NEGATIVE LOG LOSS FUNCTION OR CROSS ENTROPY

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

DON'T BE AFRAID, IT IS VERY SIMPLE. JUST BE SYSTEMATIC.

Well let's take an example.

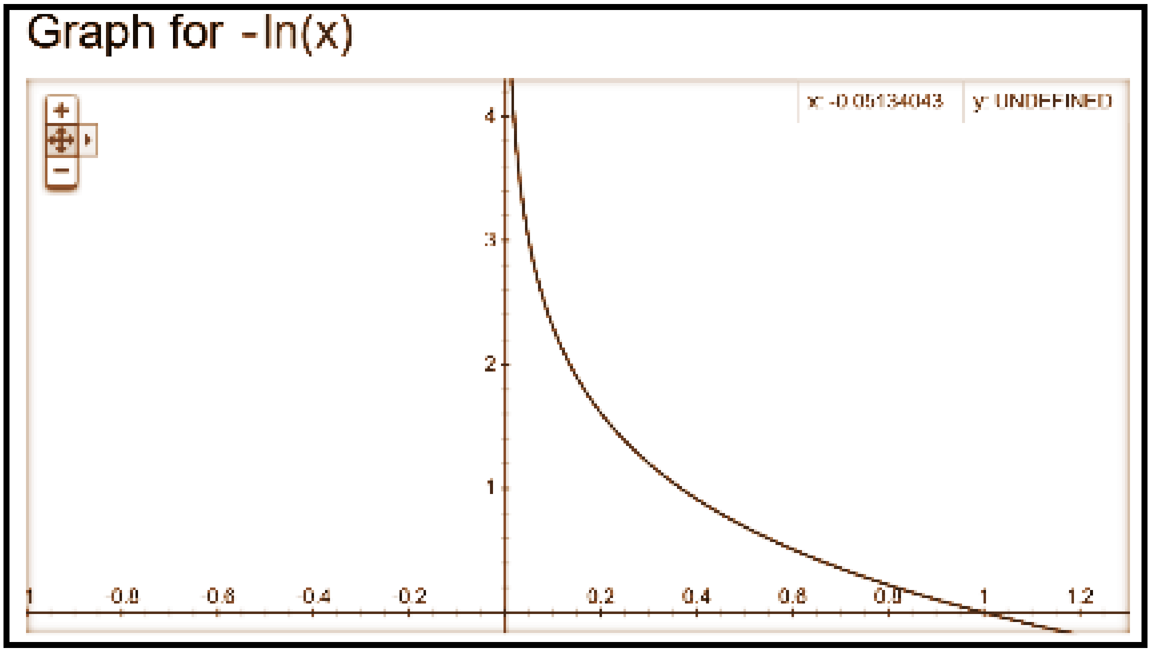
Let us suppose we want to classify **dogs and cats. Let us define 1 as cat and 0 as dogs.**

Now we define Pi as the **probability to classify a data point as a CAT. Let us say P is [0.9,0.1,0.2,0.8] . This means** for the data points 1 and 4 we are **90% sure it is a cat and 80% sure it is a cat respectively.**

For the data points **2 and 3,** we say that we are **10%sure it is a cat or 90% sure it is a dog and similarly that we are 80% sure it is a dog.**

Basis of our loss function is that - The lesser it is, the better it is. Let us analyse why the negative sign.

The below is the graph of -log(x) with base e.



Note that **we are using probabilities. They lie in range [0,1].** Now see that if we have a binary data point ,namely data point 1 here, its y value is the value **1 and the p(y) value is 0.9.** This makes the graph **tend towards 0 and thus A LOWER VALUE. This means a higher probability leads to a lower loss.** This is what we wanted. **The above was in fact the explanation of this term->**

$$y_i \cdot log(p(y_i))$$

with of course a **negative sign.**

Now, if our binary data point is **0 then the above term would be deactivated as y is zero.** But this means that the second term of sum which is $\boxed{(1 - y_i) \cdot log(1 - p(y_i))}$ gets **activated.**

Now, 1-Yi becomes **1** and (**1-p(y))** becomes actually the **probability of classifying the data point as 0. Again, it is in line with our objective of defining a loss function which was the higher the probability , the lower the loss.** Ofcourse, the negative sign has to be included here also.

Now , if we take the average of ALL THESE LOSSES , WE FINALLY GET THE NEGATIVE LOG LIKELIHOOD LOSS.

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

Neat! Not so confusing now.