Introduction to Machine Learning

Taejun Yoon
ohimfrog03@gmail.com
TNT ML
2024/09/19

TRAIN AND TEST

Contents

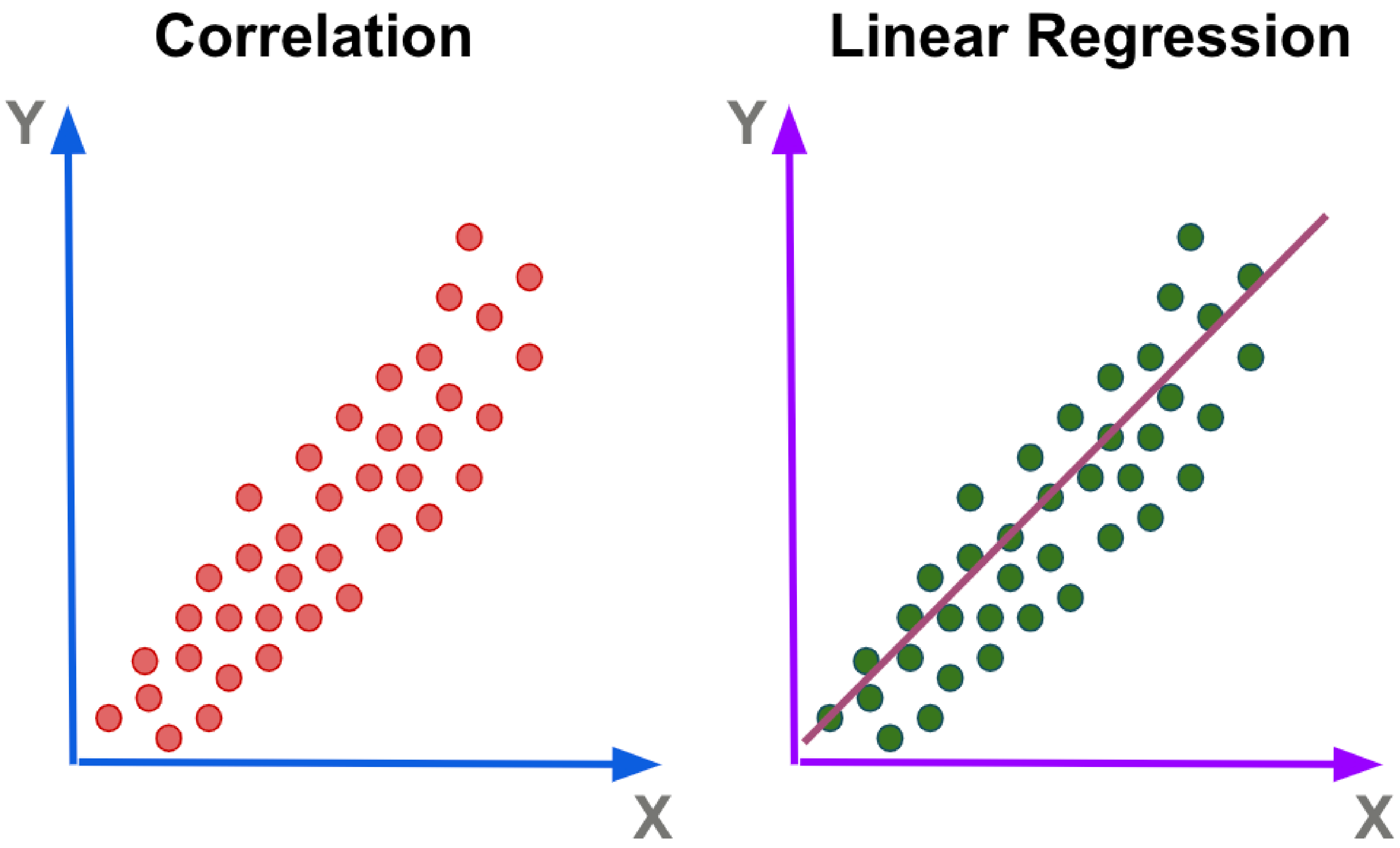
TRAIN AND TEST

What We Learned

1. Overall Machine Learning

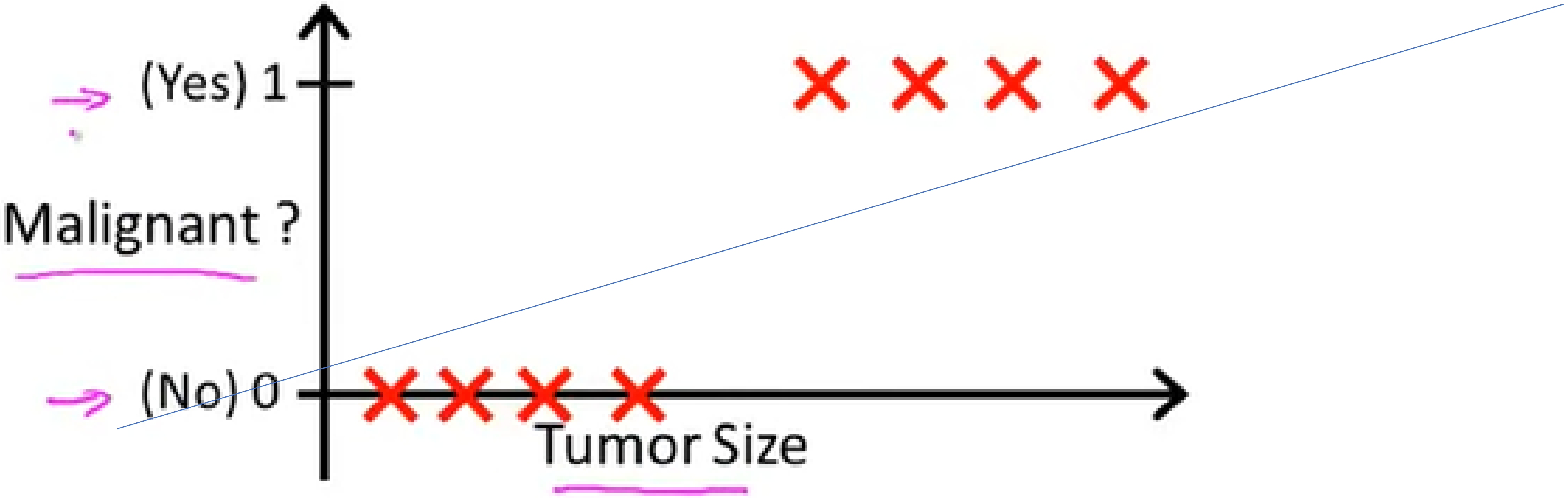
2. Regression

