CSE 574

Machine Learning

Prof: Dr. S Srihari

Project 3

Jruvika Bhimani

UBitName: jruvikam

UBitNumber: 50207613

Task:

1. Logistic Regression:

We initially calculate the activation function using:

$$a_k = \mathbf{w}_k^\top \mathbf{x} + b_k.$$

Once the activation function is calculated we calculate the output y value using:

$$p(\mathcal{C}_k|\mathbf{x}) = y_k(\mathbf{x}) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

This y value helps is knowing the error function for Stochastic Gradient Descent when compared with the expected Target value "t":

$$\nabla_{\mathbf{w}_j} E(\mathbf{x}) = (y_j - t_j)\,\mathbf{x}$$

Stochastic Gradient descent:

$$\mathbf{w}_j^{t+1} = \mathbf{w}_j^t - \eta \nabla_{\mathbf{w}_j} E(\mathbf{x})$$

we set eta value as 0.01, and start with weights = a vector of 0. after every iteration we update the weights.

These weights are then used in validating the data. we get an accuracy of 92.11.

The same weights are used on USPS data and we obtain an accuracy of 0.7

Thus, proving the "No Free Lunch Theorem" (https://en.wikipedia.org/wiki/No_free_lunch_theorem)

2. Neural Networks:

we start activation of input layer to hidden layer using the function of activation where z = h(x) i.e. z=q(q-1) and q =

$$\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + b_j^{(1)}\right)$$

the activation function for hidden to output layer is given by:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + b_k^{(2)}$$

These activated values are used to calculate the soft-max using:

$$y_k = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

Since I was getting overflow error I used ak = (ak-c) where c is max(vector-ak)

Once we get the activated values we need to reset the weights and minimize the error the error functions are given by:

$$\delta_k = y_k - t_k$$

$$\delta_j = h'(z_j) \sum_{k=1}^{K} w_{kj} \delta_k$$

where h-dash(z) is the sigmoid function.

$$\frac{\partial E}{\partial w_{ji}^{(1)}} = \delta_j x_i,$$

$$\frac{\partial E}{\partial w_{kj}^{(2)}} = \delta_k z_j$$

and the weights are updated using the Stochastic Gradient Descent:

$$\mathbf{w}^{t+1} = \mathbf{w}^{t} - \eta \nabla_{\mathbf{w}} E(\mathbf{x})$$

where, eta = 0.01 and weights are taken as random values.

3. Convolution Neural Network:

The CNN was implemented using the TensorFlow given on:
https://www.tensorflow.org/versions/r0.12/tutorials/mnist/pros/index.html#first-convolutional-layer

The accuracy = 99.24

4. USPS Data:

The images were read using the skim age library, converter to gray-scale and resized. Each resized image was vectorized and given to the logistic regression and neural network functions discussed above.

The weights were used same as that received from training of the MNIST data.

Since the accuracy received here is very low, we can prove the “No Free Lunch Theorem”.