C S 487/519 Applied Machine Learning I
Fall 2018
Project 3: Compare classifiers in scikit-learn library

October 09, 2018

Jithin Jacob Benjamin Jacob
800681973

Objective:

In this individual project I have implemented various classification algorithms and compared them as per the requirements keps forth below.

- Write classification code by utilizing several scikit-learn classifiers: (i) perceptron, (ii) support vector machine (linear and non-linear using Radial Basis Function (RBF) kernel), (iii) decision tree, (iv) K-nearest neighbor, and (v) logistic regression.
- Each classifier needs to be tested using two datasets: (1) the digits dataset offered by scikit-learn library, and (2) one dataset containing time-series instances. Example of the second dataset can be the REALDISP Activity Recognition Dataset (https://archive.ics.uci.edu/ml/datasets/ REALDISP+Activity+Recognition+Dataset).
- Properly analyze the classifiers' behavior by applying the knowledge that we discussed in class. Such analysis should include at least accuracy and running time.
- Understand the source code of DecisionTreeClassifier (You can follow the source link in http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html).
- Please denote two strategies that this classifier implements to pre-prune or post-prune the tree.
- For each strategy, please clearly identify the repository _le and the lines of code that implement such strategies.

Algorithms and Outputs:

Here in these classifications we have used two datasets, namely

- Digits Dataset
- Sonar Dataset (Contains Time Series Instance)

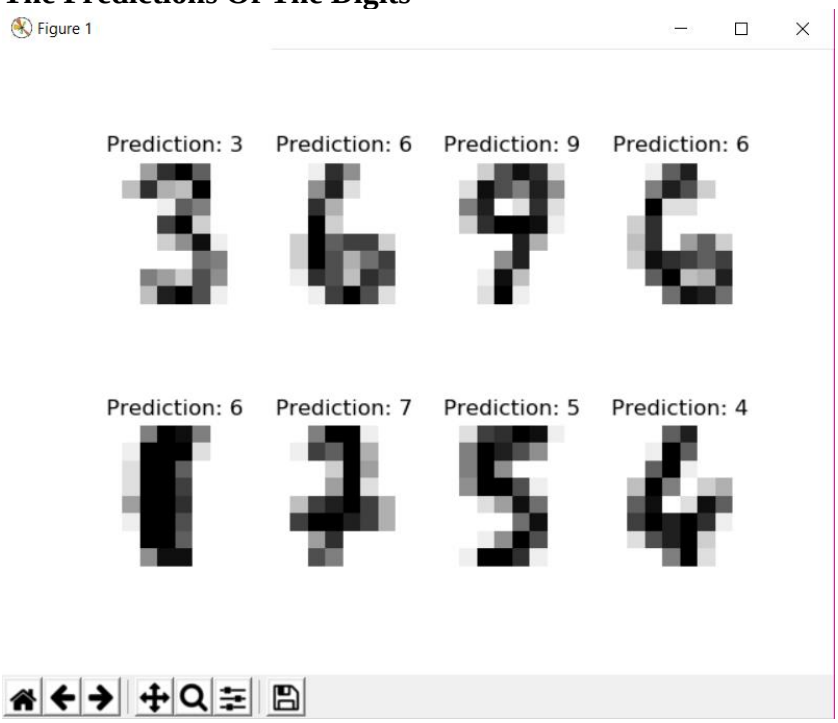
Perceptron(Digit Dataset):

The Perceptron classifier is implemented and it is been done with the digit dataset with

Accuracy : 96.87

And a total running time of 9.37 seconds.

