Logistic Regression Cost Function and Overfitting

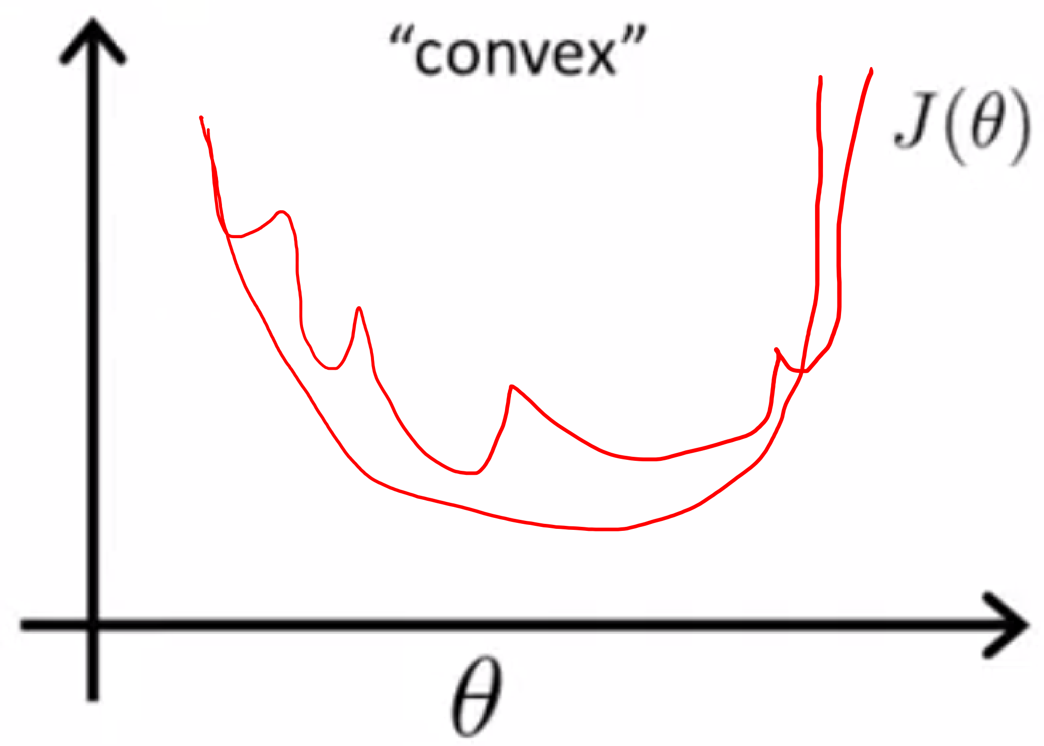
Dr. Muhammad Wasim

Cost Function for Linear Regression

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h(x) - y)^2$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}(h(x) - y)^2$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\boxed{Cost(h(x^{(i)}), y^{(i)})}$$

