Machine Learning 02

Kihyun Shin
DMSE, HBNU

TheMAD

Machine Learning Definition

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"

By Tom M. Mitchell

Machine Learning Algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised Learning

x ⟶ y

input output

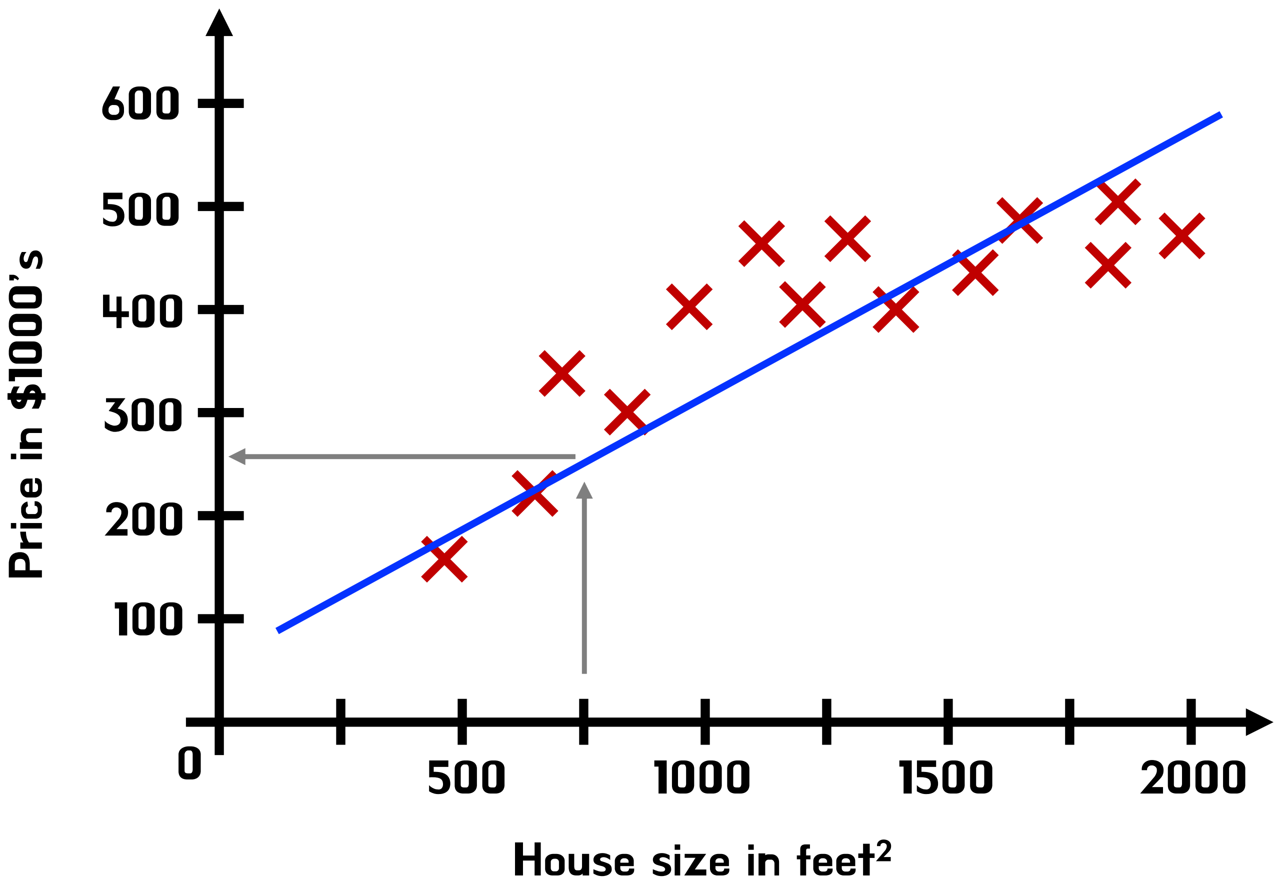
(label)

Learns from being given "right answers"

