ELL784: Assignment 2

Prashant Atkale

2021AMA2631

Question 1:

a) Binary classification

In this part, we have to first train SVM model by taking two classes out of 10 classes, and 10 features out of 25. Later trained model by taking all 25 features and two classes and compare the results for both the cases. We have to train SVM model to find out the optimum hyper parameters by performing cross validation and using various kernels. Figure 1 shows the count plot for all 10 class labels Each class is having 300 data points.

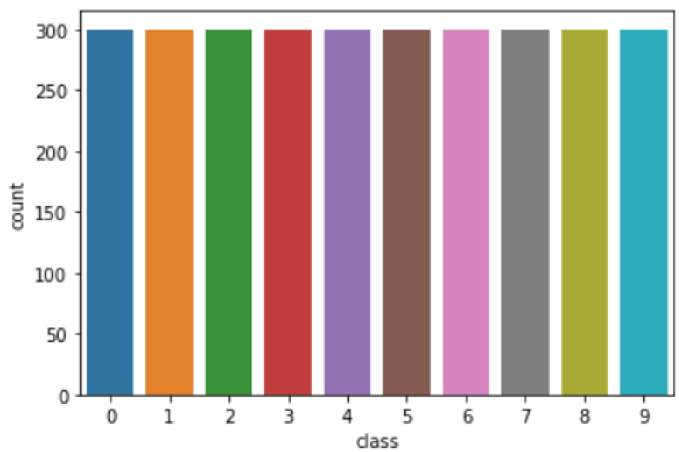


Figure 1: Count plot

Class labels ‘0’ & ‘1’:

Taken various kernels such as polynomial, RBF, Sigmoid, linear and the performed Grid search by taking the range of values of different hyperparameters such as gamma, C, degree of polynomial etc. After performing Grid search, found out the best hyperparameters for both the cases i.e., by taking 10 features first and then 25 features.

For 25 features, kernel requirement is different from what is for 10 features. With 25 features, our model is performing quite well on testing data. This highlights the need to take all the features into account to generalize well for the unseen data. While the validation score for both 10 features and 25 features are nearly same.

Best Parameters and scores for 10 features and 25 features are given in table 1:

Parameters	***10 features***	***25 features***
Kernel	linear	poly
C	0.01	0.01
gamma	scale	0.9
Degree	1	1
Validation Score	1.00	0.998
Test score	98.33	1.00

Table 1

Class labels ‘5’ & ‘9’:

When we take data for class labels 5 and 9, Validation score for 25 features is higher compared to 10 feature model. As test data is less so we can say both models are not performing well on test data.

For 25 features C =10, which tells that margin is slightly harder compared to 10 features. Some features are missed while taking first 10 features, so we must keep its margin on softer side to generalize well.

Parameters	***10 features***	***25 features***
Kernel	rbf	rbf
C	5	10
gamma	scale	0.01
Degree	1	1
Validation Score	0.977	0.994
Test score	0.975	0.976

Table 2

If we take optimized hyperparameters given in table 1, to check the performance for this class labels, then we can see validation score and test score as follows

25 features ------→ Validation score= 0.977, Test score = 0.958
10 features ------→ Validation score = 0.952, Test score = 0.933

As the optimized hyperparameters are different for this case of labels. If we are tak ing parameters from table 1 then its underfitting to the present model.

Class labels ‘3’ & ‘7’:

For this case, Model for 10 features and 25 features is performing quite well, as both have nearly same validation score. With 25 features model, C value is low compared to other model, allowing slightly more misclassifications compared to 10 feature model. Thus 25 features model is performing well on the testing data. Gamma value for 25 features is more which is adding complexity to our model by taking polynomial degree =4, compared to 10 features model.

Parameters	***10 features***	***25 features***
Kernel	rbf	poly
C	0.8	0.1
gamma	Scale	0.1
Degree	1	4
Validation Score	0.987	0.985
Test score	0.975	0.992

Table 3

If we take optimized hyperparameters given in table 2, to check the performance for this class labels, then we can see validation score and test score as follows

25 features ------→ Validation score= 0.981, Test score = 0.975
10 features ------→ Validation score = 0.979, Test score = 0.975

By comparing the scores with table 3, we can say that 10 features model is performing same in terms of validation and test score. But 25 feature model is not performing well on test data compared to table 3 values so its kind of underfitting to test data.

<u>Class labels ‘0’ & ‘1’ with first 2 features:</u>

In this case, only first two labels and features are taken. In figure 2 given below, scatterplot is shown. X-axis represents feature 1 and Y axis represents feature 2.

