# Deep Learning Assignment - 1

Name : Ambar Shingade

Roll no. : 21250

Institute : IISER Bhopal

Program : DSE

Date of submission : March 26, 2024

## Question 1

Cross-entropy loss emerges as a natural fit for logistic regression due to its foundation in maximizing the negative log likelihood of a Bernoulli distribution. This aligns seamlessly with the maximum likelihood estimation principle, a cornerstone of statistical modeling.

One of the key strengths of cross-entropy loss lies in its ability to heavily penalize the model for confidently incorrect predictions. This incentivizes the model to produce well-calibrated probabilities and minimizes confusion in classification tasks.

Compared to mean squared error (MSE), cross-entropy loss typically generates larger gradients, particularly when the predicted probability significantly deviates from the true label. This characteristic is pivotal for the optimization process during training. By yielding larger gradients, cross-entropy loss facilitates faster convergence, enabling the model to learn more efficiently and converge to the optimal solution more rapidly.

## Question 2

Option C - Both

When employing linear activation functions across all layers, the integration of either cross-entropy (CE) or mean squared error (MSE) loss yields a convex optimization problem.

Given that the activation function remains linear throughout the neural network, the entire network essentially operates as a linear function. In this context, both the Cross-Entropy (CE) and Mean Squared Error (MSE) loss functions demonstrate convexity when their inputs are linear.

## Question 3

In this project, the classification task utilized the MNIST dataset. The constructed model comprises one input layer, one output layer, and a single hidden layer. The activation function employed is ReLU (Rectified Linear Units). The number of neurons in the hidden layer was considered a hyperparameter, and an optimal value within the range of 60 to 600 was selected with a step size of 60.

Preprocessing steps were tailored specifically for the task and dataset. Since the MNIST dataset was used, all pixel values were normalized to fall within the range of 0 to 1. Additionally, the images were flattened and converted into vectors of size 28*28 = 784.

The code implementation can be accessed on GitHub via the following link:
https://github.com/ambar302003/Deep-Learning.git

## Question 4

The suitability of a model for a specific dataset hinges on several factors, including model complexity, dataset characteristics, and available computational resources.

LeNet-5: While LeNet-5 offers simplicity and efficiency, it may not achieve optimal performance compared to more intricate models like VGG or ResNet. It typically attains a test accuracy of around 30%.

AlexNet: With its deeper architecture, AlexNet demands more computational resources compared to simpler models like LeNet-5. Its performance in both training and testing tends to surpass that of LeNet-5 due to its increased complexity.

VGG: VGG exhibits strong performance in image classification tasks and excels at capturing detailed features from data. However, its deeper architecture may necessitate greater computational resources compared to shallower models such as LeNet-5.

ResNet: Renowned for its state-of-the-art performance in image classification, ResNet is highly suitable for datasets like SVHN. It effectively captures intricate features, although its deeper architecture may require substantial computational resources for training.

The code implementation can be accessed on GitHub via the following link:
https://github.com/ambar302003/Deep-Learning.git