Classification on Mnist and USPS Dataset

Kiran Prabhakar - kprabhak(50287403)

November 22, 2018

Abstract

The purpose of this project is to implement the various machine learning approaches for the classification task. The classifier combination method, using majority voting, is used to combine the classifiers to get better performance.

1 Introduction

The classification task will be to recognize 28X28 hand written image and identify it as the digit among 0,1,2,...9. In the current project the following approaches are followed to perform classification.

1. Softmax Regression.
2. Support Vector Machine.
3. Neural Networks.
4. Random Forest.
5. Majority Voting.

2 Definitions

2.1 Softmax Regression

Softmax Regression is a generalization of logistic regression that we can use for multi-class classification, under the assumption that the classes are mutually exclusive. The softmax regression replaces the sigmoid activation function with softmax function to calculate the probabilities of outputs. The equation for finding the probability is as below.

$$P(y_k(x)|z) = \frac{exp(z_k)}{\sum_m exp(z_m)}$$

Where $z = W^T X$, z_k is k^{th} class of the x^{th} data point.

2.2 Support Vector Machine (SVM)

A logistic regression classifier cannot classify the points in using a linear decision boundary if the data points are skewed. SVM gives an approach to raise the dimensions an get the data points in the space where the points are linearly separable. SVM provides a larger margin for the classes and hence it is called as Large margin classifier.

In the current project we use the methods present in sklearns library to implement SVM, where it takes two parameters C and γ.

2.3 Kernal

The hyper plane that acts as the decision boundary is decided by transforming the problem using the linear algebra function which is called Kernal or Kernal function. The technique of separating the points using a kernal is called kernal trick. The popular kernal method used are

1. Linear Kernal.
2. Polynomial Kernal.
3. Radial basis kernal.

2.4 Majority Voting

In case of majority voting classifier the predictions of different classifiers taken and the majority predictions is considered. The majority voted prediction is the output of the classifier.The technique of using the several models to reliably attain better result on the data set is called Ensemble.

2.5 Bagging

Bagging is a technique where a partition of data around 60% is taken with replacement for training the model. In the current project for each classifier we use a different bag of data and then combine them for max voting classifier.

2.6 Confusion Matrix

The confusion matrix is a table that is used to visualize the performance of the classification algorithm. The confusion matrix is drawn against the model predicted values and actual values. In the current project, as we have 10 classes, we use a 10X10 matrix to represent the confusion matrix. The diagonal elements in the confusion matrix shows the number of correctly classified samples.

3 Data Pre-processing

The classification task is performed using two data sets Mnist and USPS. The Mnist data set is used for training and for testing the test set of Mnist data and entire USPS data set is used. To make the data sets compatible with each other, we per-process the USPS data set to produce 28X28 image.

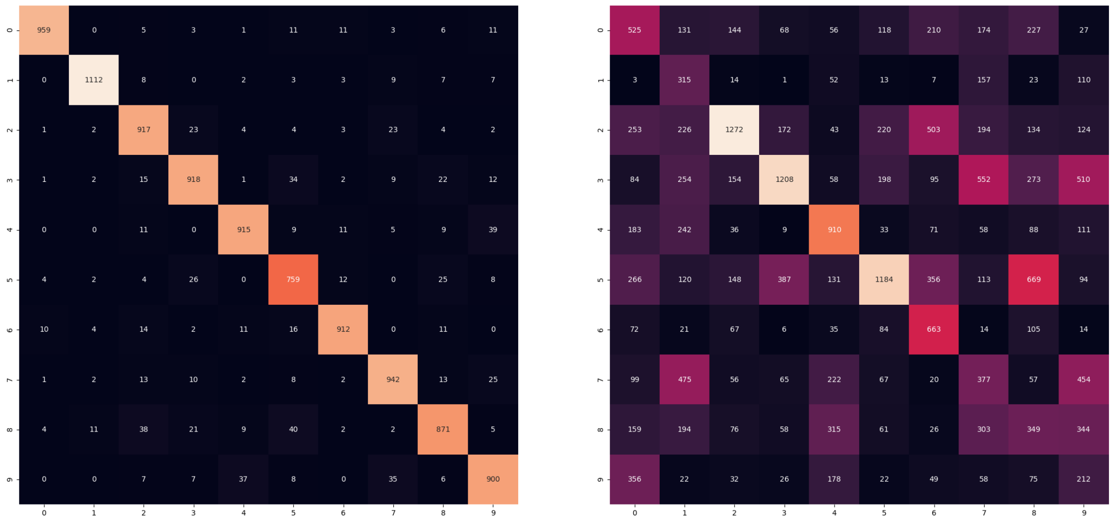
For majority voting, we obtain different bags of data by randomly shuffling and taking the 60% of data from the combined training data and target.