The Predictions Of The Digits



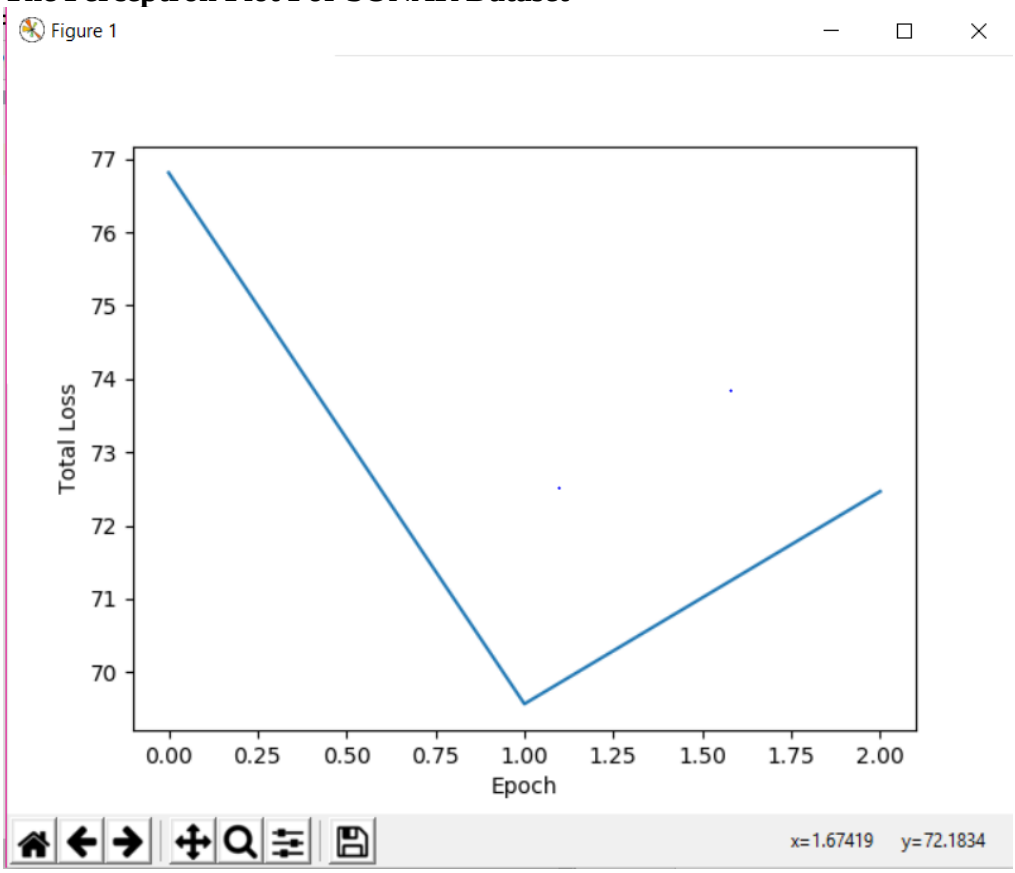
The Accuracy And Elapsed Time For Perceptron Is Displayed

```
The classification score (Accuracy): 0.96871

Classification report for classifier Perceptron(alpha=0.0001, class_weight=None, eta0=1.0, fit_intercept=True,
      max_iter=1000, n_iter=None, n_jobs=1, penalty=None, random_state=0,
      shuffle=True, tol=None, verbose=0, warm_start=False):
             precision    recall  f1-score   support

          0       1.00      0.99      0.99       140
          1       0.94      0.93      0.94       144
          2       0.99      1.00      0.99       141
          3       0.96      0.92      0.94       145
          4       0.98      0.98      0.98       147
          5       0.97      0.99      0.98       146
          6       0.98      1.00      0.99       145
          7       0.98      0.97      0.98       144
          8       0.93      0.94      0.94       140
          9       0.96      0.97      0.96       146

avg / total       0.97      0.97      0.97      1438

Confusion matrix:
[[138   0   0   0   1   0   1   0   0   0]
 [  0 134   0   2   0   0   2   0   2   4]
 [  0   0 141   0   0   0   0   0   0   0]
 [  0   1   1 134   0   3   0   2   4   0]
 [  0   1   0   0 144   0   0   0   0   2]
 [  0   2   0   0   0 144   0   0   0   0]
 [  0   0   0   0   0   0 145   0   0   0]
 [  0   0   0   0   1   0   0 140   3   0]
 [  0   3   1   1   1   2   0   0 132   0]
 [  0   1   0   2   0   0   0   1   1 141]]
Elapsed time for Perceptron:  9.37890076637268
```

Perceptron(Sonar Dataset):

The Perceptron using the SONAR dataset has acquired a accuracy of 71.04% and a total running time of 5.89 seconds.

