CSE574 Introduction to Machine Learning
Programming Assignment 3
Classification and Regression

Team Members

1. Kalyani Sunil Chikte (50247008)
2. Vijaya Harshavardhan Palla (50246589)

Table of Content

Implementation of Logistic Regression

```
[[4821    1    8    6    8   15   23    6   33    2]
 [   1 5623   27   13    3   16    2   10   41    6]
 [  34   40 4511   67   51   18   52   61  108   16]
 [  16   22  122 4614    7  133   19   43  103   52]
 [   9   20   22    6 4543   10   26   13   46  147]
 [  46   17   33  137   39 3893   83   18  106   49]
 [  24   13   28    3   23   67 4735    2   21    2]
 [  11   20   49   12   46   11    4 4962   12  138]
 [  38  107   54  124   25  125   33   20 4238   87]
 [  27   20   13   84  158   34    1  151   48 4413]]
```

Matrix 1: Confusion Matrix depicting total error w.r.t each category in the training data

```
[[975   0   1   2   0   8   5   1   6   2]
 [  0 968   4   1   2   9   0   1  13   2]
 [ 12  15 878  23  13   4  11  12  26   6]
 [  4  11  28 890   3  23   4  11  13  13]
 [  1   6   7   2 939   0   7   0   7  31]
 [  8   8   8  43  19 867  17   2  19   9]
 [  7   4   6   0   5  12 957   2   7   0]
 [  3   5   9   1  14   2   0 924   4  38]
 [ 15  26  21  27   9  25  20   4 842  11]
 [ 10   3   5  20  24   4   1  26   3 904]]
```

Matrix 2: Confusion Matrix depicting total error w.r.t each category in the validation data

```
[[ 961    0    1    3    1    5    5    2    1    1]
 [   0 1115    3    1    0    1    4    1   10    0]
 [   8   11  916   20   11    4   12   13   33    4]
 [   4    0   19  923    2   19    4   13   17    9]
 [   1    2    5    3  918    0    9    2    4   38]
 [  11    2    1   47   11  759   16    7   30    8]
 [   9    4    7    2    4   19  908    1    4    0]
 [   2   11   23    5    7    2    1  948    2   27]
 [  13   14    8   21   15   28    8   11  845   11]
 [   8    8    1   12   33   12    1   22   12  900]]
```

Matrix 3: Confusion Matrix depicting total error w.r.t each category in the testing data

Sr No	Model	Training set accuracy (%)	Validation set accuracy (%)	Testing set accuracy (%)
1	Logistic Regression (class=10, iteration=100)	92.706	91.44	91.93
2	Multi Class Logistic Regression (class=10, iteration=100)	93.448	92.48	92.55

Table 1

Support Vector Machines

Sr No	Parameters	Training set accuracy (%)	Validation set accuracy (%)	Testing set accuracy (%)	Train Time
1	Kernel: Linear Others: default	97.286	93.64	93.78	249267276
2	Kernel: RBF Gamma: 1 Others: default	100	15.48	17.14	10799971386
3	Kernel: RBF Gamma: default Others: default	94.294	94.02	94.42	429766639

