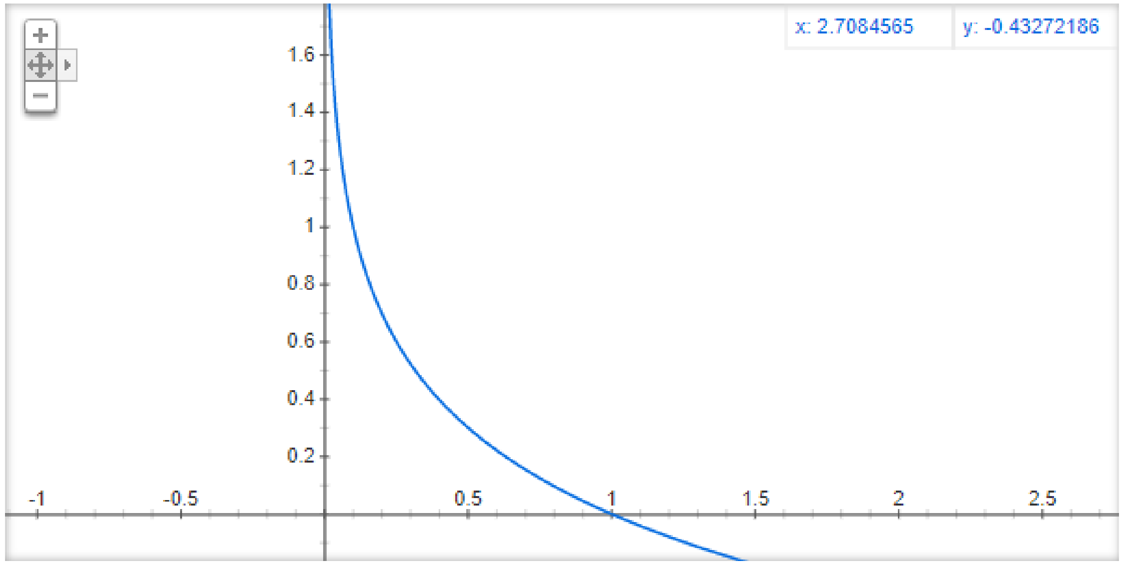
Entropy

Prithvi Shah

ML Layman Definition

- Entropy is the randomness in the information being processed in your data. It is also known as the **average** information content within the data.
- To understand entropy, intuitively, from a purely formal standpoint, lets first start with what we want information to look like –
  - Low probability events have high information and high probability events have low information
  - Information from two events x & y completely unrelated to each other - should be the sum of the individual information GAINED from each system

Graph for -log(x)



From the above 2 conditions, it is straightforward to show that the logarithm of a probability distribution is the information gained

INFORMATION:

$$h[x] = -\log_2 p(x)$$

ENTROPY:

$$H[x] = -\sum p(x)\log_2 p(x)$$

Longer Derivation of Entropy

Let's calculate <u>multiplicity</u> first (is the total number of "ways" in which different outcomes can possibly occur.)

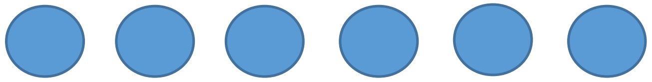
- Consider allocating N identical objects to a set of bins such that there are n_i objects in the i^{th} bin.
- The multiplicity of this system can be defined as -

$$W = \frac{N!}{\prod_i n_i!};$$

Microstates = n_i

Macrostates = $\frac{n_i}{N}$

N objects



Alternatively, we can also derive this by looking at it through combinatorics.

