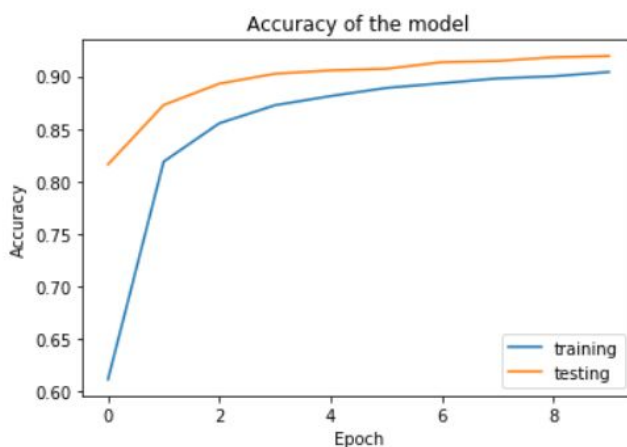


CSE 574: Intro to Machine Learning
Project 3
Rohan Hemanshu Sheth
50291746

In this project, we solve the classification problem of recognizing the digits from 0 to 9. MNIST dataset is used for training and testing. These digits are in the form of 28x28 handwritten images each having 784 features. USPS dataset is used to test the models. Four algorithms namely, multiclass logistic regression, multi-layer perceptron, Support Vector Machine and Random Forest and compare their performance on the dataset. Majority voting is used for ensemble classification. Ensemble classification method used in machine learning uses multiple classification models which are combined to obtain the best performance. Majority voting method is implemented where we choose the class that is predicted by majority of the models.

Neural Network using the given code in Project3_methods.pdf

Testing Accuracy for Mnist : 0.9034



USPS Accuracy: 0.312165608280414

Multi-layer Perceptron:

Confusion matrix for MNIST:

```
[[ 957  0  3  1  0  6  4  5  3  0]
 [  0 1120  1  4  0  2  3  1  4  0]
```

```
[ 9 2 989 7 4 1 6 6 6 2]
[ 1 0 15 962 1 17 0 8 2 4]
[ 4 1 5 1 930 1 10 2 3 25]
[ 6 2 3 31 5 824 5 1 9 6]
[ 12 2 4 0 3 7 923 1 6 0]
[ 3 4 19 9 2 2 0 975 0 14]
[ 5 5 8 12 6 13 8 6 904 7]
[ 5 3 2 11 25 7 2 12 4 938]]
```

Accuracy=0.9522952295229523

USPS:

0.29816490824541225

Confusion matrix:

```
[[ 267  0 115 118  43 414  10 967  3 63]
 [  7 118 452 209  4 283  1 917  8  1]
 [ 12 11 1283 228  6 315  8 136  0  0]
 [  4  2 140 1340  0 387  0 124  1  2]
 [ 18 33  86  55 371 290  8 1082  7 50]
 [ 20 13 172 399  4 1220  5 164  1  2]
 [ 61  7 903 126 32 483 238 146  0  4]
 [  5 19 107 572  1 222  0 1071  0  3]
 [ 19 15 293 341  8 894 20 395 13  2]
 [  4 43  84 428 10 106  0 1266 17 42]]
```

AdaboostClassifier Accuracy:

MNIST: 0.9572957295729573

USPS: 0.1464573228661433

Random Forest Classifier:

MNIST:

0.9695969596959696

Confusion matrix:

```
[[ 969  0  0  0  0  2  4  1  2  1]
 [  0 1124  1  4  0  1  2  0  2  1]
 [  7  0 999  4  3  0  4  9  6  0]
 [  0  0  9 973  0  8  0  9  8  3]
 [  1  0  2  0 959  0  5  0  2 13]
 [  5  0  0 16  2 854  6  2  5  2]
```

```
[ 5  3  0  0  5  3 938  0  4  0]
[ 1  4 20  3  1  0  0 986  2 11]
[ 2  0  6  9  5  7  3  3 932  7]
[ 8  5  2 10 10  2  1  4  6 961]]
```

USPS:

0.11605580279013951

Confusion matrix:

```
[[ 0  81  1  0 16 196  0 1706  0  0]
 [ 0 500  0  1  0  35  0 1464  0  0]
 [ 0 305  2  1  4 163  0 1524  0  0]
 [ 0  62  1  5  5 285  0 1642  0  0]
 [ 0 135  0  0 11  55  0 1799  0  0]
 [ 0 220  0  2  2 483  0 1293  0  0]
 [ 0 124  0  1  5 361  0 1509  0  0]
 [ 0 602  0  0  0  78  0 1320  0  0]
 [ 0 195  1  3  5 669  0 1127  0  0]
 [ 0 299  0  0  5  89  0 1607  0  0]]
```

