Intelligent Systems Lab

Lab No- 2

Name- Adhyyan Tripathi

Roll no -8

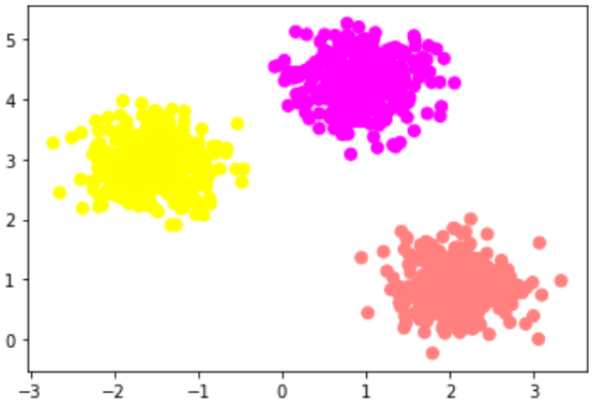
Sec C

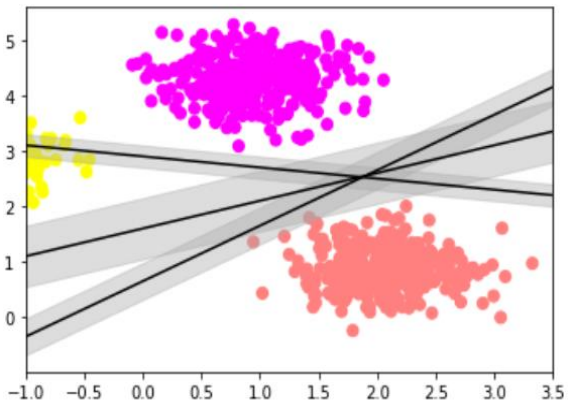
Reg no – 201700403

Q1. Data Visualization

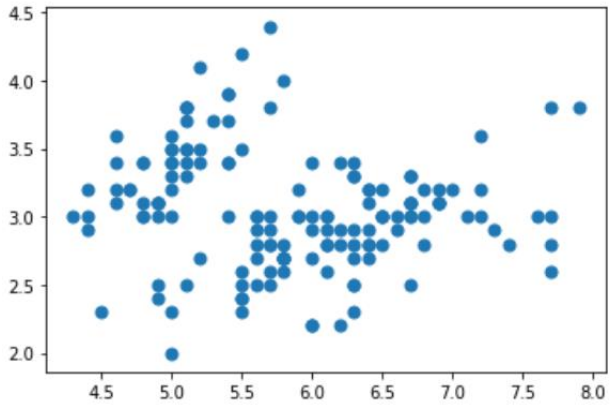
Ans- 1.1) Data Visualization of any randomly generated data-

Data generated using makeblobs (3-centers)-



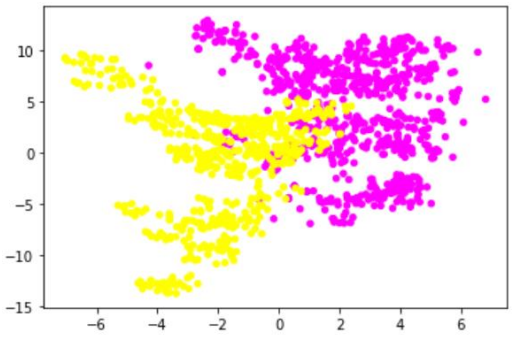


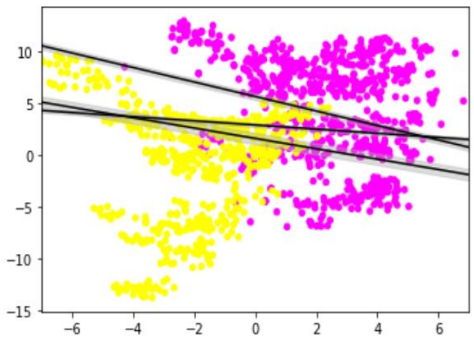
1.2) Data Visualization of Iris Dataset-



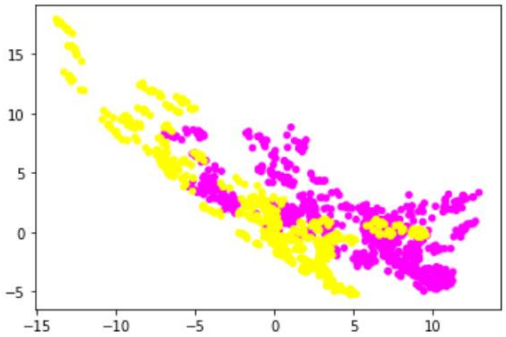
1.3) Data Visualization of Bill.csv dataset-

Column 0 vs. 1-





Column 1 vs. 2-



Q.2) Binary classification.