Regression

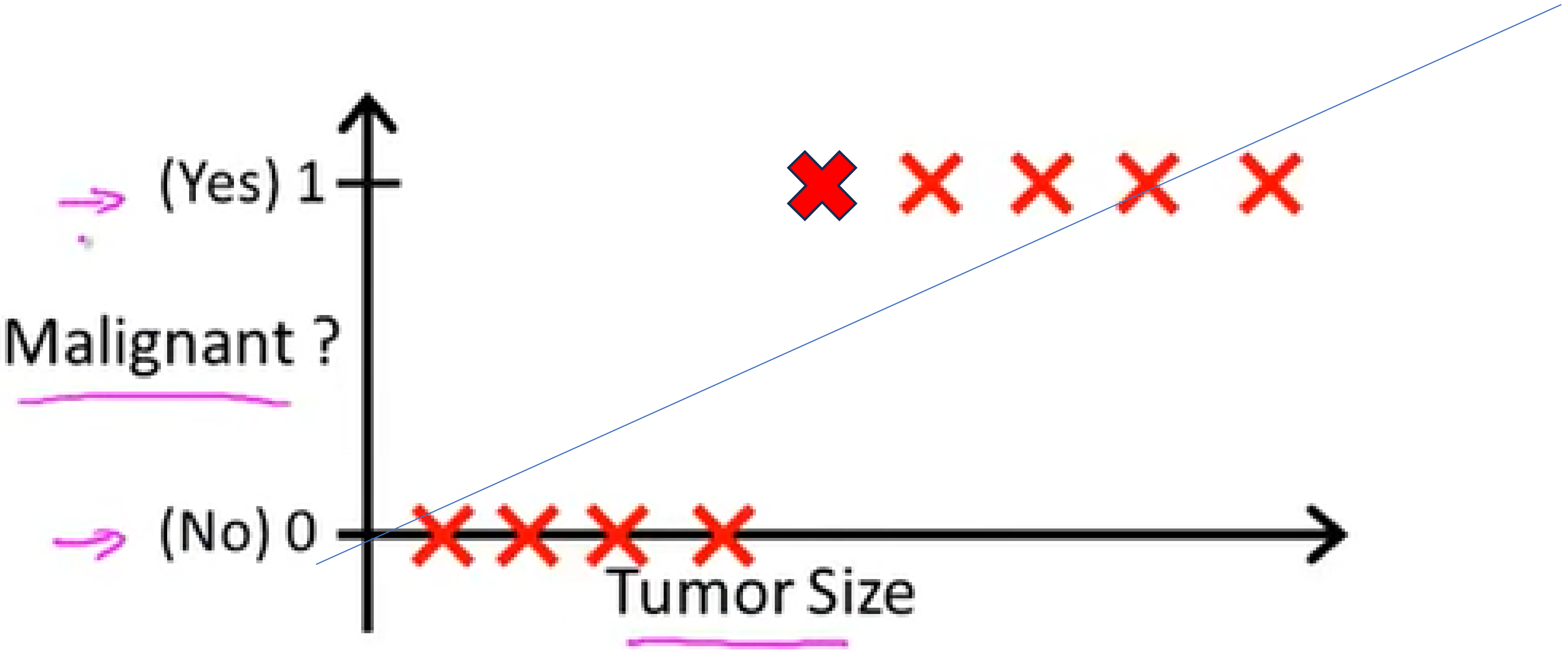


TRAIN AND TEST

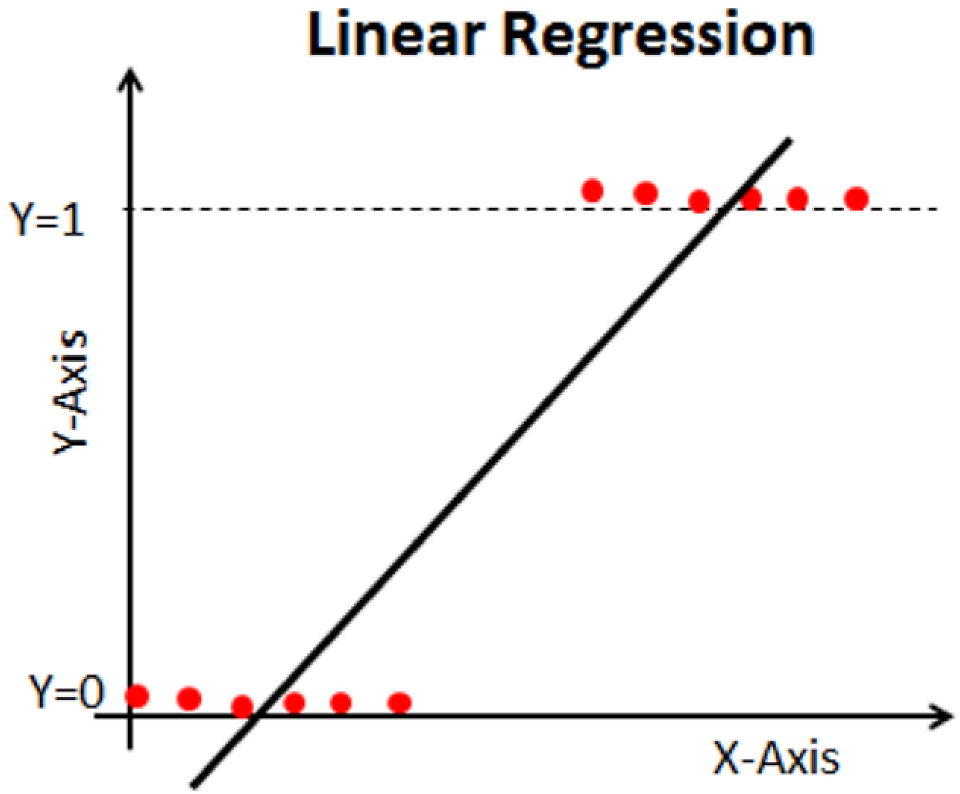
Logistic Regression



Logistic Regression



Logistic Regression



TRAIN AND TEST

Logistic Regression

$y \in \{0, 1\}$

0: “Negative Class” (e.g., benign tumor)

