1 Question 1

Which loss function, out of Cross Entropy and Mean Squared Error, works best with logistic regression because it guarantees a single best answer (no room for confusion)? Explain why this is important and maybe even show how it affects the model's training process.

In logistic regression, the Cross Entropy loss function (also known as Log Loss) works best. This is because logistic regression is commonly used for binary classification problems where the output is a probability estimate between 0 and 1. The Cross Entropy loss function is specifically designed to measure the difference between two probability distributions, making it well-suited for this scenario.

Cross Entropy loss ensures that the model's predicted probabilities are as close as possible to the actual probabilities (0 or 1 for binary classification). It penalizes the model more heavily when it is confidently wrong, leading to sharper updates during training. This is crucial because logistic regression aims to maximize the likelihood of the observed data, and using Cross Entropy loss directly optimizes this likelihood.

In contrast, Mean Squared Error (MSE) is more commonly used in regression problems where the output is a continuous value. While it could technically be used in logistic regression, it may not be as effective because it does not directly optimize the likelihood of the observed data in a binary classification scenario. MSE might lead to slower convergence or suboptimal solutions due to its emphasis on minimizing squared errors, which may not align with the goals of logistic regression.

2 Question 2

For a binary classification task with a deep neural network (containing at least one hidden layer) equipped with linear activation functions, which of the following loss functions guarantees a convex optimization problem? Justify your answer with a formal proof or a clear argument. (a) CE (b) MSE (c) Both (A) and (B) (d) None

For a binary classification task with a deep neural network equipped with linear activation functions, neither Cross Entropy (CE) nor Mean Squared Error (MSE) loss functions guarantee a convex optimization problem.

Here's the reasoning:

1. Cross Entropy Loss (CE): CE loss is commonly used for classification tasks and is convex when combined with a linear activation function for binary classification. However, when used with deep neural networks, the presence of non-linear activation functions (such as ReLU, sigmoid, tanh) in hidden layers makes the overall optimization problem non-convex. The composition of multiple non-linear functions results in a non-convex loss landscape.

2. Mean Squared Error Loss (MSE): MSE is convex only when used in linear regression problems. However, in binary classification tasks with linear activation functions, it is not suitable because it does not reflect the probabilistic nature of classification problems. Moreover, when combined with linear activation functions in deep neural networks, the presence of multiple layers introduces non-linearity, leading to a non-convex optimization problem.

3. Conclusion: Neither CE nor MSE loss functions guarantee convex optimization when used in binary classification tasks with deep neural networks containing at least one hidden layer with linear activation functions. The non-linearity introduced by hidden layers makes the optimization problem non-convex.

Therefore, the correct answer is (d) None.

-