The Perceptron Plot For SONAR Dataset

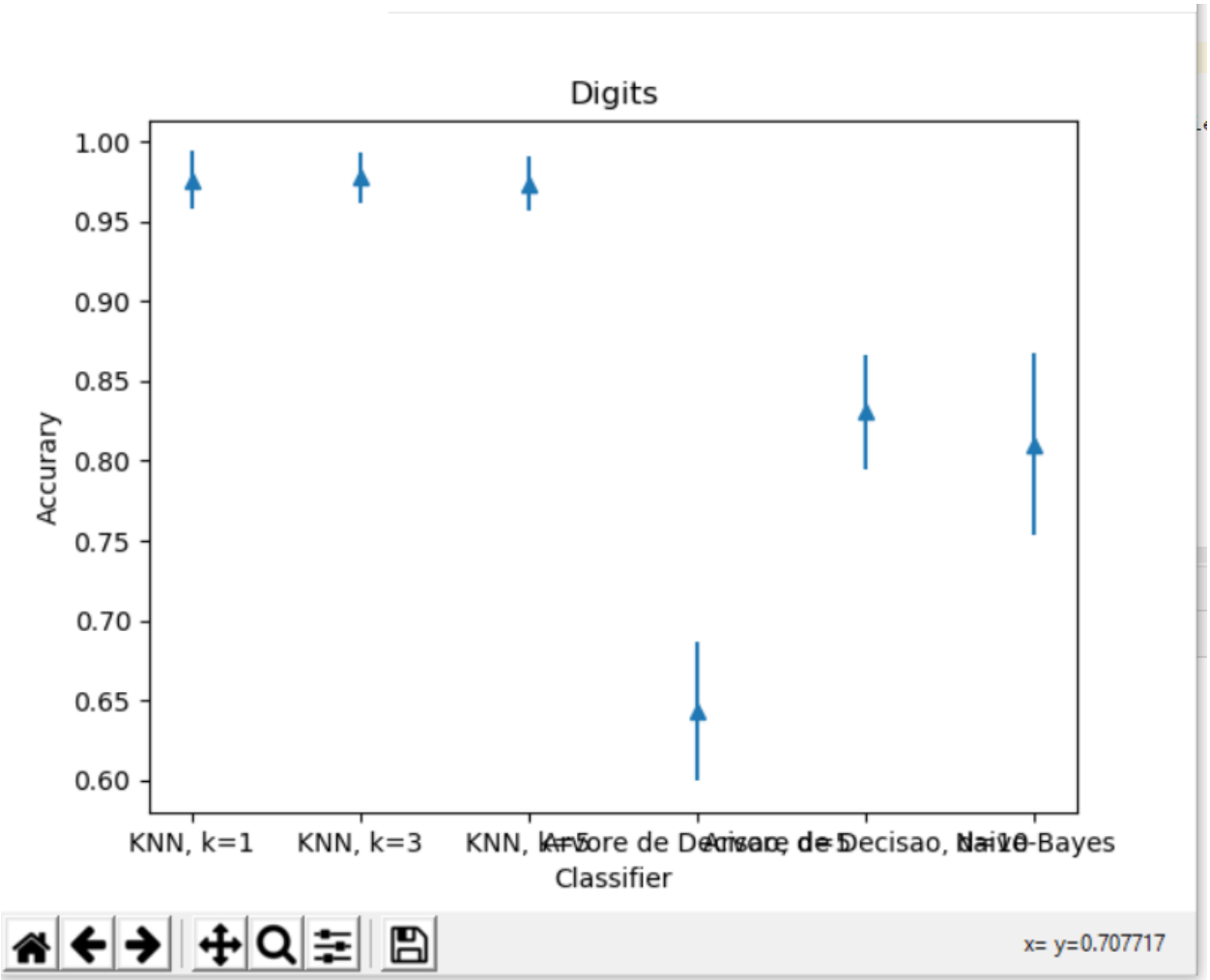


The Accuracy And The Elapsed Time for perceptron using the SONOR dataset:

```
Scores: [81.15942028985508, 69.56521739130434, 62.31884057971014]
Mean Accuracy: 71.014%
Elapsed Time Duration(s): 5.894053936004639
         0       1       2       3  ...      57      58      59  60
203  0.0187  0.0346  0.0168  0.0177 ...  0.0115  0.0193  0.0157   M
204  0.0323  0.0101  0.0298  0.0564 ...  0.0032  0.0062  0.0067   M
205  0.0522  0.0437  0.0180  0.0292 ...  0.0138  0.0077  0.0031   M
206  0.0303  0.0353  0.0490  0.0608 ...  0.0079  0.0036  0.0048   M
207  0.0260  0.0363  0.0136  0.0272 ...  0.0036  0.0061  0.0115   M

[5 rows x 61 columns]
```

Decision Tree(Digits Dataset):

The Decision Tree using the digit dataset acquired a Accuracy of 97.61% and got a elapsed running time of 6.15 seconds.



```
*** Digits ***
-> KNN, k=1
Accuracy = 0.976149386025202?0.017908413555155718
-> KNN, k=3
Accuracy = 0.97778921137986 43?0.015939574731902884
-> KNN, k=5
Accuracy = 0.97394828725469 06?0.016549601537444315
-> Arvore de Decisao, d=5
Accuracy = 0.64126767714797 44?0.04358527902632927
-> Arvore de Decisao, d=10
Accuracy = 0.82264213319844 86?0.041222005060941876
-> Naive-Bayes
Accuracy = 0.81035375835678 21?0.056655402070708565



Elapsed time 6.150292158126831

Process finished with exit code 0
```

Decision Tree(Sonar Dataset):

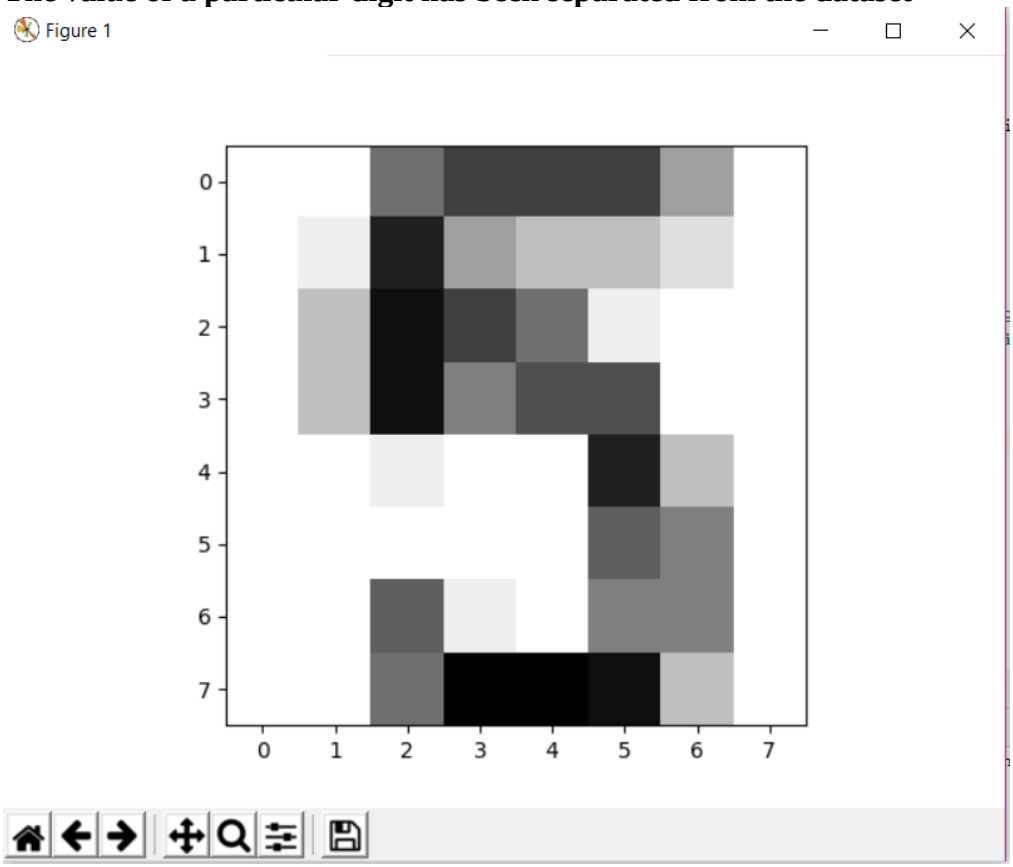
The Decision tree using the sonar dataset got a Accuracy of 60.31% and a total execution time of 0.1 seconds.

