Introduction to Machine Learning (Project 3)

Classification of MNIST and USPS Data set using logistic regression , MLP with 1 hidden Layer and Convolutional Neural Network .

Ashwin Vijayakumar (50249042)

Srivatsa Manjunath Hegde (50248870)

Introduction :

Our goal is to train different prediction models (Multiclass logistic regression , Multilayered perceptron with one hidden layer and a convolutional neural network) to predict the labels of samples from the MNIST and USPS data set .

Data Sets :

MNIST Data set :

We extract the MNIST data set using the command :

```
from tensorflow.examples.tutorials.mnist import input_data
mnist = input_data.read_data_sets("MNIST_data/", one_hot=True)
```

USPS data set :

We read the USPS data set from the local storage . We parse the folder structure to read image , downscale it to 28 * 28 resolution , normalize the values so that they fall under 0 and 1 .

a)Logistic Regression :

1)Hypeparameter tuning :

Validation Accuracy :

Sl Number	Epoch Size	Batch size	Learning Rate	Validation Accuracy ()
1	1000	100	0.5	92.18

2	500	200	0.9	92
3	5000	100	0.7	92.58
4	2000	50	0.05	91.12
5	2000	50	0.1	92
6	1000	50	0.9	91.28

From these values , we see that Epoch size = 5000 , Batch size = 100 , Learning rate = 0.7 produces a good validation accuracy .

2)Test Accuracy on MNIST data :

0.9175

3)Test Accuracy on USPS data :

0.2823

b)MLP with Single Hidden Layer :

1)Hypeparameter tuning :

Validation errors :

Sl Number	Nodes per hidden layer	Learning Rate(Adaptive learning)	Epochs	Batch size	Validation Accuracy(percentage)
1	10	0.00001	1000	100	98.22
2	10	0.0001	1000	100	99.03
3	20	0.001	2000	100	99.04
4	30	0.01	2000	50	97.02

The hyper parameter selection for single hidden layer neural network is N=20 , Learning rate = 0.001 , Epoch size = 2000 , Batch size = 100 .

2)Test Accuracy on MNIST data set:

0.9925

3)Test Accuracy on USPS data set :

0.5055

c)Convolutional Neural Network :

1)Test Accuracy on MNIST data :

0.9912

2)Test Accuracy on USPS

0.4254

Conclusion :

After It is evident from the test errors that the free lunch theorem holds in all the cases . This means that a model when trained to fit with a high accuracy on one data set , it may not have learned all the necessary details required in order to perform to the same level of accuracy on a completely new data set . In this case the MNIST(0.9912) data set will produce better test accuracies than the USPS data set (0.4254).

MLP with Single Hidden Layer (Backpropogation) :

The backpropagation algorithm is named for the way in which weights are trained , i.e , by backpropogation of the difference between expected output and actual output .

Error is calculated between the expected outputs and the outputs forward propagated from the network. These errors are then propagated backward through the network from the output layer to the hidden layer, assigning blame for the error and updating weights as they go.

The math for backpropagating error is rooted in calculus, but we will remain high level in this section and focus on what is calculated and how rather than why the calculations take this particular form.

Bayesian Logistic Regression :

The posterior distribution over the weights is given by Bayes' rule:

p(w | D) = P(D | w) p(w) P(D) ∝ P(D | w) p(w).

The normalizing constant is the integral required to make the posterior distribution integrate to one:
P(D) = Z P(D | w) p(w) dw.