Importing Required Libraries

In [78]:

```
import numpy as np
import matplotlib.pyplot as plt
import torch
import torch.nn as nn
import torchvision
import torchvision.transforms as transforms
from torch.utils.data.dataset import TensorDataset
from torch.utils.data import DataLoader
```

DATASET LOADING AND PREPARATION

In [79]:

```
train_dataset = torchvision.datasets.MNIST(root='./data', train=True, transform=
transforms.ToTensor(),  download=True)
test_dataset = torchvision.datasets.MNIST(root='./data', train=False, transform=
transforms.ToTensor())
```

In [80]:

```
loader_train = torch.utils.data.DataLoader(dataset=train_dataset,batch_size = le
n(train_dataset) ,shuffle=True)
loader_test = torch.utils.data.DataLoader(dataset=test_dataset,batch_size = len(
test_dataset) ,shuffle=True)
```

In [81]:

```
X_train,y_train=next(iter(loader_train))
X_test,y_test=next(iter(loader_test))
```

In [82]:

```
X_train=X_train.numpy()
y_train=y_train.numpy()
X_test=X_test.numpy()
y_test=y_test.numpy()
```

In [83]:

```
X_train_flattened=X_train.reshape(X_train.shape[0],X_train.shape[1]*X_train.shap
e[2]*X_train.shape[3])
X_test_flattened=X_test.reshape(X_test.shape[0],X_test.shape[1]*X_test.shape[2]*
X_test.shape[3])
```

One-Hot Encoding

In [84]:

```
train_labels_encoded = []
for i in y_train:
  A=np.array([0]*10)
  A[i]=1
  train_labels_encoded.append(A)
y_train_encoded=np.array(train_labels_encoded)
```

In [85]:

```
test_labels_encoded = []
for i in y_test:
  A=np.array([0]*10)
  A[i]=1
  test_labels_encoded.append(A)
y_test_encoded=np.array(test_labels_encoded)
```

In [86]:

```
X_train_flattened_torch=torch.from_numpy(X_train_flattened)
y_train_encoded_torch=torch.from_numpy(y_train_encoded)
```

In [87]:

```
X_test_flattened_torch=torch.from_numpy(X_test_flattened)
y_test_encoded_torch=torch.from_numpy(y_test_encoded)
```

Here we have prepared test and train dataset with flattenned images and one hot encoded labels

In [88]:

```
Train_Dataset=TensorDataset(X_train_flattened_torch,y_train_encoded_torch)
```

In [89]:

```
Test_Dataset=TensorDataset(X_test_flattened_torch,y_test_encoded_torch)
```

HELPER FUNCTIONS

In [90]:

```
def relu(z):
    return np.maximum(0,z)

def softmax(z):
    return np.exp(z)/sum(np.exp(z))
```

In [91]:

```
def dif_relu(z):
  # return np.multiply(1.0 , (z>0))
  return z>0
```

In [92]:

```
def glorot_initialisation(output_n,input_n):
  M=np.sqrt(6/(input_n+output_n))
  W=np.random.uniform(low=-M, high=M, size=(output_n,input_n))
  b=np.random.uniform(low=-M, high=M, size=(output_n,1))
  return W,b
```

In [93]:

```
def initialize_parameters(layer_dims):
    """
    Arguments:
    layer_dims -- python array (list) containing the dimensions of each layer in
our network

    Returns:
    parameters -- python dictionary containing your parameters "W1", "b1", ...,
 "WL", "bL":
                    Wl -- weight matrix of shape (layer_dims[l], layer_dims[l-
1])
                    bl -- bias vector of shape (layer_dims[l], 1)
                    Wl -- weight matrix of shape (layer_dims[l-1], layer_dims
[l])
                    bl -- bias vector of shape (1, layer_dims[l])

    Tips:
    - For example: the layer_dims for the "Planar Data classification model" wou
ld have been [2,2,1].
    This means W1's shape was (2,2), b1 was (1,2), W2 was (2,1) and b2 was (1,
1). Now you have to generalize it!
    - In the for loop, use parameters['W' + str(l)] to access Wl, where l is the
iterative integer.
    """

    np.random.seed(1390)
    parameters = {}
    L = len(layer_dims) # number of layers in the network

    for l in range(1, L):

        # M=np.sqrt(6/(self.input_n+self.output_n))
        parameters['W' + str(l)],parameters['b' + str(l)]  = glorot_initialisati
on(layer_dims[l], layer_dims[l-1])


        assert(parameters['W' + str(l)].shape == (layer_dims[l], layer_dims[l-1
]))
        assert(parameters['b' + str(l)].shape == (layer_dims[l], 1))

    return parameters
```