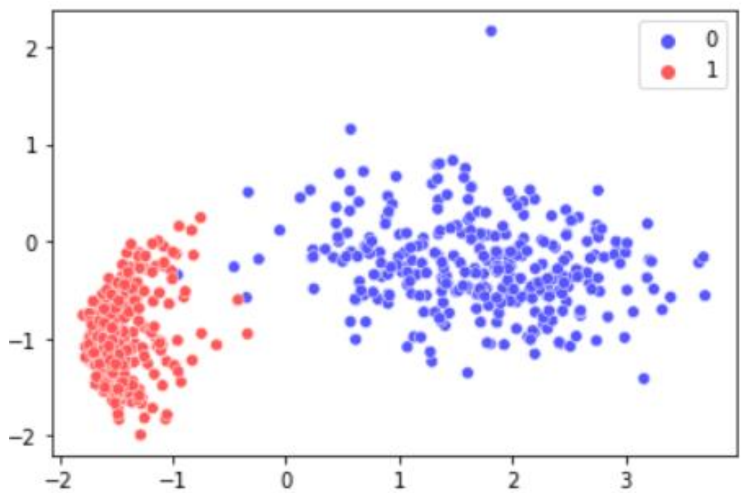


Figure 2: Scatterplot

To train this SVM model, Grid search is performed. Parameters for grid were the range of hyperparameters values and different types of kernels. Cross validation with number of folds equal to 6. After performing grid search optimized parameters were found.

Best parameters for this case are shown in table 4

Parameters	***10 features***
Kernel	Poly
C	1
gamma	Scale
Degree	1
Validation Score	0.995

Table 4

Confusion matrix and classification matrix on test data is shown below. In confusion matrix, only one data point is misclassified out of 120 data points.

```
Confusion Matrix
[[59  1]
 [ 0 60]]
```

In classification report, Precision, recall, f1-score values are given. Accuracy on test data is 0.99 which represents that our model is performing very well on unseen data.

Classification Report

	precision	recall	f1-score	support
0	1.00	0.98	0.99	60
1	0.98	1.00	0.99	60
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

From Table 4, our polynomial kernel with degree 1, is performing well. Our hyperplane is a line. Hyperplane separating two classes linearly and allowing some

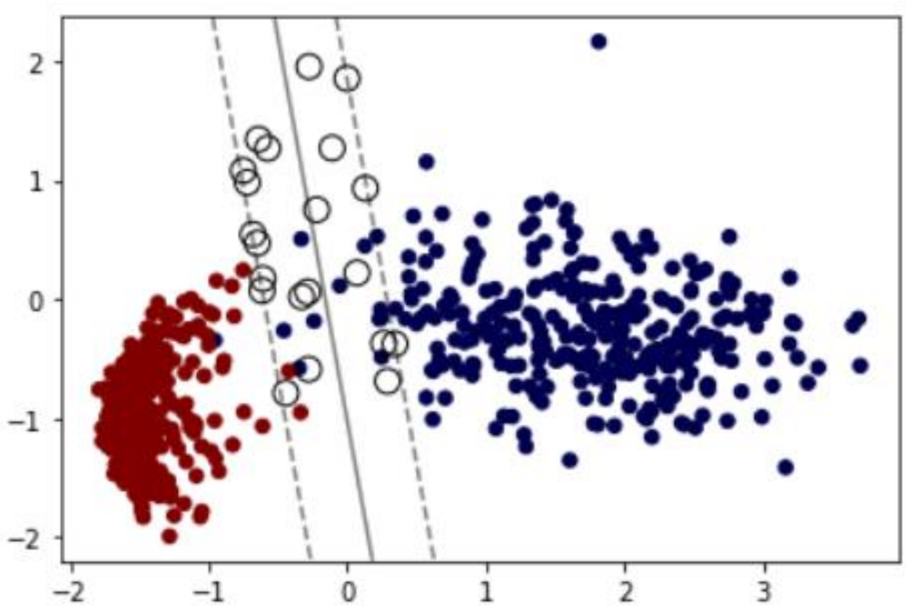


Figure 3: Separating hyperplane with soft margin

misclassifications to generalize well on unseen data points. In above figure 3 you can line separating red (class 0) and blue (class 1) classes along with soft margin.

b) Multiclass classification:

In this part we have to train our model for all 10 classes. Built a classifier for all the classes and evaluated using cross-validation. Then, studied the effects of changing the various hyperparameters and kernel function. Also finetuned hyperparameters to obtain the best possible performance.

