Sameer Verma 21236

Answer 1 - Cross Entropy loss will work better with Logistic regression due to its ability to provide a clear and meaningful measure of how well the model is performing. The goal of logistic regression is typically to find the parameters that maximize the likelihood of the observed data given the model. Cross Entropy loss directly reflects this likelihood and minimizes it during training. Additionally, because the logistic regression model outputs probabilities, Cross Entropy loss aligns well with the probabilistic nature of the problem.

In the training process, using Cross Entropy loss with logistic regression ensures that the model learns to make more accurate probability predictions for binary classification tasks. It guides the optimization process towards finding the parameters that result in the best discrimination between the classes, thus leading to better classification performance. whereas MSE is more suited for regression tasks where the output is continous as compared to classification tasks.

Answer 2 - (c) both, MSE and CE are convex functions and all linear functions are convex as well. Therefore, when combined, the objective function (CE loss or MSE loss) and the feasible region (defined by the linear transformations) both remain convex. This ensures that the optimization problem is convex.

Answer 3-

Number of Hidden Layers: The model consists of three hidden layers.

Neurons in Each Layer:

- Hidden Layer 1: 128 neurons
- Hidden Layer 2: 64 neurons
- Hidden Layer 3: 32 neurons

Activation Functions:

- Hidden Layer 1, Hidden Layer 2, and Hidden Layer 3: ReLU activation function is used.
- Output Layer: Softmax activation function is used since it's a multi-class classification problem.

Preprocessing:

- Pixel values are normalized by dividing by 255.0 to scale them between 0 and 1.
- Images are flattened from 28x28 arrays to 784-dimensional vectors using `reshape(-1, 28 * 28)`.

Hyperparameter Tuning Strategies:

- The model is compiled with the Adam optimizer, which is commonly used for its adaptive learning rate properties.
- Sparse categorical cross-entropy loss function is chosen for multi-class classification problems where the target labels are integers.
- The number of epochs for training is set to 10, which may be subject to tuning based on validation performance to prevent overfitting or underfitting.
- A validation split of 0.2 is used during training to monitor model performance on a portion of the training data.