1: “Positive Class” (e.g., malignant tumor)

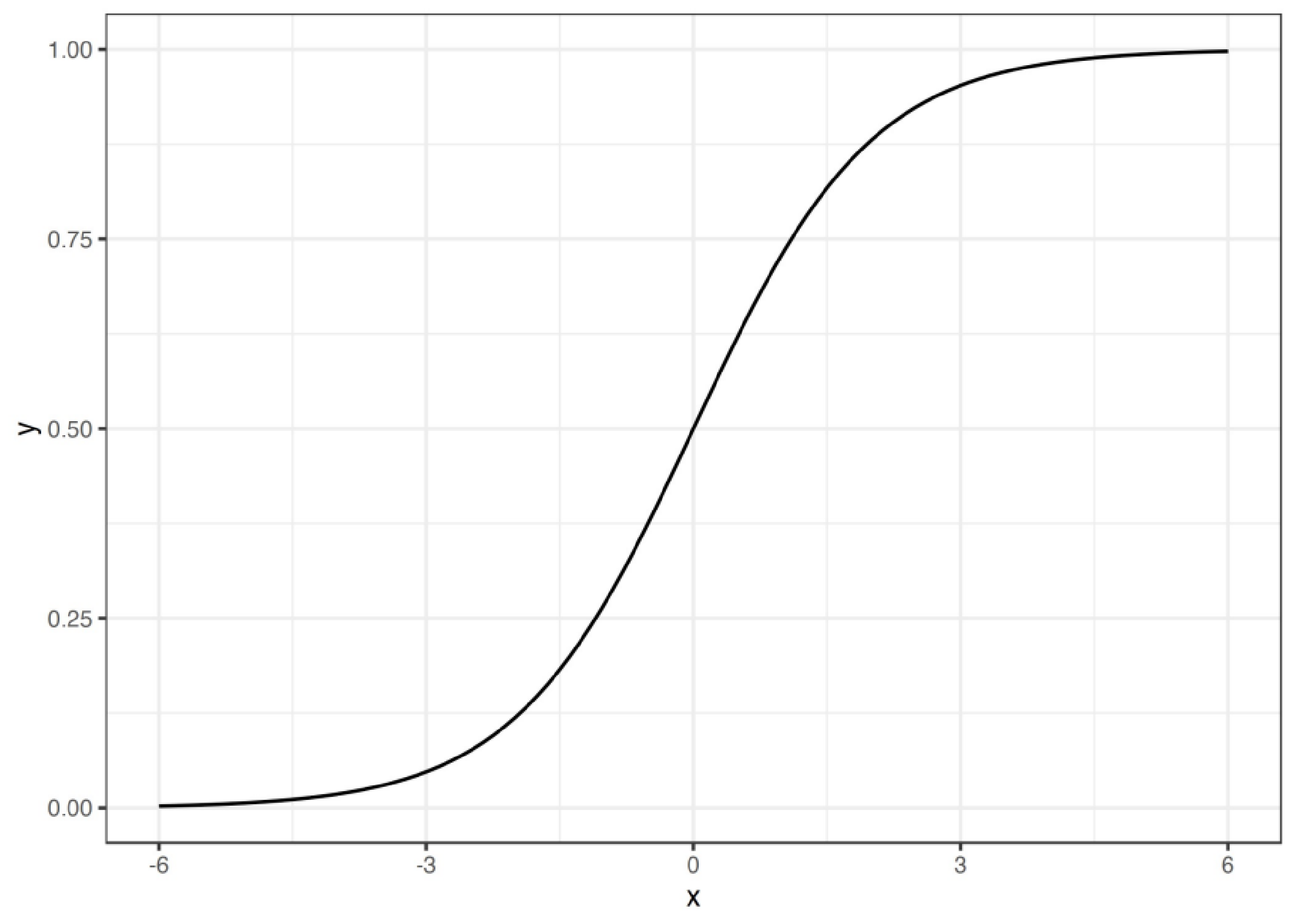
TRAIN AND TEST

Logistic Regression

1. Classification

2. Then, why it’s a regression?

Logistic Regression

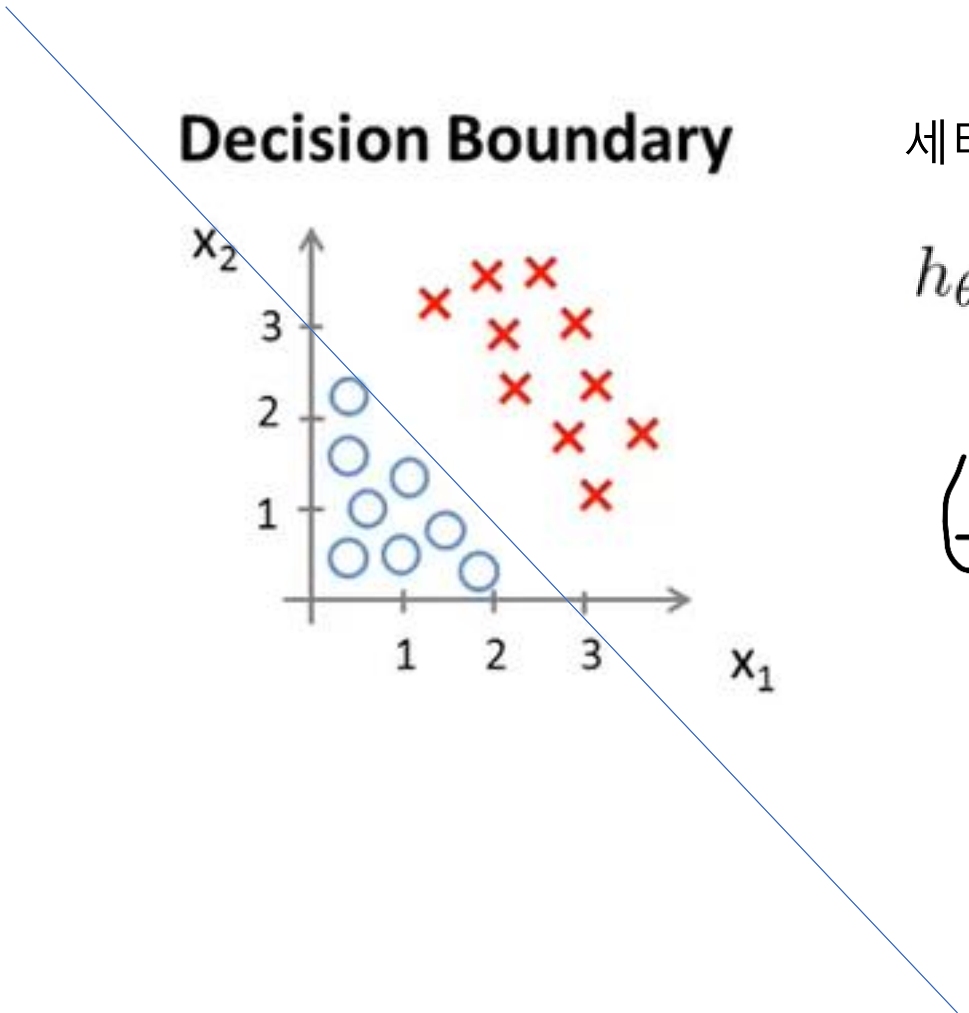


Logistic Function 수식

$$y = \frac{1}{1 + e^{-x}}$$

