## **Sigmoid Neuron and Cross Entropy**

How does it all tie up to the Sigmoid Neuron

1. Consider the Example:
   1. A signboard with the text **Mumbai**
   2. A random variable X which maps the signboard to: Text, No-Text
   3. The distributions are as follows

|  |  |  |
| --- | --- | --- |
| **X** | **y (We don’t know initially)** | **ŷ (Predicted using sigmoid)** |
| T | 1 | 0.7 |
| NT | 0 | 0.3 |

* 1. Previously, we were using **Squared-error Loss** =
  2. Now, we have a better metric, one that is grounded in probability theory (**KL- Divergence**)
  3. KLD(y||ŷ) =
  4. We aim to minimize loss by KLD with respect to the parameters w, b
  5. From KLD equation, we can see that yi doesn’t depend on w, b. So therefore, we are really only trying to minimize the first term, i.e. the cross-entropy
  6. So in practice, we can treat the second term as a constant, and the equation would really be
  7. The second term cancels out and we are left with which is the same as
  8. It can be called where c can take the value 0 or 1 which correspond to NT and T