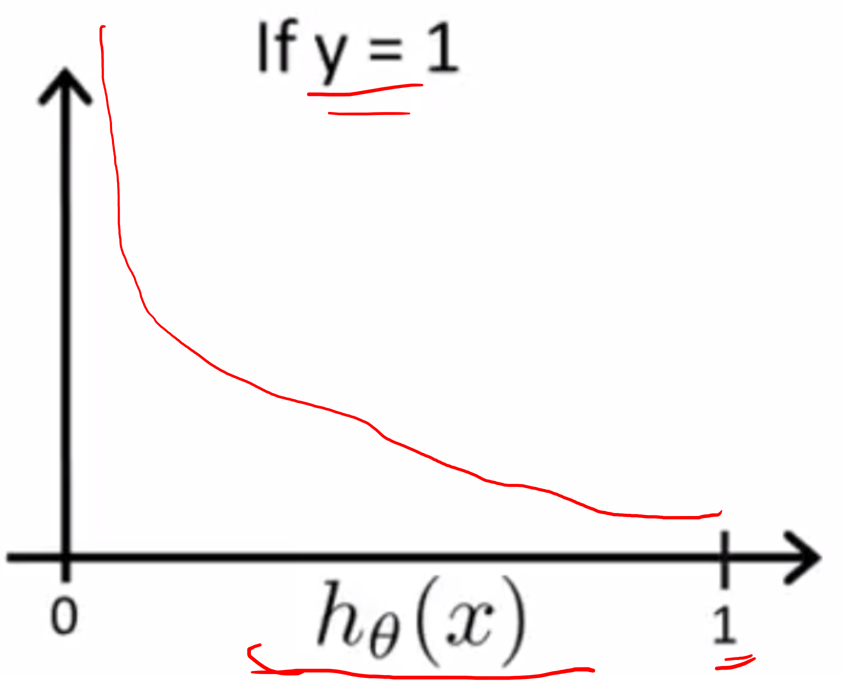


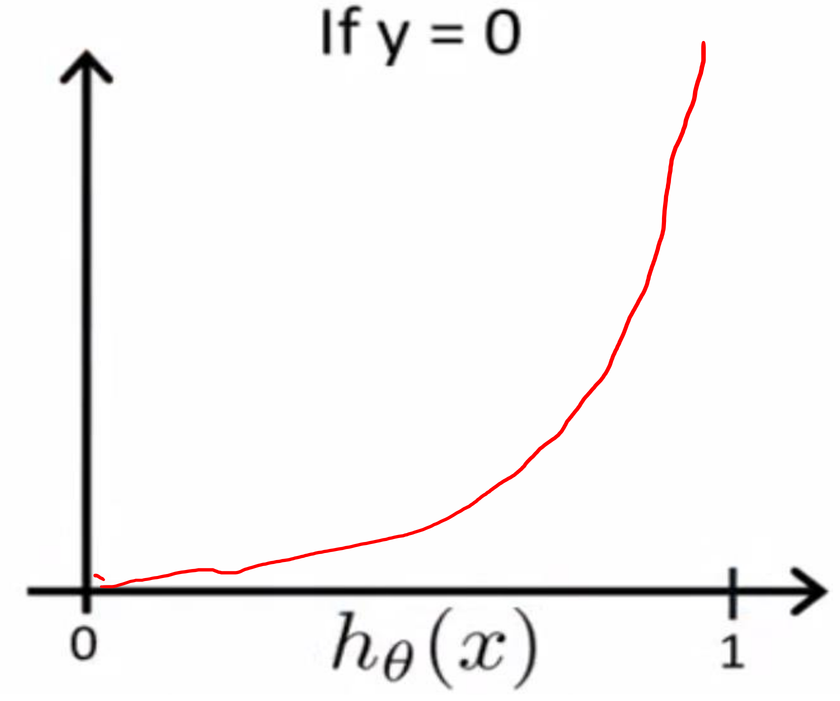
Cost Function for Logistic Regression – Binary Cross Entropy

$$Cost(h(x), y) = \begin{cases} -\log(h(x)) & if\ y = 1 \\ -\log(1 - h(x)) & if\ y = 0 \end{cases}$$

$+\infty$

Cost





Binary Cross Entropy Loss

$$Cost(h(x), y) = \begin{cases} -\log(h(x)) & if\ y = 1 \\ -\log(1 - h(x)) & if\ y = 0 \end{cases}$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(h\left(x^{(i)}\right), y^{(i)}\right)$$

0 1 − 1

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} \cancel{y^{(i)} \log h\left(x^{(i)}\right)} + \left(1 - y^{(i)}\right)\log\left(1 - h\left(x^{(i)}\right)\right)\right]$$

Finding the Parameter Values – Gradient Descent

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log(1 - h\left(x^{(i)}\right))]$$

- Gradient Descent

Repeat until convergence {

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

After taking derivative:

Repeat until convergence {

$$\theta_j = \theta_j - \alpha \underline{\sum_{i=1}^{m} h\left(x^{(i)}\right) - y^{(i)})x_j^{(i)}}$$

}

$\uparrow \alpha$

$\downarrow c$

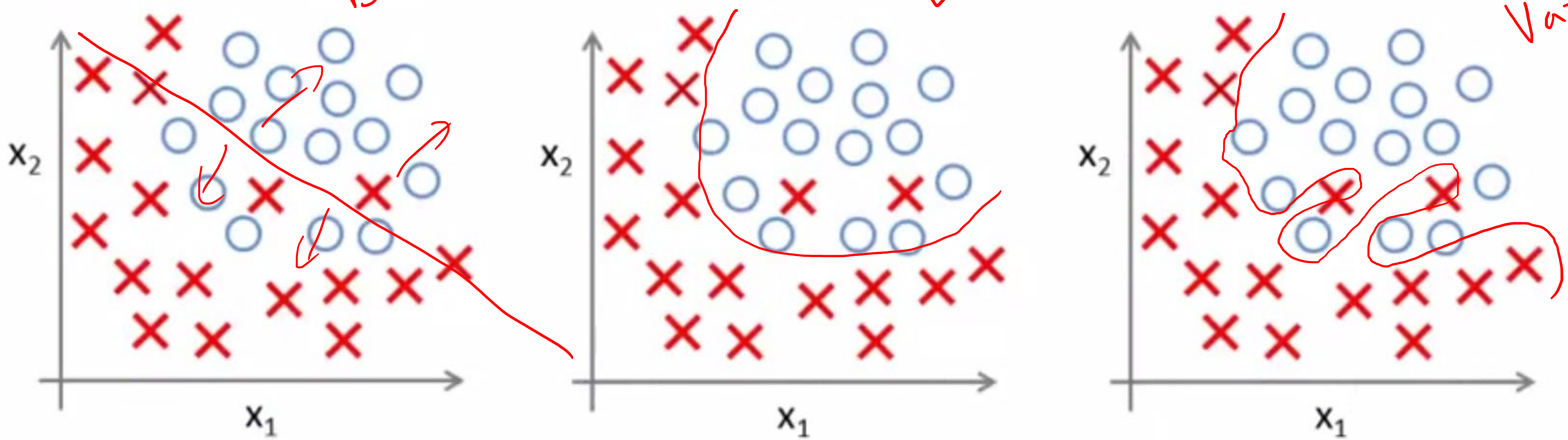
Overfitting in Logistic Regression

underfitting

bias $\uparrow$

overfitting

Variance



$h_\theta(x) = g(\theta_0 \underline{+ \theta_1 x_1 + \theta_2 x_2})$

(g = sigmoid function)

$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2$
$+ \theta_3 x_1^2 + \theta_4 x_2^2$
$+ \theta_5 x_1 x_2)$

$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2$
$+ \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2$
$+ \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$