범위가 0과 1사이에 존재

Decision Boundary



세타는 파라미터

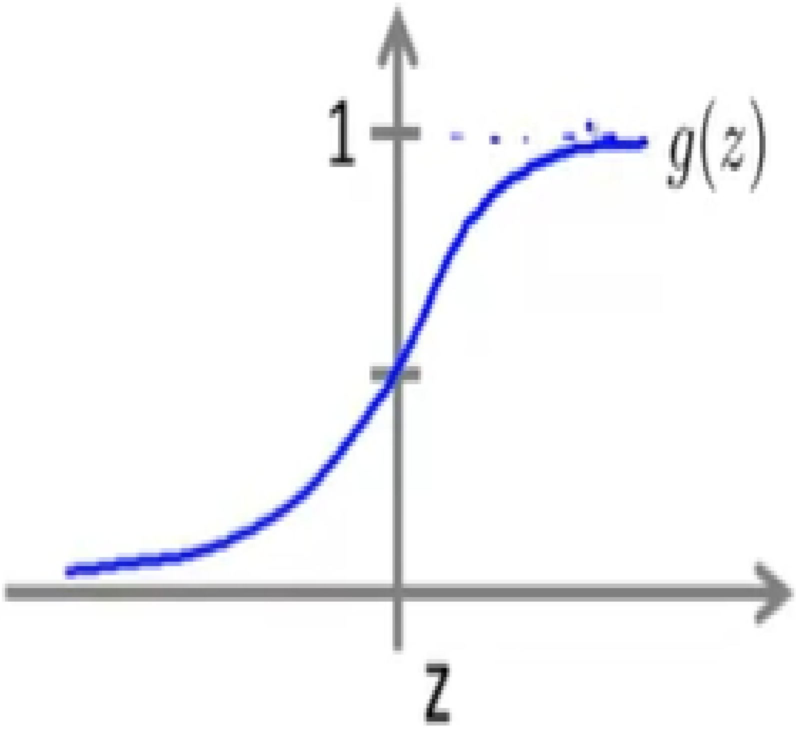
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

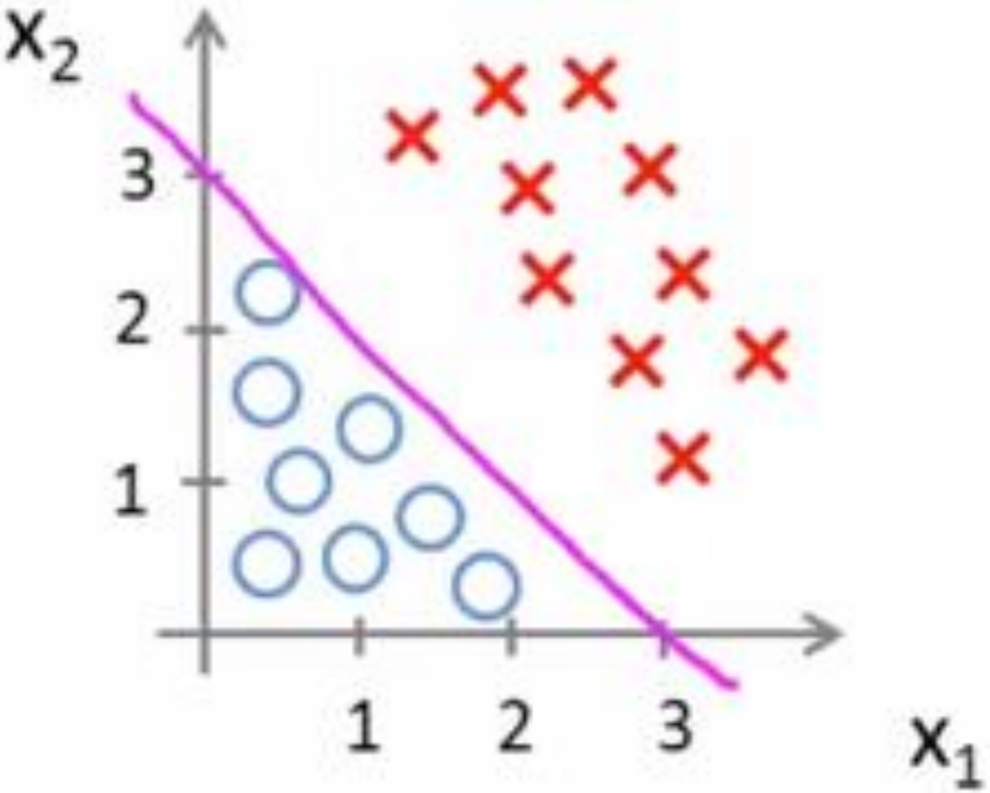
$$\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \rightarrow -3 + x_1 + x_2 \geq 0.$$