In [94]:

```
def forward_propagation(X, parameters):
    """

    SHAPE OF X = 784,samples(i.e. 64 for a batch)
    Implements the forward propagation (and computes the loss) presented in Figu
re 2.

    Arguments:
    X -- input dataset, of shape (input size, number of examples)
    parameters -- python dictionary containing your parameters "W1", "b1", "W2",
"b2", "W3", "b3","W4","b4":
                    W1 -- weight matrix of shape (500,784)
                    b1 -- bias vector of shape (500,1)
                    W2 -- weight matrix of shape (250,500)
                    b2 -- bias vector of shape (250,1)
                    W3 -- weight matrix of shape (100,250)
                    b3 -- bias vector of shape (100,1)
                    W4 -- weight matrix of shape (10,100)
                    b4 -- bias vector of shape (10,1)


    Returns:
    loss -- the loss function (vanilla logistic loss)
    """

    cache={}


    cache['Z2']=np.dot(parameters['W1'],X)+parameters['b1']
    cache['A2']=relu(cache['Z2'])

    cache['Z3']=np.dot(parameters['W2'],cache['A2'])+parameters['b2']
    cache['A3']=relu(cache['Z3'])

    cache['Z4']=np.dot(parameters['W3'],cache['A3'])+parameters['b3']
    cache['A4']=relu(cache['Z4'])

    cache['Z5']=np.dot(parameters['W4'],cache['A4'])+parameters['b4']
    cache['A5']=softmax(cache['Z5'])

    A_last=cache['A5']

    cache['W1']=parameters['W1']
    cache['b1']=parameters['b1']

    cache['W2']=parameters['W2']
    cache['b2']=parameters['b2']

    cache['W3']=parameters['W3']
    cache['b3']=parameters['b3']

    cache['W4']=parameters['W4']
    cache['b4']=parameters['b4']

   # cache = {"W1": W1, "b1": b1,"Z2": Z2, "A2": A2,
   #             "W2": W2, "b2": b2,"Z3": Z3, "A3": A3,
   #             "W3": W3, "b3": b3,"Z4": Z4, "A4": A4,
   #             "W4": W4, "b4": b4,"Z5": Z5, "A5": A5}
```

```
    return A_last, cache
```

In [95]:

```
# Calculating the loss function using the cross entropy
"""Arguments:
    A -- post-activation, output of forward propagation
    Y -- "true" labels vector, same shape as A

    Returns:
    cost - value of the cost function
"""
def compute_cost(A, Y,cache,lambd):
    #A is predicted
    #Y is actual
    m = Y.shape[1]
    logprobs = np.multiply(-np.log(A),Y) + np.multiply(-np.log(1 - A), 1 - Y)
    cost = 1./m * np.nansum(logprobs)

    l2_regularisation_cost=(lambd/(2*m))*(np.sum(np.square(cache['W1']))+np.sum(
np.square(cache['W2']))+np.sum(np.square(cache['W3']))+np.sum(np.square(cache['W
4'])))

    regularized_total_cost=cost +l2_regularisation_cost
    return regularized_total_cost
```