4 Softmax regression

The objective of the softmax regression is to determine the weight matrix for the classifier. As there are 784 features and 10 different classes, the softmax classifier produces 784X10 weight matrix. In the current project, below are the hyper parameters tuned to perform the classification.

epochs	Learning Rate	Mnist Accuracy	USPS Accuracy
1000	1	91.05	34.54
1000	0.5	92.05	35.07

The confusion matrix for the classifier is as follows



(a) Mnist Testing

(b) USPS testing

Figure 1: Softmax Regression Confusion Matrix

For Softmax regression the mini batch gradient descent technique is used to find the weights, where at each batch the weights are updated by finding the derivative of the cost function. This is carried out for the specified number of epochs. The below mentioned equation is used to update the wights.

$$w_j^{t+1} = w_j^t - \eta \bigtriangledown_{wj} E(X)$$

To improve the speed of the computation, vectorized input and weight matrices are taken in to consideration.

Also at each batch, the softmax output is calculated, as mentioned in the section 2.1, to predict the class

5 Support Vector Machine (SVM)

To implement the SVM, we use the standard library function SVC (Support Vector Classifier) to calculate the model for the training data. The library method accepts two parameters regularizer (C) and γ.

The smaller value of C will cause the optimizer to look for larger margin by paying the cost of mis-classifying few points, which is also a positive case to avoid variance.

The smaller the value of γ the margin is derived by considering the points which are far away from the plausible margin.

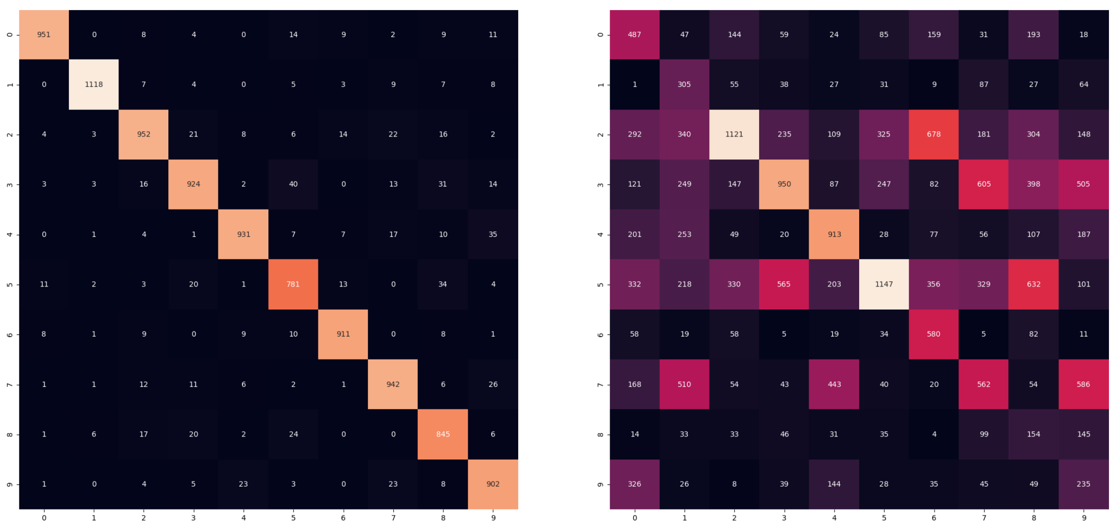
Two SVM classifiers using different kernals, which is helpful later in ensemble classifier.

In the current project, we use the below mentioned hyper parameters on the data set and the results are obtained as mentioned

Kernal	γ	Mnist Accuracy	USPS Accuracy
Linear	0.01	92.57	32.27
rbf	1	18.13	10.01
rbf	default	93.05	38.51
rbf	0.01	97.82	48.12

From the above results it is evident that for the $\gamma = 1$ the model performs very poorly and so a lower value of γ should be considered.