Support Vector Classifier with kernel='rbf', C=2, gamma = 0.05

Mnist Test accuracy: 0.9827

USPS accuracy: 0.2614130706535327

Confusion matrix MNIST:

```
[[ 974  0  1  0  0  1  1  1  2  0]
 [  0 1128  3  1  0  1  0  1  1  0]
 [  4  0 1015  1  1  0  0  6  5  0]
 [  0  0  1 997  0  3  0  5  4  0]
 [  0  1  3  0 964  0  4  0  2  8]
 [  2  0  1  7  1 872  3  1  4  1]
 [  5  2  0  0  2  3 945  0  1  0]
 [  0  3  9  1  1  0  0 1004  2  8]
 [  2  0  1  6  1  2  0  2 958  2]
 [  4  4  2  8  7  2  0  6  6 970]]
```

Confusion matrix USPS:

```
[[ 226  0 1564  2 26 35  2  0 79 66]
 [ 78 257 713 172 262 77 12 337 88  4]
 [  8  0 1944  6  2 20  1  6 11  1]
 [  4  0 1193 725  0 41  0  0 37  0]
 [  6  0 1045 18 522 96  0 56 252  5]
 [ 15  0 1305 16  1 626  0  0 37  0]
 [ 78  0 1534  2 10 61 290  0 22  3]
 [ 17  6 1435 129  6 134  0 220 52  1]
 [  7  0 1387 14  4 221  0  0 367  0]
 [  1  0 1508 79 26 29  0 39 267 51]]
```

Support Vector Classifier with kernel='linear', gamma= default:

Mnist Test accuracy: 0.9154

USPS accuracy: 0.2605630281514076

Confusion matrix MNIST:

```
[[ 958  0  2  1  0  7  8  2  2  0]
 [  0 1113  4  1  0  1  4  1 11  0]
 [ 10  9 912 22 11  4 12 10 39  3]
 [  5  2 20 915  3 20  5 12 19  9]
 [  1  3  5  3 914  0 10  3  6 37]
 [ 10  3  1 38 11 761 20  8 31  9]
 [  9  4  6  2  6 20 908  1  2  0]
 [  2  9 21  6  6  1  1 945  4 33]
 [  8 13  8 23 13 39  8 15 834 13]
 [  6  8  2 15 35 10  0 26 13 894]]
```

Confusion matrix USPS:

```
[[ 313  1 381 311  59 174 115 486  59 101]
 [ 44 279 664 159 354 99 29 279 72 21]
 [ 70 46 1286 100 47 171 149 83 21 26]
 [ 49 38 473 749 17 483 33 80 44 34]
 [ 53 54 189 117 594 179 73 550 133 58]
 [ 42 24 839 218 23 645 80 94 27  8]
 [ 98 12 732 114 55 332 510 79 15 53]
 [127 87 231 534 115 150 25 607 85 39]
 [175 31 148 658 123 398 113 199 115 40]
 [ 40 51 156 550 114 75 16 709 176 113]]
```

Support Vector Classifier with kernel='rbf', gamma= default:

Mnist Test accuracy: 0.9435

USPS accuracy: 0.38541927096354817

Confusion matrix MNIST:

```
[[ 967  0  1  0  0  5  4  1  2  0]
 [  0 1120  2  3  0  1  3  1  5  0]
 [  9  1 962  7 10  1 13 11 16  2]
 [  1  1 14 950  1 17  1 10 11  4]
 [  1  1  7  0 937  0  7  2  2 25]
 [  7  4  5 33  7 808 11  2 10  5]
 [ 10  3  4  1  5 10 924  0  1  0]
 [  2 13 22  5  7  1  0 954  4 20]
 [  4  6  6 14  8 24 10  8 891  3]
 [ 10  6  0 12 33  5  1 14  6 922]]
```

Confusion matrix USPS:

```
[[ 573  2 428 19 285 248 73 44  6 322]
 [ 110 429 285 137 273 180 46 501 22 17]
 [ 128 18 1402 59 39 198 61 57 23 14]
 [ 76  3 186 1123 11 483  5 70 27 16]
 [ 18 67 91 14 1167 267 22 194 69 91]
 [ 108 17 257 102 25 1367 60 43 15  6]
 [ 197  7 489 24 98 394 748 13  7 23]
 [ 50 225 457 265 57 416 15 452 41 22]
 [ 73 25 209 193 87 1006 95 41 244 27]
 [ 26 166 228 278 213 165  8 499 214 203]]
```

Ensemble Classification:

VotingClassifier with SVM (kernel=linear), Multi-layer Perceptron and Random Forest Classifier

MNIST: 0.976

USPS: 0.3074536531

Logistic Regression (Multiclass Logistic Regression with softmax layer) :

MNIST:

Training Accuracy: 85.874

Validation Accuracy: 87.56

Test Accuracy: 87.05

USPS Test Accuracy: 33.621

The free lunch theorem states that any two optimization algorithms are equivalent when their performance is averaged across all possible problems. In terms of machine learning, a highly optimized model might not work well for any set of new data. This holds true for the above implementations. As can be observed, MNIST accuracies are far higher than those of USPS accuracies. The confusion matrix is better for MNIST than USPS for all the above implementations because of the high difference between diagonal and non-diagonal elements where diagonal elements have larger values compared to non-diagonal elements.