Ans- 2.1) Linear SVM on Bill.csv dataset-

For various values of C-

```
[[159   2]
 [  0 114]]
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       161
           1       0.98      1.00      0.99       114

    accuracy                           0.99       275
   macro avg       0.99      0.99      0.99       275
weighted avg       0.99      0.99      0.99       275
```

C=1, Accuracy=99%

```
[[160   0]
 [  0 115]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       160
           1       1.00      1.00      1.00       115

    accuracy                           1.00       275
   macro avg       1.00      1.00      1.00       275
weighted avg       1.00      1.00      1.00       275
```

C=5, Accuracy=100%

```
[[140   3]
 [  3 129]]
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       143
           1       0.98      0.98      0.98       132

    accuracy                           0.98       275
   macro avg       0.98      0.98      0.98       275
weighted avg       0.98      0.98      0.98       275
```

C=15, Accuracy=98%

From above results we conclude that when we use linear kernel with a small value of Cost("C") parameter the model will try to find a large margin that separates hyperplane even if it misclassifies more points.

Conversely, if the value of C is high the model will use a small margin to separate hyperplane.

For various values of Gamma-

```
[[162   2]
 [  2 109]]
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       164
           1       0.98      0.98      0.98       111

    accuracy                           0.99       275
   macro avg       0.98      0.98      0.98       275
weighted avg       0.99      0.99      0.99       275
```

Gamma=10, Accuracy=99%

```
[[158   1]
 [  0 116]]
              precision    recall  f1-score   support

           0       1.00      0.99      1.00       159
           1       0.99      1.00      1.00       116

    accuracy                           1.00       275
   macro avg       1.00      1.00      1.00       275
weighted avg       1.00      1.00      1.00       275
```

Gamma=50, Accuracy=100%

```
[[158   1]
 [  0 116]]
              precision    recall  f1-score   support

           0       1.00      0.99      1.00       159
           1       0.99      1.00      1.00       116

    accuracy                           1.00       275
   macro avg       1.00      1.00      1.00       275
weighted avg       1.00      1.00      1.00       275
```

Gamma=100, Accuracy=100%

From above outputs we conclude that, when we increase the value of gamma, we get higher accuracy. In case of gamma=100 the model tries to converge more towards the cluster of data. This leads to almost 0 error in classification and almost 100% accuracy, which results in overfitting of the model.

Using different types of kernel –

```
[[150   0]
 [  1 124]]
              precision    recall  f1-score   support

           0       0.99      1.00      1.00       150
           1       1.00      0.99      1.00       125

    accuracy                           1.00       275
   macro avg       1.00      1.00      1.00       275
weighted avg       1.00      1.00      1.00       275
```

Type=Polynomial, Accuracy=100%

```
[[159   0]
 [  0 116]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       159
           1       1.00      1.00      1.00       116

    accuracy                           1.00       275
   macro avg       1.00      1.00      1.00       275
weighted avg       1.00      1.00      1.00       275
```

Type=RBF, Accuracy=100%

```
[[126  32]
 [ 41  76]]
              precision    recall  f1-score   support

           0       0.75      0.80      0.78       158
           1       0.70      0.65      0.68       117

    accuracy                           0.73       275
   macro avg       0.73      0.72      0.73       275
weighted avg       0.73      0.73      0.73       275
```

Type=Sigmoid, Accuracy=73%

When we use different kernels on the same dataset (Bill.csv) we get different accuracy for every kernel type.

Comparing all the kernels with c=1 and gamma=1, we found that, polynomial and RBF gives us 100% accuracy, but only RBF classified every single point correctly.