Figure 4 shows the correlation of all features with respect to the class labels

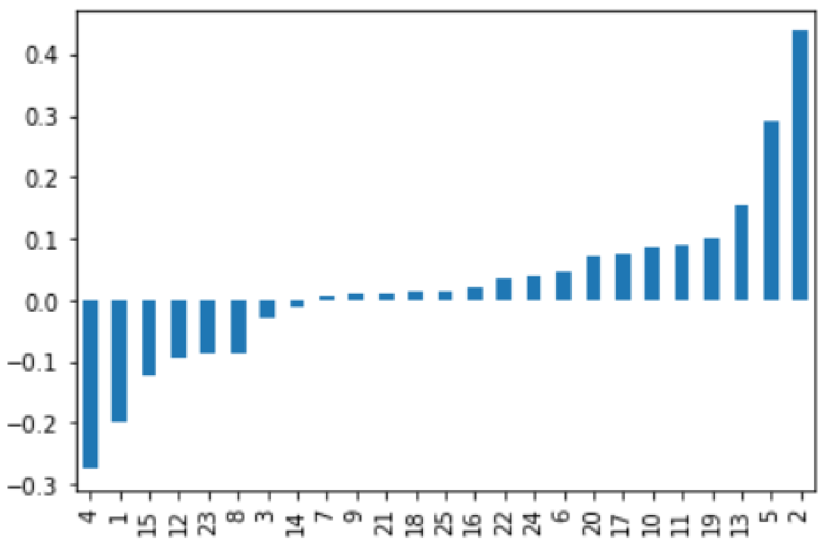


Figure 4: Correlation with respect to class labels

This problem is further divided into two cases

i) First 10 features and all 10 classes
ii) All 25 features and 10 classes

We have to compare the results for these two cases as well,

RBF Kernel

RBF kernel is taken, and results are compared for the validation, test accuracy as well as for the best hyperparameters. Hyperparameters for RBF kernel are C and gamma value. Range of C and gamma values are taken trained model with cross validation with 6 folds for both 10 and 25 features case.

From Figure 5,6 we can see that validation accuracy for the model trained using 10 features is less compared to that of 25 features. Same for the test accuracy.

Parameters	*10 features*	*25 features*
Kernel	rbf	rbf
C	1	10
gamma	Scale	Scale
Degree	1	1
Validation Score	0.889	0.952
Test score	0.890	0.947

Table 5

From table 5, we can compare parameters for both the cases. With 25 features, our model is giving far better results than for the 10 features. Here we are considering all the features so model complexity has risen which leads to decrease in the accuracy for the models compared to binary classified models in previous case.

.

Figure 5: First 10 features

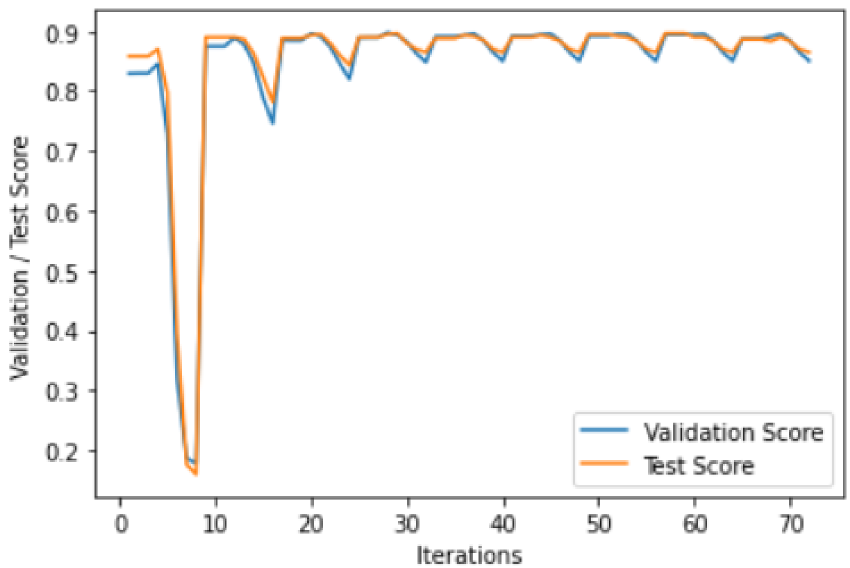
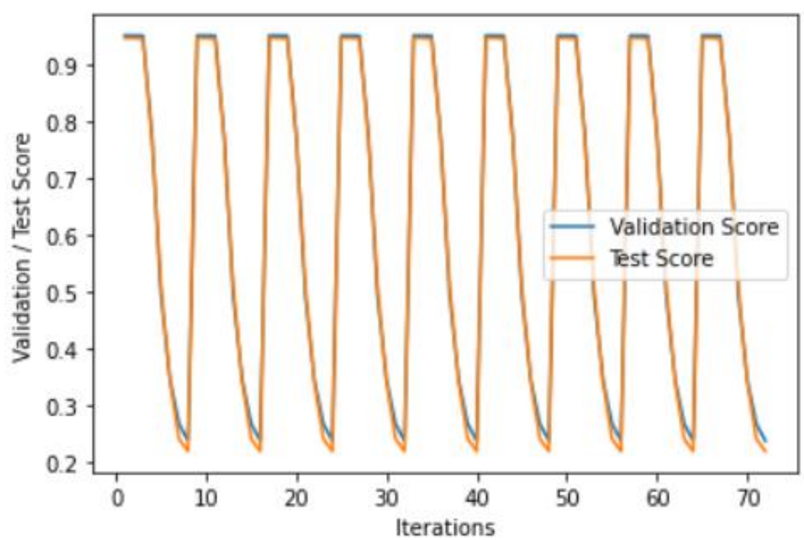


Figure 5: 25 features



POLY Kernel

Polynomial kernel function is taken with hyperparameters of C and polynomial degree value. Optimized hyperparameters by using grid search CV are shown in Table 6 below

Parameters	***10 features***	***25 features***
Kernel	poly	poly
C	5	1
gamma	Scale	Scale
Degree	3	3
Validation Score	0.873	0.939
Test score	0.866	0.932

Table 6

With 10 features, model with poly kernel is not performing well compared to 25 features model. Validation and test accuracy in for 10 features is less compared to 25 feature model. If we use hyperparameters of 10 feature model for 25 feature model, then its performance is like what 25 feature model is performing now.

