HW5

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
   - For the number of hidden layers and number of hidden nodes per layer, please check the HYPERPARAMETER Selection section below
   - I am using 10 output neurons. For example, the value 1 is represented by [010000000]
   - I am using sigmoid activation function for all the neurons except the output neuron
   - I am converting the output neuron with highest local field value as digit 1 and rest as 0 using “argmax”
   - I am using an energy function of form **summation (di – yi)/n**
   - I am using learning rates of 10 and 15 because with high learning rate, the algorithm is oscillating around the minima.
   - And in the feedback graph, I am increasing the learning rate by 1.05 times (self.eta*1.05) after every layer to tackle the VANISHING gradient problem.

2. I standardized the data.
3. Design Process:
   a. I am using a generalizable algorithm.
   b. First, I built an algorithm without matrices. It worked for homework 4 but it was not scalable.
   c. So, next I developed an algorithm using numpy arrays. But with this algorithm, the MSE is continuously increasing with every epoch.
   d. After carefully crosschecking the intermediate values with values I calculated myself, I found that instead of temporarily storing the gradient vectors of each layer, I was updating the weights in the same iteration. So, since we need the original weights for the next layer not the updated ones, the algorithm was not giving optimal weights at each epoch.
   e. Then, the algorithm was working well in terms of MSE but not misclassifications. I was just using sigmoid function at the output neurons. Since, we might get more than one 1’s in the output neurons, the misclassifications remained same even though the MSE was decreasing with every epoch.
   f. So, for the next algorithm, instead of using the sigmoid activation functions at the output neurons, I just used the “argmax” function which gives the index of maximum value. I am changing the value at the output of argmax, in an numpy array of zeros of size 10, to 1.
   g. Then, I forgot to include the learning rate (self.eta) parameter in the algorithm. I included it now.
   h. This is my final algorithm.

4. **Tanh activation Function:** (for all layers except the output layer)

```
def tanh(self, s):
        return np.tanh(s)
```

5. **Derivative of tanh:** (for all layers except the output layer)

```
def sigmoidPrime(self, s):
        return s * (1 - s)
```

6. **Activation Function for the output layer in the feedforward graph**

```
  def af(self,s):
    t = 1/(1+np.exp(-s))
    max = np.argmax(t[0])
    y = np.zeros(self.numOutputs)
    y[max] = 1
    return np.asarray(y)
```

7. **Derivative of sigmoid for the output layer**

```
 def sigmoidPrime(self, s):
        return s * (1 - s)
```

8. **Activation function for the whole input data**

```
  def af_predict(self,s):
    t = 1/(1+np.exp(-s))
    p =[]
    for i in range(len(t)):
      max = np.argmax(t[i])
      y = np.zeros(len(t[i]))
      y[max] = 1
      p.append(y)
    return np.asarray(p)
```

9. **Misclassifications Function:**

This is used to calculate the mean squared error corresponding to the given weights.

Algorithm:

```
def misclassifications(self, x,y):
    count = 0
    for i in range(y.shape[0]):
        if np.any(x[i]-y[i]):
            count += 1
    return count
```

10. I created a **NeuralNetworks** class which has predict, feedforward, backward and train functions.
11. 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.
12. **Function call:**

    Ex:

    ```
    nLayers = 1
    nNodes = 2
    nOut = 1
    eta = 10
    iter = 10
    ep = 0.2
    w, epoch, obj, mis = NeuralNetwork(x, d, nLayers, nNodes, nOut, eta, iter, ep).train()
    ```
13. I am using a list with name "**weights**" to store weights (including biases). I am uniformly choosing weights between -2 and 2. For each layer, I am using a list of size = (number of nodes in the current layer x number of nodes in the previous layer) in the list "weights".

```
self.weights=[np.random.uniform(low = -2, high = 2, size = (self.nHiddenNodes, self.nInputNodes))]
for i in range(self.nHiddenLayers-1):
self.weights.append(np.random.uniform(low =-2, high = 2, size =(self.nHiddenNodes, self.nHiddenNodes+1)))          # +1 is for biases
self.weights.append(np.random.uniform(low =-2, high = 2, size = (self.numOutputs, self.nHiddenNodes+1)))          # +1 is for biases
```

