# Deep Learning Assignment

Vikas Rajpoot(21306)

March 27, 2024

## Question 1

In logistic regression, the cross-entropy loss function is preferred over mean squared error (MSE) for several reasons.

Logistic regression is commonly used for binary classification problems, where the target variable has two possible outcomes. The logistic function (also known as the sigmoid function) is used to model the probability that a given input belongs to one of the classes.

The lower the cross-entropy loss, the better the model's predictions align with the true labels. In contrast, MSE doesn't directly relate to the probabilities. It's more suited for regression tasks where the output is continuous and measuring the squared difference between predicted and actual values is meaningful.

Cross-entropy loss leads to smoother gradients during optimization compared to MSE, especially when used in conjunction with the logistic activation function. In logistic regression, optimizing cross-entropy loss through techniques like gradient descent tends to converge faster and more reliably to the global optimum due to its convex nature.

Cross-entropy loss guarantees a single best solution due to its convexity. There's no ambiguity in the optimization process, and the model converges to a unique set of parameters that minimize the loss function. With MSE, especially in logistic regression where it's not the ideal loss function, there might be multiple local minima, leading to potential confusion and suboptimal solutions.

## Question 2

In a binary classification task with a deep neural network equipped with linear activation functions, the loss function that guarantees a convex optimization problem is (b) Mean Squared Error (MSE).

In binary classification, let y denote the target label (either 0 or 1) and $\hat{y}$ denote the predicted output of the model. As the output of the model is a linear function of the inputs, the MSE loss function becomes a quadratic function of the parameters (weights and biases). Quadratic functions are convex, meaning they have a single global minimum. Thus, using MSE as the loss function guarantees convexity in the optimization problem.

While Cross Entropy loss is commonly used for binary classification, it's not guaranteed to result in a convex optimization problem when combined with linear activation functions. The nonlinearity introduced by the logarithm function makes the loss function non-convex, and the optimization problem may have multiple local minima.

## Question3

The FNN architecture consists of three hidden layers followed by an output layer. The specifications are as follows:

Number of Hidden Layers: 3 Neurons per Layer: First Hidden Layer: 512 neurons Second Hidden Layer: 256 neurons Third Hidden Layer: 128 neurons Activation Functions: ReLU (Rectified Linear Unit) Softmax activation function is used for the output layer to produce class probabilities.

3. Preprocessing of Input Images: For the CIFAR-10 dataset, the following preprocessing steps are applied:

Normalization: Flattening: Each 32x32x3 image (RGB channels) is flattened into a 1D array of length 32x32x3=3072. This converts the input data into a format suitable for fully connected layers.

4. Hyperparameter Tuning Strategies: Hyperparameter tuning plays a vital role in optimizing the performance of the neural network. Several strategies are considered for tuning hyperparameters:

Batch Size: Grid search or random search can be used to explore different batch sizes. Number of Epochs: The number of epochs defines how many times the entire training dataset is passed through the network. Early stopping can be employed to prevent overfitting and determine the optimal number of epochs.

Regularization: Techniques like L2 regularization or dropout can be applied to prevent overfitting and improve generalization performance.

Network Architecture: Experiment with different architectures, including the number of hidden layers, neurons per layer, and activation functions. Techniques like model ensembling can also be explored to combine multiple models for improved performance.

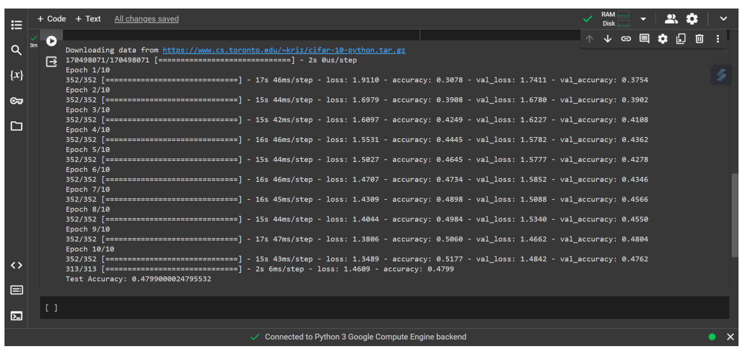


Figure 1: Accuracy

## Question4