SML Assignment 4
Kesar Shrivastava
2019051

Question 1

- In the gradient boosting classifier, we train a weak model and then train the subsequent models on the reside: y - F(x); where F is the model trained on dataset x and y is the label.
- Here, the models used are decision tree regression stumps (max_depth = 1).
- The labels are updated using the learning rate: 0.1
- Since we needed to ignore the iteration-wise accuracy, the **final training and testing accuracy are 0.1703 and 0.1706 respectively**.
- **The regression outputs are rounded to the nearest to get the class labels.**

Question 2

- The number of layers used in the training is five.
  Input layer : (input = 784, output = 512, activation = sigmoid)
  Hidden layer: (input = 512, output = 128, activation = sigmoid)
  Hidden layer: (input = 128, output = 64, activation = sigmoid)
  Hidden layer: (input = 64, output = 32, activation = sigmoid)
  Output layer: (input = 32, output = 10, activation = softmax)
- The hyperparameters assumed other than those mentioned in the question are:
  trainable=True, use_bias=True, activation=`sigmoid` for all the layers other than the output layer where activation=`softmax`
- **Batch size=700, number of epochs=28 and learning rate=0.01**
- The softmax activation gives probabilities for all the ten classes. The class with maximum probability is assigned to the sample point.
- The testing accuracy: 0.8203

Question 3

- The autoencoder is built as given in the instructions.
- **The number of output nodes in the last layer of the network is 784** because it is the number of dimensions that the flattened image contains.
- **The loss used is Mean Squared Error**. Here we are encoding the image to a compressed form and then decoding it to reconstruct the original image, hence we use MSE because we are trying to predict the floating-point numbers.
- **The learning rate used in autoencoder is 0.001**
- The decoder consists of the last three layers mentioned in the instructions.
- In the classification model, the first three layers are kept as trainable=False because they have already been trained in the autoencoder.
- **The output layer contains 10** nodes because we have 10 classes and the activation used is softmax. The last two layers are kept as trainable=True because they are needed to be trained.
- **The learning rate used in the classification problem is 0.001**
- The testing accuracy: 0.8163

Question 4

- Here, implementing a bagging model with three decision trees.
- Three different datasets are made by randomly sampling with replacements from the original dataset.
- The three decision trees are trained over the three different datasets and the final prediction is done using majority voting.
- If the output is three different classes, the output of the first decision tree is considered.
- The testing accuracy: 0.9099