

# ELL409 : Assignment 2

Nikhil Shenoy 2016BB10022

## Binary Classification Analysis (Using All Features = 25) :

*Some key points :*

1. SVC from sklearn.svm (which implements libsvm) has been used
2. Binary classification for labels 4 and 7 will be performed initially, 600 datapoints with equal number of labels

### ***Type of Kernel : Linear Kernel***

Hyperparameters :

1. C (Regularization Parameter)

Using Grid Search to get optimal value of hyperparameters :

K-fold : 6

Range of C : 50 values linearly spaced between (1e-3, 0.08)

***Results :***

Max Cross Validation Score : 1

Optimal Value of C : 0.075

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

Test Accuracy Using the above optimal values : 98.66%

Variation of Hyperparameters :

SVC parameters :

Range of C : 0.001 to 0.08

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

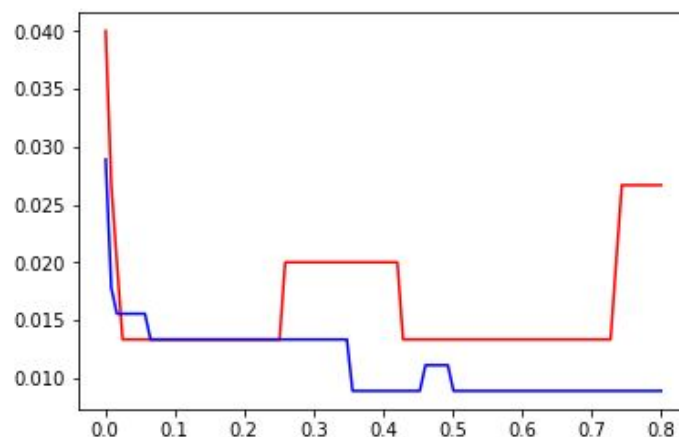


Fig. Train error (Blue) and Test Error (Red) vs C - values

## ***Type of Kernel : Gaussian Kernel***

Hyperparameters :

1. C (Regularization Parameter)
2. Gamma ( $1/\sigma^2$ )

Using 2D Grid Search to get optimal value of hyperparameters :

K-fold : 6

Range of C : 50 values linearly spaced between ( $1e-3$ , 0.08)

Range of Gamma : 50 values linearly spaced between ( $1e-3$ ,  $1e-1$ )

***Results :***

Max Cross Validation Score : 0.9983334

Optimal Value of C : 0.08

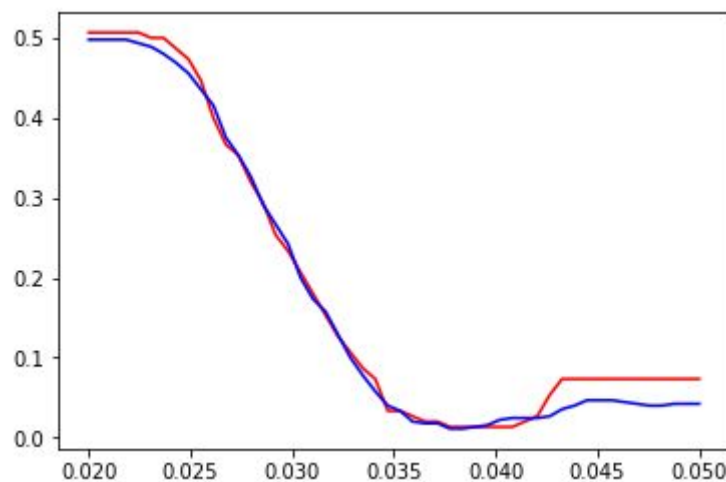
Optimal Value of Gamma : 0.071714

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

Test Accuracy Using the above optimal values : 94.667%

Variation of Hyperparameters :

