamma
3. Degree of Polynomial (p)

Using 3D Grid Search to get optimal value of hyperparameters :

K-fold : 6

Range of C : 50 values linearly spaced between (0.01, 0.0001)

Range of Gamma : 50 values linearly spaced between (0.1, 0.001)

Range of p : 0-5

Results :

Max Cross Validation Score : 100%

Optimal Value of C : 0.00583158

Optimal Value of Gamma : 0.0791579

Optimal Value of Degree : 3

Train-Test Split : 75% Train, 25% Test

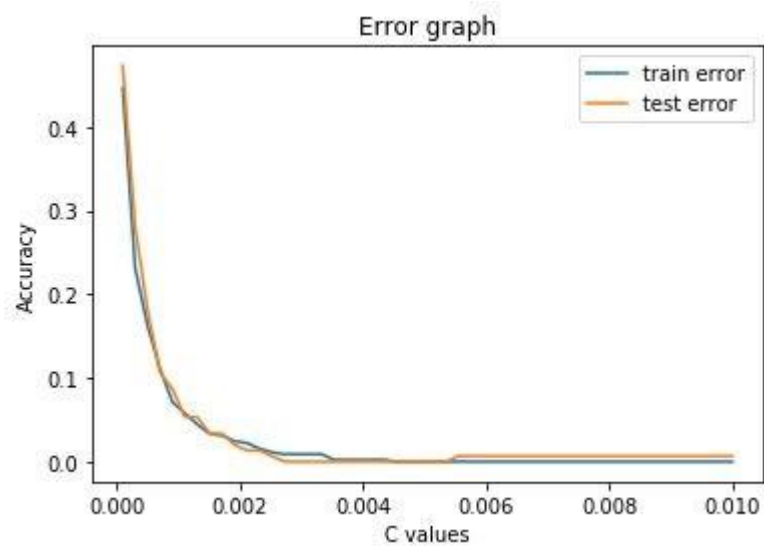
Test Accuracy Using the above optimal values : 92%

Variation of Hyperparameters :

1. Using optimal Gamma and Varying C :

Optimal value of gamma : 0.0791579

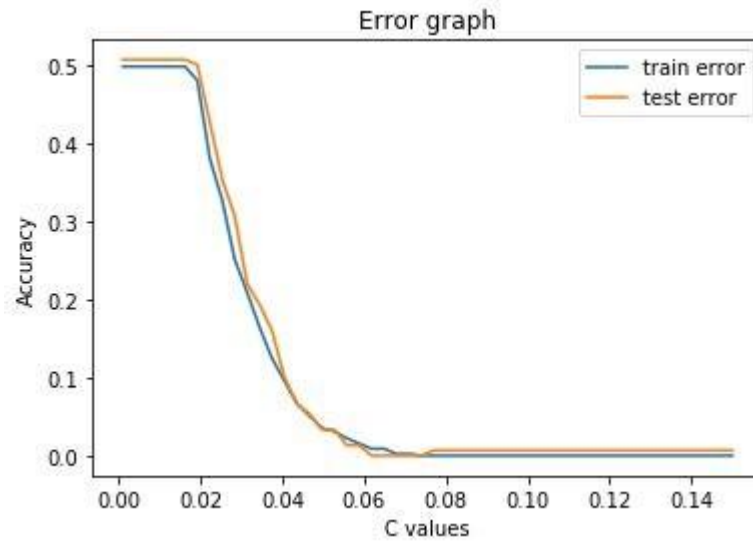
Optimal Value of Degree : 3



2. Using Optimal C and varying gamma :

Optimal value of C : 0.00583158

Optimal Value of Degree : 3



Multiclass Classification

One-versus-one Classification-

Linear Kernel

Hyperparameters is

1. C (Regularization parameter)

Using k-fold cross validation with 10 folds and ranging values of c from 0.1 to 0.01