- If bin 1 were to have 1 object, bin 2 were to have 2 objects and so on, then number of ways of placing an object in the i^{th} bin is –

$${}^{N}C_1 * {}^{(N-1)}C_2 * {}^{(N-3)}C_3 * \ldots {}^{(N-i)}C_i$$

Simplified =

$$\frac{N!}{1!\,(N-1)!} \cdot \frac{(N-1)!}{2!\,(N-3)!} \cdot \frac{(N-3)!}{3!\,(N-6)!} \ldots$$

This is simplified further as -

$$= \frac{N}{1!} \cdot \frac{(N-1)(N-2)}{2!} \cdot \frac{(N-3)(N-4)(N-5)}{3!}$$

$$= \frac{N!}{\prod_i n_i!} = W$$

Longer Derivation of Entropy (cont)

- Entropy is the log of the multiplicity scaled by an appropriate constant

$$H = \frac{1}{N}\log W \Rightarrow \frac{1}{N}\log\frac{N!}{\prod n_i!}$$

To derive this, we must use, Stirling's approximation.

$$\log N! = N\ log\ N\ - N$$

Lets derive Stirling's approximation first – on the right -

$\log N! = \ \sum_k \ \log k$; k = 1 to N

From RHS,

Log N! = $\int_1^N \log x\ .\ dx$ = N log N –N

∫f(x).g(x).dx = f(x).∫ g(x).dx - ∫(f'(x) .∫ g(x).dx).dx

∫ logx.dx = ∫ logx.1.dx

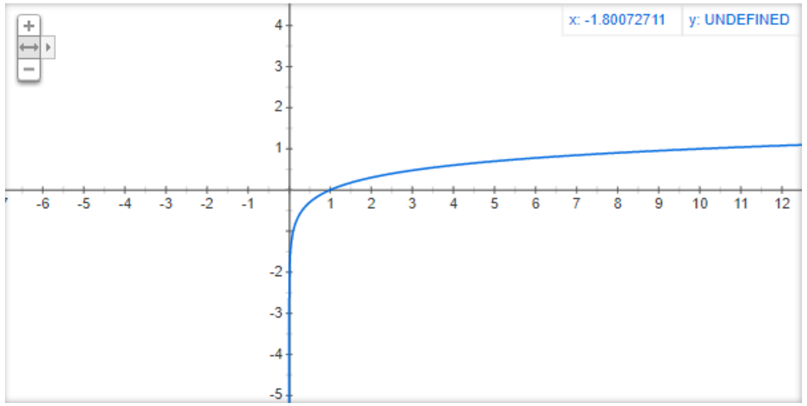
= logx. ∫1.dx - ∫ ((logx)'.∫ 1.dx).dx

=logx.x -∫ (1/x .x).dx

=xlogx - ∫ 1.dx

=x logx - x + C

Graph of log x



The value of log x is the point on the blue curve for every value of x. The integral of log x is the area under the curve. We can approximate this integral by using Reimann sums i.e. discrete M buckets under the curve

$$\int_1^N \log x\ .\ dx = \sum_1^M \log x_k\ .\Delta x$$

$$\Delta x\ = \frac{M}{(N-1)}\ ;\ \ x_k\ = \ 1 + (k).\Delta x = 1 + k\left(\frac{M}{N-1}\right)$$

$$\int_1^N \log x\ .\ dx = \ \sum_1^N \log k\ \ = RHS$$

Longer Derivation of Entropy (cont)

$$H = \frac{1}{N}\log W \Rightarrow \frac{1}{N}\log\frac{N!}{\prod n_i!}$$

$= \frac{1}{N}(\log N! - \sum_i \log n_i!)$

$= \frac{1}{N}\ [(N\log N - N) - \sum_i (n_i \log n_i - n_i)\ ]$

= log N -1 - $\frac{1}{N}\sum_i n_i \log n_i + \frac{1}{N}\sum_i n_i$; $(p_i = \frac{n_i}{N})$

= log N - $\frac{1}{N}\sum_i p_i N \log p_i N$

= log N - $\frac{1}{N}\sum_i p_i N \log N$ - $\frac{1}{N}\sum_i p_i N \log p_i$

= - $\sum_i p_i \log p_i$ = ENTROPY