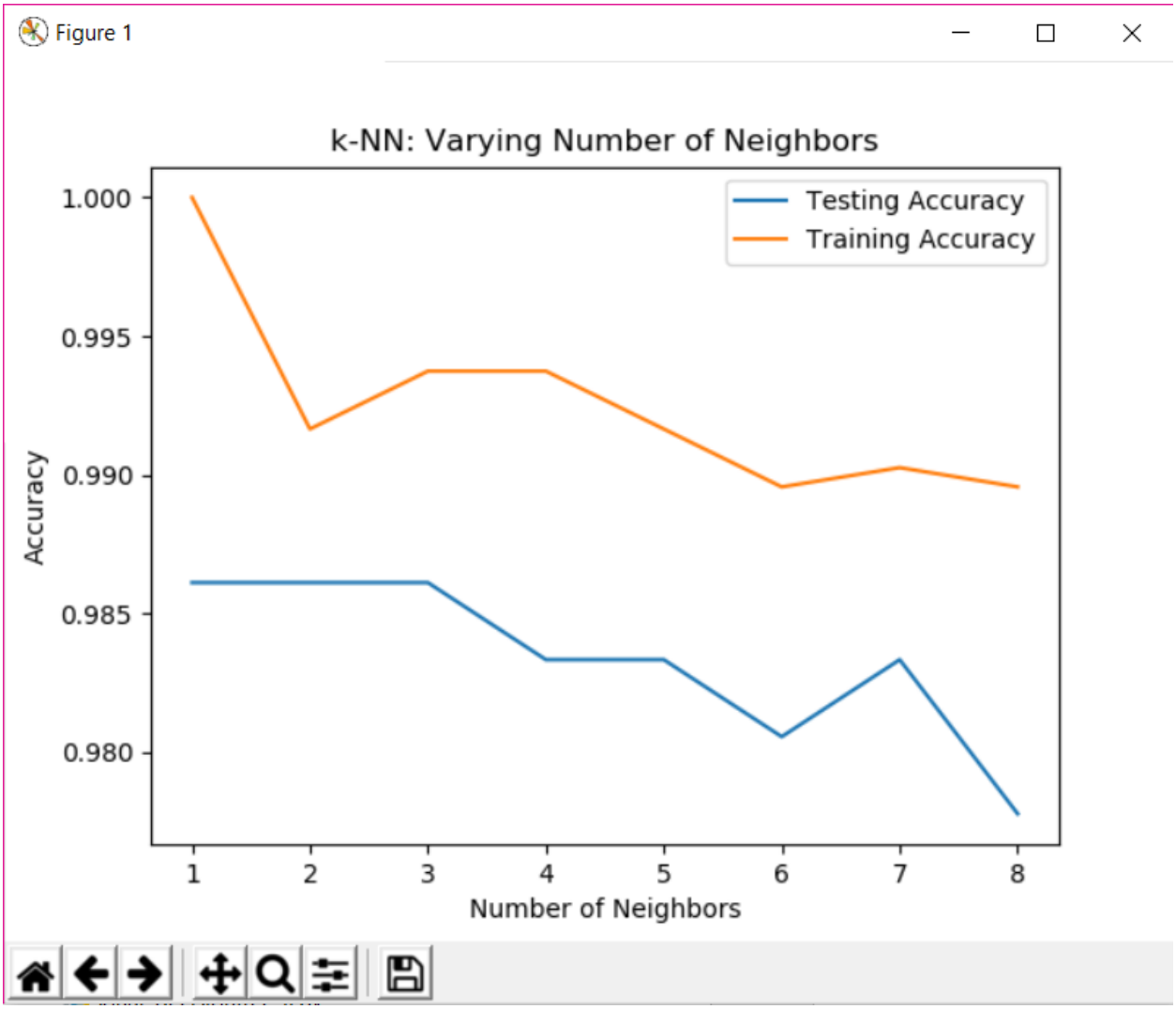
```
Dataset Lenght::  208
Dataset Shape::  (208, 61)
['M' 'M' 'R' 'R' 'M' 'R' 'M' 'R' 'M' 'M' 'M' 'M' 'R' 'R' 'R' 'M' 'M' 'M'
 'R' 'R' 'R' 'R' 'M' 'R' 'M' 'M' 'R' 'R' 'R' 'M' 'M' 'R' 'R' 'R' 'M' 'M'
 'R' 'M' 'R' 'M' 'R' 'R' 'R' 'M' 'M' 'M' 'M' 'M' 'R' 'R' 'R' 'R' 'M' 'M'
 'M' 'R' 'R' 'M' 'R' 'R' 'R' 'R' 'M']
['M' 'M' 'R' 'R' 'M' 'R' 'M' 'M' 'M' 'M' 'R' 'M' 'M' 'R' 'R' 'R' 'M' 'M'
 'R' 'M' 'M' 'M' 'M' 'M' 'M' 'M' 'M' 'R' 'R' 'R' 'M' 'R' 'M' 'M' 'M' 'M'
 'M' 'M' 'M' 'M' 'R' 'R' 'M' 'R' 'R' 'M' 'R' 'R' 'R' 'M' 'R' 'M' 'M' 'M'
 'R' 'M' 'R' 'R' 'R' 'M' 'R' 'R' 'M']
Accuracy is  60.317460317460316
Elapsed Time (s): 0.10887622833251953
```