$$y = 1$$

$$x_1 + x_2 \geq 3$$

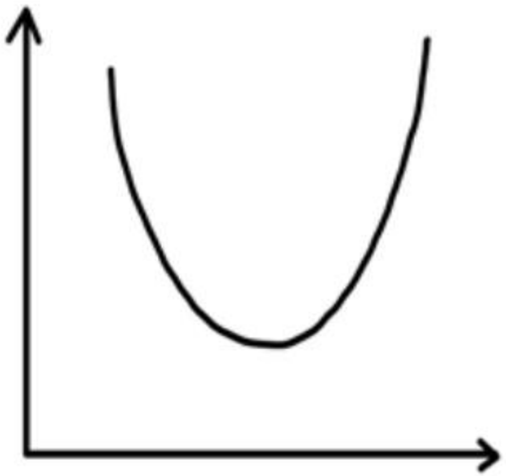
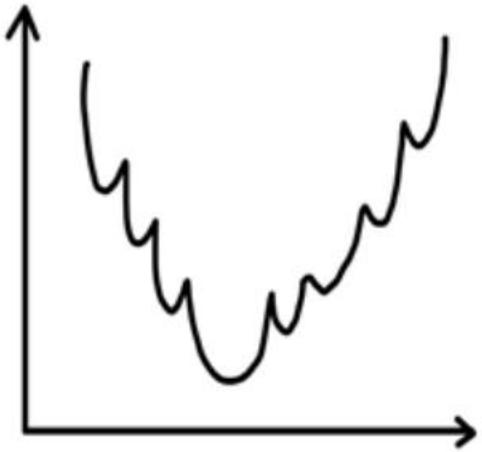
Decision Boundary





Loss Function for Logistic Regression

$$\frac{1}{N}\sum_{i=1}^{n}(h(x_i)-y_i)^2$$

<table>
<tr><td colspan="2">$$J(w) = \frac{1}{N}\sum_{i=1}^{n}\mathrm{cost}(h_w(x_i), y_i) = \frac{1}{N}\sum_{i=1}^{n}(h_w(x_i)-y_i)^2$$</td></tr>
<tr><td>선형회귀분석
$h_w(x) = w^T x$</td><td>로지스틱 회귀분석
$h(x) = \frac{1}{1+\exp(w^T x)}$</td></tr>
<tr><td></td><td></td></tr>
</table>

Simplified Loss Function

$$\mathrm{cost}(h_w(x), y) = -y\log(h_w(x)) - (1-y)\log(1-h_w(x))$$

TRAIN AND TEST

Loss Function for Logistic Regression

$$\mathrm{cost}(h_w(x), y) = -y\log(h_w(x)) - (1-y)\log(1-h_w(x))$$

cost(hθ(x),y)=−ylog(hθ(x))−(1−y)log(1−hθ(x))

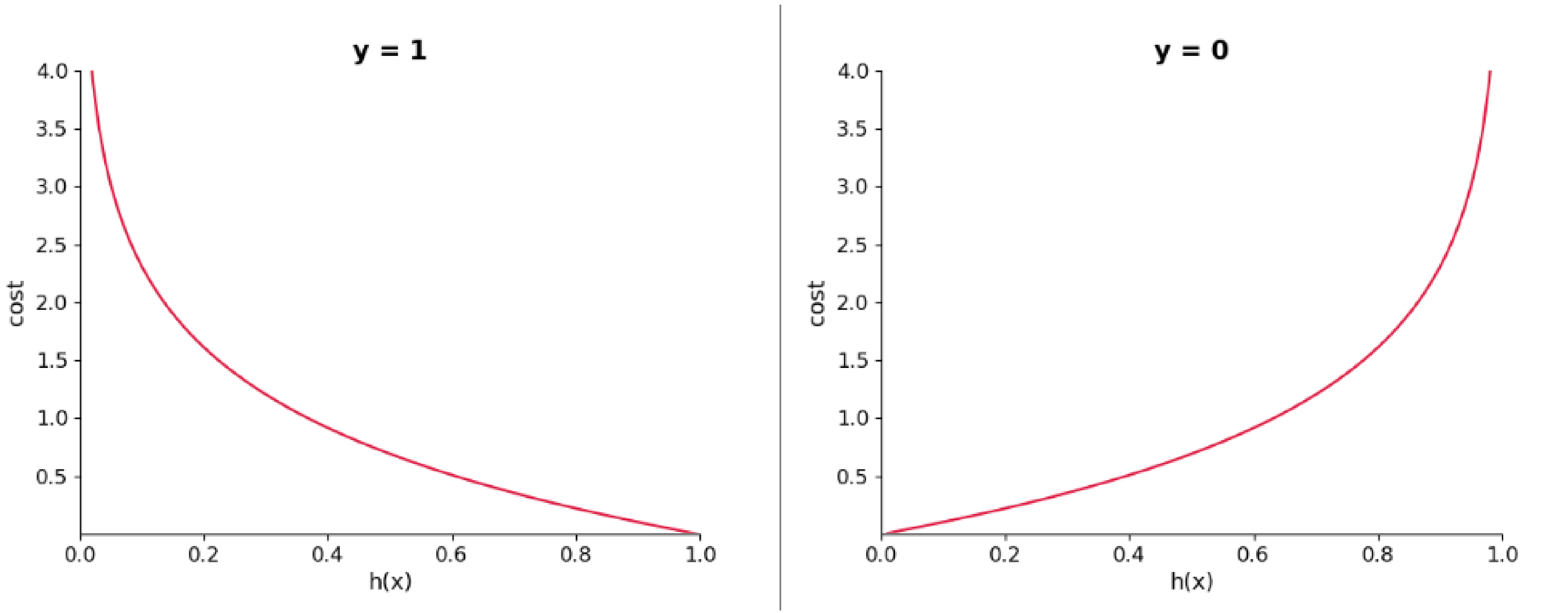
If y=1,cost(hθ(x),y)=−log(hθ(x))y=1,cost(hθ(x),y)=−log(hθ(x))
If y=0,cost(hθ(x),y)=−log(1−hθ(x))y=0,cost(hθ(x),y)=−log(1−hθ(x))

