

Data Set (whole data given)

$$x_i, y_i \quad i = 1 \dots m_{total}$$

training set

(%70 of data set randomly selected)

test set

(%30 of data set randomly selected)

$$(X_{training}(i), y_{training}(i)) \\ i = 1, \dots, m_{training}$$

$$(X_{test}(i), y_{test}(i)) \\ i = 1, \dots, m_{test}$$

train your hypothesis
minimizing training
error

$$J(\theta)$$

$\theta_{trained}$

calculate test set error

$$J_{test}(\theta) = \frac{1}{2m_{test}} \sum_{i=1}^{m_{test}} \left(h_{\theta}(x_{test}^{(i)}) - (y_{test}^{(i)}) \right)^2$$

$$J_{test}(\theta) = -\frac{1}{m_{test}} \sum_{i=1}^{m_{test}} \left[y_{test}^{(i)} \log(h_{\theta}(x_{test}^{(i)})) + (1 - y_{test}^{(i)}) \log(1 - h_{\theta}(x_{test}^{(i)})) \right]$$

$$test_error = \frac{1}{m_{test}} \sum_{i=1}^{m_{test}} \left(err(h_{\theta}(x_{test}^{(i)}), (y_{test}^{(i)})) \right)$$

$$err(h_{\theta}(x), y) = \begin{cases} 1, & \text{if } h_{\theta} \geq 0.5, y = 0 \text{ or } h_{\theta} < 0.5, y = 1 \\ 0, & \text{otherwise} \end{cases}$$