Questions 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.

Answer: The Cross Entropy loss function is typically preferred over Mean Squared Error (MSE) for logistic regression. The reason for this preference lies in the nature of logistic regression and the characteristics of these loss functions.

Logistic regression is commonly used for binary classification tasks, where the goal is to predict whether an instance belongs to one class or another. The logistic regression model applies a logistic function (sigmoid) to the linear combination of input features, producing probabilities between 0 and 1. These probabilities are then thresholded to make binary predictions.

Cross Entropy loss, also known as Log Loss, is specifically designed for classification tasks and is particularly well-suited for logistic regression. It measures the dissimilarity between the predicted probabilities and the actual labels. By penalizing incorrect predictions more severely, it encourages the model to output high probabilities for the correct class and low probabilities for the incorrect class. This property is crucial for classification tasks because it directly aligns with the objective of minimizing misclassifications.

On the other hand, Mean Squared Error (MSE) is more commonly associated with regression tasks, where the goal is to predict continuous values. While it can be used for logistic regression, it is less appropriate because it doesn't directly address the probabilistic nature of the predictions. MSE loss tends to penalize large deviations heavily, which may not be ideal for classification problems where we're interested in the probability distribution of the classes.

In terms of the training process, using Cross Entropy loss typically leads to faster convergence and better performance compared to MSE in logistic regression. This is because Cross Entropy loss provides stronger gradients when the predicted probabilities are far from the true labels, leading to more effective updates to the model parameters during training. Moreover, Cross Entropy loss guarantees a single best answer due to its focus on probabilistic predictions, reducing ambiguity and confusion in the optimization process.

In summary, for logistic regression, Cross Entropy loss is preferred over Mean Squared Error because it is tailored for classification tasks, aligns with the probabilistic nature of the predictions, and leads to better training convergence and performance.