Max Cross Validation Score : 89.26%

Optimal Value of C : 0.0667

Now using train test split with 75% training set and 25% test set

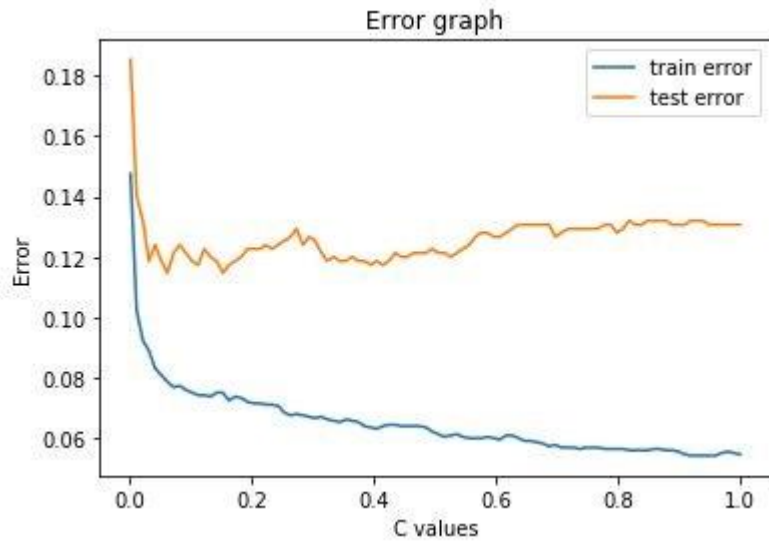
Train Accuracy using optimal values : 92.222%

Test Accuracy Using optimal values : 88.4%

Visualization of variation of hyperparameters

Range of C : 0.001 to 1

Train-Test Errors Plot with variation in C (75% split) :



RBF Kernel

Hyperparameters is

1. C (Regularization parameter)
2. Gamma

Grid search was performed to find the optimal values with k-fold cross validation k-fold=10

10 values of C between 3 to 0.4

10 values of Gamma between 0.1 to 0.001

Max Cross Validation Score : 0.9470

Optimal Value of C : 2.13333

Optimal Value of Gamma : 0.034

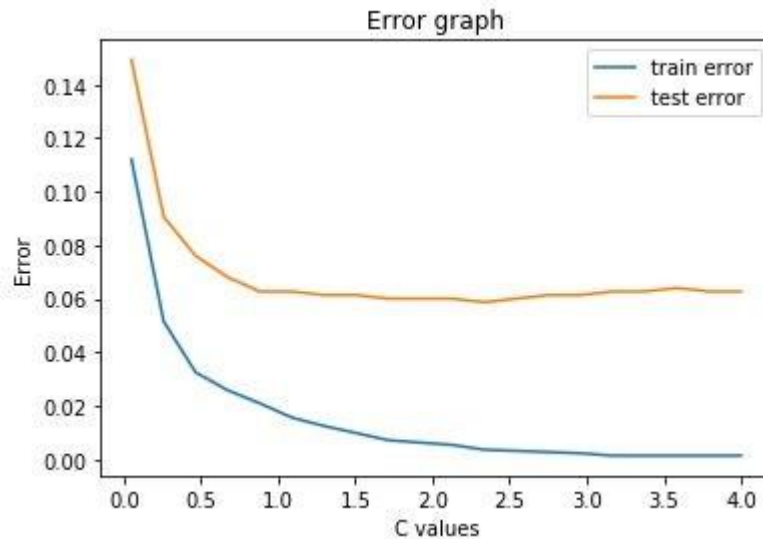
Train-Test Split : 75% Train, 25% Test

Train Accuracy : 99.46%

Test Accuracy : 94%

Visualization of variation of hyperparameters Using
optimal Gamma and Varying C :

Gamma : 0.034



Polynomial Kernel

Hyperparameters is

1. C (Regularization parameter)
2. Gamma
3. Degree of Polynomial (p)

Using 3D Grid Search to get optimal value of hyperparameters :

K-fold : 10

Range of C : 10 values linearly spaced between (0.04, 0.8)

Range of Gamma : 10 values linearly spaced between (0.1, 0.001)

Range of p : 0-5

Results :