Logistic regression cost function

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

TRAIN AND TEST

Loss Function for Logistic Regression



TRAIN AND TEST

Gradient Descent for Logistic Regression

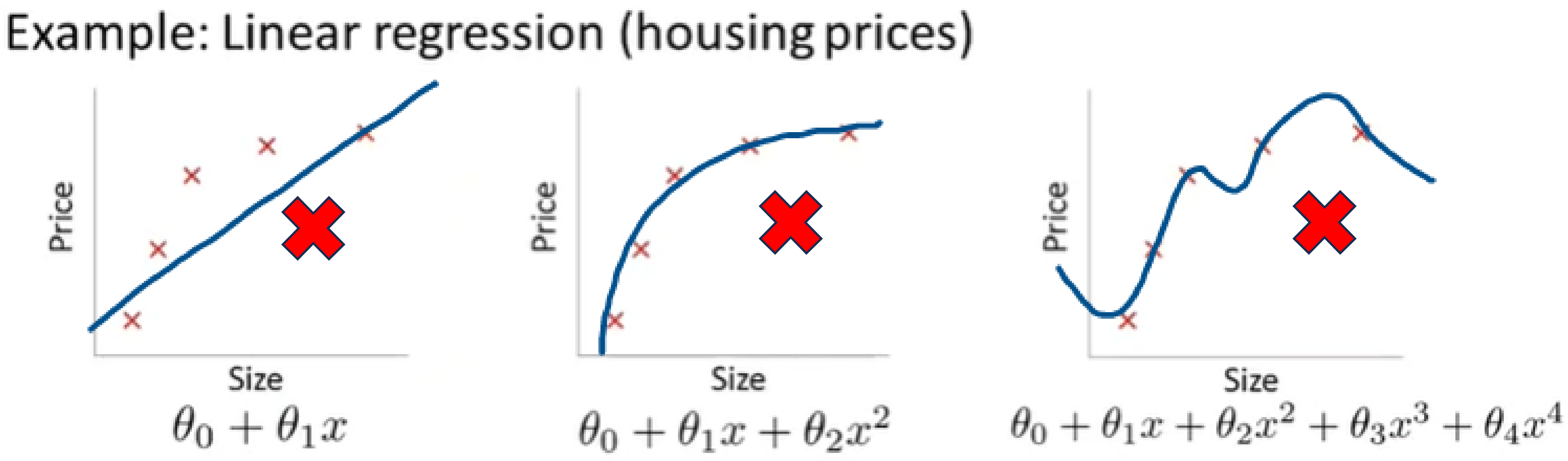
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (\text{동시업데이트 } j = 0, 1, \ldots, n)$$

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

TRAIN AND TEST

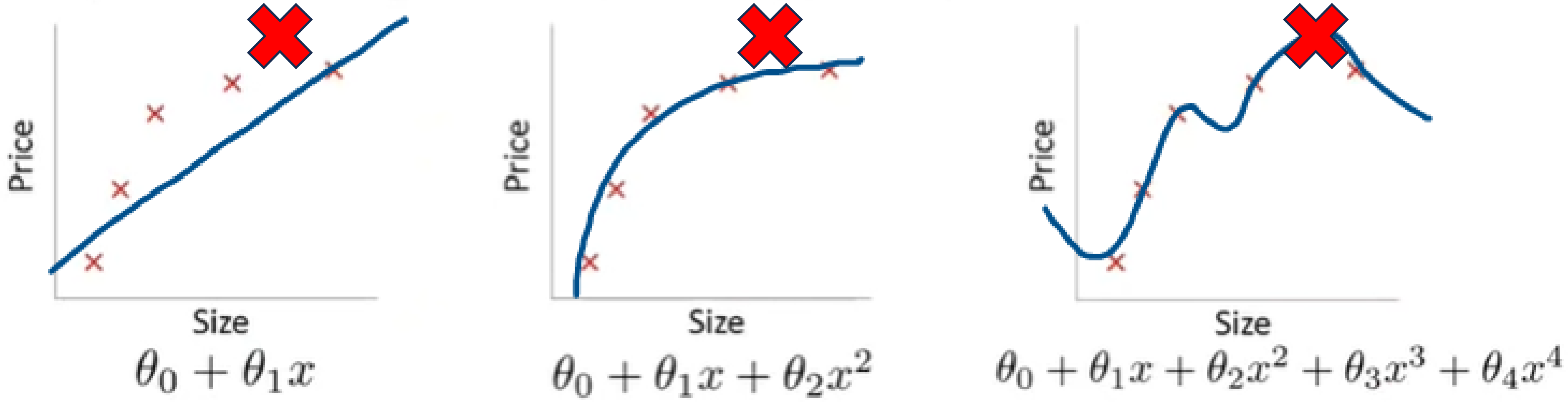
Overfitting



TRAIN AND TEST

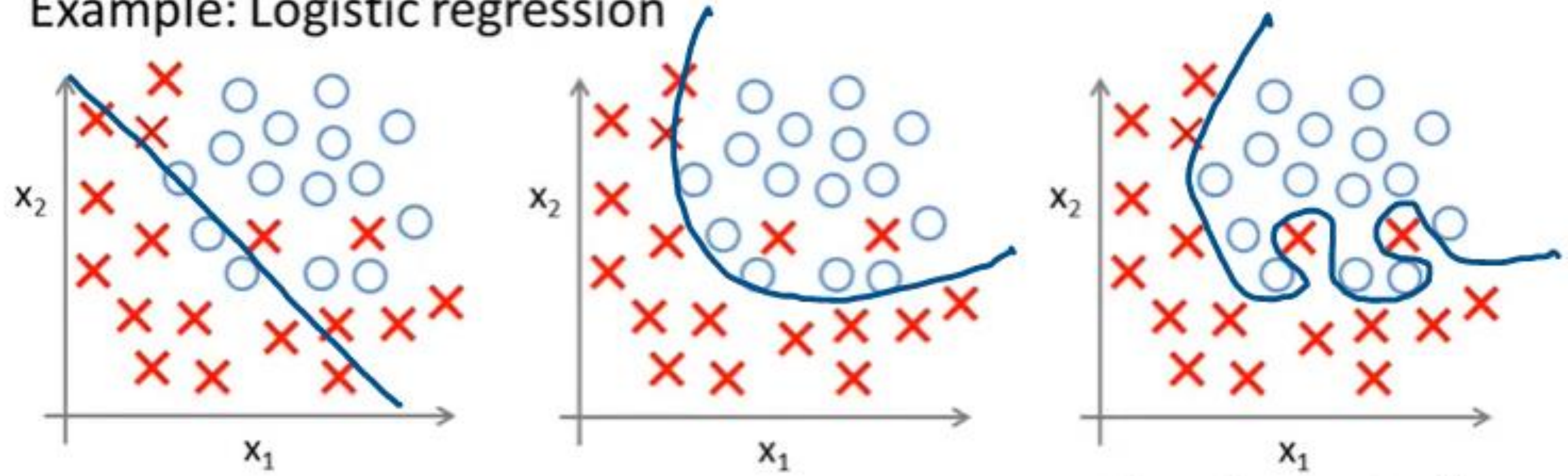
Overfitting

Example: Linear regression (housing prices)