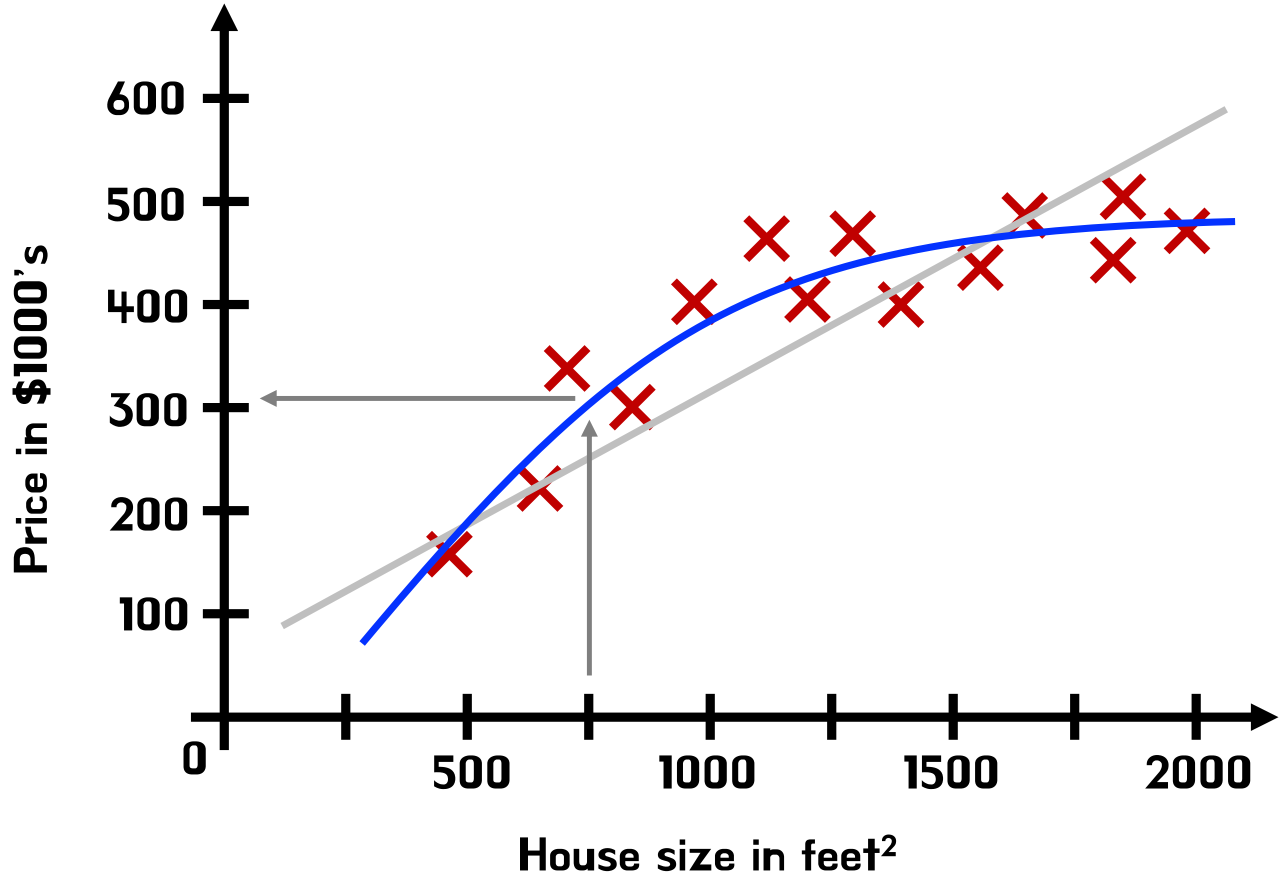
Supervised Learning

Input (x)	Output (y)	Application
email	spam? (0/1)	Spam filtering
audio	text transcripts	Speech recognition
english	spanish	Machine translation
Ad, user info.	Click ? (0/1)	Online advertising
Image, radar info.	Position of other cars	Self-driving car
Image of phone	Defect ? (0/1)	Visual inspection

Regression: Housing price prediction



Regression: Housing price prediction

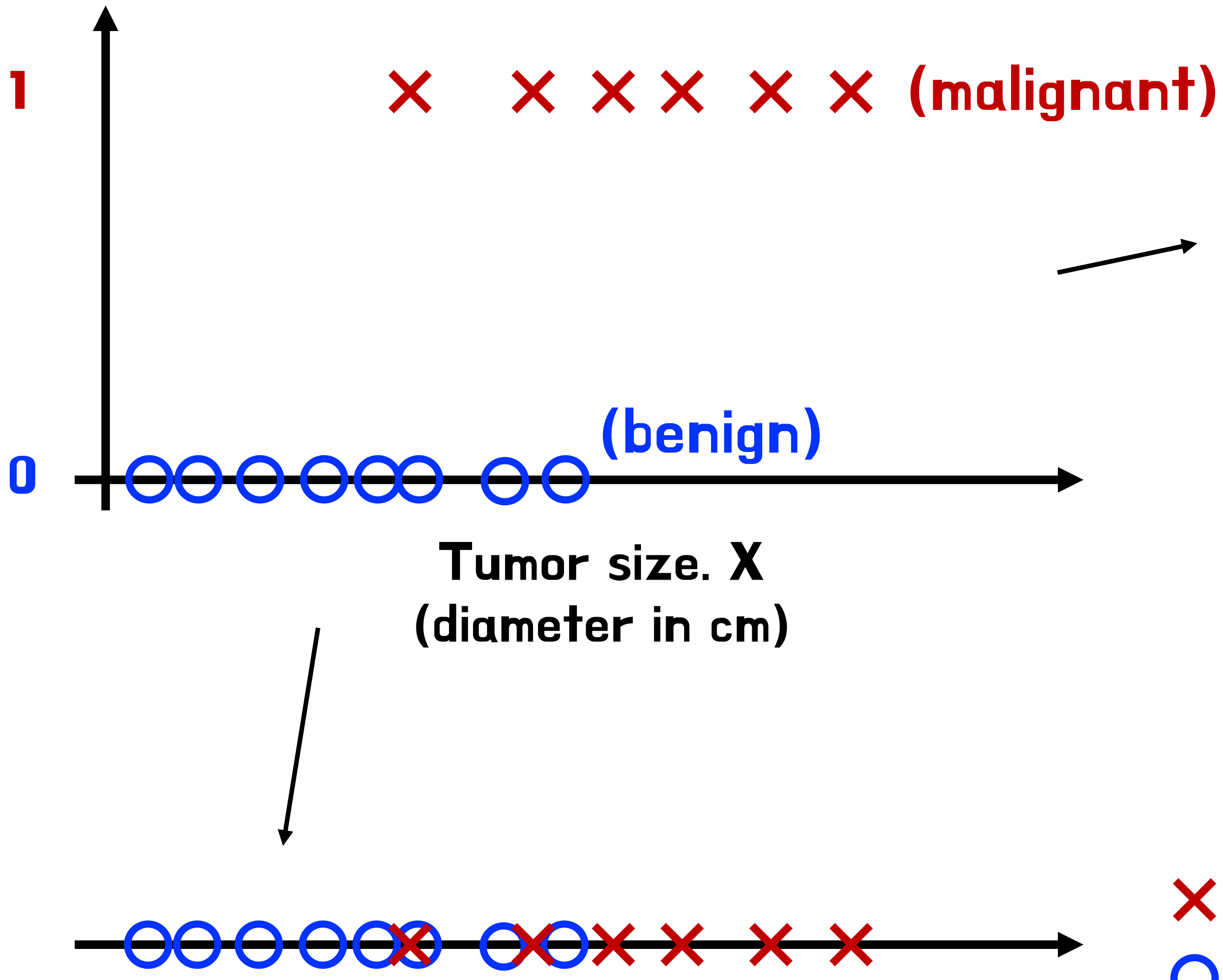


Regression

predict a number

(infinitely many possible outputs)

