Key Differences:

- **MSE**:
  - **For large errors**: The gradient is large, so the weight updates are large, speeding up convergence initially.
  - **For small errors**: The gradient becomes small as the error shrinks, leading to slower convergence because the updates become smaller.
- **Log Loss**:
  - **For large errors** (incorrect predictions): The gradient is very large, leading to **faster convergence** initially by making large corrections to the weights.
  - **For small errors** (near the optimal solution): The gradient is small, but it still provides **faster convergence** because the model is making **small, precise updates** as it gets closer to the correct classification.

In Summary:

- **MSE**: Slows down as the error reduces (smaller updates for small errors).
- **Log Loss**: Continues to update effectively as the model moves toward a correct prediction (small updates, but effective convergence).

Revised Table:

Aspect	MAE (Mean Absolute Error)	MSE (Mean Squared Error)	Log Loss (Binary Cross-Entropy)
Gradient Descent	Constant gradients (± 1)	Gradients depend on the size of the error (larger errors = larger updates)	Gradients are large for incorrect predictions, small for correct ones
Loss Function	Absolute error between actual and predicted values	Squared error between actual and predicted values	Binary cross-entropy between actual and predicted probabilities
Cost Function	Average of absolute errors	Average of squared errors	Average of binary cross-entropy loss
Convergence for Small Errors	Slower convergence (constant gradient)	Slower convergence (smaller gradients for smaller errors)	Faster convergence (small gradients near the minimum for correct predictions)
Convergence for Large Errors	Slower convergence (constant gradient)	Faster convergence (larger gradients for larger errors)	Faster convergence (steep gradients for incorrect predictions, larger updates)

Aspect	MAE (Mean Absolute Error)	MSE (Mean Squared Error)	Log Loss (Binary Cross-Entropy)
Used For	Regression (robust to outliers)	Regression (sensitive to outliers)	Binary classification (e.g., NLP, sentiment analysis)

Conclusion:

While both **MSE** and **Log Loss** exhibit faster convergence for large errors, **MSE** slows down significantly for small errors, while **Log Loss** continues to make effective progress with small, precise updates, thus allowing for **more efficient convergence near the optimal solution**.