K-Nearest Neighbour(Digits Dataset):

The KNN classifier using the digits dataset has a total Accuracy of 98.33% and a total elapsed time of 145 seconds.

The value of a particular digit has been separated from the dataset

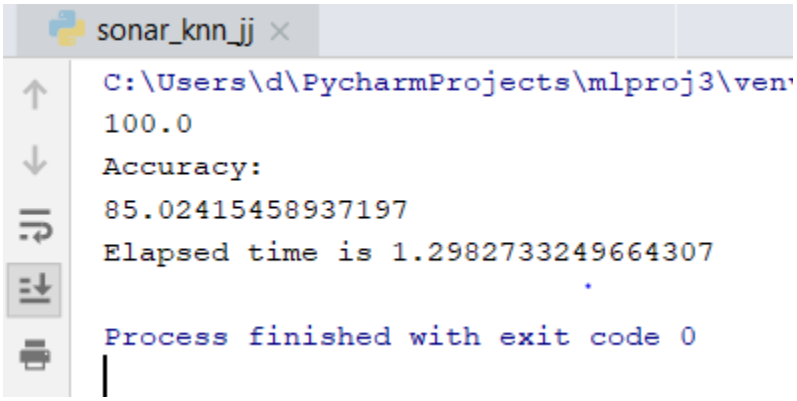




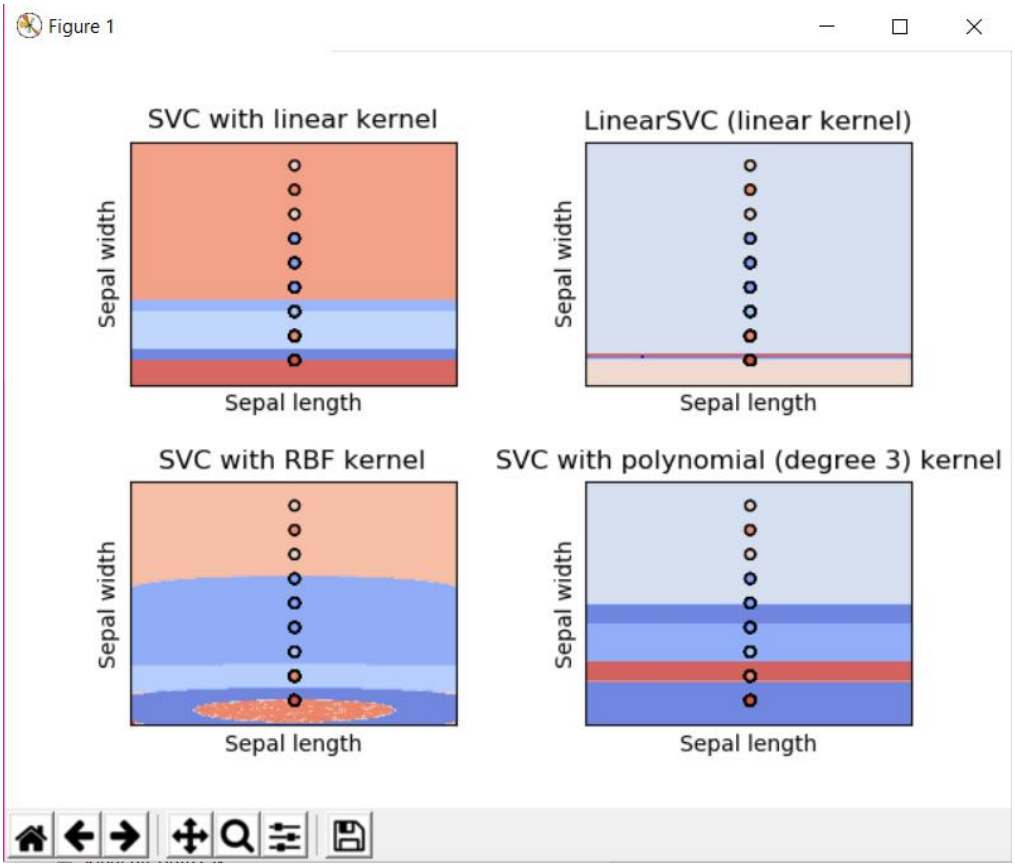
digits_knn_jj

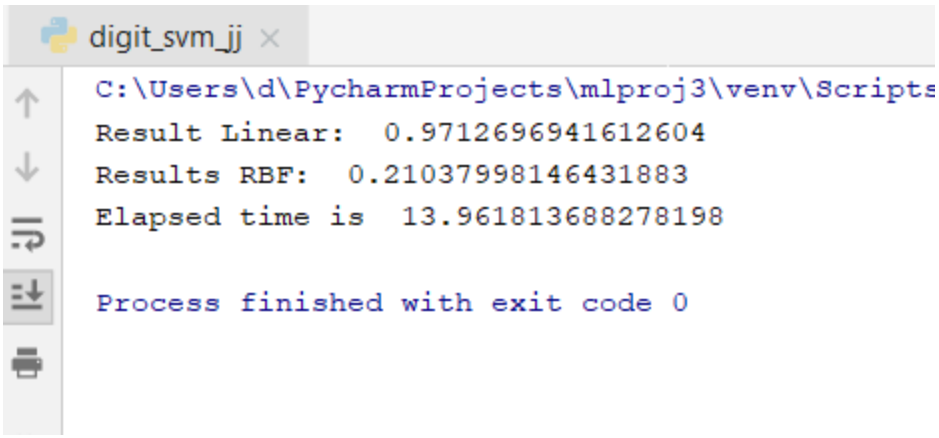
```
C:\Users\d\PycharmProjects\mlproj3\v
dict_keys(['data', 'target', 'target
(1797, 8, 8)
(1797, 64)
0.9833333333333333
Elapsed time is  145.25028467178345