The confusion matrices for the last two setting in the above table is as follows

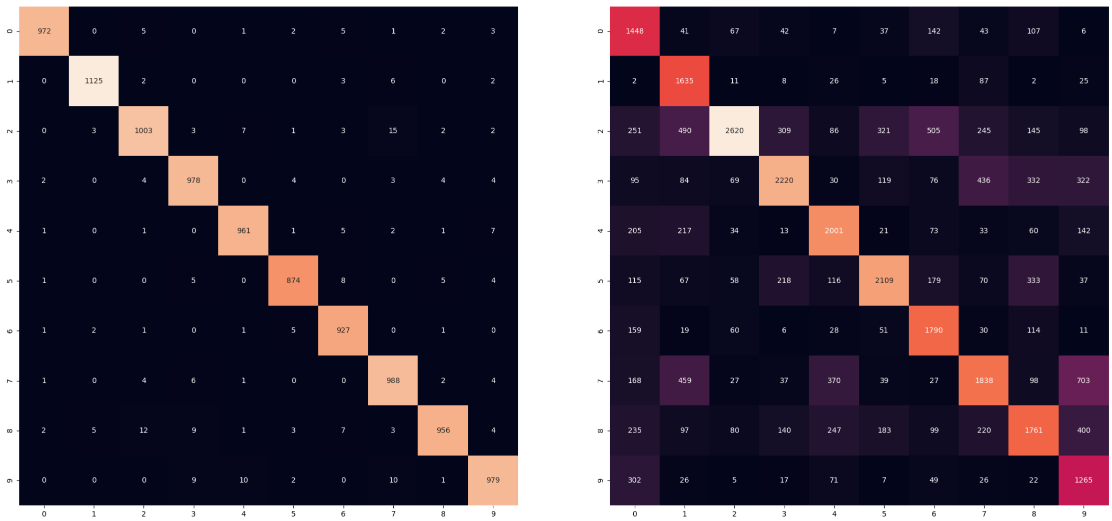


(a) Mnist Testing

(b) USPS testing

Figure 2: SVM with Linear Kernal Confusion Matrix

The confusion matrix for RBF kernal is as below



(a) Mnist Testing

(b) USPS testing

Figure 3: SVM with RBF Kernal Confusion Matrix

5.1 Neural Networks

In the current project deep neural network is used to perform classification. The Sequential model form Keras library is used to obtain neural network model.

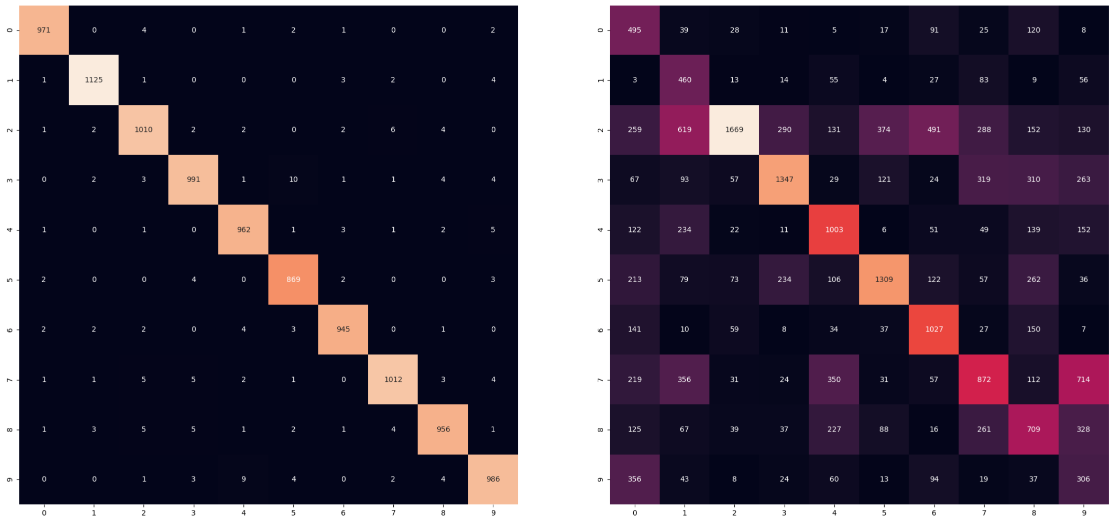
As the number of pixels for an image is equal to the features, we take 784X1 as the dimension for input layer. The output predicted by the model will be equal to the number of classes, which is 10X1.

The neural network is constructed on below setting and the accuracy is obtained as mentioned in the below table.

Input layer	Optimizer	Activation	Mnist Accuracy	USPS Accuracy
784	Adamax	Relu & Softmax	98.27	45.98

The testing targets are one hot encoded which is useful to compare the outputs predicted by the model. So the number of nodes in the output layer is 10, which corresponds to each class.

The confusion matrix for the above neural network classifier is as below



(a) Mnist Testing (b) USPS testing

Figure 4: Neural Network Confusion Matrix

5.2 Random Forest

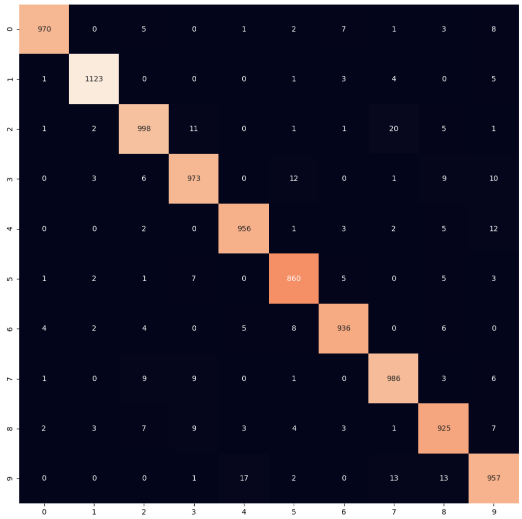
To implement Random forest classification **RandomForestClassifier** form sklearns.ensemble is used. To construct the classifier library accepts the parameter ”n_estimators” which is equivalent to number of trees in the random forest.

