Classification Problem :-

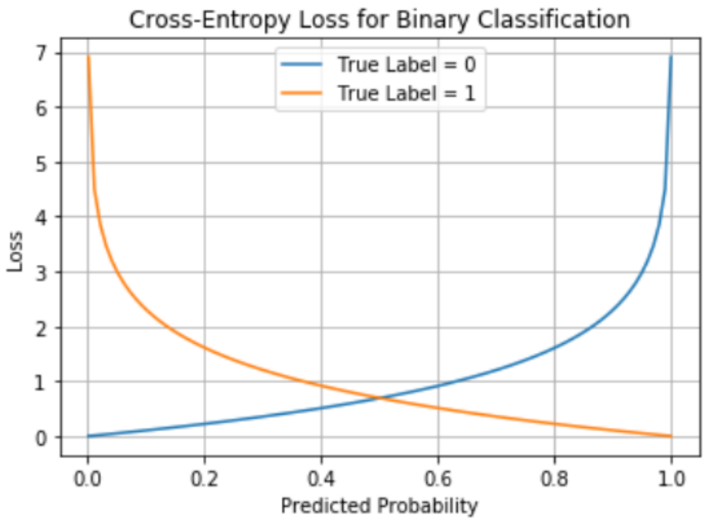
Let's take a look at binary cross entropy loss function,

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

Here Y_i is the ground truth and $\hat{Y}_i$ is the predicted label.

If Y_i = 1, then $L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n} \log \hat{Y}_i$

If Y_i = 0,then $L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n} \log(1 - \hat{Y}_i))$



This can be inferred from the graph above:-

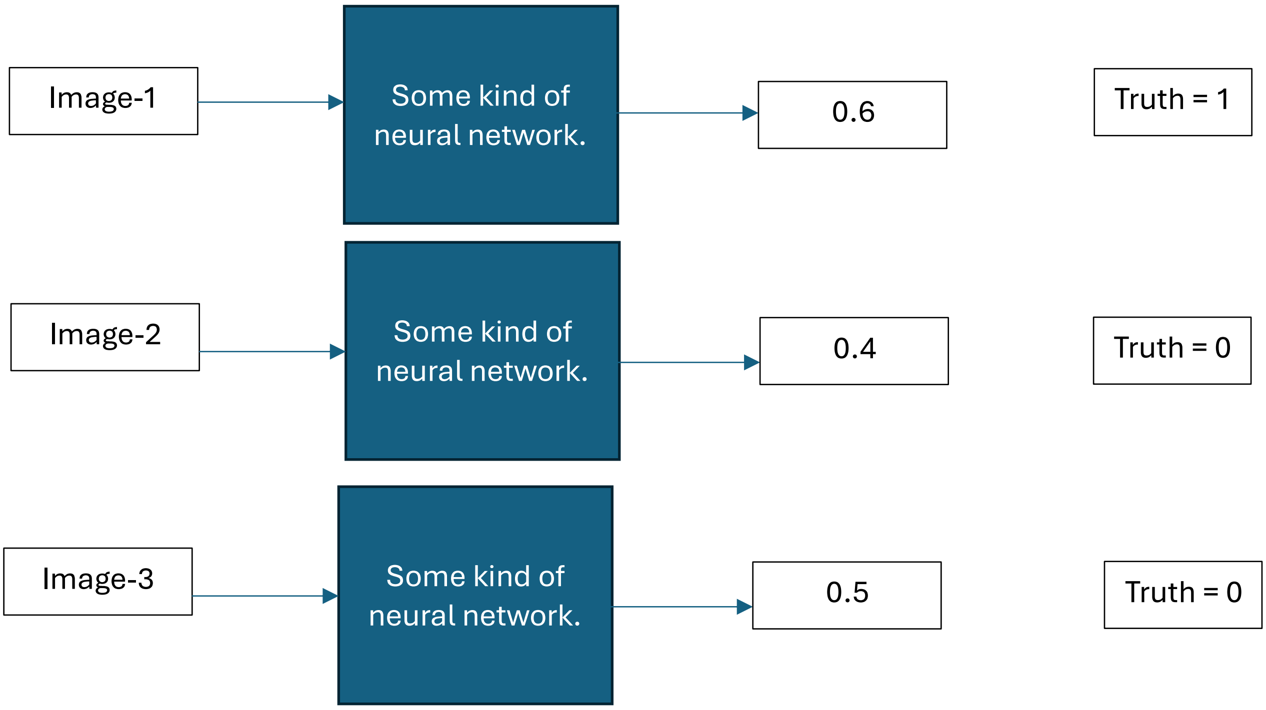
When **True Label = 0 if we predict 1 we get a very large loss and if we predict 0 we get 0 loss.**
When **True Label = 1 if we predict 1 we get a 0 loss and if we predict 1 we get a very large loss.**

