Project-3 : Classification

CSE-574: Machine Learning

Sneha Krishnamurthy

UBID: snehakri

Person Number: 50208395

IMPLEMENTATION

Hyper Parameter Tuning

Hyper-parameters η is used to train the data for Logistic Regression and Convolution Network.
Hyper-parameters M and η are used to train the data for Single Layer Neural Networks.

M and η Optimization

The values of M which is the number of hidden layers in the neural network and η which is the learning rate are selected using the Grid Search method where a subset of possible values of M and η are taken and for each pair in the Cartesian product of the subsets, the weight vector is calculated on the training data-set and the **Accuracy** is calculated on the validation set. The pair with the Highest Accuracy value is finally selected to find the error for the Test data.

Logistic Regression

The MNIST dataset is trained using the Logistic regression model to get the weight W for the training set. This W is then used on the validation set to find the accuracy for the chosen learning rate η. The above iteration is run for different values of η and the value giving highest accuracy is fixed.
This value of η is then used to find the accuracy for the Test Data set.

The same value for η is used to find the accuracy for the USPS data-set which gives a low accuracy showing that without training on any data set results cannot be obtained proving the "no free lunch theorem".

Back Propagation Neural network

The MNIST data-set is trained using the Back propagation neural network model to get the weight W1 and W2 for the hidden layer and output layer respectively using the training set. Multiple epochs are run to find the optimal W1 and W2(approximate number of epochs: 60-80). These W1 and W2 are then used on the validation set to find the accuracy for the chosen learning rate η and Value of M. The above iteration is run

for different values in the Cartesian product of η and M, and the value giving highest accuracy is fixed.
This value of η and M is then used to find the accuracy for the Test Data set.

The same value for η and M are used to find the accuracy for the USPS data-set which gives a low accuracy showing that without training on any data set results cannot be obtained proving the "no free lunch theorem".

Convolution Neural Network

The Tensor flow package is used to train the Convolution network for MNIST data-set. The hyper parameter η is is tuned to get accuracy and the value giving the highest accuracy is fixed.

Hyper Parameter Chosen

- The number of nodes in the hidden layer, M are varied between 800 and 1000 as the hidden layer should contain more nodes than the input layer which has 784 nodes.
- η values are varied between $\{0,1\}$ at an interval of .002 and the value giving the highest accuracy for each model respectively is chosen.

PERFORMANCE MATRICES MNIST

Logistic Regression

The weight W was initialized as matrix of 0s of size 10*784
The bias vector b is initialized to 1s and is of size 10*1

ACCURACY

Validation Dataset: 91.4
Test Dataser: 90.7

η = .0068

Back Propagation Neural Network

The weights W1 and W2 are chosen as randomized values using np.random.randn function and are of size:
W1=784*M
W2=M*10
Bias vectors b1 and b2 are initialized to 1s and are of size:
b1=M*1
b2=10*1

ACCURACY

Validation Dataset: 97.4
Test Dataset: 97.3

η = .002
M = 1000

Convolution Neural Network

ACCURACY

Test Dataset: 99.26

$\eta = .0001$

PERFORMANCE MATRICES USPS

Logistic Regression

Test Dataser: 10.7

Back Propagation Neural Network

ACCURACY

Test Dataset: 17.3

Convolution Neural Network

ACCURACY

Test Dataset: 19.26