In [96]:

```
def backward_propagation(X,Y,cache,lambd):

  m=X.shape[1]

  grads={}

  grads['dZ5']=cache['A5']-Y
  grads['dW4']= 1./m * np.dot(grads['dZ5'],cache['A4'].T) + (lambd*cache['W4'])/
m
  grads['db4']= 1./m * np.sum(grads['dZ5'],axis=1,keepdims=True)

  grads['dA4']=np.dot(cache['W4'].T,grads['dZ5'])
  grads['dZ4']=np.multiply(grads['dA4'],dif_relu(cache['Z4']))
  grads['dW3']=1./m * np.dot(grads['dZ4'],cache['A3'].T) + (lambd*cache['W3'])/m
  grads['db3']=1./m * np.sum(grads['dZ4'],axis=1,keepdims=True)

  grads['dA3']=np.dot(cache['W3'].T,grads['dZ4'])
  grads['dZ3']=np.multiply(grads['dA3'],dif_relu(cache['Z3']))
  grads['dW2']=1./m * np.dot(grads['dZ3'],cache['A2'].T)  + (lambd*cache['W2'])/
m
  grads['db2']=1./m * np.sum(grads['dZ3'],axis=1,keepdims=True)

  grads['dA2']=np.dot(cache['W2'].T,grads['dZ3'])
  grads['dZ2']=np.multiply(grads['dA2'],dif_relu(cache['Z2']))
  grads['dW1']=1./m * np.dot(grads['dZ2'],X.T)  + (lambd*cache['W1'])/m
  grads['db1']=1./m * np.sum(grads['dZ2'],axis=1,keepdims=True)


  return grads
```

In [97]:

```
def update_parameters(parameters,grads,learning_rate):

  updated_parameters={}

  updated_parameters['W1']=parameters['W1']-learning_rate*grads['dW1']
  updated_parameters['b1']=parameters['b1']-learning_rate*grads['db1']

  updated_parameters['W2']=parameters['W2']-learning_rate*grads['dW2']
  updated_parameters['b2']=parameters['b2']-learning_rate*grads['db2']

  updated_parameters['W3']=parameters['W3']-learning_rate*grads['dW3']
  updated_parameters['b3']=parameters['b3']-learning_rate*grads['db3']

  updated_parameters['W4']=parameters['W4']-learning_rate*grads['dW4']
  updated_parameters['b4']=parameters['b4']-learning_rate*grads['db4']

  return updated_parameters
```

In [98]:

```
#Finding the accuracy of the parameter at the output
"""
    Arguments:
    y_actual -  given in the dataset / also called as the ground truth
    y_pred - generated from the neural network , after a series of forward and b
ackprop

    Returns:
    accuracy =  finding the matches of the prdicted vs the actual
"""
def calculate_accuracy(y_actual,y_pred):
    accuracy = np.count_nonzero(np.argmax(y_pred,axis=0)==np.argmax(y_actual,axi
s=0))/y_actual.shape[1]
    return accuracy
```

In [99]:

```
def predict(X,Y,parameters):

    """
    This function is used to predict the results of a  n-layer neural network.

    Arguments:
    X -- data set of examples you would like to label
    Y -- data set of examples
    parameters -- parameters of the trained model

    Returns:
    ypred -- predictions for the given dataset X
    """

    y_pred,cache=forward_propagation(X,parameters)
    return y_pred
```

MODEL TRAINING

In [100]:

```
def model(Train_Dataset,layer_dimensions,total_epochs=15,Batch_Size=64,learning_
rate=0.01,lambd=0.8):

  costs=[]
  accuracy=[]

  parameters=initialize_parameters(layer_dimensions)
  num_iterations=len(Train_Dataset)//Batch_Size


  #Train_Dataset=TensorDataset(X_training,Y_training)
  for epoch in range(total_epochs):
    for iteration in range(num_iterations):
      Data_Loader=torch.utils.data.DataLoader(dataset=Train_Dataset,batch_size=6
4, shuffle=True)

      data_iter=iter(Data_Loader)
      Data=next(data_iter)
      X,y=Data #X.shape=(batch_size,784) y.shape=(batch_size,10)
      X=X.numpy()
      y=y.numpy()
      a5,cache=forward_propagation(X.T,parameters)
      cost=compute_cost(a5,y.T,cache,lambd)
      gradients=backward_propagation(X.T,y.T,cache,lambd)
      parameters=update_parameters(parameters,gradients,learning_rate)

      if iteration%200==0:
        print("epoch: ",epoch+1,"/",total_epochs, "  iteration= ",iteration+1,
"/",num_iterations, "  Loss: ",cost)
      accuracy.append(calculate_accuracy(y.T,a5))
      costs.append(cost)

  return accuracy,costs,parameters
```