2.2.1) Dataset with multiple classes with all 25 features- (Train_Data.csv)

For this dataset I selected class 2 and class 6, then applied the model on 500 data points, by randomly shuffling the dataframe.

The results for various parameters and kernels are given below-

1. **Using linear kernel with various parameters-**

```
[[42  1]
 [ 8 49]]
              precision    recall  f1-score   support

           2       0.84      0.98      0.90        43
           6       0.98      0.86      0.92        57

    accuracy                           0.91       100
   macro avg       0.91      0.92      0.91       100
weighted avg       0.92      0.91      0.91       100

rmse= 1.2
```

C=1, Gamma = 1

```
[[53  2]
 [ 5 40]]
              precision    recall  f1-score   support

           2       0.91      0.96      0.94        55
           6       0.95      0.89      0.92        45

    accuracy                           0.93       100
   macro avg       0.93      0.93      0.93       100
weighted avg       0.93      0.93      0.93       100

rmse= 1.0583005244258363
```

C=20, Gamma =1

```
[[54  1]
 [ 3 42]]
              precision    recall  f1-score   support

           2       0.95      0.98      0.96        55
           6       0.98      0.93      0.95        45

    accuracy                           0.96       100
   macro avg       0.96      0.96      0.96       100
weighted avg       0.96      0.96      0.96       100

rmse= 0.8
```

C=1, Gamma=100

- When using the linear kernel with C=1 and gamma=1 the rmse is 1.2 with an accuracy of 91%. This configuration neither underfits nor overfits the model.
- When using the linear kernel with C=20 and gamma=1 the rmse is 1.05 with an accuracy of 93%. This configuration is found out to be a good fit for this kernel.
- When using the linear kernel with C=1 and gamma=100 the rmse is 0.8 with an accuracy of 96%. This configuration is a best fit.

2. <u>Using polynomial kernel with various parameters-</u>

```
[[50  1]
 [ 1 48]]
              precision    recall  f1-score   support

           2       0.98      0.98      0.98        51
           6       0.98      0.98      0.98        49

    accuracy                           0.98       100
   macro avg       0.98      0.98      0.98       100
weighted avg       0.98      0.98      0.98       100

rmse= 0.565685424949238
```

C=1, Gamma=1

```
[[49  0]
 [ 0 51]]
              precision    recall  f1-score   support

           2       1.00      1.00      1.00        49
           6       1.00      1.00      1.00        51

    accuracy                           1.00       100
   macro avg       1.00      1.00      1.00       100
weighted avg       1.00      1.00      1.00       100

rmse= 0.0
```

C=20, Gamma=1

```
[[54  0]
 [ 1 45]]
              precision    recall  f1-score   support

           2       0.98      1.00      0.99        54
           6       1.00      0.98      0.99        46

    accuracy                           0.99       100
   macro avg       0.99      0.99      0.99       100
weighted avg       0.99      0.99      0.99       100

rmse= 0.4
```

C=1, Gamma=100

- When using the polynomial kernel with C=1 and gamma=1 the rmse is 0.56 with an accuracy of 98%. This configuration is the best fit.
- When using the polynomial kernel with C=20 and gamma=1 the rmse is 0 with an accuracy of 100%. This configuration is found out to be overfit.

- When using the polynomial kernel with C=1 and gamma=100 the rmse is 0.4 with an accuracy of 99%. This configuration is overfitting the model.

3. Using RBF kernel with various parameters-

```
[[48  0]
 [52  0]]
              precision    recall  f1-score   support

           2       0.48      1.00      0.65        48
           6       0.00      0.00      0.00        52

    accuracy                           0.48       100
   macro avg       0.24      0.50      0.32       100
weighted avg       0.23      0.48      0.31       100

rmse= 2.8844410203711917
```

C=1, Gamma=1

```
[[54  0]
 [46  0]]
              precision    recall  f1-score   support

           2       0.54      1.00      0.70        54
           6       0.00      0.00      0.00        46

    accuracy                           0.54       100
   macro avg       0.27      0.50      0.35       100
weighted avg       0.29      0.54      0.38       100

rmse= 2.7129319932501073
```

