QS-1: Define binary cross entropy as a cost function

Answer: Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1. Because for categorical prediction or binary classification, the predicted output must be bound between [0,1] and this function calculate the loss in this bound as the result is either 0 or 1.
If we try to train a binary classification then there is a good chance to use binary cross-entropy / log loss as a loss function.

Loss Function: Binary Cross-Entropy / Log Loss

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log (1 - \hat{Y}_i))$$

QS-2: Dierence between adam optimizer and gradient descent

Answer:
Gradient Descent:
Suppose we have a convex cost function and our goal is to minimize its value and find the value of the parameters (x,y) for which f(x,y) is minimum. What the gradient descent algorithm does is, start at a specific point (initially guessed) on the curve and use the first derivative or gradient of the function in that particular point

to check if is it the minimum (close to 0). Then with a learning rate or an extensive value of propagation, we find the direction of the steepest descent and take a small step in that direction and keep iterating till our value starts converging.

Adam Optimizer:

Adaptive Moment Estimation is an algorithm for optimization techniques for gradient descent. The method is really efficient when working with large problems involving a lot of data or parameters. It requires less memory and is efficient. Intuitively, it is a combination of the ‘gradient descent with the momentum algorithm and the ‘RMSP’ algorithm.