In [114]:

```
layer_dimensions=[784,500,250,100,10]
Train_accuracy,Train_costs,Trained_parameters=model(Train_Dataset,layer_dimensio
ns,15,64,0.01,0.8)
```

```
epoch:  1 / 15   iteration=  1 / 937   Loss:  10.24665749963704
epoch:  1 / 15   iteration=  201 / 937   Loss:  8.564261611610512
epoch:  1 / 15   iteration=  401 / 937   Loss:  7.420605826921399
epoch:  1 / 15   iteration=  601 / 937   Loss:  6.885346259516561
epoch:  1 / 15   iteration=  801 / 937   Loss:  6.403292172120739
epoch:  2 / 15   iteration=  1 / 937   Loss:  6.452327503465423
epoch:  2 / 15   iteration=  201 / 937   Loss:  6.089522975943435
epoch:  2 / 15   iteration=  401 / 937   Loss:  5.962707372903211
epoch:  2 / 15   iteration=  601 / 937   Loss:  5.4670342284631905
epoch:  2 / 15   iteration=  801 / 937   Loss:  5.202636592920128
epoch:  3 / 15   iteration=  1 / 937   Loss:  4.77820843767748
epoch:  3 / 15   iteration=  201 / 937   Loss:  4.841851990559459
epoch:  3 / 15   iteration=  401 / 937   Loss:  4.696018944101629
epoch:  3 / 15   iteration=  601 / 937   Loss:  4.488708957460135
epoch:  3 / 15   iteration=  801 / 937   Loss:  4.28934694401448
epoch:  4 / 15   iteration=  1 / 937   Loss:  4.028385348035927
epoch:  4 / 15   iteration=  201 / 937   Loss:  4.0758647084738895
epoch:  4 / 15   iteration=  401 / 937   Loss:  3.974530131459747
epoch:  4 / 15   iteration=  601 / 937   Loss:  3.6465236357136117
epoch:  4 / 15   iteration=  801 / 937   Loss:  3.637218839290589
epoch:  5 / 15   iteration=  1 / 937   Loss:  3.734181223285472
epoch:  5 / 15   iteration=  201 / 937   Loss:  3.5855167756247734
epoch:  5 / 15   iteration=  401 / 937   Loss:  3.2157957371924164
epoch:  5 / 15   iteration=  601 / 937   Loss:  3.0251044494455823
epoch:  5 / 15   iteration=  801 / 937   Loss:  3.0387374471144772
epoch:  6 / 15   iteration=  1 / 937   Loss:  2.92617117637501
epoch:  6 / 15   iteration=  201 / 937   Loss:  2.851057199354434
epoch:  6 / 15   iteration=  401 / 937   Loss:  2.7548506878929024
epoch:  6 / 15   iteration=  601 / 937   Loss:  2.6273646131420714
epoch:  6 / 15   iteration=  801 / 937   Loss:  2.655667992753477
epoch:  7 / 15   iteration=  1 / 937   Loss:  2.51058362023355
epoch:  7 / 15   iteration=  201 / 937   Loss:  2.3837788372842788
epoch:  7 / 15   iteration=  401 / 937   Loss:  2.3898939307404827
epoch:  7 / 15   iteration=  601 / 937   Loss:  2.243522821889658
epoch:  7 / 15   iteration=  801 / 937   Loss:  2.3308796751895793
epoch:  8 / 15   iteration=  1 / 937   Loss:  2.1649825337329487
epoch:  8 / 15   iteration=  201 / 937   Loss:  2.0946083964101474
epoch:  8 / 15   iteration=  401 / 937   Loss:  2.2828127545878028
epoch:  8 / 15   iteration=  601 / 937   Loss:  1.8249833646968407
epoch:  8 / 15   iteration=  801 / 937   Loss:  1.7898704587402183
epoch:  9 / 15   iteration=  1 / 937   Loss:  1.8255597629277764
epoch:  9 / 15   iteration=  201 / 937   Loss:  1.9323719681470295
epoch:  9 / 15   iteration=  401 / 937   Loss:  1.7352646031430177
epoch:  9 / 15   iteration=  601 / 937   Loss:  1.8296360457139014
epoch:  9 / 15   iteration=  801 / 937   Loss:  1.6330053947461676
epoch:  10 / 15   iteration=  1 / 937   Loss:  1.6274586381869076
epoch:  10 / 15   iteration=  201 / 937   Loss:  1.6261009944552596
epoch:  10 / 15   iteration=  401 / 937   Loss:  1.4480423532469877
epoch:  10 / 15   iteration=  601 / 937   Loss:  1.5667137942252174
epoch:  10 / 15   iteration=  801 / 937   Loss:  1.444136211820898
epoch:  11 / 15   iteration=  1 / 937   Loss:  1.3858226251440717
epoch:  11 / 15   iteration=  201 / 937   Loss:  1.6059569652673968
epoch:  11 / 15   iteration=  401 / 937   Loss:  1.3160241285108274
epoch:  11 / 15   iteration=  601 / 937   Loss:  1.2525268225964128
epoch:  11 / 15   iteration=  801 / 937   Loss:  1.3333353130193764
epoch:  12 / 15   iteration=  1 / 937   Loss:  1.3257485188478042
epoch:  12 / 15   iteration=  201 / 937   Loss:  1.209194326008392
epoch:  12 / 15   iteration=  401 / 937   Loss:  1.1110561832529737
epoch:  12 / 15   iteration=  601 / 937   Loss:  1.238720947743997
epoch:  12 / 15   iteration=  801 / 937   Loss:  1.2932704654218936
epoch:  13 / 15   iteration=  1 / 937   Loss:  1.464326811340972
```

