

## Logistic Regression: Cost Function

To train the parameters  $w$  and  $b$ , we need to define a cost function.

Recap:

$$\hat{y}^{(i)} = \sigma(w^T x^{(i)} + b), \text{ where } \sigma(z^{(i)}) = \frac{1}{1 + e^{-z^{(i)}}}$$

$x^{(i)}$ the $i$ -th training example
--

Given  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ , we want  $\hat{y}^{(i)} \approx y^{(i)}$

Loss (error) function:

The loss function measures the discrepancy between the prediction ( $\hat{y}^{(i)}$ ) and the desired output ( $y^{(i)}$ ). In other words, the loss function computes the error for a single training example.

$$L(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{2} (\hat{y}^{(i)} - y^{(i)})^2$$

$$L(\hat{y}^{(i)}, y^{(i)}) = -(y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}))$$

- If  $y^{(i)} = 1$ :  $L(\hat{y}^{(i)}, y^{(i)}) = -\log(\hat{y}^{(i)})$  where  $\log(\hat{y}^{(i)})$  and  $\hat{y}^{(i)}$  should be close to 1
- If  $y^{(i)} = 0$ :  $L(\hat{y}^{(i)}, y^{(i)}) = -\log(1 - \hat{y}^{(i)})$  where  $\log(1 - \hat{y}^{(i)})$  and  $\hat{y}^{(i)}$  should be close to 0

Cost function

The cost function is the average of the loss function of the entire training set. We are going to find the parameters  $w$  and  $b$  that minimize the overall cost function.

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})]$$