Process finished with exit code 0
```

K- Nearest Neighbour(Sonar Dataset):
The KNN using the sonar dataset has a Accuracy of 85.02% and a total elapsed time of 1.29 seconds.

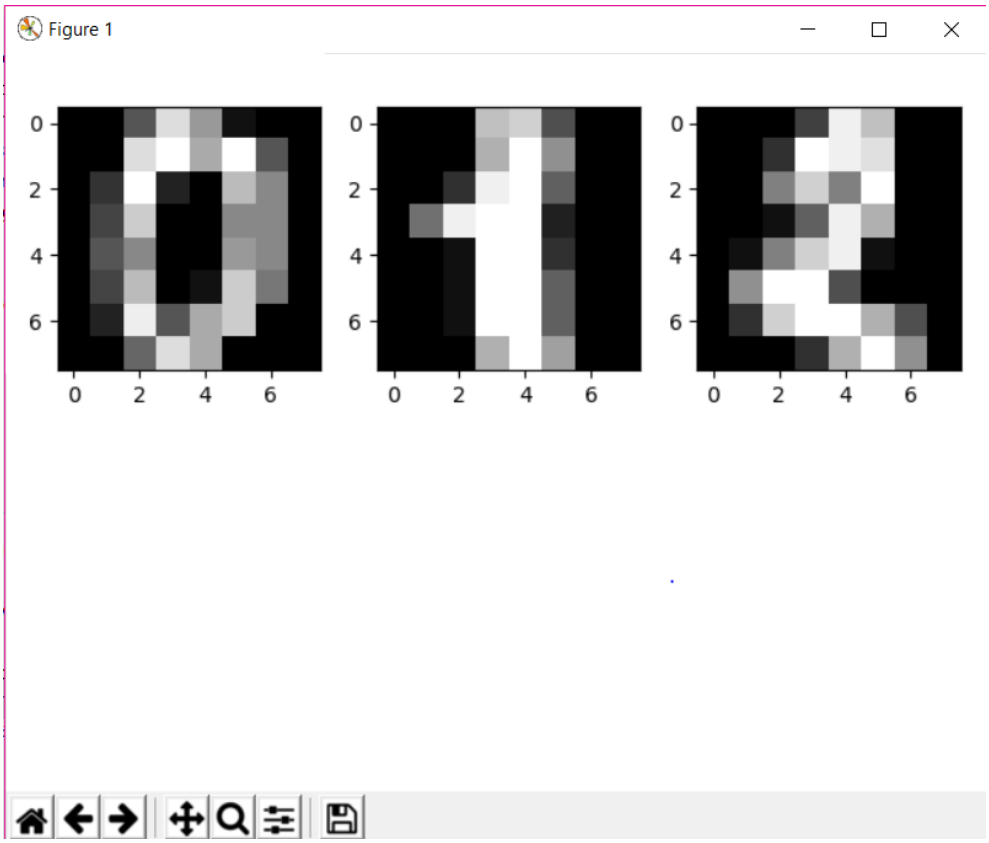


SVM(Digits Dataset):
SVM using the digits dataset has a the following values
Linear Kernal Accuracy: 97.12%
RBF Kernel Accuracy: 21.03%
Elapsed time: 13.96 seconds





LOGISTIC REGRESSION(DIGITS DATASET):
The logistic regression using the digits dataset has a Accuracy 89.74% and a total elapsed time of 55.93 seconds.



