1. Network Topology:

* For simplicity, I am using same number of neurons for all the HIDDEN Layers
* There are 784 neurons in the input layer
* I am using only one output neuron
* I am using sigmoid activation function for all the neurons except the output neuron

1. I am using a learning rate of 100
2. I standardized the data.
3. **Activation Function:**

# Activation Function

def af(t):

return tanh(t)

1. **Derivative of tanh(v)**

def af\_derivative(x):

return 1-(tanh(x)\*\*2)

1. **mse Function:**

This isused to calculate the mean squared error corresponding to the given weights.

Algorithm:

Mse(data, weights):

C = 0

N = no of inputs

For each input i in data:

Previous = i

For each layer “L”:

initialize an empty list “u”

initialize an empty list “p”

for each node “n” in “L”:

t = weights[L]['weights'][n] X Previous

append t to the list “u”

for each point “j” in the list “u”:

append af(j) to the list “p”

append 1 to the list “p”

Previous = p

y = weights[nLayers]['weights'][0] X Previous

C = C + (d[i] - y)^2

return c/N

1. I created a **NeuralNetworks** class which has feedforward, backpropagation and train functions.
2. The NeuralNetworks class takes the data, labels, number of hidden layers, number of nodes per hidden layer, learning rate, and maximum iterations as input parameters.
3. **Function call:**

nLayers = 1

nNodes = 24

nOut = 1

eta = 10

iter = 100000

ep = 0.2

w, obj, epoch = NeuralNetwork(x, d, nLayers, nNodes, nOut, eta, iter, ep).train()

1. I used a list with name “**weights**” to store weights and gradient values. I am uniformly choosing weights between -1 and 1.
2. I used dictionaries of python with keys for each layer: “weights”, “s”, and “g” inside the list. The key “s” has the values before the weights of the layer in the feedback graph. The key “g” has the gradient descent values for all the weights in a layer.

weights = [{"weights":np.random.uniform(low =-4, high = 4, size = (self.nNodes, len(self.data[0])))}]

for i in range(self.nLayers):

self.weights.append({"weights":np.random.uniform(low =-4, high = 4, size = (self.nNodes,self.nNodes+1))}) # 1 is for Biases

if self.nLayers >= 0:

self.weights.append({"weights":np.random.uniform(low =-4, high = 4, size = (self.numOutputs,self.nNodes+1))}) # 1 is for Biases

Ex: For number of hidden layers = 1 and num of nodes per hidden layer = 2.

[{'weights': array([[0.78888993, 0.36114814, 0.84949542],

[0.05266805, 0.89136156, 0.4386164 ]]),

's': [-0.003484950437606789, -0.008284032450920822],

'g': [[0.0017424752188033945, 0.0017424752188033945, 0.0017424752188033945],

[0.004142016225460411, 0.004142016225460411, 0.004142016225460411]]},

{'weights': array([[0.22397066, 0.34915997, 0.12250444]]),

's': [-0.14772231324694815],

'g': [[0.06505347805517415, 0.05904723432620448, 0.07386115662347408]]}]

1. **Train function:**

* This function uses the gradient descent vector from the backpropagation function to update the weights every epoch until maximum epochs are reached.
* If after an epoch, the mse is more than that of the mse of previous epoch, I am reducing the learning rate to 0.9 \* learning rate
* If learning rate falls below 0.0001 then I am running the algorithm with new weights and original learning rate

Algorithm:

Train(self):

temp\_eta = self.temp\_eta # Temporarily storing the initial learning rate

temp2 = mse(data, labels, weights, nLayers, nNodes) # MSE with the initial weights

print(temp2)

e = 0

initialize an empty list obj # List to store MSE after each epoch

initialize an empty list epoch # List to store epochs

cos = 100000000

while cos >= ep and e <= maxIt:

prev = cos

for i in range(len(data)):

self.backprop(labels[i], data[i]) # Call for backpropagation function

for m in range(len(weights)):

for j in range(len(weights[m]['weights'])):

for k in range(len(weights[m]['weights'][j])):

weights[m]['weights'][j][k] -= self.eta \* weights[m]['g'][j][k] #Updation of weights

cos = mse(data, self.labels, weights, nLayers,nNodes)

if cos > prev:

self.eta = 0.9\*self.eta # Reducing the learning rate

initialize an empty list obj

initialize an empty list epoch

e = 0

cos = 100000000

if self.eta <= 0.00001: # Starting from the beginning with different weights

self.eta = temp\_eta

weights = [{"weights":np.random.rand(nNodes, len(data[0]))}]

for i in range(nLayers-1):

weights.append({"weights":np.random.rand(nNodes,nNodes+1)})

if nLayers >= 0:

weights.append({"weights":np.random.rand(numOutputs,nNodes+1)})

elif cos <= prev:

epoch.append(e)

obj.append(cos)

e += 1

return weights, obj, epoch

1. **Feedforward function:**

For the given input, this records the local field values and values after the activation function of every neuron in each layer in the feedforward graph.

Algorithm:

feedforward(self,x=[]):

prev = x

initialize an empty list “r”

append “prev” to the list “r”

initialize an empty list “t”

for j in range(numLayers):

initialize an empty list “l”

initialize an empty list “s”

for m in range(numNodes):

append (weights[j]["weights"][m]) X prev to the list “s”

for k in s:

append af(k) to the list “l”

append 1 to the list “l” # This is for the bias of the neurons except the ones in the first layer

prev = np.asarray(l)

append “s” to the list “t”

append “l” to the list “r”

initialize an empty list s

p = (weights[numLayers]["weights"][0]) X l

append “p” to the list “s”

append “s” to the list “t”

return t, r, p

1. **Backpropagation function:**

For the given input, this records the values before the weights of all the layers in the backward (feedback) graph and the corresponding gradient descent vector.

Algorithm:

Backpropagation (labels, data):

initialize an empty list “der”

t,r,q = self.feedforward(data) # Call for the feedforward function

for i in range(numLayers):

initialize an empty list “a”

for m in range(numNodes):

append af\_derivative(t[i][m]) to the list “a”

append “a” to the list “der”

initialize an empty list “diff”

initialize an empty list “s”

append af\_derivative(t[nLayers][0]) to the list “s”

append d-q to the list “diff”

append “s” to the list “der”

for i in reversed(range(len(weights))):

layer = weights[i]

initialize an empty list “errors”

if i != len(weights)-1:

for j in range(len(layer['weights'])):

error = 0

for k in range(len(weights[i + 1]['weights'])):

error += (weights[i + 1]['weights'][k][j] \* weights[i + 1]['s'][k])

append error to the list “errors”

else:

append “diff[0]” to the list “errors”

initialize an empty list “layer['s']”

for j in range(len(layer['weights'])):

append errors[j]\*der[i][j] to the list “layer['s']”

# Finding the corresponding gradient vector

for j in range(len(weights)):

layer = weights[j]

initialize an empty list layer['g']

for k in range(len(weights[j]['weights'])):

initialize an empty list “s”

for m in range(len(weights[j]['weights'][k])):

append ((-(r[j][m])\*weights[j]['s'][k])\*2)/len(data) to the list “s” #Gradient values

append “s” to the list “layer['g']”

return 0.0