PadhAI: Information Theory and Cross Entropy

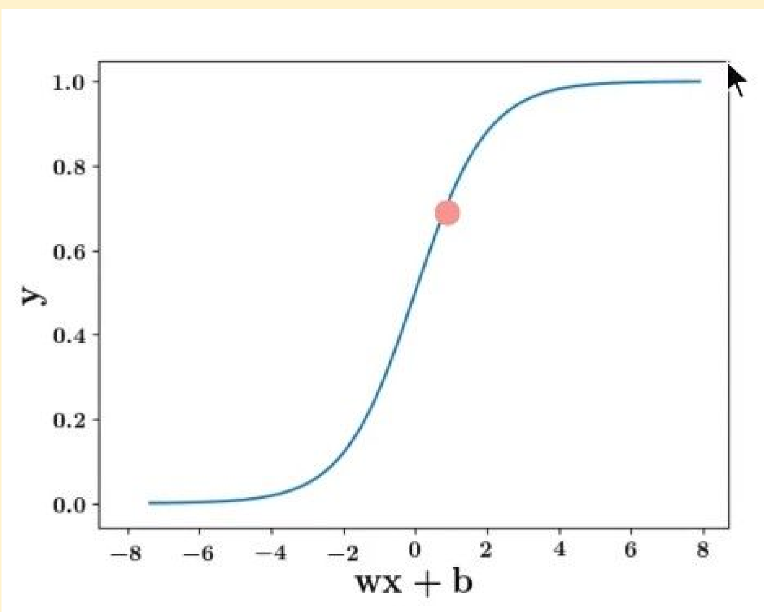
One Fourth Labs

Using Cross Entropy With Sigmoid Neuron

What does the cross entropy loss function look like

1. Consider an example in the scope of our final project
2. Look at the following signboard

3. $x = image, \quad y = [0, 1]$ (True distribution, where 1 corresponds to Text)
4. $\hat{y} = \frac{1}{1+e^{(-(w.x + b))}}$
5. This corresponds to $\hat{y}$ =0.7



6. Thus, the predicted distribution is $\tilde{y} = [0.3, \; 0.7]$ (where 0.7 corresponds to Text)
7. The Loss function is $L(\theta) = -\Sigma_i y_i log \tilde{y}_i$ where $i \in \{0, 1\}$
8. $L(\theta) = -((y_o \, log \; \tilde{y}_o) + (y_1 log \; \tilde{y}_1))$

---

9. $L(\theta) = -((y_o \log(1-\tilde{y}_1)) + (y_1 \log \tilde{y}_1))$ (from probability axioms, $y_0 = 1 - y_1$)
10. Consider two examples side by side

Training Data	Image	$L(\theta) = -((y_o \log(1-\tilde{y}_1)) + (y_1 \log\ ))$	Loss function
y = [0, 1] $\hat{y} = 0.7$ $\tilde{y} = [0.3, 0.7]$ (Text)		$L(\theta) = -(0 * \log(0.3)) + (1 * \log(0.7)))$ $L(\theta) = -\log(0.7)$	$L(\theta) = -\log(\hat{y})$ When true output is 1
y = [1, 0] $\hat{y} = 0.2$ $\tilde{y} = [0.8, 0.2]$ (No-Text)		$L(\theta) = -(1 * \log(0.8)) + (0 * \log(0.2)))$ $L(\theta) = -\log(0.8)$	$L(\theta) = -\log(1-\hat{y})$ When true output is 0

11. The Loss function can be expressed as follows
    a. $L(\theta) = -\log(\hat{y})$ if y = 1
    b. $L(\theta) = -\log(1-\hat{y})$ if y = 0
    c. Combining them and removing the if conditions:
    d. $L(\theta) = -[(1-y)\log(1-\hat{y}) + y\log(\hat{y})]$
       i. When y = 1, the first term becomes 0
       ii. When y = 0, the second term becomes 0