ASSIGNMENT 3
Report on Multi Layer Perceptron classifier

Submitted By: Anima Prasad (21CS60R66)
Arkapravo Ghosh (21CS60R64)

1. Dataset Analysis and Preprocessing
   - The dataset used here is Letter Recognition Data Set: https://archive.ics.uci.edu/ml/datasets/letter+recognition.
   - In this dataset, there are total 17 attributes. The first attribute named "lettr" specifies the alphabet from A-Z and is a classifier attribute. The remaining 16 attributes are the feature attributes in the dataset. Therefore, the number of nodes in input layer is 16 as number of input features is 16 and number of nodes in output layer is 26 as total number of possible classes is 26.
   - The input dataset is spilt into train and test set in the ratio 80:20. Stratification is also done to get a good spilt.
   - The given dataset is already scaled using standard scalar. The values after scaling are between -1 to 1.
   - Using torch -> Dataloader, input train set is divided into mini batches.
   - The hyper parameters needed are: learning rate, activation function, optimizer, batch size, epochs.
   - The learning rate defines how quickly a network updates its parameters. Here, the learning rates used are: 0.1, 0.08, 0.05, 0.01, 0.001, 0.0001, 0.00001.
   - Activation functions are used to introduce nonlinearity to models, which allows deep learning models to learn nonlinear prediction boundaries. Here, "ReLU" activation function is used for hidden layer and "LogSoftmax" is used for output layer as it is a multi-class prediction.
   - Number of epochs is the number of times the whole training data is fit to the network while training. Here, number of epochs is 1000.
   - Mini batch size is the number of sub samples given to the network after which parameter update happens. Here, the batch size is 128 for the training set.
   - Optimizers are algorithms or methods used to change the attributes of the neural network such as weights to reduce the losses. As mentioned in question, stochastic gradient decent method is used while optimizing.

2. MLP Classifier
   - A class named "NeuralNetwork" is created. This class takes layer information as input. This information contains number of layers in network and number of neurons in each layer.