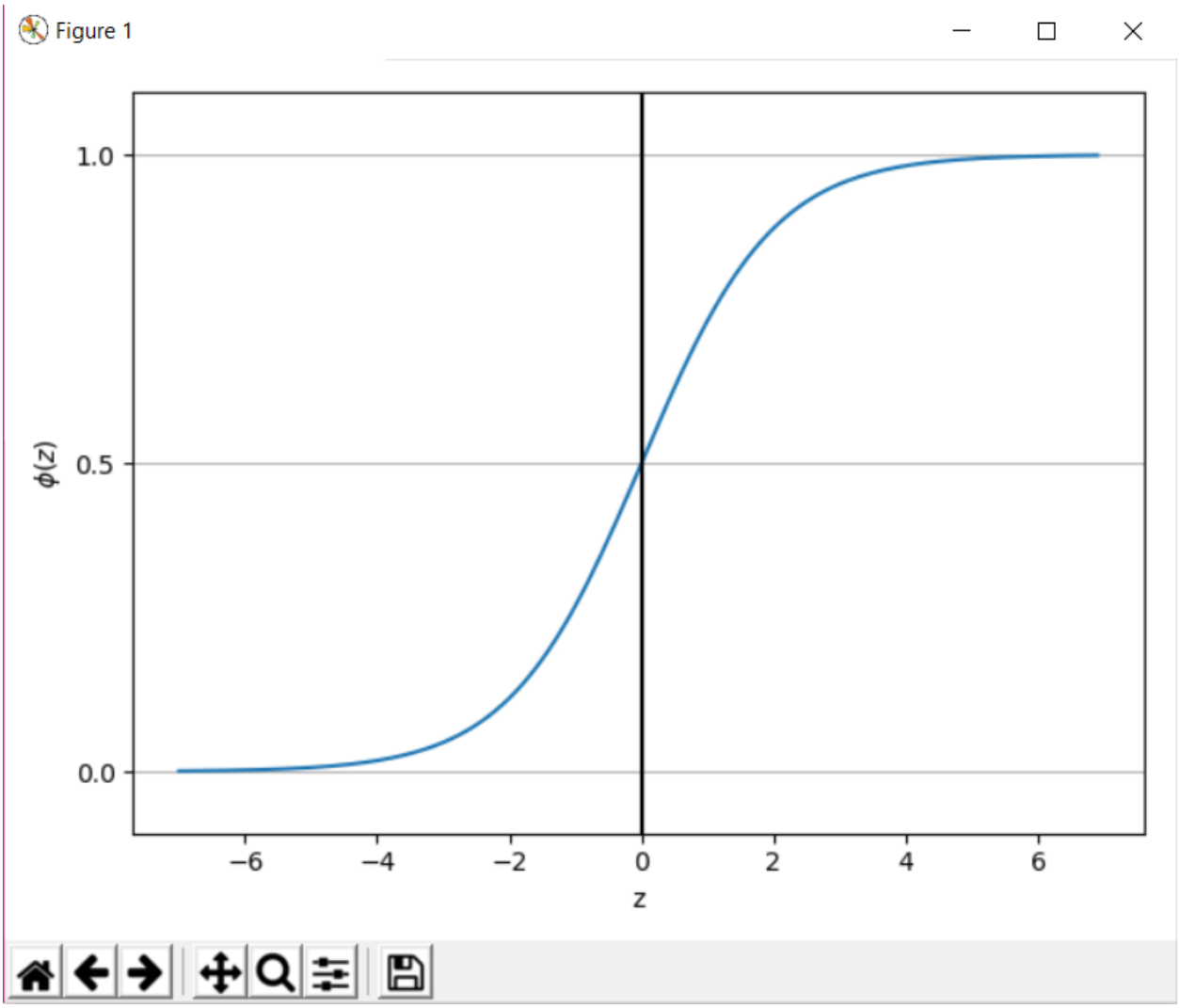
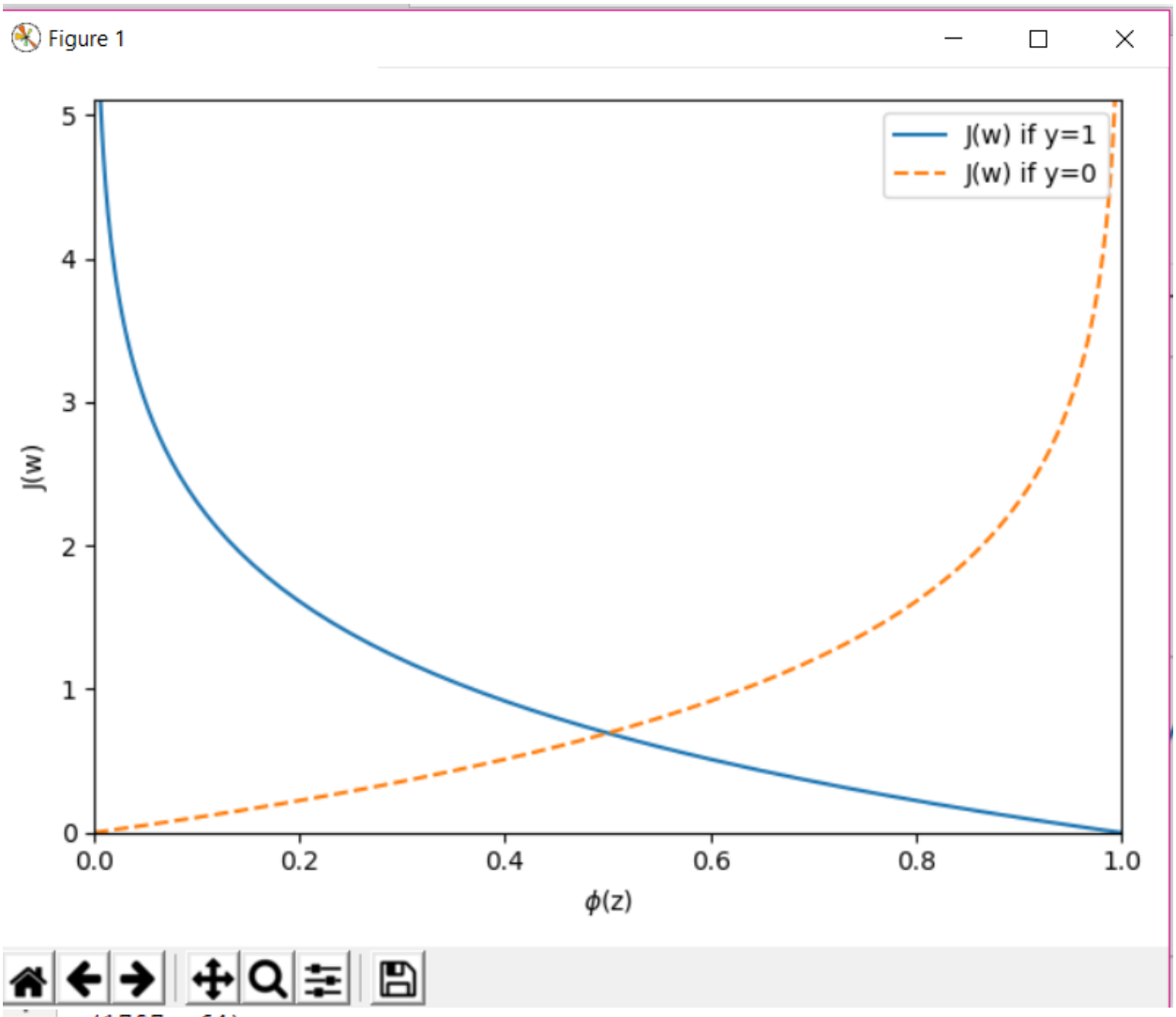