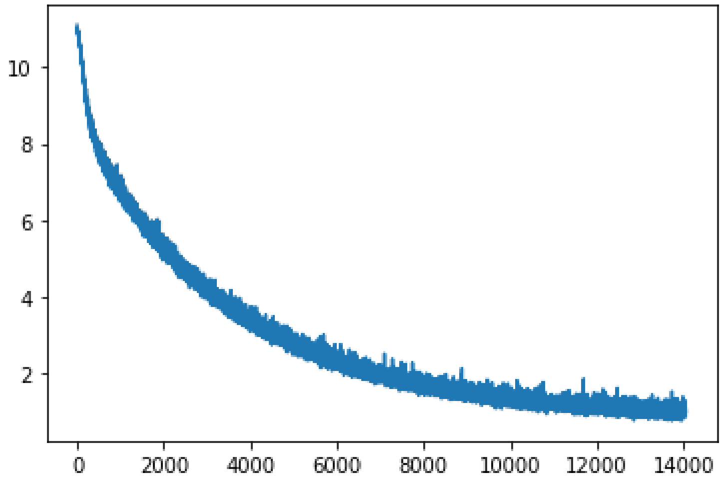
```
epoch:  13 / 15   iteration=  201 / 937   Loss:  1.272144386531624
epoch:  13 / 15   iteration=  401 / 937   Loss:  1.3448446867930093
epoch:  13 / 15   iteration=  601 / 937   Loss:  1.1313718587661445
epoch:  13 / 15   iteration=  801 / 937   Loss:  1.2057377839806718
epoch:  14 / 15   iteration=  1 / 937   Loss:  1.0369554623507273
epoch:  14 / 15   iteration=  201 / 937   Loss:  1.0573854059126497
epoch:  14 / 15   iteration=  401 / 937   Loss:  1.2270571120069174
epoch:  14 / 15   iteration=  601 / 937   Loss:  0.8787962443751072
epoch:  14 / 15   iteration=  801 / 937   Loss:  1.1267776343406106
epoch:  15 / 15   iteration=  1 / 937   Loss:  1.1392730653798524
epoch:  15 / 15   iteration=  201 / 937   Loss:  1.1172321624453316
epoch:  15 / 15   iteration=  401 / 937   Loss:  0.8915757681278184
epoch:  15 / 15   iteration=  601 / 937   Loss:  0.9160363899972161
epoch:  15 / 15   iteration=  801 / 937   Loss:  0.9858567924604414
```