Let's say we are classifying whether image is fake or real,

Data looks something like this

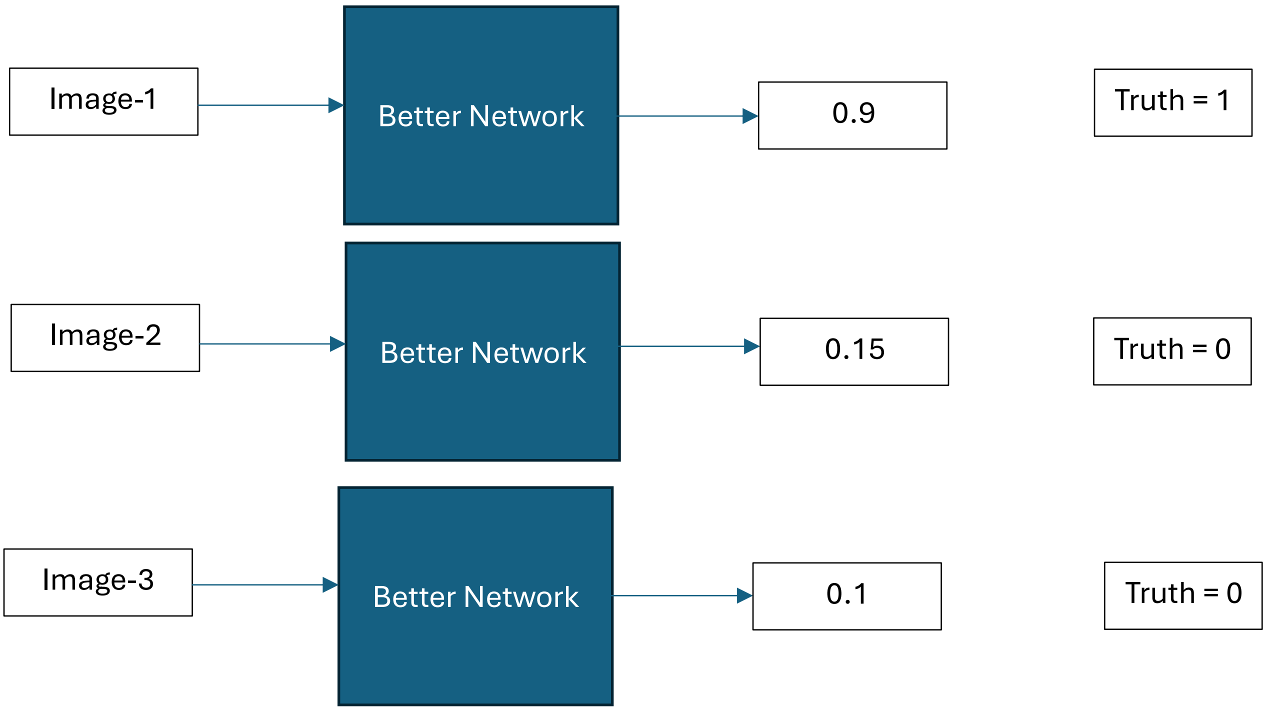
Data	**Label**
Image-1	1
Image-2	0
Image-3	0

Let's feed images into our imaginary network initially as you can see the output is not that good:-



BCE (initially) = $-\frac{1}{3}((1 * \log(0.6) + 0 * \log(0.4) + 0 * \log(0.3) + \ (0 * \log(1-0.6) + 1 * \log(1-0.4) + 1 * \log(1-0.3))$ = 0.199

For a better classifier :-



BCE (for this better network) = $-\frac{1}{3}((1 * \log(0.9) + 0 * \log(0.15) + 0 * \log(0.1) + \ (0 * \log(1-0.9) + 1 * \log(1-0.15) + 1 * \log(1-0.1))$ = **0.054**

BCE decreases for better network , so while training the weights and biases changes to decrease our BCE hence giving better classification.