C=20, Gamma=1

```
[[ 0 64]
 [ 0 36]]
              precision    recall  f1-score   support

           2       0.00      0.00      0.00        64
           6       0.36      1.00      0.53        36

    accuracy                           0.36       100
   macro avg       0.18      0.50      0.26       100
weighted avg       0.13      0.36      0.19       100

rmse= 3.2
```

C=1, Gamma=100

- When using the RBF kernel with C=1 and gamma=1 the rmse is 2.88 with an accuracy of 48%. This configuration is an underfit.
- When using the RBF kernel with C=20 and gamma=1 the rmse is 2.71 with an accuracy of 54%. This configuration is found out to be the bestfit.
- When using the RBF kernel with C=1 and gamma=100 the rmse is 3.2 with an accuracy of 36%. This configuration is underfitting the model.

4. Using sigmoid kernel with various parameters-

```
[[27 21]
 [19 33]]
              precision    recall  f1-score   support

           2       0.59      0.56      0.57        48
           6       0.61      0.63      0.62        52

    accuracy                           0.60       100
   macro avg       0.60      0.60      0.60       100
weighted avg       0.60      0.60      0.60       100

rmse= 2.5298221281347035
```

C=1, Gamma=1

```
[[35 20]
 [21 24]]
              precision    recall  f1-score   support

           2       0.62      0.64      0.63        55
           6       0.55      0.53      0.54        45

    accuracy                           0.59       100
   macro avg       0.59      0.58      0.58       100
weighted avg       0.59      0.59      0.59       100

rmse= 2.5612496949731396
```

C=20, Gamma=1

```
[[36 18]
 [23 23]]
              precision    recall  f1-score   support

           2       0.61      0.67      0.64        54
           6       0.56      0.50      0.53        46

    accuracy                           0.59       100
   macro avg       0.59      0.58      0.58       100
weighted avg       0.59      0.59      0.59       100

rmse= 2.5612496949731396
```

C=1, Gamma=100

- When using the sigmoid kernel with C=1 and gamma=1 the rmse is 2.53 with an accuracy of 60%. This configuration is the best fit.
- When using the sigmoid kernel with C=20 and gamma=1 the rmse is 2.56 with an accuracy of 59%. This configuration is the best fit.
- When using the sigmoid kernel with C=1 and gamma=100 the rmse is 2.56 with an accuracy of 59%. This configuration is the best fit.

2.2.2) Dataset with multiple classes with 10 features- (Train_Data.csv)

1. Using linear kernel with various parameters-

```
[[45  3]
 [ 4 48]]
              precision    recall  f1-score   support

           2       0.92      0.94      0.93        48
           6       0.94      0.92      0.93        52

    accuracy                           0.93       100
   macro avg       0.93      0.93      0.93       100
weighted avg       0.93      0.93      0.93       100

rmse= 1.0583005244258363
```

C=1, Gamma=1

```
[[44  4]
 [ 4 48]]
              precision    recall  f1-score   support

           2       0.92      0.92      0.92        48
           6       0.92      0.92      0.92        52

    accuracy                           0.92       100
   macro avg       0.92      0.92      0.92       100
weighted avg       0.92      0.92      0.92       100

rmse= 1.131370849898476
```

C=20, Gamma=1

```
[[51  4]
 [ 1 44]]
              precision    recall  f1-score   support

           2       0.98      0.93      0.95        55
           6       0.92      0.98      0.95        45

    accuracy                           0.95       100
   macro avg       0.95      0.95      0.95       100
weighted avg       0.95      0.95      0.95       100

rmse= 0.8944271909999159
```

C=1, Gamma=100

- When using the linear kernel with C=1 and gamma=1 the rmse is 1.05 with an accuracy of 93%. This configuration is neither overfit nor underfit.
- When using the linear kernel with C=20 and gamma=1 the rmse is 1.13 with an accuracy of 92%. This configuration is neither overfit nor underfit.
- When using the linear kernel with C=1 and gamma=100 the rmse is 0.89 with an accuracy of 95%. This configuration is a best fit.

