**PROJECT REPORT**

**Machine Learning - CSE 574**

**Handwritten digit recognition – Logistic Regression and Neural Networks**

Harshith Kumar Ramadev

50134007

**INTRODUCTION AND VARIANTS OF CLASSIFICATION:**

Hand written digit recognition is formulated by many machine learning algorithms like logistic regression, neural networks, Gaussian processes and so on. The field of deep learning is famous because it enables automatic feature extraction. Usually, in the field of machine learning the features that we extract are based on a specific knowledge of the data, i.e., the Deep learning extracts these features on its own. The machine learning algorithms named logistic regression and neural networks are implemented as follows:

**a) LOGISTIC REGRESSION:**

This algorithm uses a soft-max function in combination with a gradient descent algorithm in order to determine the globally optimal weights for which the corresponding Error value is minimal. This method processes a (19978,513) input matrix (including bias) and also considers a random weight vector of size (513,10) where 10 denotes the number of classes here. This is a multi-class logistic regression problem where the digits 0-9 are taken as each class.

Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables, which are usually (but not necessarily) continuous, by using probability scores as the predicted values of the dependent variable.

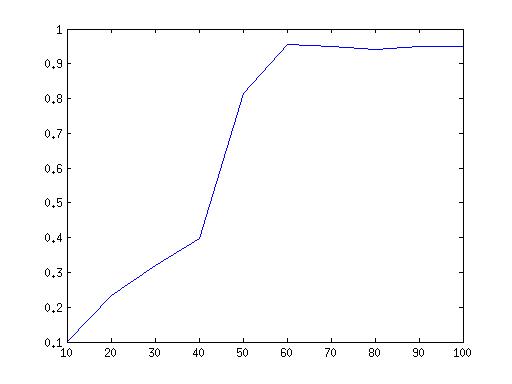
**EVALUATION:**

This classification algorithm is evaluated on the given dataset which contains the input feature space between 0-9. This (19978,513) vector is given as input along with a weight vector which is generated at random. This uses a gradient descent algorithm with a learning rate of “0.001” to iteratively minimize error during training and compute the optimum value of weight W. This value of W is used in future for prediction of incoming handwritten numerals.

This uses a 1 of K coding scheme to compute the target vector. For example, the target vector is assigned a value of 1 and nine zeros if the digit to be recognized to be zero and so on. This constitutes the formation of the target vector. The output is a (19978,10) matrix with values for each of the ten classes.

The absolute error is computed using the Cross entropy function and the Gradient of Error function, the derivative of the sigmoid is used to minimize the weights using the gradient descent approach. These weights are used to compute the accuracy and error rate with respect to the testing data.

The following graph plots between Accuracy and number of iterations:



**OBSERVATIONS:**

Results:

Accuracy - 97.07%

Error\_rate\_testing = 2.93%

**b) NEURAL NETWORKS:**

This classifier contains input, hidden and output layers. Weights are assigned for the transition from the input to hidden and hidden to output layer. These are denoted by W1 and W2.

The activation function from the input to the hidden layer and from the hidden layer to output is given by the “Sigmoid function”. The Sigmoid function is applied on the product of the input matrix and the weights W1 assigned at random. Also, the input bias of 1 is added to the input matrix. The error and other parameters computed indicate the number of neurons in the hidden layer.

The weights W1 and W2 are iteratively minimized using the gradient descent algorithm as in case of logistic regression and the output is calculated using the weight from the hidden to output layer W2. The output comes with an activation function and all the other processes associated with training the neural network. The network is trained with M=200, output nodes = 10 and run for 100 iterations in order to minimize error value.

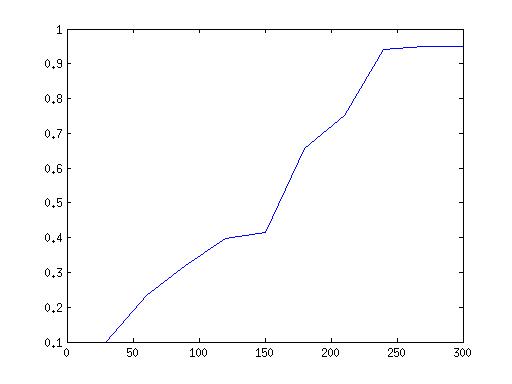
**OBSERVATIONS:**

Results:

1. Accuracy - 91.93%

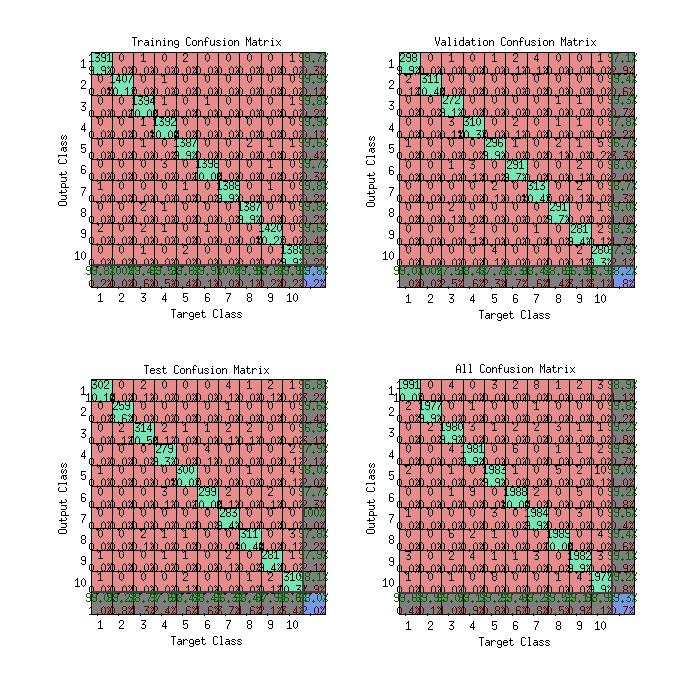
2. Error rate\_testing= 8.07%

Accuracy plotted against number of iterations:

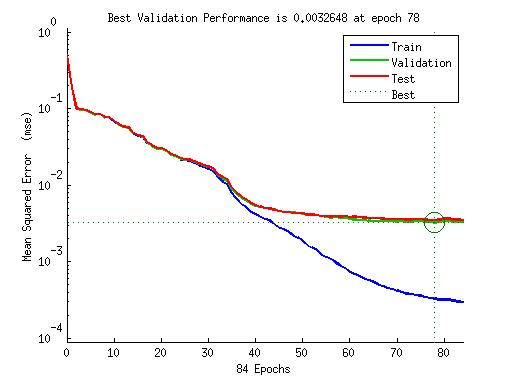


The Support Vector Machine package is also used to compute the Accuracy and Error for the data.

We used the SVM model as Neural Network, for which we used the “nnstart” command and pass the parameters as input matrix including the validation and testing data set. The below confusion graphs are obtained during the analysis:



Below is the illustration of Performance analysis when the iterations and mean square error is evaluated for Training, Testing and Validation data sets.



**COMPARITIVE ANALYSIS:**

The results are matched against as follows: SVM > Neural network > Logistic regression. SVM is a powerful classifier and gives a high accuracy. The NN model will give a high accuracy when the numbers of hidden layers are increased and the learning rate is decreased (kept below the threshold value) respectively.