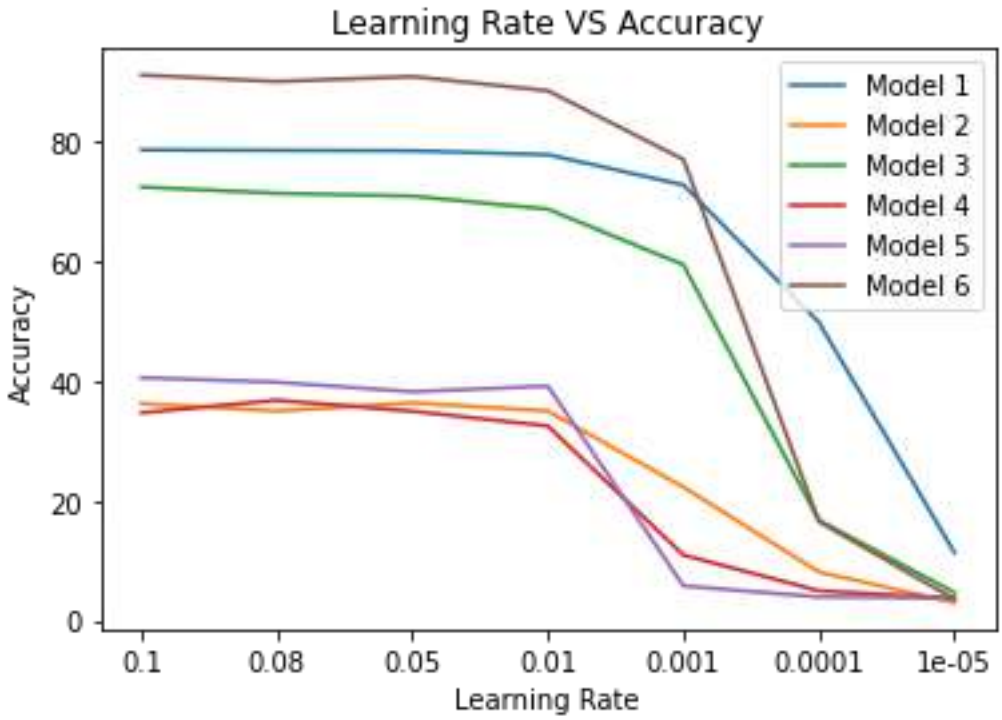
- This above class is used to create the different linear units along with the required activation functions (hidden layers). This class forms the basic architecture of the neural network. “ReLU” activation function is applied to hidden layers. “LogSoftmax” is applied to the output layer of this multi class prediction problem as “NLLLoss” does not work with “Softmax” activation function.
- Then, a class named “Model” is created. This class takes layer information, learning rate, number of epochs as input. It internally creates object for above class “NeuralNetwork” by passing argument layer information. The model so returned is trained on training data set using a negative log likelihood loss function “torch -> NLLLoss” and SGD optimizer.
- For given 5 types of architectures with varying number of hidden layers and number of neurons in hidden layers, accuracy is calculated on test data set for given number of epochs. For every architecture, 5 different learning rates are explored.
- Accuracies of the Model 1 at all learning rates:
    - Learning rate: 0.1 Test Accuracy for Model 1: 78.625
    - Learning rate: 0.08 Test Accuracy for Model 1: 78.55
    - Learning rate: 0.05 Test Accuracy for Model 1: 78.475
    - Learning rate: 0.01 Test Accuracy for Model 1: 77.825
    - Learning rate: 0.001 Test Accuracy for Model 1: 72.775
    - Learning rate: 0.0001 Test Accuracy for Model 1: 49.825
    - Learning rate: 1e-05 Test Accuracy for Model 1: 11.4
- Accuracies of the Model 2 at all learning rates:
    - Learning rate: 0.1 Test Accuracy for Model 2: 36.25
    - Learning rate: 0.08 Test Accuracy for Model 2: 35.0
    - Learning rate: 0.05 Test Accuracy for Model 2: 36.375
    - Learning rate: 0.01 Test Accuracy for Model 2: 35.0
    - Learning rate: 0.001 Test Accuracy for Model 2: 22.3
    - Learning rate: 0.0001 Test Accuracy for Model 2: 8.125
    - Learning rate: 1e-05 Test Accuracy for Model 2: 3.0
- Accuracies of the Model 3 at all learning rates:
    - Learning rate: 0.1 Test Accuracy for Model 3: 72.425
    - Learning rate: 0.08 Test Accuracy for Model 3: 71.375
    - Learning rate: 0.05 Test Accuracy for Model 3: 70.9
    - Learning rate: 0.01 Test Accuracy for Model 3: 68.725
    - Learning rate: 0.001 Test Accuracy for Model 3: 59.425
    - Learning rate: 0.0001 Test Accuracy for Model 3: 16.75
    - Learning rate: 1e-05 Test Accuracy for Model 3: 4.7
- Accuracies of the Model 4 at all learning rates:
    - Learning rate: 0.1 Test Accuracy for Model 4: 34.725
    - Learning rate: 0.08 Test Accuracy for Model 4: 36.8
    - Learning rate: 0.05 Test Accuracy for Model 4: 34.975
    - Learning rate: 0.01 Test Accuracy for Model 4: 32.5

- Learning rate: 0.001 Test Accuracy for Model 4: 10.95
- Learning rate: 0.0001 Test Accuracy for Model 4: 4.975
- Learning rate: 1e-05 Test Accuracy for Model 4: 3.725