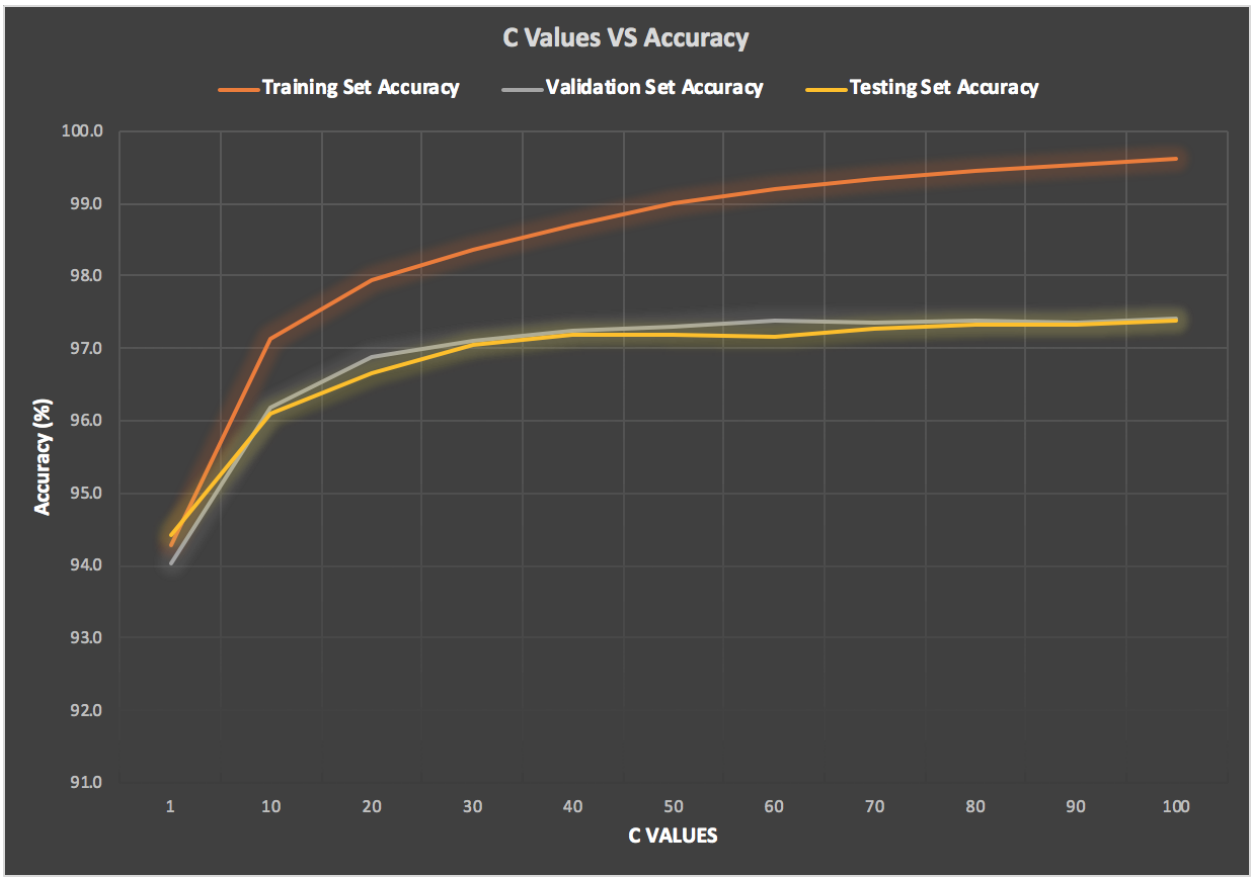
Table2: Accuracy for SVM

Conclusion:

- When gamma is equal to 1, accuracy for validation set and testing set is minimum as model tries to overfit.
- The best performance can be obtained when gamma is default.
- Default value of gamma is 1/(number of features) i.e 1/716 (for this data set), which is much less than 1.
- Hence, accuracy when gamma=1, is lesser than that when gamma is default, even if training accuracy is 100%
- From table 1 and 2 we can see that SVM performs better than logistic regression on any data set.
- Logistic regression learns the line (plane) which separates data whereas, SVM learns the best line (plane) one with maximum margin.
- Use linear kernel when number of features is larger than number of observations. [1]
- Use gaussian kernel when number of observations is larger than number of features. [1]

Sl.No	Parameters	Training set accuracy (%)	Validation set accuracy (%)	Testing set accuracy (%)	Train Time
1	Kernel: RBF C: 1 Others: default	94.294	94.02	94.42	428910042
2	Kernel: RBF C: 10 Others: default	97.132	96.179	96.1	227241428
3	Kernel: RBF C: 20 Others: default	97.952	96.89	96.67	201808526
4	Kernel: RBF C: 30 Others: default	98.372	97.1	97.04	193462859
5	Kernel: RBF C: 40 Others: default	98.706	97.23	97.19	191571823
6	Kernel: RBF C: 50 Others: default	99.002	97.31	97.19	190368030
7	Kernel: RBF C: 60 Others: default	99.196	97.38	97.16	189390584
8	Kernel: RBF C: 70 Others: default	99.339	97.36	97.26	189503135
9	Kernel: RBF C: 80 Others: default	99.438	97.39	97.33	189547055
10	Kernel: RBF C: 90 Others: default	99.542	97.36	97.34	189135037
11	Kernel: RBF C: 100 Others: default	99.612	97.41	97.39	194664599

Table 3: Train, Validation, Test set accuracy for different C values



Graph: C values versus Train, Validation, Test set accuracy

C Parameter:

- C Penalty parameter the error term.
- C tried to balance between maximum margin and wrong classification.
- As a result as C value increases, accuracy for train, validation and test set increases.
- However, there is risk of overfitting with high C values but it cannot be seen in the above graph.

Extra Credit: Multi Class Logistic Regression

```
[[4786    1   12    7   11   33   30    7   32    4]
 [   1 5592   26   17    6   19    2   13   58    8]
 [  23   45 4503   72   58   24   59   53  108   13]
 [  14   18   95 4654    4  148   15   39  105   39]
 [   8   20   21    7 4576    6   42   13   24  125]
 [  39   13   36  117   34 3963   68   18  102   31]
 [  23   11   29    1   24   52 4758    2   16    2]
 [   8   16   49   18   34    9    4 4989   14  124]
 [  22   75   51  103   16  113   23   16 4387   45]
 [  17   18    9   55  126   30    2  134   42 4516]]
```

Matrix 4: Confusion Matrix depicting total error w.r.t each category for the training data

[[975 0 1 3 2 7 3 2 6 1]
[0 972 3 2 1 5 0 2 13 2]
[10 13 896 22 13 4 11 9 18 4]
[1 7 23 902 3 28 2 12 13 9]
[1 4 8 3 941 1 10 2 7 23]
[9 4 6 37 17 884 14 2 22 5]
[9 2 4 1 7 12 957 1 6 1]
[2 3 9 0 9 1 0 931 3 42]
[13 17 19 27 9 20 19 2 868 6]
[4 3 5 14 19 4 1 24 4 922]]

Matrix 5: Confusion Matrix depicting total error w.r.t each category for the validation data

[[960 0 0 3 0 6 6 4 1 0]
[0 1110 3 2 0 2 4 2 12 0]
[6 8 924 16 10 3 14 8 39 4]
[4 1 20 914 0 25 3 10 26 7]
[1 1 6 2 921 0 9 4 9 29]
[10 2 2 37 10 773 15 6 30 7]
[9 3 4 2 7 15 914 3 1 0]
[1 9 19 6 6 2 0 952 2 31]
[9 8 6 26 9 23 10 8 868 7]
[11 8 0 10 28 5 0 20 8 919]]

Matrix 6: Confusion Matrix depicting total error w.r.t each category for the training data

References:

[1]https://stats.stackexchange.com/questions/73032/linear-kernel-and-non-linear-kernel-for-support-vector-machine