2. <u>Using polynomial kernel with various parameters-</u>

```
[[51  0]
 [ 0 49]]
              precision    recall  f1-score   support

           2       1.00      1.00      1.00        51
           6       1.00      1.00      1.00        49

    accuracy                           1.00       100
   macro avg       1.00      1.00      1.00       100
weighted avg       1.00      1.00      1.00       100

rmse= 0.0
```

C=1, Gamma=1

```
[[43  0]
 [ 3 54]]
              precision    recall  f1-score   support

           2       0.93      1.00      0.97        43
           6       1.00      0.95      0.97        57

    accuracy                           0.97       100
   macro avg       0.97      0.97      0.97       100
weighted avg       0.97      0.97      0.97       100

rmse= 0.6928203230275509
```

C=20, Gamma=1

```
[[55  0]
 [ 0 45]]
              precision    recall  f1-score   support

           2       1.00      1.00      1.00        55
           6       1.00      1.00      1.00        45

    accuracy                           1.00       100
   macro avg       1.00      1.00      1.00       100
weighted avg       1.00      1.00      1.00       100

rmse= 0.0
```

C=1, Gamma=100

- When using the polynomial kernel with C=1 and gamma=1 the rmse is 0 with an accuracy of 100%. This configuration is an overfit.
- When using the polynomial kernel with C=20 and gamma=1 the rmse is 0.69 with an accuracy of 97%. This configuration is found out to be the bestfit.

- When using the polynomial kernel with C=1 and gamma=100 the rmse is 0 with an accuracy of 100%. This configuration is overfitting the model.

3. Using RBF kernel with various parameters-

```
[[ 0 60]
 [ 0 40]]
              precision    recall  f1-score   support

           2       0.00      0.00      0.00        60
           6       0.40      1.00      0.57        40

    accuracy                           0.40       100
   macro avg       0.20      0.50      0.29       100
weighted avg       0.16      0.40      0.23       100

rmse= 3.0983866769659336
```

C=1, Gamma=1

```
[[49  0]
 [51  0]]
              precision    recall  f1-score   support

           2       0.49      1.00      0.66        49
           6       0.00      0.00      0.00        51

    accuracy                           0.49       100
   macro avg       0.24      0.50      0.33       100
weighted avg       0.24      0.49      0.32       100

rmse= 2.85657137141714
```

C=20, Gamma=1

```
[[51  0]
 [49  0]]
              precision    recall  f1-score   support

           2       0.51      1.00      0.68        51
           6       0.00      0.00      0.00        49

    accuracy                           0.51       100
   macro avg       0.26      0.50      0.34       100
weighted avg       0.26      0.51      0.34       100

rmse= 2.8
```

C=1, Gamma=100

- When using the RBF kernel with C=1 and gamma=1 the rmse is 3.09 with an accuracy of 40%. This configuration is an underfit.
- When using the RBF kernel with C=20 and gamma=1 the rmse is 2.85 with an accuracy of 49%. This configuration is found out to be an underfit.
- When using the RBF kernel with C=1 and gamma=100 the rmse is 2.8 with an accuracy of 51%. This configuration is best fit.

4. Using sigmoid kernel with various parameters-

```
[[32 24]
 [17 27]]
              precision    recall  f1-score   support

           2       0.65      0.57      0.61        56
           6       0.53      0.61      0.57        44

    accuracy                           0.59       100
   macro avg       0.59      0.59      0.59       100
weighted avg       0.60      0.59      0.59       100

rmse= 2.5612496949731396
```

C=1, Gamma=1

```
[[33 16]
 [28 23]]
              precision    recall  f1-score   support

           2       0.54      0.67      0.60        49
           6       0.59      0.45      0.51        51

    accuracy                           0.56       100
   macro avg       0.57      0.56      0.56       100
weighted avg       0.57      0.56      0.55       100

rmse= 2.6532998322843198
```

C=20, Gamma=1

```
[[51  0]
 [49  0]]
              precision    recall  f1-score   support

           2       0.51      1.00      0.68        51
           6       0.00      0.00      0.00        49

    accuracy                           0.51       100
   macro avg       0.26      0.50      0.34       100
weighted avg       0.26      0.51      0.34       100

rmse= 2.8
```

C=1, Gamma=100

- When using the sigmoid kernel with C=1 and gamma=1 the rmse is 2.56 with an accuracy of 59%. This configuration is the bestfit.
- When using the sigmoid kernel with C=20 and gamma=1 the rmse is 2.65 with an accuracy of 56%. This configuration is found out to be an underfit.
- When using the sigmoid kernel with C=1 and gamma=100 the rmse is 2.8 with an accuracy of 51%. This configuration is under fit.