Classification: Breast cancer detection

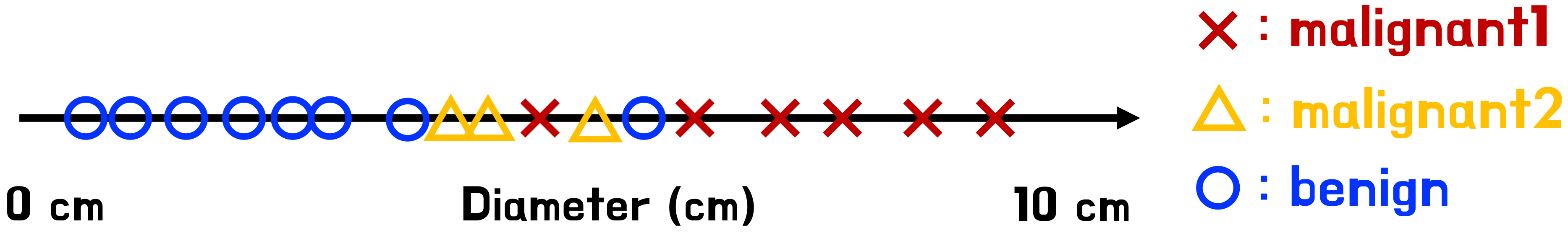


Size	Diagnosis
2	0
5	1
1	0
7	1
6	1
⋮	⋮

× : malignant

○ : benign

Classification: Breast cancer detection

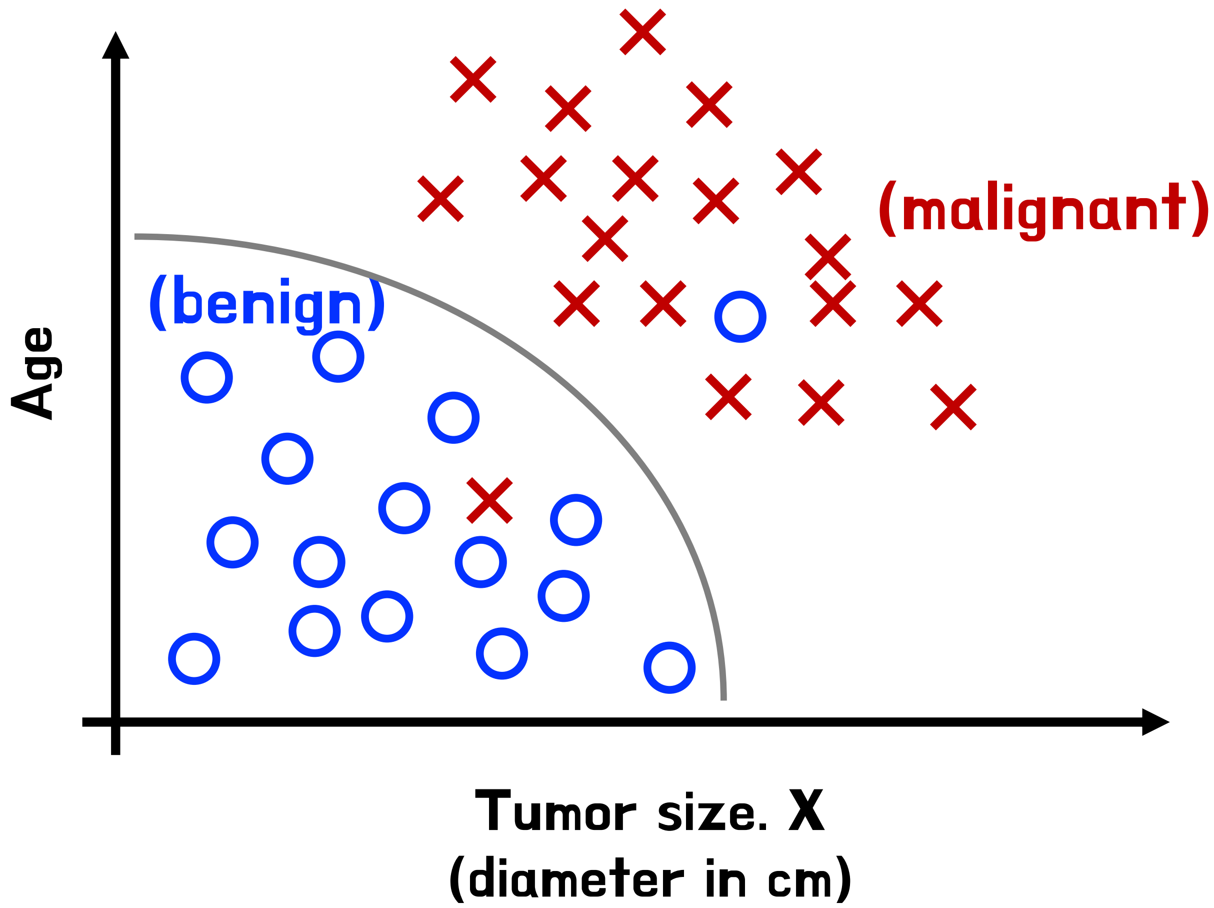


Classification

predict categoreis

(small number of possible outputs)

Two or more inputs



Many more inputs can be also available
(thickness, uniformity of size, shape)

Supervised learning

Supervised Learning

Learns from being given "right answers"

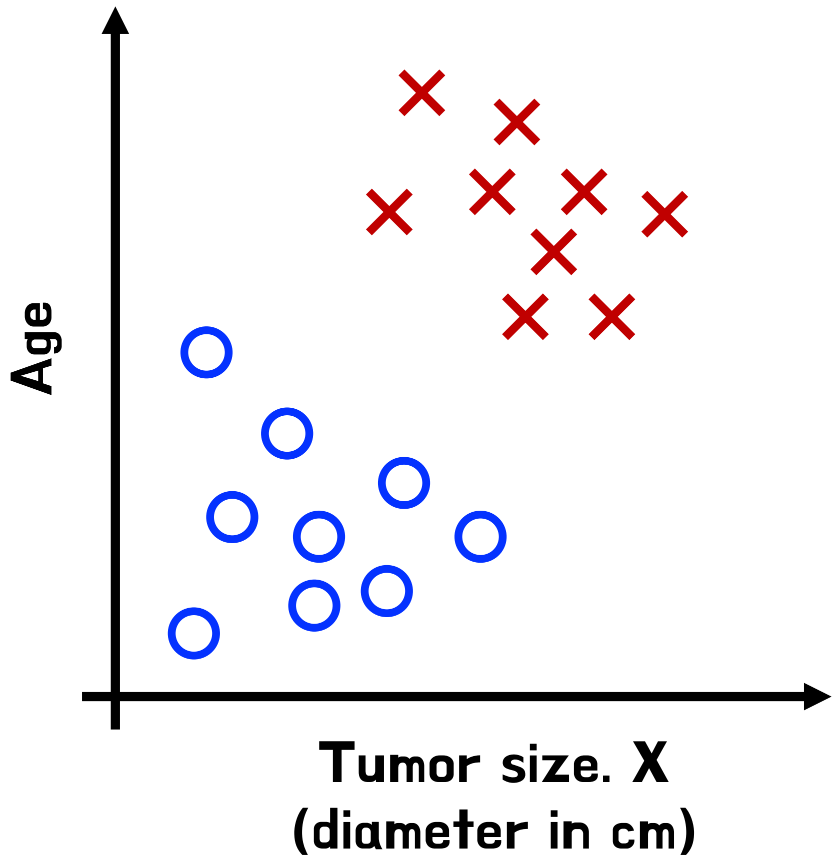
Regression	Classification
Predict a number	Predict categories
Infinitely many possible outputs	Small number of possible outputs