Figure 7,8 shows performance of sigmoid kernel for both cases.

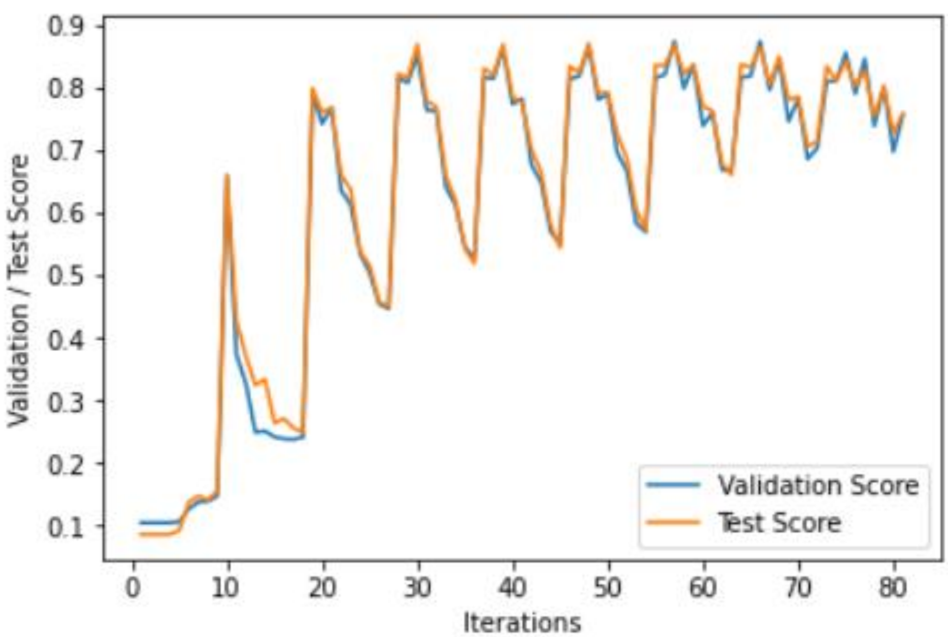


Figure 7: First 10 features

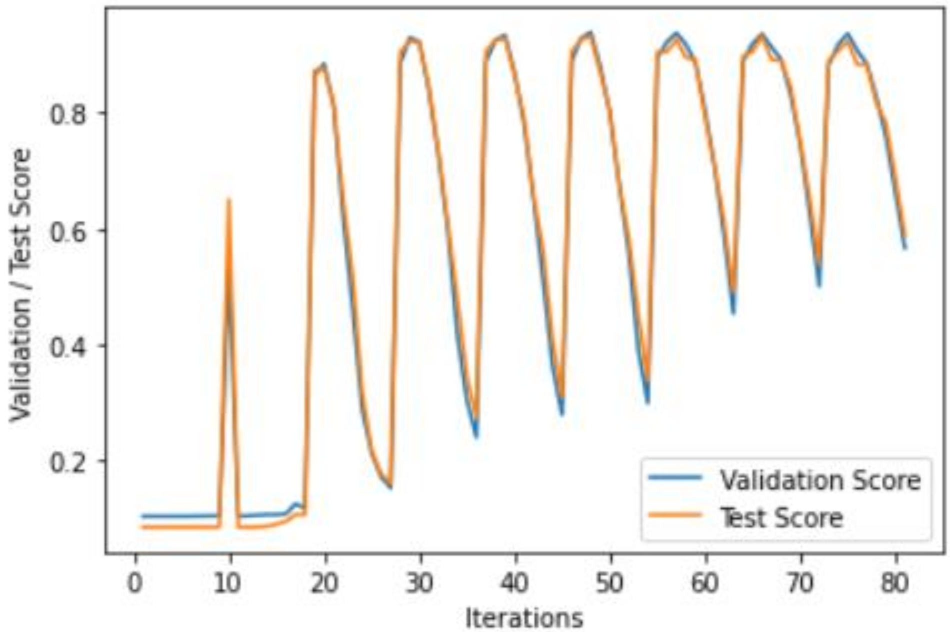


Figure 8: 25 features

SIGMOID Kernel

SIGMOID kernel is taken, and results are compared for the validation accuracy, test accuracy as well as for the best hyperparameters. Hyperparameters for RBF kernel are C and gamma value. Range of C and gamma values are taken. Model is then trained with cross validation of 6 folds for both 10 and 25 features case.

Optimum hyperparameters for both the cases are shown in table 7

Parameters	***10 features***	***25 features***
Kernel	sigmoid	sigmoid
C	50	50
gamma	0.001	0.001
Validation Score	0.814	0.891
Test score	0.821	0.907

Table 7

C and gamma values are same for both the models. But because of less information by taking 10 features, model couldn't generalize well. So, model for 10 features is performing poorly on both validation and test data. While it's performing relatively well for 25 features data.