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

```
[array([[0.78888993, 0.36114814, 0.84949542],
```

```
[0.05266805, 0.89136156, 0.4386164 ]]), array([[0.22397066, 0.34915997, 0.12250444]])]
```

14. Train function:

- This function calls the backward function which updates the weights every epoch until maximum epochs are reached.
- And this finds the mse and number of misclassifications for both training and testing datasets using updated weights after each epoch.
- 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

Algorithm:

```
def train (self):
    e = 0
    obj_training =[]
    obj_testing = []
    epoch = []
    mis_training = []
    mis_testing =[]
    epoch.append(e)
    pred_train = self.predict(self.data)
    pred_test = self.predict(self.test)
    obj_training.append(((np.linalg.norm(self.labels - pred_train))**2)/len(self.data))
    obj_testing.append(((np.linalg.norm(self.testlabels - pred_test))**2)/len(self.test))
    mis_training.append(self.misclassifications(self.labels, pred_train))
    mis_testing.append(self.misclassifications(self.testlabels, pred_test))
    mse = 100000000000
    while e <= self.maxIt:
        prev = mse
        e += 1
        for i in range(len(self.data)):
            self.backward(self.data[i].reshape(1,self.nInputNodes), self.labels[i])
        pred_train = self.predict(self.data)
        pred_test = self.predict(self.test)
        mse = ((np.linalg.norm(self.labels - pred_train))**2)/len(self.data)
        mse_test = ((np.linalg.norm(self.testlabels - pred_test))**2)/len(self.test)
        obj_training.append(mse)
        obj_testing.append(mse_test)
        epoch.append(e)
        mis_training.append(self.misclassifications(self.labels, pred_train))
        mis_testing.append(self.misclassifications(self.testlabels, pred_test))
        if mse >= prev:
            self.eta = 0.9*self.eta
        prev = mse
    return self.weights, epoch, obj_training, mis_training, obj_testing, mis_testing
```

15. Predict Function: This finds the output labels for the whole input data

```
def predict (self, data=[]):
    prev = data.T
    for i in range(len(self.weights)-1):
        temp = (np.matmul(self.weights[i], prev))
        temp2 = self.sigmoid(temp)
        prev = temp2
    temp_f = np.matmul(self.weights[self.nHiddenLayers], prev)
    temp4 = self.af_predict(np.transpose(temp_f))
    return temp4
```

16. Feedforward function: This finds the output label for the given input datapoint

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:

```
def feedforward (self, x):
    self.r = []
    self.r.append(x)
    prev = x.T
    for i in range(len(self.weights)-1):
        temp = np.matmul(self.weights[i], prev)
        temp2 = self.sigmoid(temp)
        self.r.append(temp2)
        prev = temp2
    temp_f = np.matmul(self.weights[self.nHiddenLayers], prev)
    temp4 =self.af(temp_f.T)
    return temp4
```

17. Backward 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.
- Then it updates the weights using the gradient descent vector.

Algorithm:

```
def backward (self, x, y):
    self.delta = []
    self.gradient=[]
    o = self.feedforward(x)
    self.out_error = y – o
    for i in reversed(range(len(self.weights))):
        if i == 0 and i != len(self.weights)-1:
            temp2 = self.weights[i+1].T.dot(temp)
            temp3 = temp2*self.sigmoidPrime(self.r[i+1])
            temp4 = (temp3.dot(self.r[i]))*2/len(self.data)
            self.gradient.append(self.eta*1.05*temp4)
            temp = temp3
            self.delta.append(temp3)

        elif i > 0 and i<len(self.weights)-1:
            temp2 = self.weights[i+1].T.dot(temp)
            temp3 = temp2*self.sigmoidPrime(self.r[i+1])
            temp4 = (temp3.dot(self.r[i].T))*2/len(self.data)
            self.gradient.append(self.eta*1.05*temp4)
            temp = temp3
            self.delta.append(temp3)

          elif i == len(self.weights)-1:
            temp = self.out_error.reshape(self.numOutputs,1)
            temp2 = self.r[i].dot(temp.T)*2/len(self.data)
            self.delta.append(temp)
            self.gradient.append(self.eta*temp2.T)

        self.gradient = self.gradient[::-1]
        for i in range(len(self.weights)):
          self.weights[i] += self.gradient[i]
```

18. Hyperparameter Selection:

- I am using three lists each for number of hidden layers, number of nodes per hidden layer and learning rates
- I am building 8 Neural network models using the above
- I am running the algorithm for 50 epochs for now.