1. Using optimal Gamma and Varying C :  
Gamma : 0.071714



*Fig. Train error (Blue) and Test Error (Red) vs C - values*

2. Using Optimal C and varying gamma :

C = 0.08

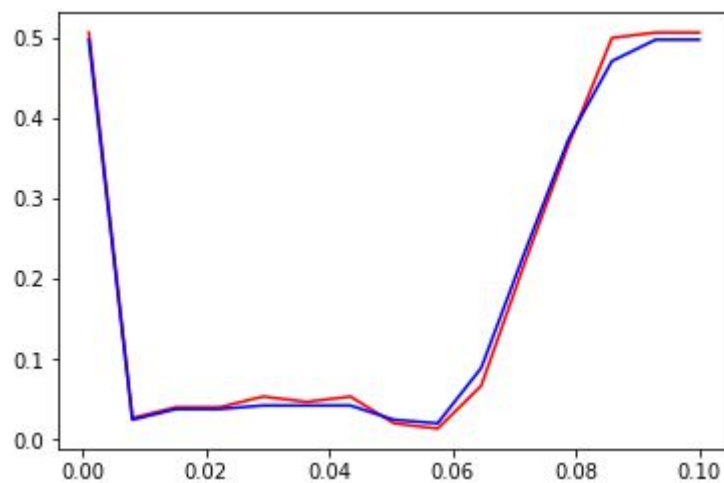


Fig. Train error (Blue) and Test Error (Red) vs C - values

## Type of Kernel : Polynomial Kernel

Hyperparameters :

1. C (Regularization Parameter)
2. Gamma ( $1/\sigma^2$ )
3. Degree of Polynomial (p)

Using 2D Grid Search to get optimal value of hyperparameters :

K-fold : 6

Range of C : 50 values linearly spaced between (1e-3, 0.08)

Range of Gamma : 50 values linearly spaced between (1e-3, 1e-1)

**Results :**

Max Cross Validation Score : 1

Optimal Value of C : 0.005157

Optimal Value of Gamma : 0.07878

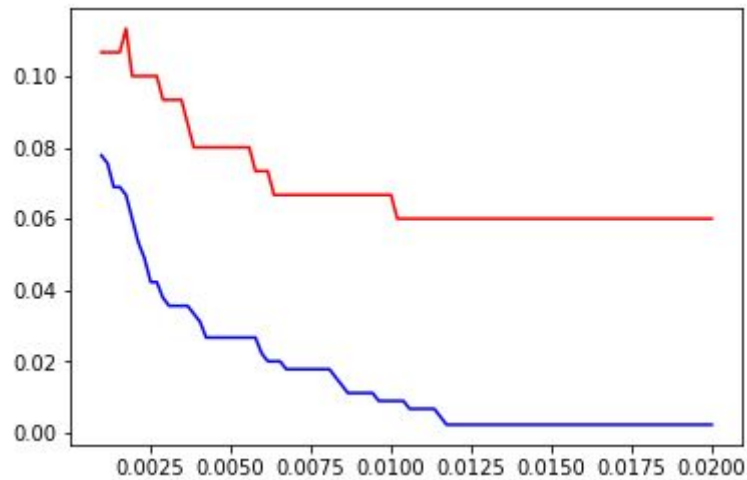
Optimal Value of Degree : 5

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

Test Accuracy Using the above optimal values : 92%

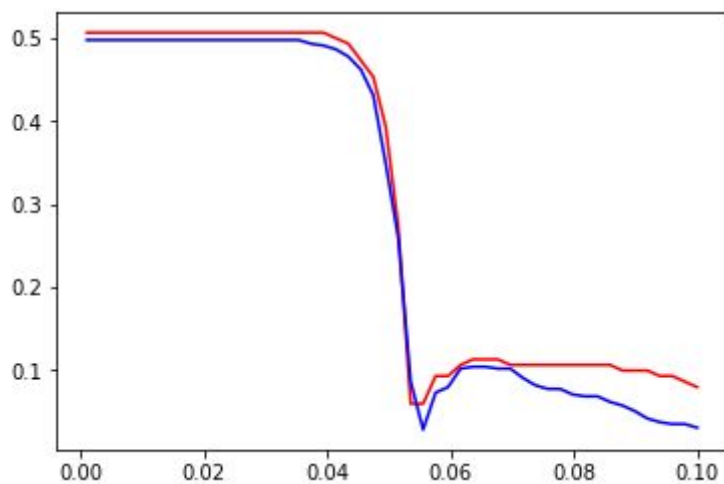
Variation of Hyperparameters :