Figure 9,10 shows the variation of accuracy by taking range of hyperparameters values.

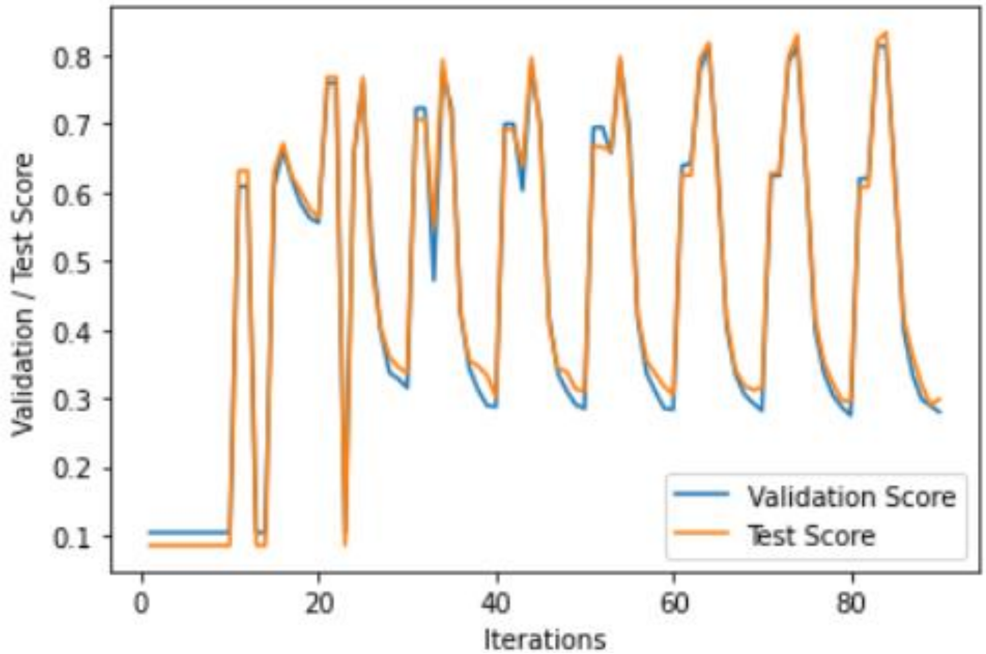


Figure 9: First 10 features

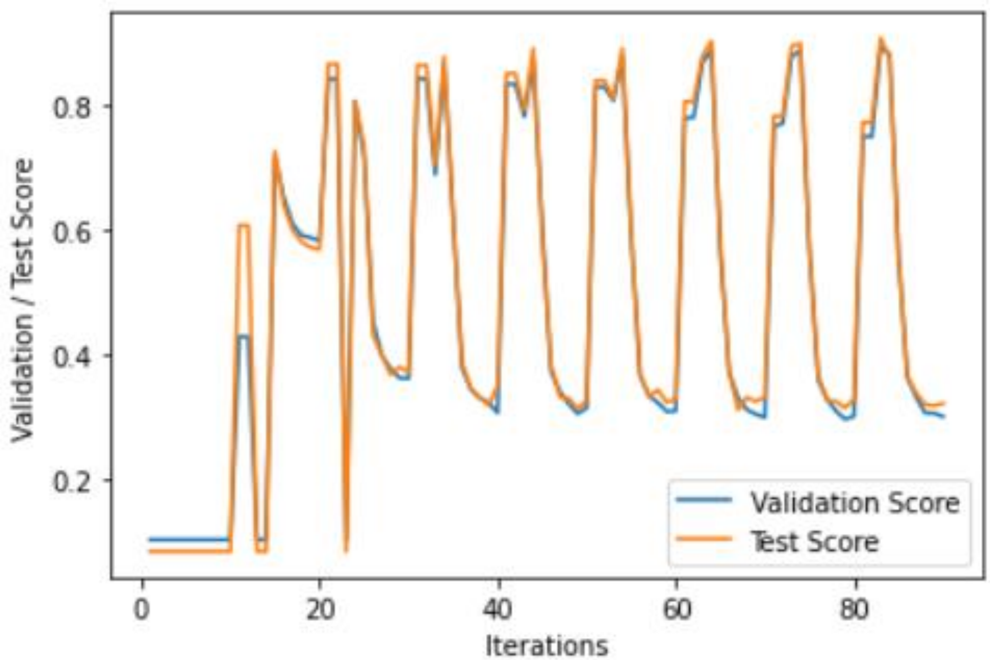


Figure 10: 25 features

Linear kernel

Linear kernel function takes C as hyperparameter value. Range of C values are taken and trained our model with cross validation. Performance of 25 feature model is high compared to 10 feature model.

In figure 11,12 variation of accuracy for range of hyperparameter values are shown. Table 8 shows the optimum hyperparameters after performing GridSearchCV.