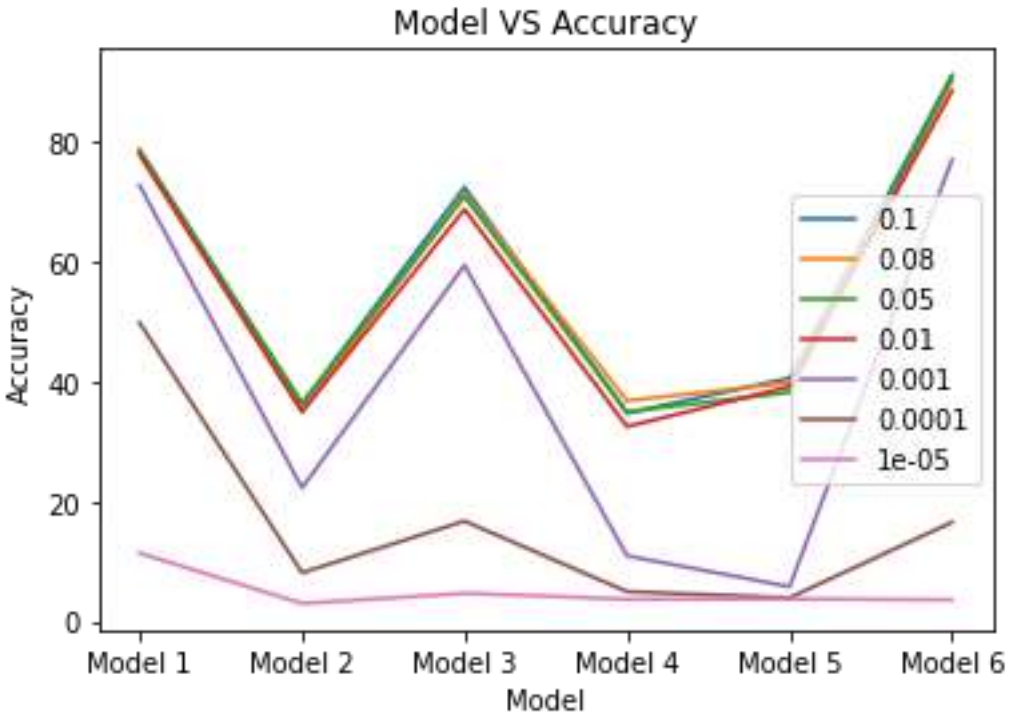
- Accuracies of the Model 5 at all learning rates:
  - Learning rate: 0.1 Test Accuracy for Model 5: 40.575
  - Learning rate: 0.08 Test Accuracy for Model 5: 39.825
  - Learning rate: 0.05 Test Accuracy for Model 5: 38.225
  - Learning rate: 0.01 Test Accuracy for Model 5: 39.175
  - Learning rate: 0.001 Test Accuracy for Model 5: 5.8
  - Learning rate: 0.0001 Test Accuracy for Model 5: 3.95
  - Learning rate: 1e-05 Test Accuracy for Model 5: 3.775

- Accuracies of the Model 6 at all learning rates:
  - Learning rate: 0.1 Test Accuracy for Model: 91.125
  - Learning rate: 0.08 Test Accuracy for Model: 90.05
  - Learning rate: 0.05 Test Accuracy for Model: 90.875
  - Learning rate: 0.01 Test Accuracy for Model: 88.5
  - Learning rate: 0.001 Test Accuracy for Model: 77.025
  - Learning rate: 0.0001 Test Accuracy for Model: 16.55
  - Learning rate: 1e-05 Test Accuracy for Model: 3.6

3. Learning rate Vs Accuracy Vs Model
   - First plot describes the accuracies corresponding to different learning rates for all six models.