1. Using optimal Gamma and Varying C :  
Optimal value of gamma : 0.07878  
Optimal Value of Degree : 5



*Fig. Train error (Blue) and Test Error (Red) vs C - values*

- Using Optimal C and varying gamma :  
 Optimal value of C : 0.005157  
 Optimal Value of Degree : 5



*Fig. Train error (Blue) and Test Error (Red) vs C - values*

## CVX Optimization Analysis :

*Some key points :*

- Binary classification between two randomly chosen classes will be performed
- 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)

**Type of Kernel : Linear Kernel**

Optimal Value of C : 0.075

Number of Support Vectors : 54 out of 450 points

Accuracy on Test Set : 98.66%

### ***Type of Kernel : Gaussian Kernel***

Optimal Value of C : 0.8

Optimal Value of Gamma :

Number of Support Vectors : 130 out of 450 points

Accuracy on Test Set : 91.33%

### ***Type of Kernel : Polynomial Kernel***

Optimal Value of C : 0.8

Optimal Value of Degree : 5

Optimal Value of Gamma :

Accuracy on Test Set : 97.33%

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

Hyperparameters and Results After changing Features :

Type of Kernel	C (Regularization Parameter)		Gamma Value ( $\gamma$ )		Degree (p)		Accuracy using Train-Test Data	
	25 Features	10 Features	25 Features	10 Features	25 Features	10 Features	25 Features	10 Features
Linear	0.075	0.054	None	None	None	None	98.66%	96.67%
Gaussian	0.08	0.0758	0.07171	0.0929	None	None	94.67%	97.33%
Poly	0.00515	0.0633	0.07878	0.1	5	6	94.667%	94%

*Inference* : The regularization parameter (C) reduces for linear and gaussian kernel on reducing the number of features. But for Polynomial kernel the value of regularization parameter increases.

Therefore, we can say that the removal of features is not adding much of a difference in the two scenarios.

### For Different Pairs of Target classes :

Target Classes	C (Regularization Parameter)		Gamma Value ( $\gamma$ )		Accuracy using Train-Test Data	
	Linear Kernel	Gaussian Kernel	Linear Kernel	Gaussian Kernel	Linear Kernel	Gaussian Kernel
A = 4 B = 7	0.075	0.08	None	0.07171	98.66%	94.67%
A = 1 B = 2	0.0225	0.0675	None	0.0505	96.67%	94.67%
A = 5 B = 9	0.0408	0.0675	None	0.0505	98.67%	96%

*Inference* : The accuracy remains almost the same on changing the target classes for binary classification using libsvm. This shows the the target classes are similarly distributed.

### Multi-Class Analysis (Using all 25 features) :

Some key points :

1. Dataset contains 10 target classes and 3000 points
2. SVC from sklearn.svm is being used for classification
3. The values in the table below have been constructed using k-fold cross validation ( $k = 10$ ) and performing a grid search based on the hyperparameters used by the kernel

### Decision Surface Shape : One vs One

Target Classes	C	Gamma Value ( $\gamma$ )	Degree of Polynomial ( $p$ )	Train Accuracy	Test Accuracy
Linear Kernel	0.2831	None	None	93.28%	90.4%

Gaussian Kernel	0.158	0.121	None	99.06%	93.6%
Polynomial Kernel	0.11	0.0925	3	99.556%	95.34%

### ***Decision Surface Shape : One vs Rest***

Target Classes	C	Gamma Value ( $\gamma$ )	Degree of Polynomial (p)	Train Accuracy	Test Accuracy
Linear Kernel	0.385	None	None	93.6%	90.13
Gaussian Kernel	0.108	0.1684	None	99.066%	93.6%
Polynomial Kernel	0.1	0.0925	3	99.55%	95.33%

*Inference* : The gaussian kernel performs exactly the same in case of the two decision surface shapes whereas there are minor changes in the polynomial and linear kernel.

### **Kaggle Dataset Analysis :**

Some key points :

1. Dataset contains 10 target classes and 10000 points
2. SVC from sklearn.svm is being used for classification

Choice of Hyperparameters	Value
Kernel Type	Gaussian
C (Regularization Parameter)	0.8 (Based on Minimum testing error)
Gamma Value	4.23e-7
K-fold	10

The testing accuracy achieved was 96.4%, based on a 75% Train-Test Split.