Parameters	***10 features***	***25 features***
Kernel	Linear	Linear
C	0.8	0.1
gamma	Scale	Scale
Validation Score	0.817	0.893
Test score	0.833	0.911

Table 8

The accuracy of 10 feature model is less compared to 25 feature model. C values for both the cases are different. If C value of 10 feature model is taken for 25 feature model, then validation score is coming nearly same as that of present model.

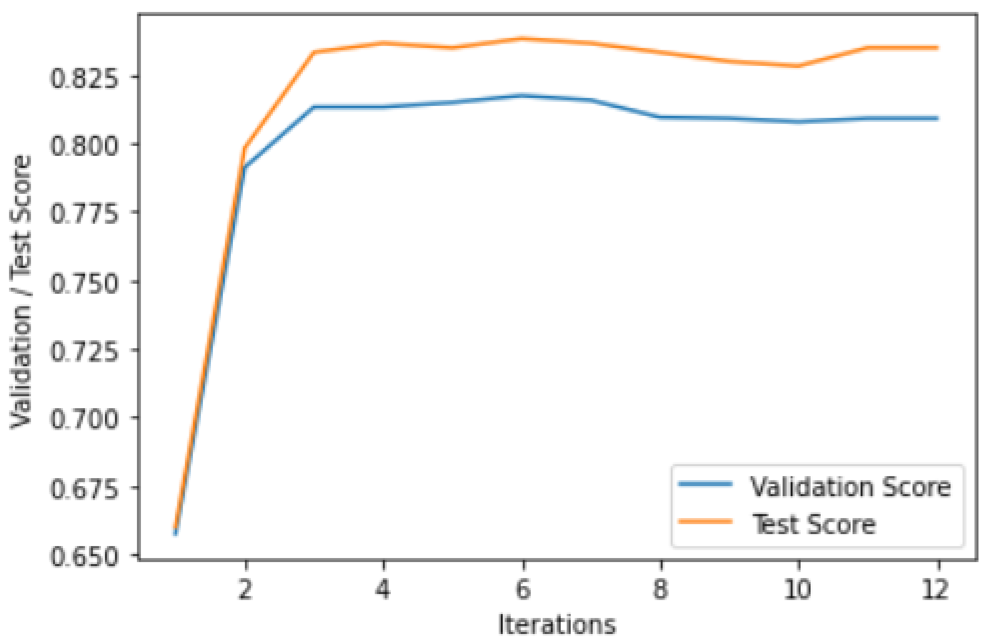


Figure 11: 10 features

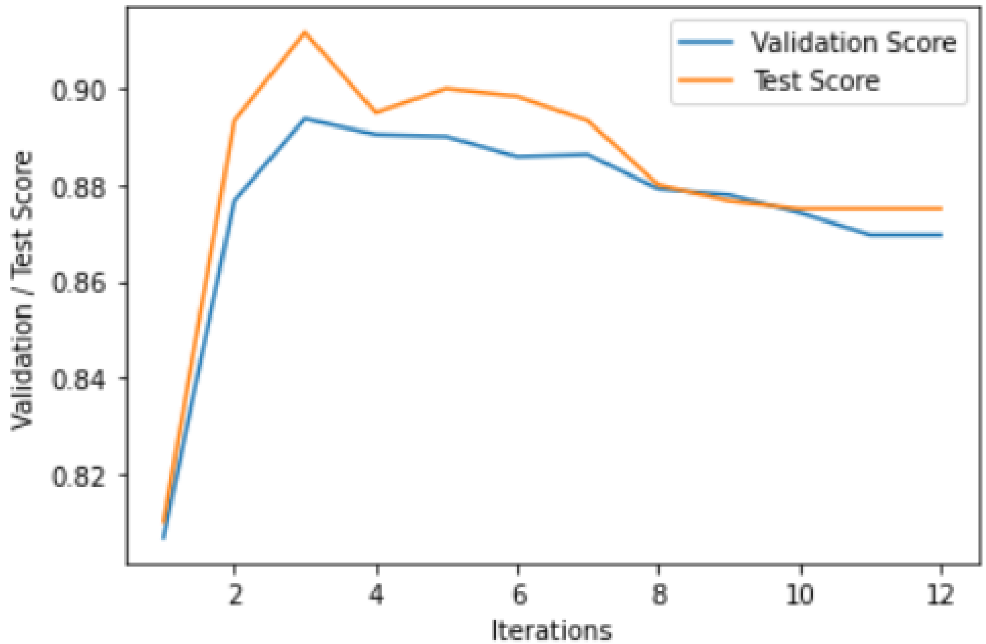


Figure 12: 25 features

<u>Optimized parameters for multiclass classification problem</u>

Following table 9 shows the best hyperparameters and accuracy after fine tuning.

