Deep Learning Assignment 01

Name: Bhabesh Prasad Rana
Roll No.: 21043
Department: DSE
Date of Submission: March 27, 2024

1 Question 1

The Cross Entropy Loss function works better with logistic regression compared to Mean Squared Error (MSE). This is because Cross Entropy Loss is specifically designed for classification tasks, such as logistic regression, where the goal is to predict probabilities for each class and minimize the difference between predicted probabilities and actual class labels.It penalizes the model more heavily for confidently incorrect predictions, which encourages the model to output well-calibrated probabilities. This is important in classification tasks because it helps reduce confusion by assigning higher confidence to correct predictions and lower confidence to incorrect predictions.

During the model training process, using Cross Entropy Loss leads to faster convergence compared to MSE, especially when the predicted probabilities are far from the true labels. This is because Cross Entropy Loss tends to produce larger gradients in such cases, which accelerates the optimization process and helps the model learn more efficiently. As a result, the model trained with Cross Entropy Loss is more likely to converge to the optimal solution faster compared to using MSE.

2 Question 2

Option C - Both

When all layers in a neural network employ linear activation functions, using either Cross-Entropy (CE) or Mean Squared Error (MSE) loss yields a convex optimization problem. In such a scenario, the neural network effectively behaves as a linear model since the linearity propagates throughout the network. Both CE and MSE loss functions become convex when applied to linear inputs.

If every layer in a neural network employs a linear activation function, the combination of such activation functions with either Cross-Entropy (CE) or Mean Squared Error (MSE) loss transforms the optimization problem into a convex one. This occurs because the entire neural network essentially operates as a linear model due to the linearity of its activation functions. Consequently, both CE and MSE loss functions exhibit convexity when dealing with linear inputs.

3 Question 3

For the task of classifying the MNIST dataset, a neural network model was devised comprising one input layer, one output layer, and a solitary hidden layer. Rectified Linear Units (ReLU) served as the chosen activation function. The number of neurons in the hidden layer was treated as a hyper-parameter, with an optimal range selected between 60 and 600, using increments of 60. The preprocessing is task and dataset specific which means that different methods is required for different datasets,. In the case of MNIST, pixel values were normalized to fall within the range of 0 to 1. Furthermore, the images were flattened and converted into vectors, each of size $28 \times 28 = 784$.

4 Question 4

1. LeNet-5: Its simplicity and efficiency make it applicable for the SVHN dataset. However, its performance may not match more complex models such as VGG or ResNet.

2. AlexNet: With a deeper architecture, AlexNet demands more computational resources compared to simpler counterparts like LeNet-5. Its superior test and train scores are assigned to its complexity.

3. VGG: Given its robust performance in image classification tasks, VGG proves suitable for the SVHN dataset. But, its deeper design may require increased computational resources compared to shallower alternatives like LeNet-5.

4. ResNet: Renowned for its state-of-the-art performance in image classification, ResNet emerges as a highly viable choice for the SVHN dataset. However, its deeper architecture may demand essential computational resources during training.