<u>Comparison with 25 features and 10 features results-</u>

```
[[42  1]
 [ 8 49]]
              precision    recall  f1-score   support

           2       0.84      0.98      0.90        43
           6       0.98      0.86      0.92        57

    accuracy                           0.91       100
   macro avg       0.91      0.92      0.91       100
weighted avg       0.92      0.91      0.91       100

rmse= 1.2
```

25 Feat. (linear)

```
[[45  3]
 [ 4 48]]
              precision    recall  f1-score   support

           2       0.92      0.94      0.93        48
           6       0.94      0.92      0.93        52

    accuracy                           0.93       100
   macro avg       0.93      0.93      0.93       100
weighted avg       0.93      0.93      0.93       100

rmse= 1.0583005244258363
```

10 Feat. (linear)

```
[[50  1]
 [ 1 48]]
              precision    recall  f1-score   support

           2       0.98      0.98      0.98        51
           6       0.98      0.98      0.98        49

    accuracy                           0.98       100
   macro avg       0.98      0.98      0.98       100
weighted avg       0.98      0.98      0.98       100

rmse= 0.565685424949238
```

25 Feat. (poly)

```
[[51  0]
 [ 0 49]]
              precision    recall  f1-score   support

           2       1.00      1.00      1.00        51
           6       1.00      1.00      1.00        49

    accuracy                           1.00       100
   macro avg       1.00      1.00      1.00       100
weighted avg       1.00      1.00      1.00       100

rmse= 0.0
```

10 Feat. (poly)

```
[[48  0]
 [52  0]]
              precision    recall  f1-score   support

           2       0.48      1.00      0.65        48
           6       0.00      0.00      0.00        52

    accuracy                           0.48       100
   macro avg       0.24      0.50      0.32       100
weighted avg       0.23      0.48      0.31       100

rmse= 2.8844410203711917
```

25 Feat. (RBF)

```
[[ 0 60]
 [ 0 40]]
              precision    recall  f1-score   support

           2       0.00      0.00      0.00        60
           6       0.40      1.00      0.57        40

    accuracy                           0.40       100
   macro avg       0.20      0.50      0.29       100
weighted avg       0.16      0.40      0.23       100

rmse= 3.0983866769659336
```

10 Feat. (RBF)

```
[[27 21]
 [19 33]]
              precision    recall  f1-score   support

           2       0.59      0.56      0.57        48
           6       0.61      0.63      0.62        52

    accuracy                           0.60       100
   macro avg       0.60      0.60      0.60       100
weighted avg       0.60      0.60      0.60       100

rmse= 2.5298221281347035
```

25 Feat. (sigmoid)

```
[[32 24]
 [17 27]]
              precision    recall  f1-score   support

           2       0.65      0.57      0.61        56
           6       0.53      0.61      0.57        44

    accuracy                           0.59       100
   macro avg       0.59      0.59      0.59       100
weighted avg       0.60      0.59      0.59       100

rmse= 2.5612496949731396
```

10 Feat. (sigmoid)

By comparing the above rmse values and accuracy with each other, the results are very similar to each other.

Again, by comparing the results when we consider any other pair of classes-

```
[[48  0]
 [52  0]]
              precision    recall  f1-score   support

           2       0.48      1.00      0.65        48
           6       0.00      0.00      0.00        52

    accuracy                           0.48       100
   macro avg       0.24      0.50      0.32       100
weighted avg       0.23      0.48      0.31       100

rmse= 2.8844410203711917
```

With class 2 and 6

```
[[50  0]
 [50  0]]
              precision    recall  f1-score   support

           1       0.50      1.00      0.67        50
           4       0.00      0.00      0.00        50

    accuracy                           0.50       100
   macro avg       0.25      0.50      0.33       100
weighted avg       0.25      0.50      0.33       100

rmse= 2.1213203435596424
```

with class 1 and 4

We conclude that the difference is almost negligible, hence we get almost the same results when we change the pair of classes.