The below table shows the accuracy obtained for various values of n_estimators

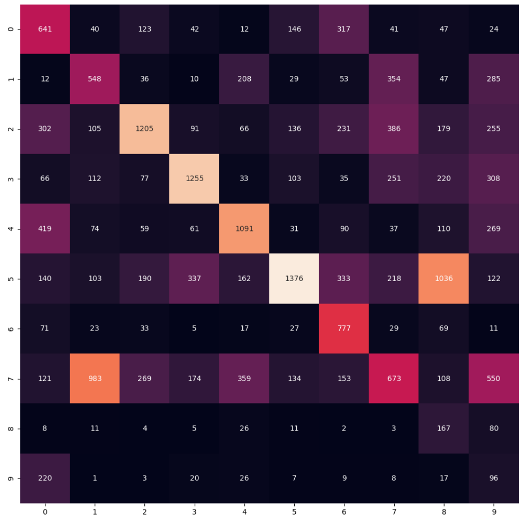
n_estimators	Mnist Accuracy	USPS Accuracy
10	94.71	31.38
100	96.84	39.14
1000	97.03	39.76

Thus from the above results it is evident that that the models accuracy doesn't improve much even after increasing the hyper parameter. Thus the optimal value for n_estimators is taken as 100.

The below confusion matrix is plotted for the both Mnist and Usps testing data.



(a) Mnist Testing



(b) USPS testing

Figure 5: Random forest Confusion Matrix

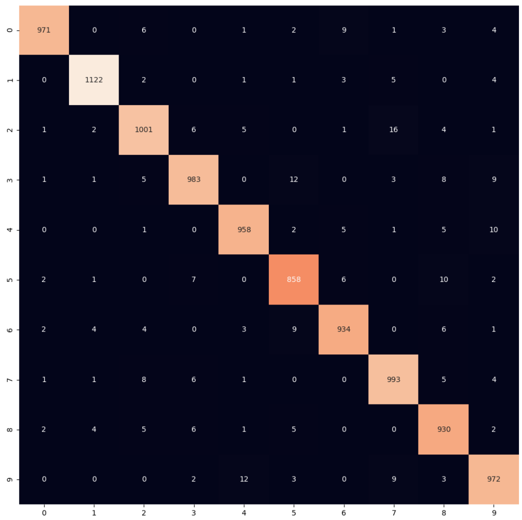
5.3 Majority Voting

To perform majority voting independent bagged data obtained from the same training data set is taken with replacement.

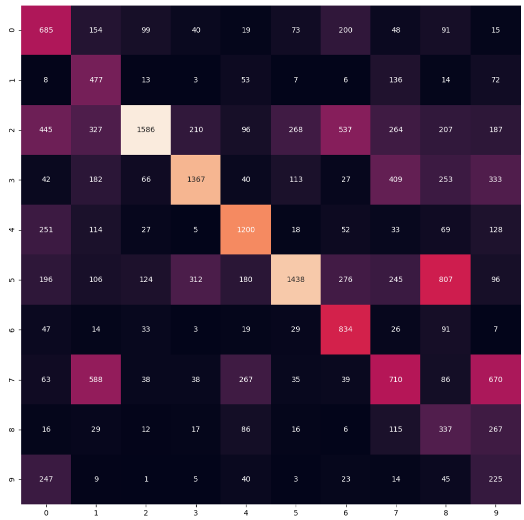
A different 60% obtained after randomly sampling the training data is used to train different models.The weights from Softmax and models form other classifiers is applied on testing. The majority voting is performed as described in the section 2.4.

The output of majority voting is 98.61% on Mnist testing data and 46.29% USPS test data, which is slightly better than Neural network classifier and comparatively better enough than other classifier's

The confusion matrix for the majority voting classifier is as below



(a) Mnist Testing



(b) USPS testing

Figure 6: Majority voting Confusion Matrix

6 Conclusion

Based on the four classifiers used in the project, below are the observations that can be noticed.

- The model that performs best on Mnist data set **doesn't** ensure that the model also performs best on the USPS data set. This behavior of different models **supports** "No Free Lunch Theorem"
- By examining the confusion matrices individual classifiers, it is evident that the performance of Dense Neural networks is better on both Mnist and USPS data.
- However the majority voting technique using the classifier combination yields higher accuracy on the testing data.

Thus by performing classification using various classifiers and comparing the results, the best classifier for a data set can be obtained.

References

[1] Project description slides

[2] https://towardsdatascience.com

[3] https://medium.com