Overfitting

Example: Logistic regression



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

(g = sigmoid function)

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

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

How to deal with overfitting?

1. Reduce the number of features

2. Regularization

4. K-fold Cross Validation

5. Dropout

6. Early Stopping

7. Data Augmentation

8. Ensemble – Bagging/Boosting

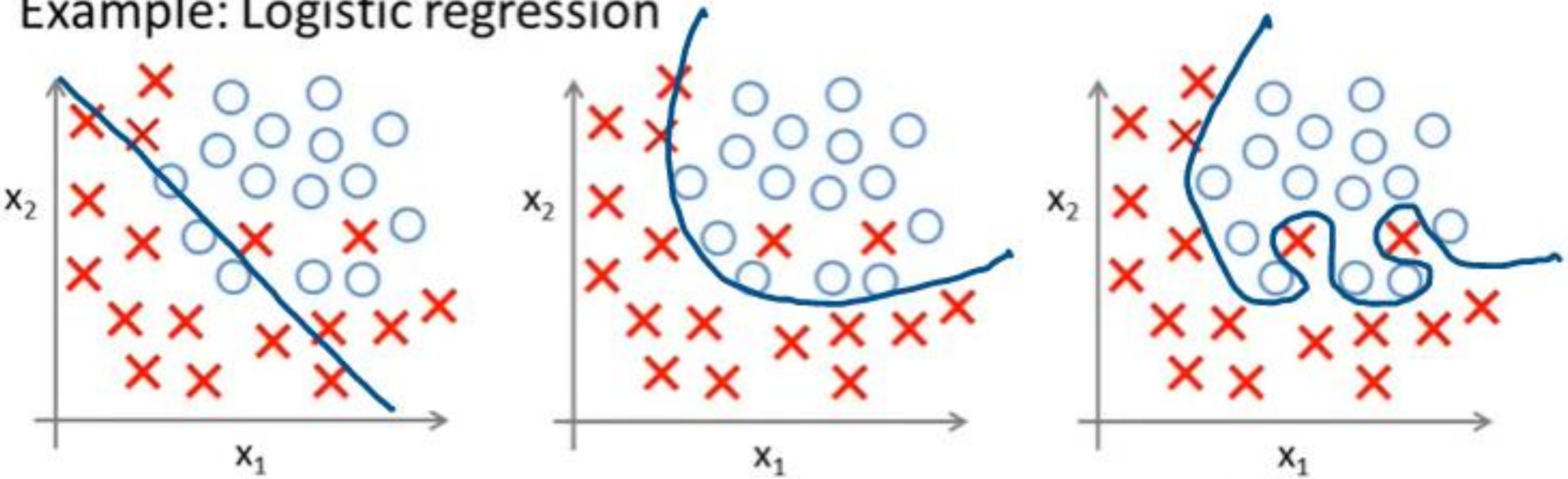
9. Collecting more data

TRAIN AND TEST

How to deal with overfitting?

1. Reduce the number of features

Example: Logistic regression



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

(g = sigmoid function)

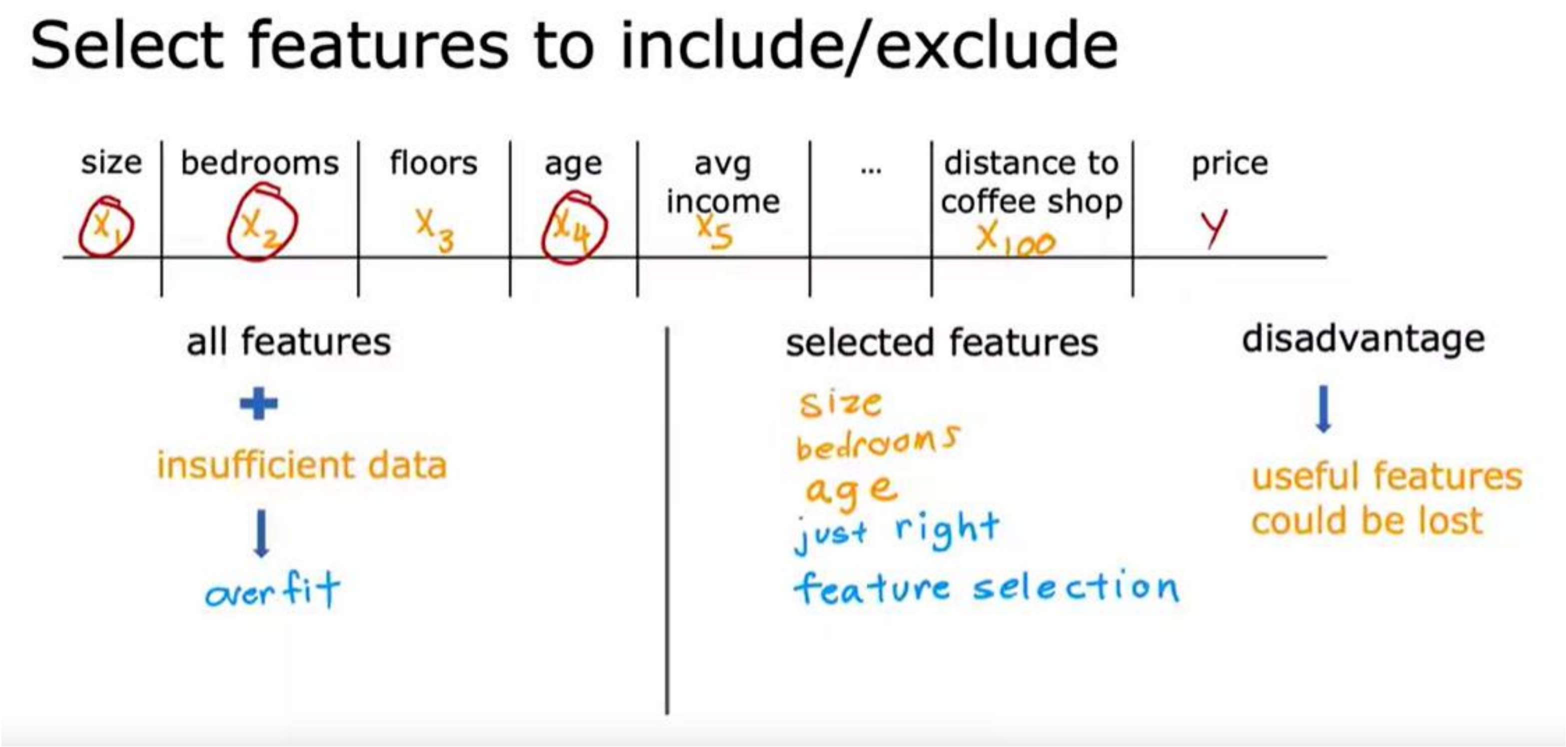
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$