In [102]:

```
plt.plot(Train_costs)
```

Out[102]:

```
[<matplotlib.lines.Line2D at 0x7fb97f6a7810>]
```

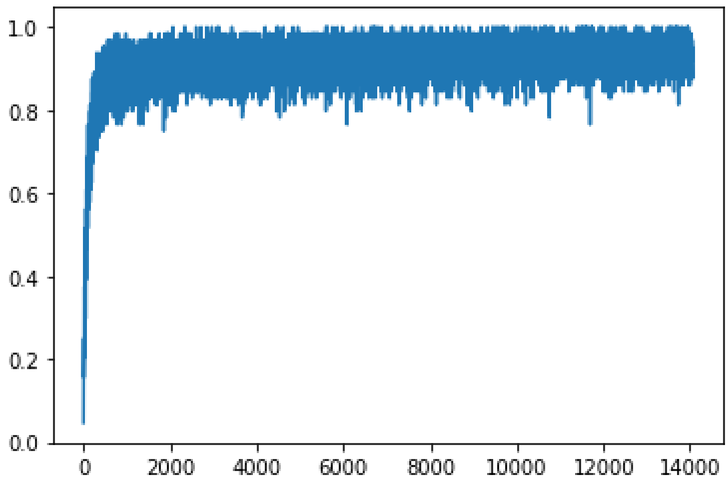


In [103]:

```
plt.plot(Train_accuracy)
```

Out[103]:

```
[<matplotlib.lines.Line2D at 0x7fb97f687950>]
```



In [104]:

```
Train_accuracy[-1]
```

Out[104]:

```
0.921875
```

In [105]:

```
Test_Data_Loader=torch.utils.data.DataLoader(dataset=Test_Dataset,batch_size=len
(Test_Dataset),shuffle=True)
data_iter=iter(Test_Data_Loader)
Test_Data=next(data_iter)
X,y=Test_Data
X=X.numpy()
y=y.numpy()
y_predicted,cache_out=forward_propagation(X.T,Trained_parameters)
Test_accuracy=calculate_accuracy(y.T,y_predicted)
```