Unsupervised learning

Supervised learning

Learn from data labeled

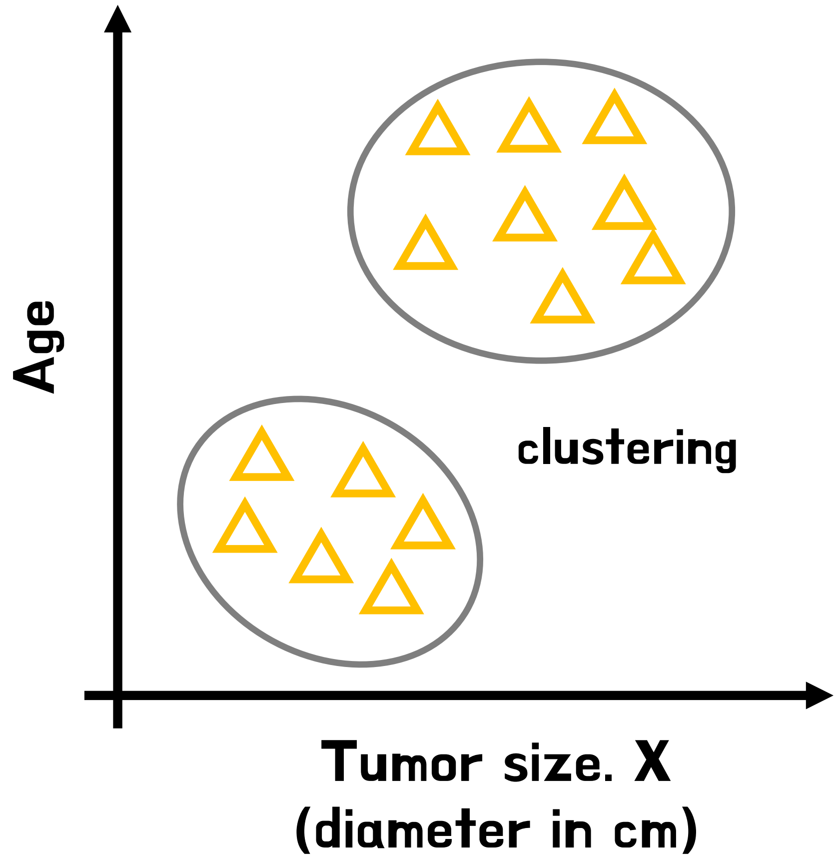
With the “right answers”



Unsupervised learning

Find something interesting

in unlabeled data



Unsupervised learning

Giant panda gives birth to rare twin cubs at Japan's oldest zoo

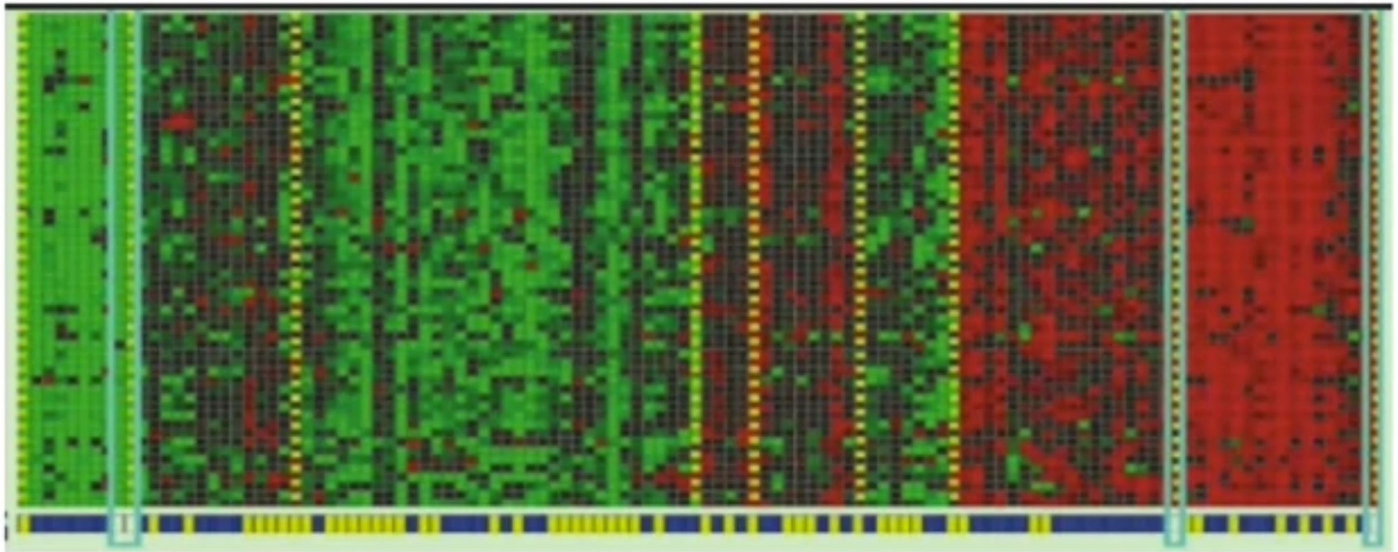
USA TODAY · 6 hours ago

- Giant panda gives birth to twin cubs at Japan's oldest zoo
  CBS News · 7 hours ago
- Giant panda gives birth to twin cubs at Tokyo's Ueno Zoo
  WHBL News · 16 hours ago
- A Joyful Surprise at Japan's Oldest Zoo: The Birth of Twin Pandas
  The New York Times · 1 hour ago
- Twin Panda Cubs Born at Tokyo's Ueno Zoo
  PEOPLE · 6 hours ago

View Full Coverage

Unsupervised learning

Genes (each row)



Individuals (each column)

Unsupervised learning

Data only comes with inputs x, but not output labels y.

Algorithm has to find structure in the data.