Parameters	***10 features***	***25 features***
Kernel	rbf	rbf
C	1	10
gamma	Scale	Scale
Degree	1	1
Validation Score	0.889	0.952
Test score	0.890	0.947

Table 9

Confusion matrix for 25 features is shown below

```
Confusion matrix
[[70  0  0  0  0  0  1  0  0  0]
 [ 0 63  0  0  0  0  0  0  1  0]
 [ 0  0 47  1  1  0  0  1  1  0]
 [ 0  0  1 62  0  1  0  1  0  0]
 [ 0  0  2  1 62  0  0  0  0  0]
 [ 0  0  0  2  1 52  1  0  1  1]
 [ 0  0  0  0  0  0 54  0  0  0]
 [ 0  1  1  0  1  0  0 53  0  1]
 [ 0  0  0  1  0  1  1  1 53  0]
 [ 0  0  0  1  1  0  0  3  1 52]]
```

Question 2:

In this part of the assignment, we have to train model by using cross validation along with hyperparameter tuning, different kernel functions.

Taken polynomial, radial basis function, sigmoid, linear kernel, then performed grid search to find out the optimum hyperparameters.

<u>RBF Kernel:</u>

Table 9 shows the optimized hyperparameters using RBF kernel. Figure 13 below shows the variation of validation/ test accuracy with different hyperparameters.

Parameters	***25 features***
Kernel	rbf
C	5.4
gamma	0.08
Validation Score	0.971
Test score	0.974

Table 9

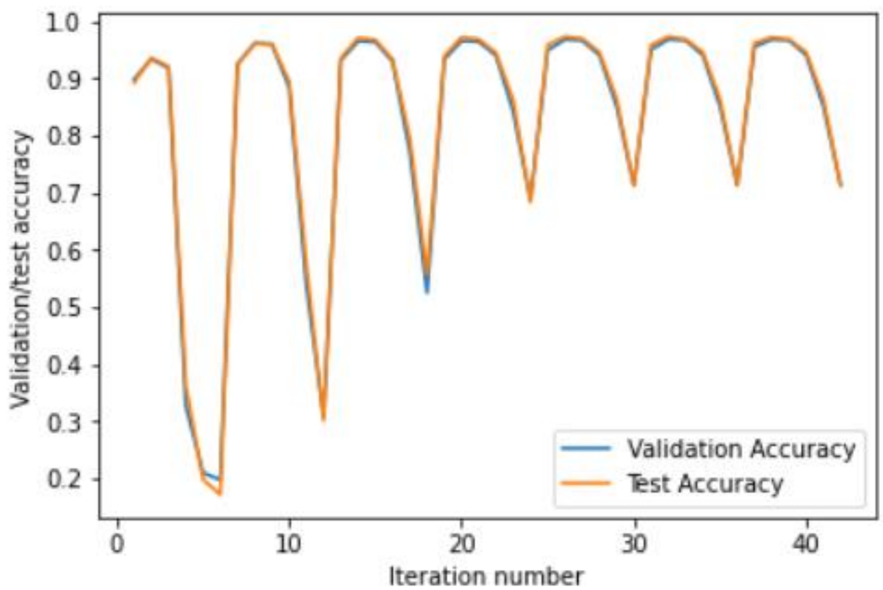


Figure 13: Accuracy plot for RBF kernel

Fine tuning has been done to obtain the required hyperparameter which maximizes the performance.

POLY Kernel:

Table 10 shows the optimized hyperparameters using POLY kernel. Figure 14 below shows the variation of validation/ test accuracy with different hyperparameters setting.

Parameters	*25 features*
Kernel	poly
C	0.1
Degree	3
Validation Score	0.955
Test score	0.959

Table 10

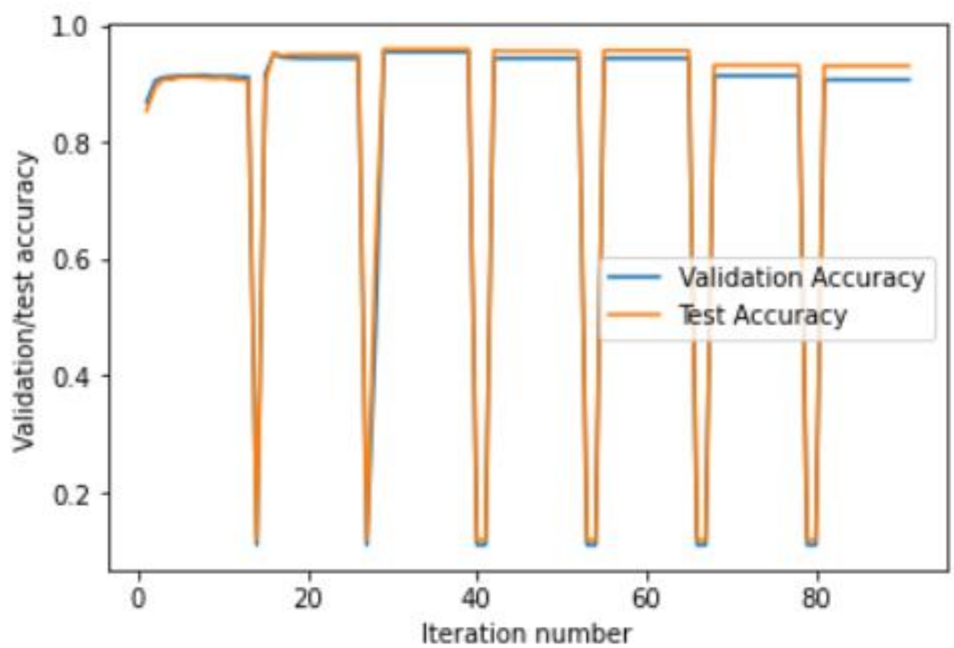


Figure 14: Accuracy plot for POLY kernel

SIGMOID Kernel

Parameters	*25 features*
Kernel	sigmoid
C	5
gamma	0.08
Validation Score	0.968
Test score	0.974