Figure 1
1.0
0.5
0.0
ϕ(z)
−6
−4
−2
0
2
4
6
z




```
(1797, 64)
Author @jithin
<function accuracy_score at 0x00000214B7D2C400>
0.8974042027194067
Predicted :  [1]
Actual :  1
[[4.75045461e-18 9.99447460e-01 7.00809699e-10 3.72475330e-09
  2.15616661e-06 1.35167550e-09 5.71303497e-10 1.95595337e-13
  5.50377100e-04 5.64607392e-10]]
Class labels: [0 1 2 3 4 5 6 7 8 9]
Labels counts in y: [178 182 177 183 181 182 181 179 174 180]
Labels counts in y_train: [124 127 124 128 127 127 127 125 122 126]
Labels counts in y_test: [54 55 53 55 54 55 54 54 52 54]
Elapsed Time (s): 55.92375826835632
```

SVM AND LOGISTIC REGRESSION(SONAR DATASET):

SVM Accuracy: 86.74%
LR Accuracy: 71.42%
Total elapsed time: 5.17 seconds

```
[8 rows x 60 columns]
60
M    111
R     97
dtype: int64
Elapsed Time for LR (s): 18.893187522888184
LR : 0.782721 (0.093796)
SVM : 0.608824 (0.118656)
ScalerLR : 0.734191 (0.095885)
ScalerSVM : 0.836397 (0.088697)
:0.8674698795180723 {'C': 1.5, 'kernel': 'rbf'}
             precision    recall  f1-score   support

          M       0.92      0.85      0.88        27
          R       0.76      0.87      0.81        15

avg / total       0.86      0.86      0.86        42

Elapsed Time for SVM (s): 5.171906471252441
Accuracy for LR
0.7142857142857143
[[19  8]
 [ 4 11]]
             precision    recall  f1-score   support

          M       0.83      0.70      0.76        27
          R       0.58      0.73      0.65        15

avg / total       0.74      0.71      0.72        42
```

Final Analysis Of Classifiers:
The various classifiers have been implemented using the digits dataset and the sonar dataset and their respective accuracy and the elapsed time during the execution have been found

ACCURACY:

DIGITS DATASET:

Perceptron Accuracy: 96.87%
Decision Tree Accuracy: 97.61%
KNN Accuracy: 98.33%
Logistic Regression Accuracy: 89.74%
SVM Accuracy:
Linear: 97.12%
RBF: 21.03%

SONAR DATASET:

Perceptron Accuracy: 71.04%
Decision Tree Accuracy: 60.31%
KNN Accuracy: 85.02%
Logistic Regression Accuracy: 71.42%
SVM Accuracy: 86.74%

ELAPSED TIME(seconds):

DIGITS DATASET:

Perceptron Elapsed Time: 9.36
Decision Tree Elapsed Time: 6.15
KNN Elapsed Time: 145
Logistic Regression Elapsed Time: 55.93
SVM Elapsed Time: 13.96

SONAR DATASET:

Perceptron Elapsed Time: 5.89
Decision Tree Elapsed Time: 0.1
KNN Elapsed Time: 1.29
Logistic Regression Elapsed Time: 5.17
SVM Elapsed Time: 5.17