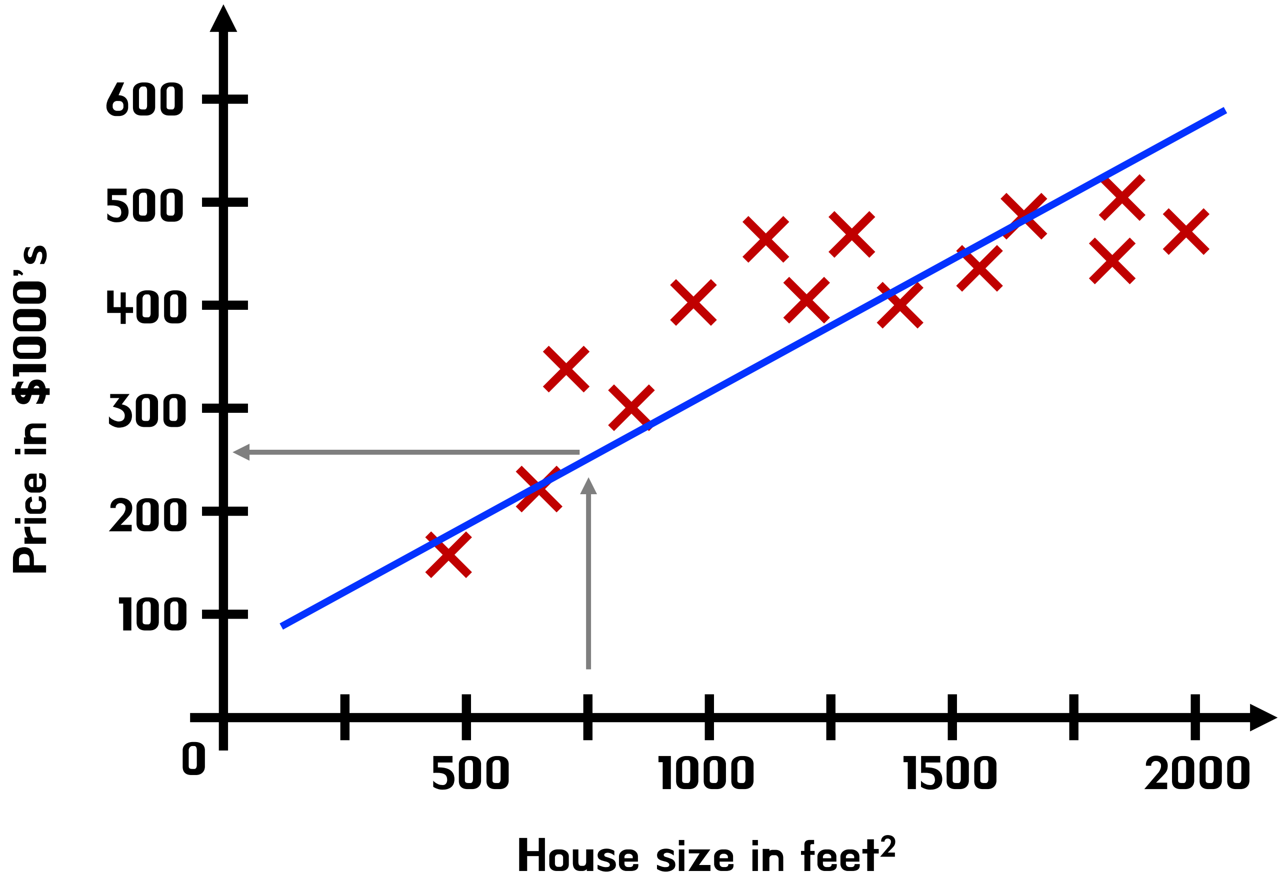
Clustering
Group similar data points together.

Dimensionality reduction (차원축소)
Compress data using fewer numbers.

Anomaly detection
Find unusual data points.

Linear Regression
(선형 회귀)

Linear regression

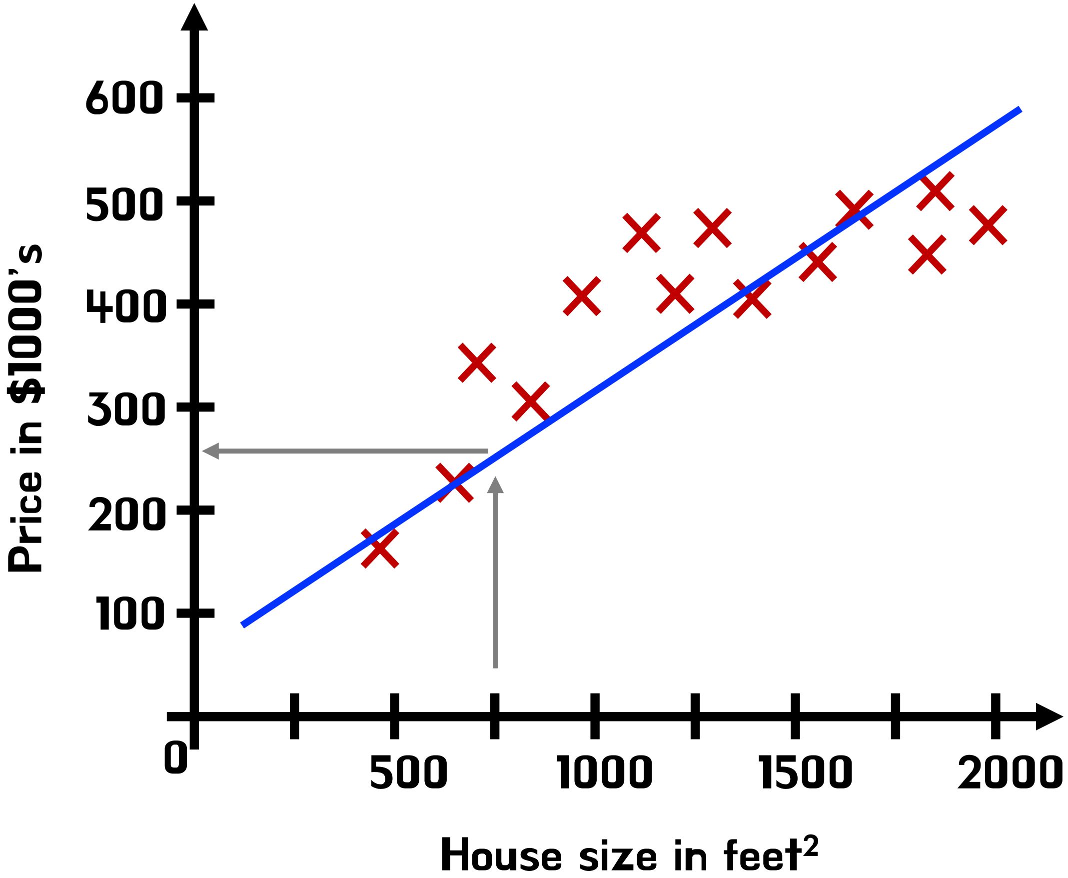


Regression model
Predicts numbers
(infinitely many possible outputs)

Supervised learning model
Data has "right answers"

Classification model
Predicts categories
(small number of possible outputs)

Linear regression



Data table

Size in feet^2	Price in $1000's
900	310
700	320
990	400
1550	406
1700	490
⋮	⋮

Terminology

Training set : Data used to train the model

	x Size in $feet^2$	y Price in $1000's
(1)	900	310
(2)	700	320
(3)	990	400
(4)	1550	406
(5)	1700	490
⋮	⋮	⋮

Test set : Data used to test the model

	Size in $feet^2$	Price in $1000's
(a)	980	310
(b)	1300	390
(c)	2000	500

Notation:

x = "input" variable
feature

y = "output" variable
"target" variable

m = number of training examples

(x, y) = single training example

$(x^{(i)}, y^{(i)})$ = i^{th} training example

x = 900, y = 400

(x, y) = (900, 400)

