Deep Learning Assignment 1

Name: Prasanna Kharode
Registration No./Roll No.: 21144
Institute/University Name: IISER Bhopal
Program/Stream: DSE
Date of Submission: March 26, 2024

1 Question 1

Cross Entropy Loss is a natural choice for logistic regression as we get cross entropy loss by maximizing negative log likelihood of bernoulli disribution, cross-entropy loss aligns well the maximum likelihood estimation principle commonly used in statistical modeling.
Cross Entropy Loss penalizes the model heavily for confidently incorrect predictions, encouraging well-calibrated probabilities and reducing confusion in classification tasks.
Cross-entropy loss tends to produce larger gradients compared to MSE, especially when the predicted probability is far from the true label. This is crucial for the optimization process during training. Larger gradients lead to faster convergence, which means the model learns more efficiently and reaches the optimal solution faster.

2 Question 2

Option C - Both
If the activation function is linear for all layers, then the combination of linear activation functions with either cross-entropy (CE) or mean squared error (MSE) loss results in a convex optimization problem.
If the activation function is linear the entire neural network becomes a linear function. Both the Cross-Entropy (CE) and Mean Squared Error (MSE) loss functions are convex when their inputs are linear.

3 Question 3

Here MNIST dataset was used for classification.
The model constructed consist of 1 input layer, 1 output layer and a single hidden layer. The activation function used is relu(Rectified Linear Units). Number of neurons in hidden layer was treated as a hyper-parameter and an optimal value between 60-600 was selected with a step size of 60. The preprocessing is task and dataset specific which means that differnt methods would be required for different datasets, since we were working on MNIST dataset - all pixel values were normalized between 0-1 and also the image to flattened and was converted into a vector or size 28*28 = 784. The code implementation is available on GitHub at the following link: GITHUB LINK

4 Question 4

The suitability of a model for a particular dataset depends on various factors such as model complex ity, dataset characteristics, and computational resources.

LeNet-5 : It can be used for the SVHN dataset due to its simplicity and efficiency. However, it may not achieve great performance compared to more complex models like VGG or ResNet. It achieves a test accuracy of 30% .

AlexNet : It’s deeper architecture requires more computational resources compared to simpler models like LeNet-5. Its test and train score are better than LeNet5, this is because it is a more complex model.

VGG : VGG can be well suited for the SVHN dataset due to its strong performance on image classifica tion tasks and its ability to capture intricate features from the data. However, its deeper architecture may require more computational resources compared to shallower models like LeNet-5.

ResNet : ResNet can be highly suitable for the SVHN dataset due to its state-of-the-art perfor mance on image classification tasks and its ability to effectively capture intricate features. However, its deeper architecture may require significant computational resources for training.

The code implementation is available on GitHub at the following link: GITHUB LINK