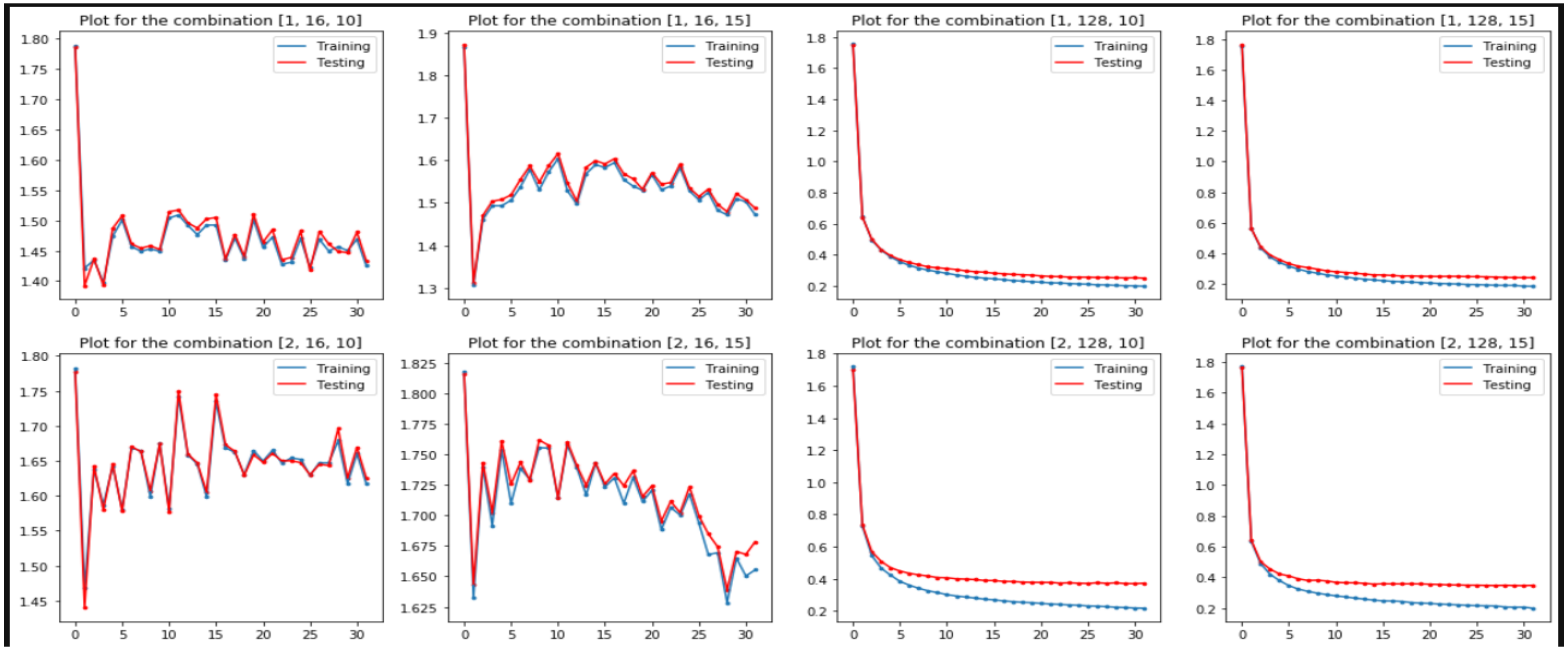
```
hid = [1,2]
hidnodes= [64,128]

# I read some where that the number of hidden nodes in the powers of 2 would result in better performance

lrate = [10,15]
a =[]
b =[]
c=[]
q =[]
iter = 50
numOutputs = 10
prec = 1.6
for i in hid:
```

```
for j in hidnodes:
    for k in lrate:
        #print([i,j,k])
        NN = Neural_Network(xtrain, d, xtest, dt, i, j, numOutputs, k, iter, prec)
        w, epoch, obj_training, mis_training, obj_testing, mis_testing = NN.train()
        a.append(obj_training)
        b.append(mis_training)
        c.append(obj_testing)
        q.append(mis_testing)
```

I generated the plots using matplotlib. Please check my Jupyter notebook.