Terminology

Training set

↓

Learning algorithm

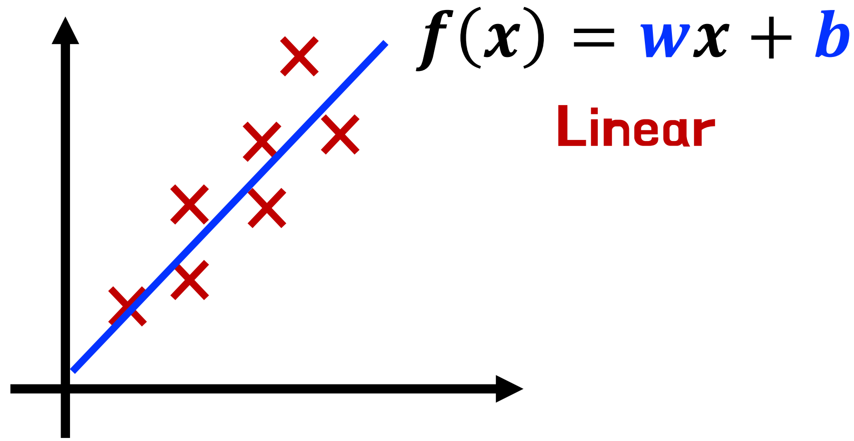
↓

$x \rightarrow f \rightarrow \hat{y}$

- x: **feature**
- f: **model**
- $\hat{y}$: **"y-hat"** — **Prediction** (estimated y)

How to represent f?

$$f_{w,b}(x) = f(x) = wx + b$$



Linear regression with one variable.
Univariate linear regression.

Parameters (Hyper parameters)

Training set

(feature) (targets)

Size in feet^2	Price in $1000's
900	310
700	320
990	400
1550	406
1700	490
⋮	⋮

$$f(x) = wx + b$$

w, b : parameters

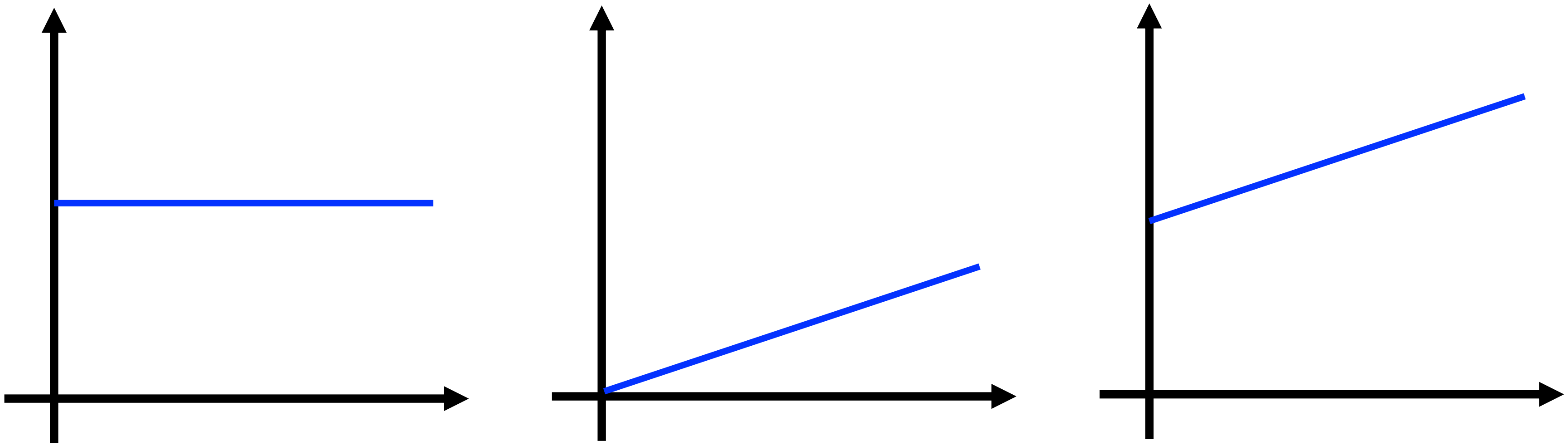
w → Weights

b → Bias

What do w, b do ?

Cost function

$$f(x) = wx + b$$



$w = 0$

$b = 1.5$

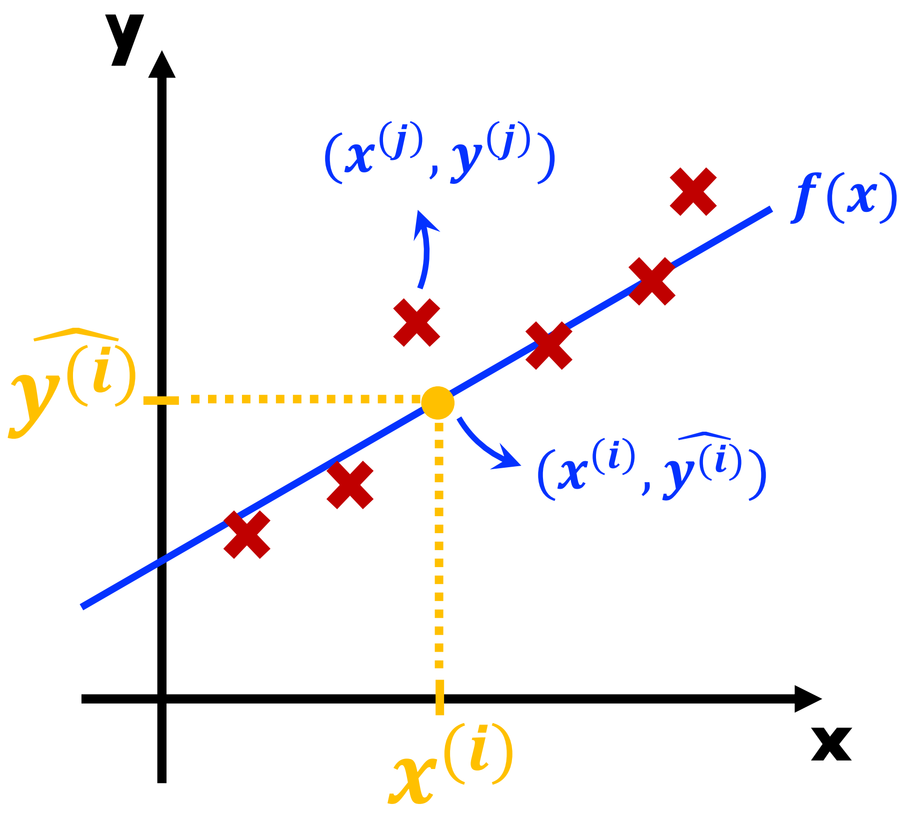
$w = 0.5$

$b = 0$

$w = 0.5$

$b = 1$

Cost function



$$\widehat{y^{(i)}} = f\left(x^{(i)}\right) = wx^{(i)} + b$$

Cost function:
Squared error cost function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(\widehat{y^{(i)}} - y^{(i)}\right)^2$$

error

m: Number of training examples

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2$$

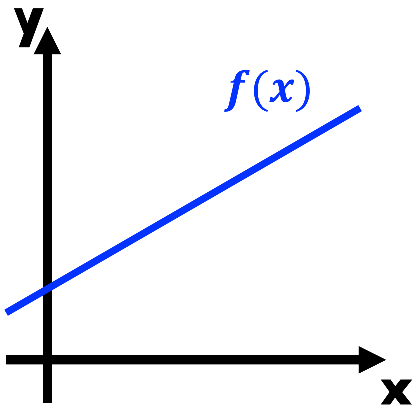
$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)^2$$