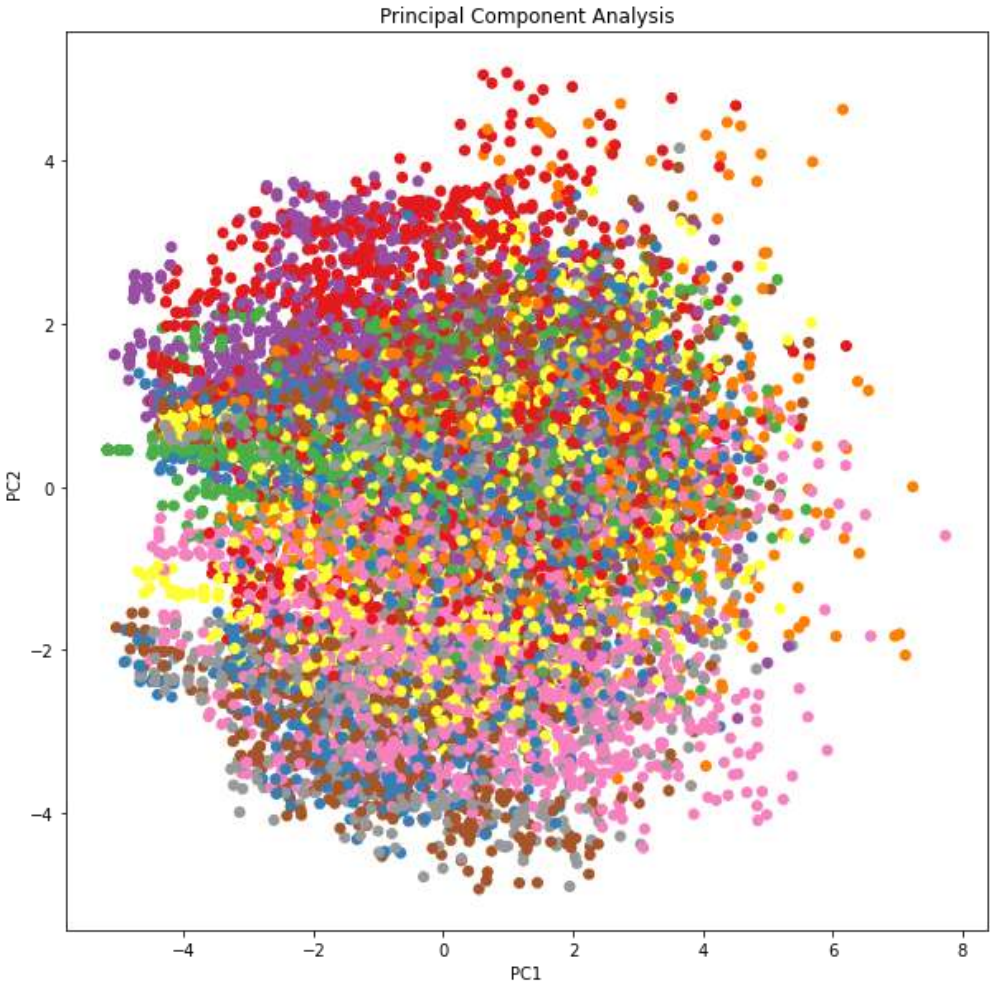


- Second plot describes the accuracies corresponding to each model at all learning rates.



4. Best Model
    - The best model for given dataset is Model 6. The model has 2 hidden layers with 14 and 22 nodes in each layer respectively. It is providing an accuracy of 91.125% with a learning rate of 0.1. The hyperparameters are:-
        - Epoch = 1000
        - Activation Function of hidden layer – Relu
        - Activation Function at output layer – Log Softmax
        - Optimizer – SDG Optimizer

5. Principal component Analysis
    - PCA is an unsupervised statistical technique that is used to reduce the dimensions of the dataset.
    - Here, the dimension of input feature set is reduced from 16 to 2.
    - Dimension reduction is performed using sklearn -> PCA.
    - The fit () method is invoked for training dataset, and training dataset is transformed using transform () method.
    - The test data set is also transformed using transform () method.
    - The transformed training data set is divided into mini batches using torch -> Dataloader.
    - As a result, each data in data set is represented by two attributes/features (PC1, PC2).
    - Below plot shows the data and its corresponding PC1 and PC2 attribute values. Data belonging to same class is represented by same color.



6. MLP on reduced feature space
    - Best learning rate of model 1 : 0.1, model 2 : 0.05, model 3 : 0.1, model 4 : 0.08, model 5 : 0.1, model 6 : 0.1 according to results from task 2.
    - All 5 models are trained over reduced feature train data set and accuracy is computed over reduced feature test data set.
      Model 1 - Learning rate: 0.1 Test Accuracy for Model 1: 15.6
      Model 2 - Learning rate: 0.05 Test Accuracy for Model 2: 14.825
      Model 3 - Learning rate: 0.1 Test Accuracy for Model 3: 16.225
      Model 4 - Learning rate: 0.08 Test Accuracy for Model 4: 15.425
      Model 5 - Learning rate: 0.1 Test Accuracy for Model 5: 15.3
      Model 6 - Learning rate: 0.1 Test Accuracy for Model 6: 21.425