Max Cross Validation Score : 93.8%

Optimal Value of C : 0.546667

Optimal Value of Gamma : 0.045

Optimal Value of Degree : 3

Train-Test Split : 75% Train, 25% Test

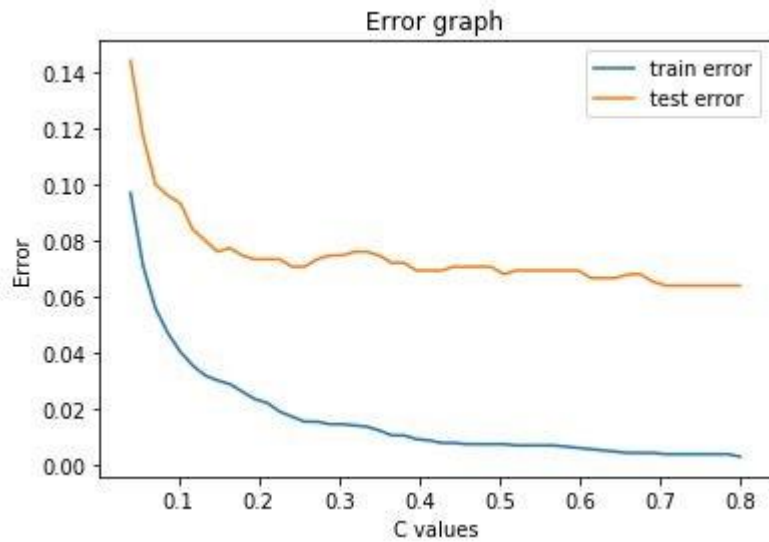
Training Accuracy : 99.28%

Test Accuracy : 93.06%

Variation of Hyperparameters :

1. Using optimal Gamma and Varying C :
Optimal value of gamma : 0.045

Optimal Value of Degree : 3



One-versus-many Classification- Linear Kernel-

Max Cross Validation Score : 89.26%

Optimal Value of C : 0.0659

Now using train test split with 75% training set and 25% test set

Train Accuracy using optimal values : 92.222%

Test Accuracy Using optimal values : 88.3%

RBF Kernel-

Max Cross Validation Score : 94.70%

Optimal Value of C : 2.14

Optimal Value of Gamma : 0.034

Train-Test Split : 75% Train, 25% Test

Train Accuracy : 99.46%

Test Accuracy : 94%

Polynomial Kernel

Max Cross Validation Score : 93.8%

Optimal Value of C : 0.547

Optimal Value of Gamma : 0.045

Optimal Value of Degree : 3

Train-Test Split : 75% Train, 25% Test

Training Accuracy : 99.28%

Test Accuracy : 93.06%

Change in number of features (to first 10 features) :

Hyperparameters and Results After changing Features :

Type of Kernel	C (Regularization Parameter)		Gamma Value (γ)		Degree (p)		Accuracy using Train-Test Data	
	25 Features	10 Features	25 Features	10 Features	25 Features	10 Features	25 Features	10 Features
Linear	0.00408 755	0.00070 6122	None	None	None	None	100%	100%
Gaussian	0.00857 286	0.00758	0.01663 16	0.00092 9	None	None	97.33%	96.33%
Poly	0.00583 158	0.00433	0.07915 79	0.06	3	3	92%	91.8%

C decreases when we take only 10 features as compared to when we took 25 features

For Different Pairs of Target classes :

Target Classes	C (Regularization Parameter)		Gamma Value (γ)		Accuracy using Train-Test Data	
	Linear Kernel	Gaussian Kernel	Linear Kernel	Gaussian Kernel	Linear Kernel	Gaussian Kernel
A = 0 B = 1	0.00408	0.00857	None	0.01663	100%	97.33%

A = 6 B = 7	0.0009081	0.0167333	None	0.012	99.33%	100%
----------------	-----------	-----------	------	-------	--------	------

We see that as we change the pair, the choice of the best kernel changes from linear to gaussian which gives 100% accuracy in the 2end pair.

CVX Optimization Analysis :

Hyperparameters from the respective kernel results have been used

Soft_threshold : 1e-4 (The values above which we consider the point to be a support vector)

Linear Kernel

Optimal Value of C : 0.00408755

Accuracy on Test Set : 100%

RBF Kernel

Optimal Value of C : 0.0000857286

Optimal Value of Gamma : 0.0166316

Accuracy on Test Set : 100%

Polynomial Kernel

Optimal Value of C : 0.00018315

Optimal Value of Degree : 3

Optimal Value of Gamma : 0.0791579

Accuracy on Test Set : 100%