19. Results of Hyperparameter Selection:

From the plots, I got the continuous decrease in the MSE with number of hidden nodes = 128



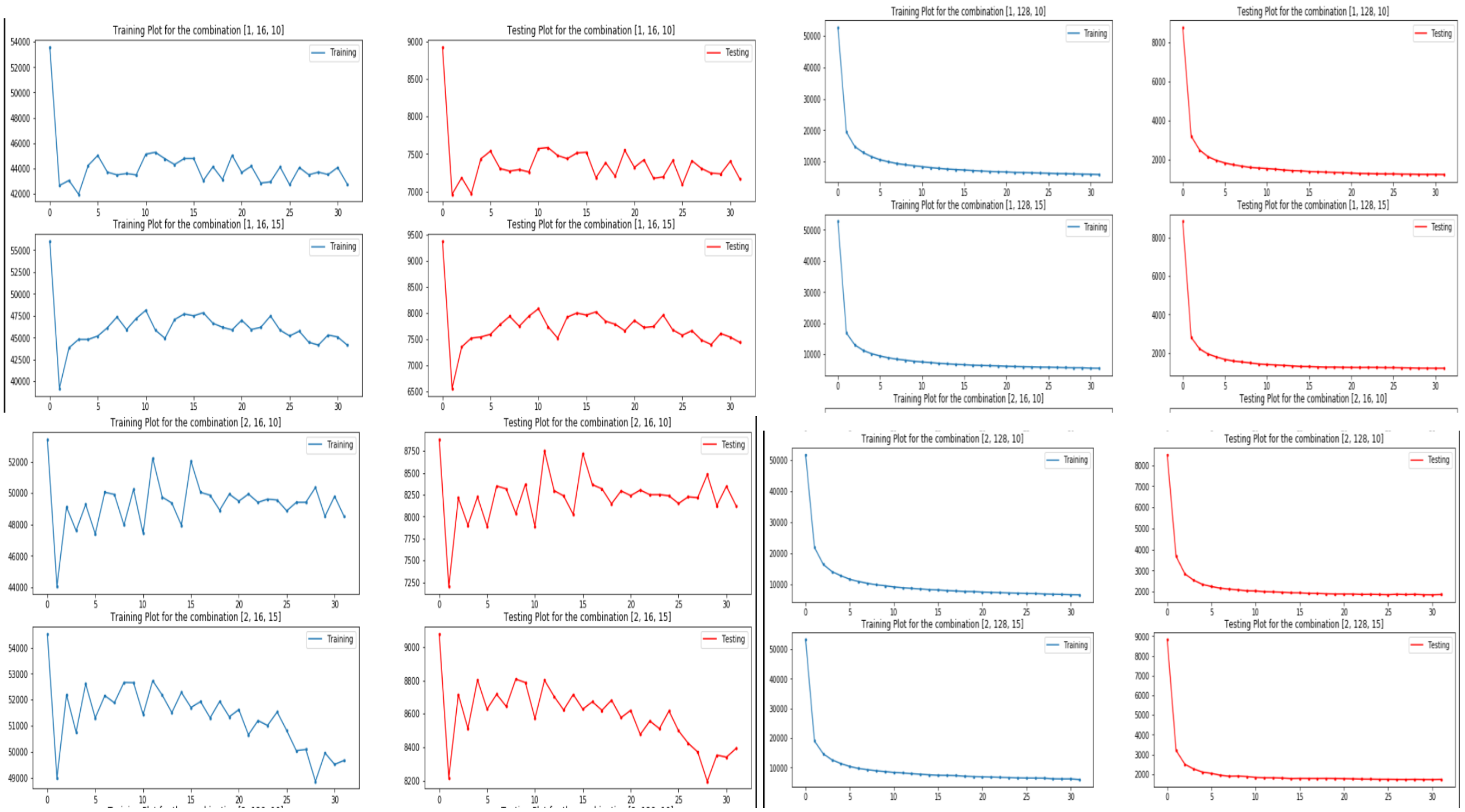
HW5

and less testing misclassifications with the following combination:

Number of hidden layers: 1

Number of nodes per hidden layer: 128

Learning Rate = 10



20. Conclusion:

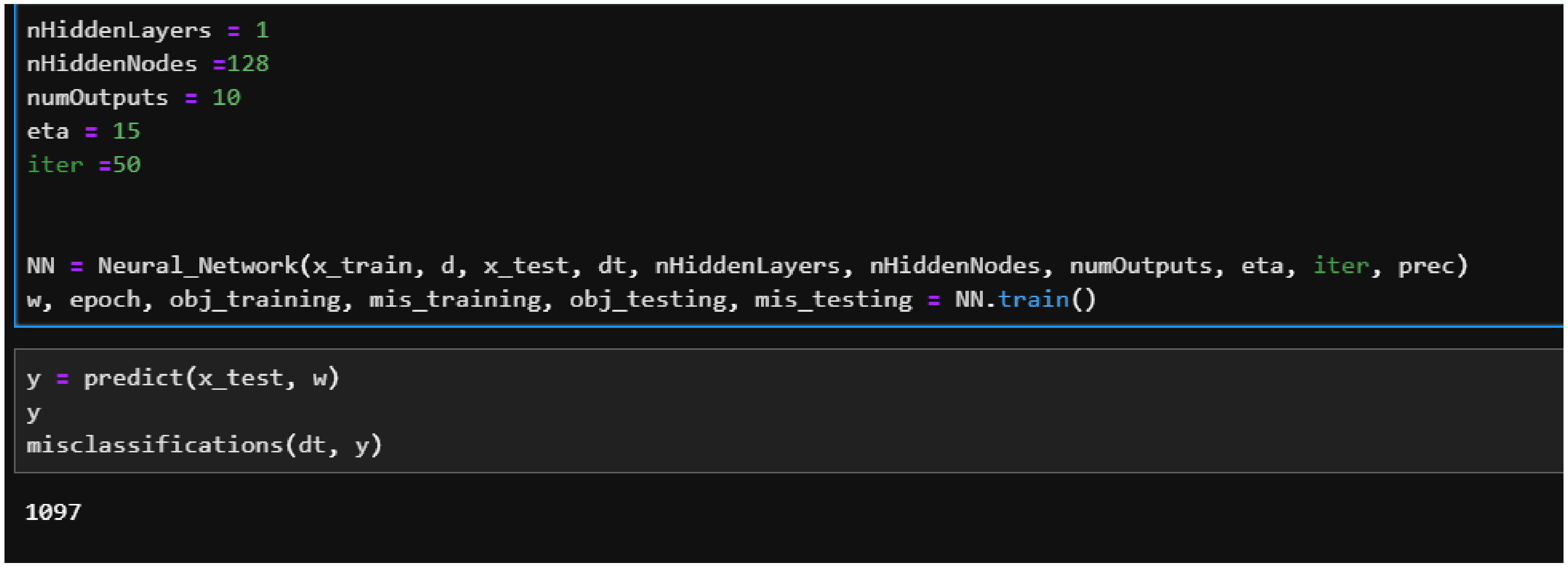
Now to further improve the accuracy, I am running the algorithm for 30 epochs with

Number of hidden layers: 1

Number of nodes per hidden layer: 128

Learning Rate = 10

And I got the accuracy of 89 %



```
nHiddenLayers = 1
nHiddenNodes =128
numOutputs = 10
eta = 15
iter =50


NN = Neural_Network(x_train, d, x_test, dt, nHiddenLayers, nHiddenNodes, numOutputs, eta, iter, prec)
w, epoch, obj_training, mis_training, obj_testing, mis_testing = NN.train()
```

```
y = predict(x_test, w)
y
misclassifications(dt, y)
```

```
1097
```