Table 11

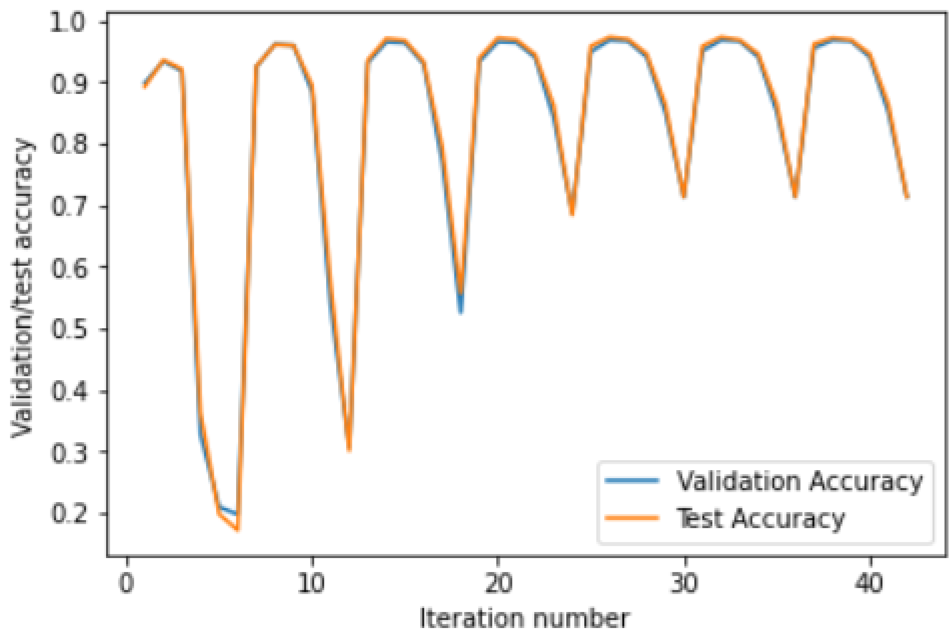


Figure 15: Accuracy plot for SIGMOID kernel

Table 11 shows the optimized hyperparameters using SIGMOID kernel. Figure 15 below shows the variation of validation/ test accuracy with different hyperparameters setting.

Linear Kernel

Table 12 shows the optimized hyperparameters using LINEAR kernel. Figure 16 below shows the variation of validation/ test accuracy with different hyperparameters setting.

Parameters	***25 features***
Kernel	linear
C	0.5
Validation Score	0.915
Test score	0.911

Table 12

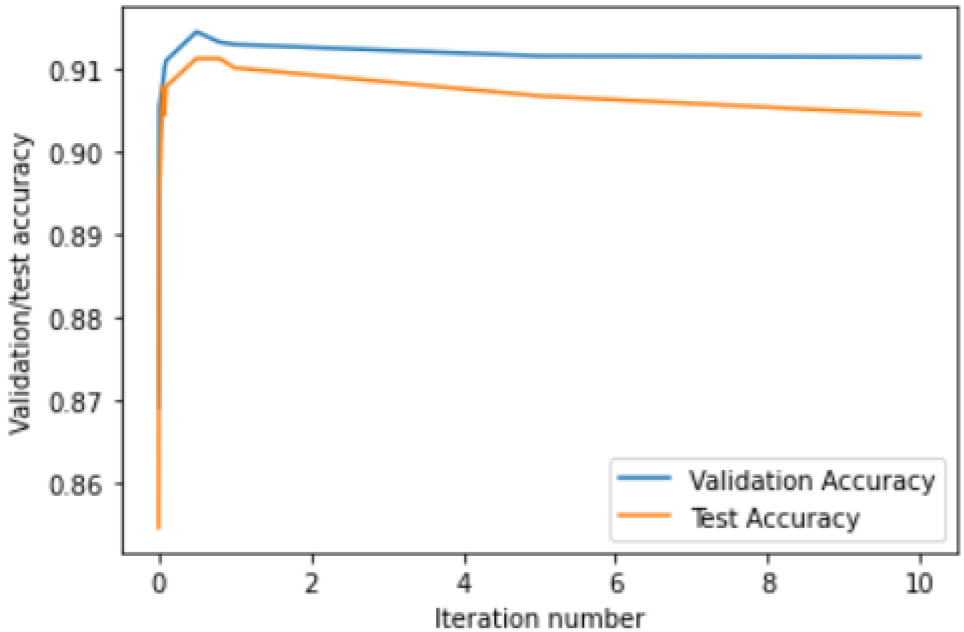


Figure 16: Accuracy plot for Linear kernel