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

TRAIN AND TEST

How to deal with overfitting?

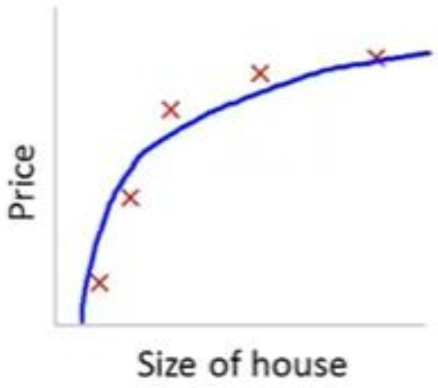
1. Reduce the number of features



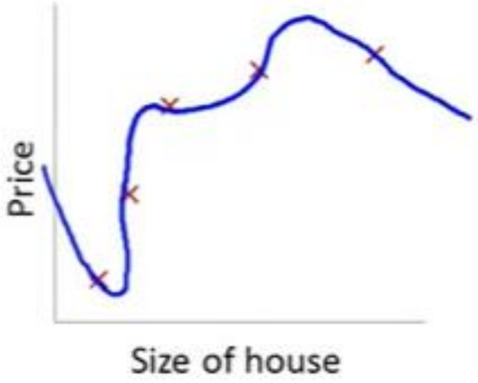
How to deal with overfitting?

2. Regularization

Intuition



$\theta_0 + \theta_1 x + \theta_2 x^2$



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

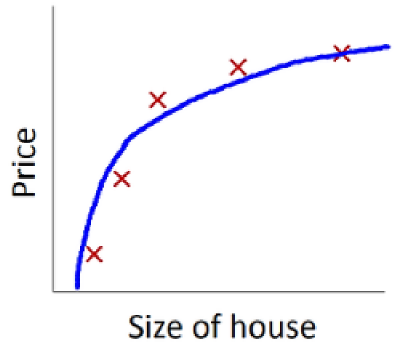
Suppose we penalize and make θ_3, θ_4 really small.

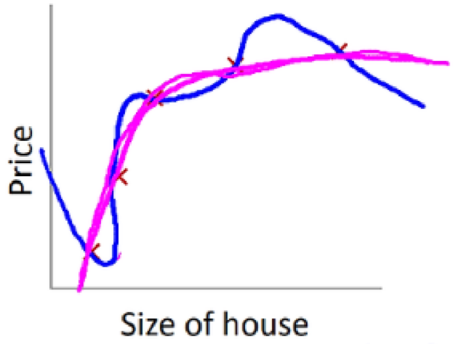
TNT TRAIN AND TEST

How to deal with overfitting?

2. Regularization

Intuition





$$\theta_0 + \theta_1 x + \theta_2 x^2 \qquad \theta_0 + \theta_1 x + \theta_2 x^2 + \cancel{\theta_3 x^3} + \cancel{\theta_4 x^4}$$

Suppose we penalize and make θ_3, θ_4 really small.

$$\min_\theta \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + 1000\,\theta_3^2 + 1000\,\theta_4^2$$

$\theta_3 \approx 0 \quad \theta_4 \approx 0$

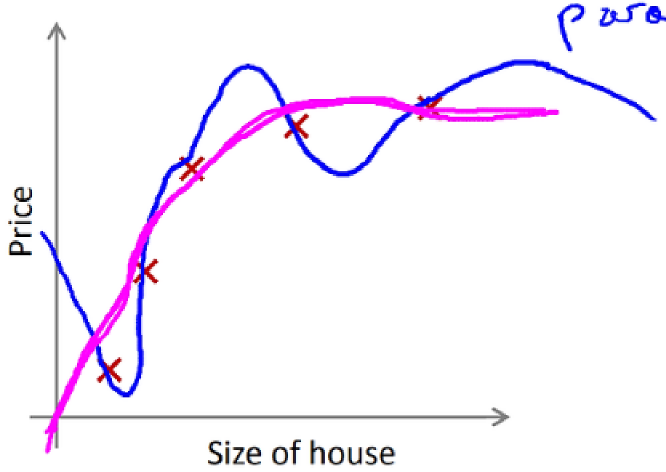
How to deal with overfitting?

2. Regularization

Regularization.

$$J(\theta) = \frac{1}{2m}\left[\underbrace{\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2}_{} + \underbrace{\lambda}_{\text{regularization parameter}} \sum_{j=1}^{n}\theta_j^2\right]$$

$$\min_\theta J(\theta)$$



Regularized Logistic Regression

$$J(\theta) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right)\log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} {\theta_j}^2$$

TRAIN AND TEST

Overall Review

- Logistic Regression
- Decision Boundary
- Loss Function for Logistic Regression
- Simplified Loss Function
- Overfitting
- Regularized Logistic Regression

TRAIN AND TEST

TRAIN AND TEST