In [106]:

```
Test_accuracy
```

Out[106]:

```
0.9368
```

In [107]:

```
Y_Predicted=np.array(np.argmax(y_predicted,axis=0))
```

In [108]:

```
Y_Actual=np.array(np.argmax(y.T,axis=0)  )
```

CONFUSION MATRIX AND CLASSIFICATION REPORT

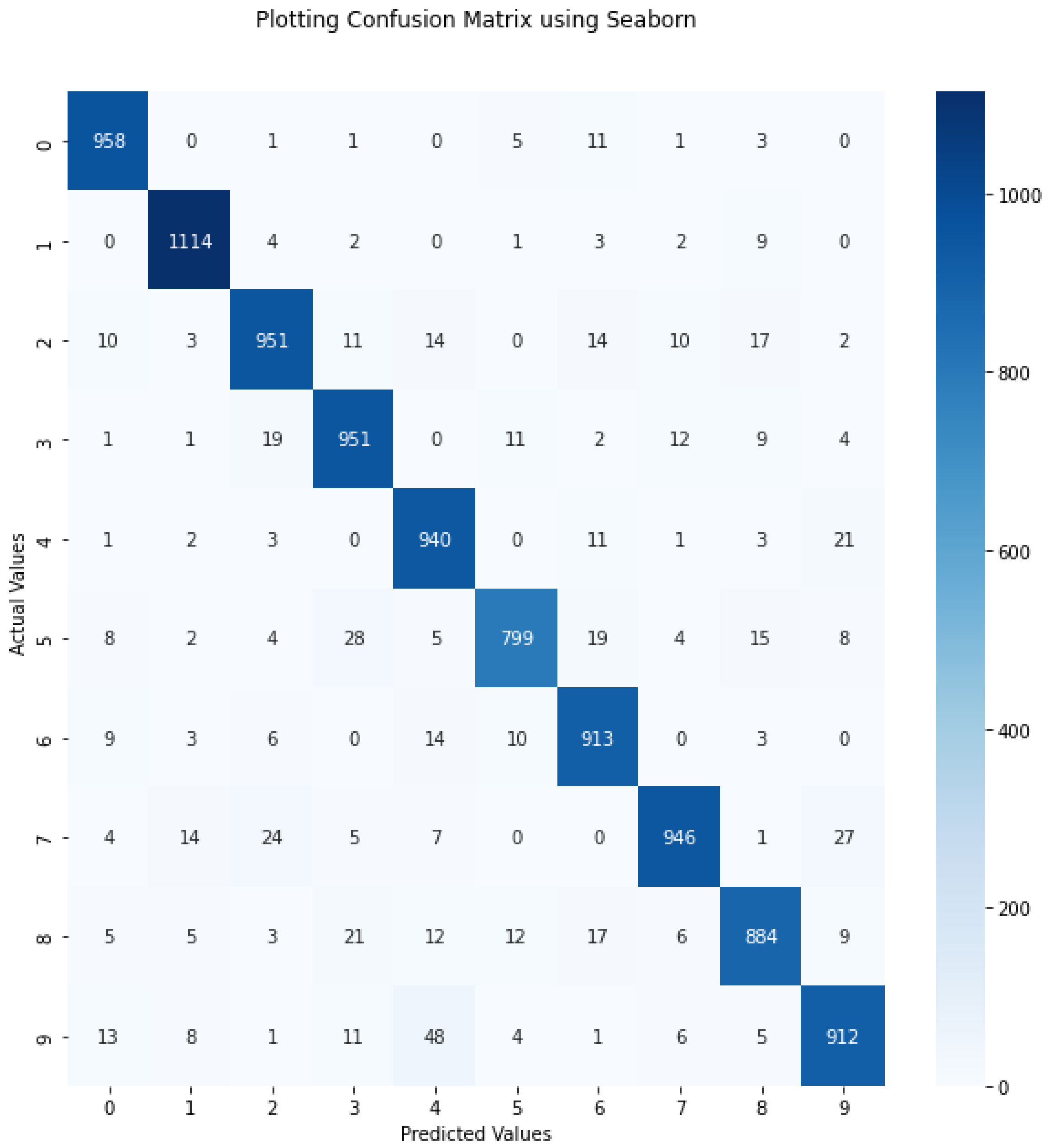
In [109]:

```
from sklearn.metrics import confusion_matrix
import seaborn as sns
plt.figure(figsize=(10,10))
conf_matrix = (confusion_matrix(Y_Actual, Y_Predicted, labels=np.unique(Y_Actual
)))

# Using Seaborn heatmap to create the plot
fx = sns.heatmap(conf_matrix, annot=True, cmap='Blues',fmt='d')

# labels the title and x, y axis of plot
fx.set_title('Plotting Confusion Matrix using Seaborn\n\n');
fx.set_xlabel('Predicted Values')
fx.set_ylabel('Actual Values ');


plt.show()
```



In [110]:

```
from sklearn.metrics import classification_report
print(classification_report(Y_Actual, Y_Predicted))
```

```
              precision    recall  f1-score   support

           0       0.95      0.98      0.96       980
           1       0.97      0.98      0.97      1135
           2       0.94      0.92      0.93      1032
           3       0.92      0.94      0.93      1010
           4       0.90      0.96      0.93       982
           5       0.95      0.90      0.92       892
           6       0.92      0.95      0.94       958
           7       0.96      0.92      0.94      1028
           8       0.93      0.91      0.92       974
           9       0.93      0.90      0.92      1009

    accuracy                           0.94     10000
   macro avg       0.94      0.94      0.94     10000
weighted avg       0.94      0.94      0.94     10000
```

In [111]:

```
unique_p, counts_p = np.unique(Y_Predicted, return_counts=True)
```

In [112]:

```
unique_p
```

Out[112]:

```
array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
```

In [113]:

```
counts_p
```

Out[113]:

```
array([1009, 1152, 1016, 1030, 1040,  842,  991,  988,  949,  983])
```

REPORTING ACCURACY OF MODEL

TRAIN ACCURACY: 92.18%

TEST ACCURACY: 93.68%

lambda=0.8

Inference: This model performs outstands all unregularised models with sigmoid, tanh, and relu activations. This shows that regularisation can help in model improvement.