 TheMAD

Cost function

Model:

$$f(x) = wx + b$$



Parameters:

$$w, b$$

Cost function:

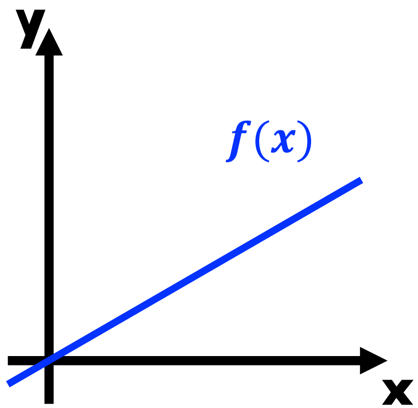
$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Goal:

$$\underset{w,b}{\text{minimize}}\, J(w,b)$$

Simplified Model:

$$f(x) = wx$$



Parameters:

$$w$$

Cost function:

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Goal:

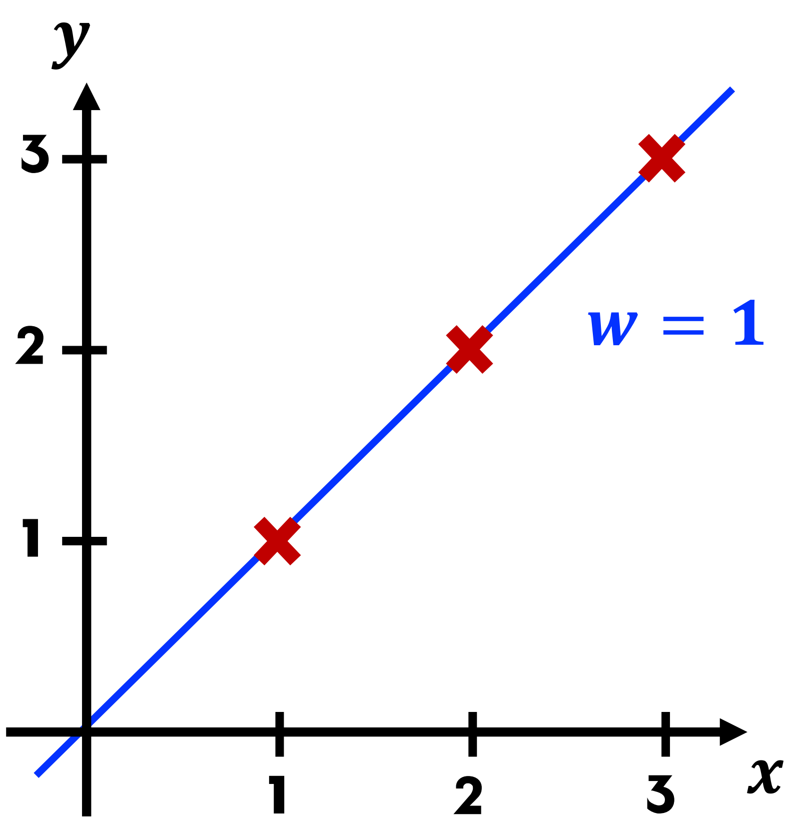
$$\underset{w}{\text{minimize}}\, J(w)$$

Cost function

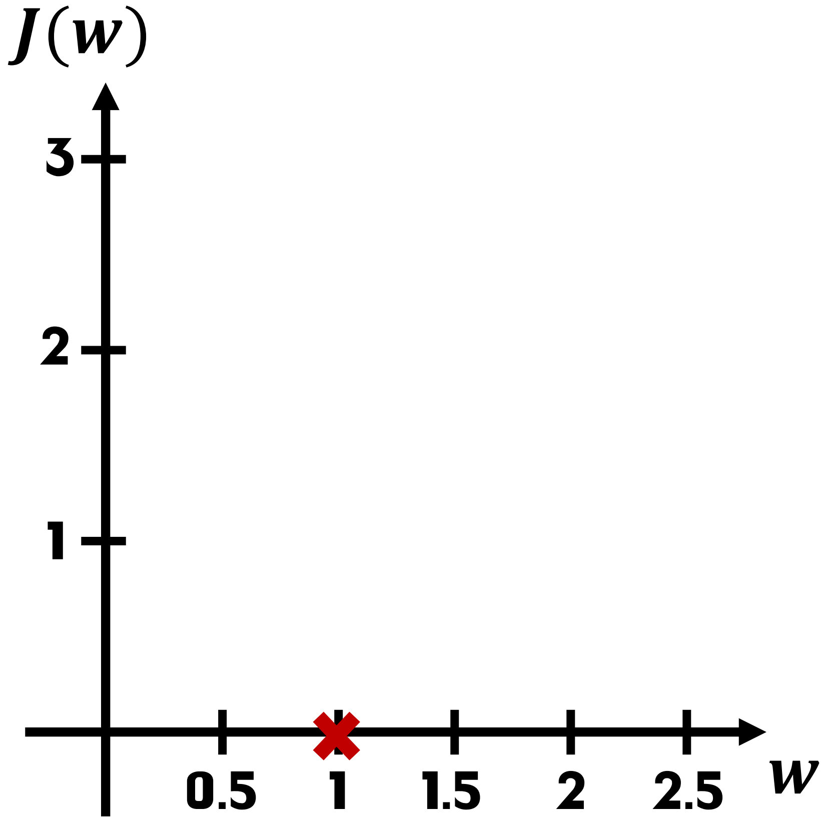
$f_w(x)$

$J(w)$

For fixed w, function of x



Function of w



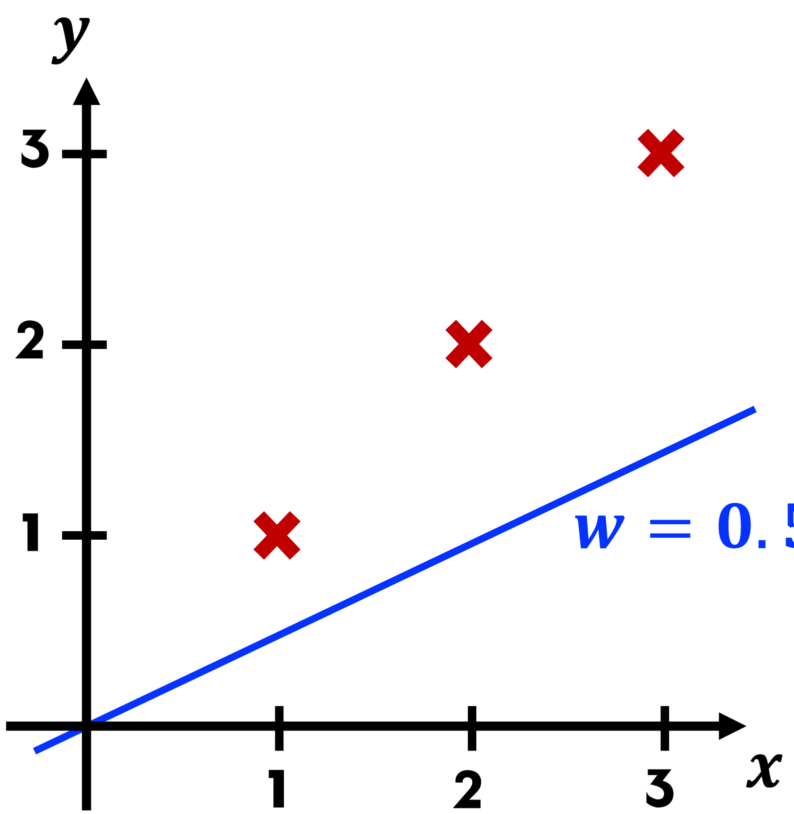
$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{2m}\sum_{i=1}^{m}(wx^{(i)} - y^{(i)})^2 = \frac{1}{2m}\left(0^2 + 0^2 + 0^2\right) = 0$$

TheMAD

Cost function

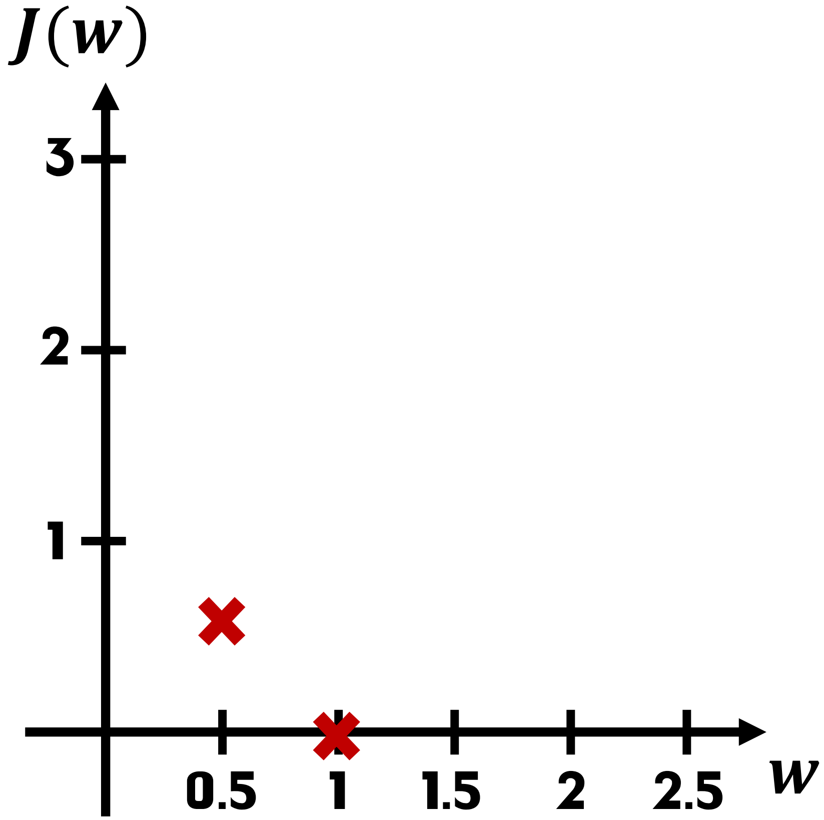
$f_w(x)$

For fixed w, function of x



$J(w)$

Function of w

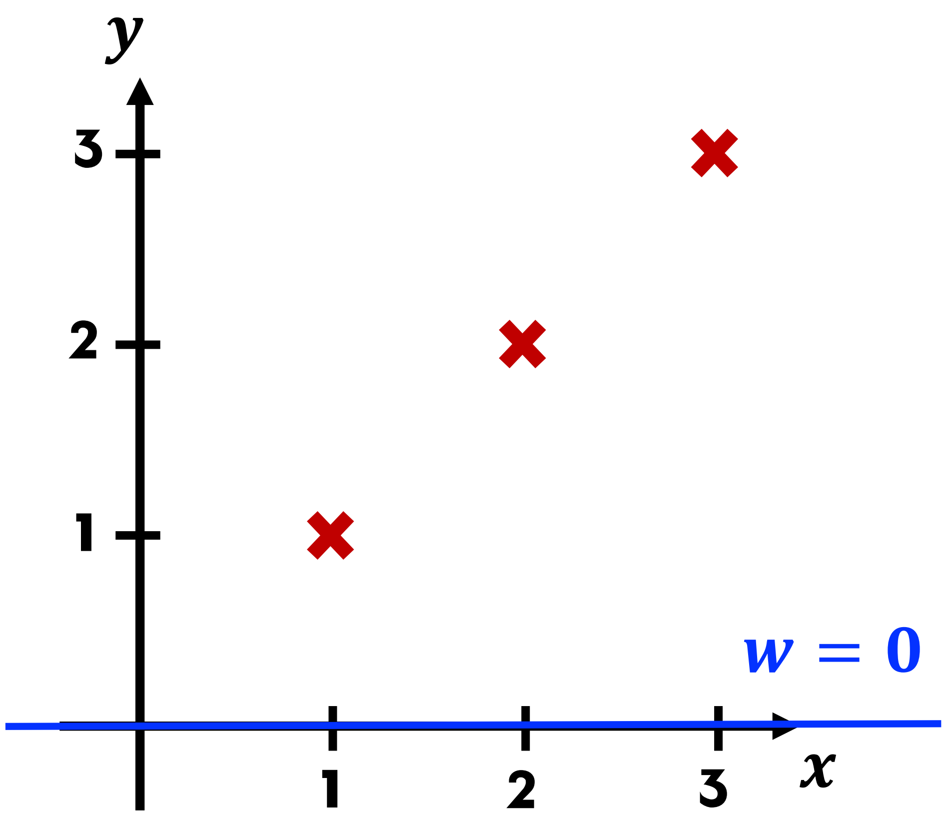


$$J(w) = \frac{1}{2m}\left((0.5-1)^2+(1-2)^2+(1.5-3)^2\right) = \frac{3.5}{6} = 0.58$$

Cost function

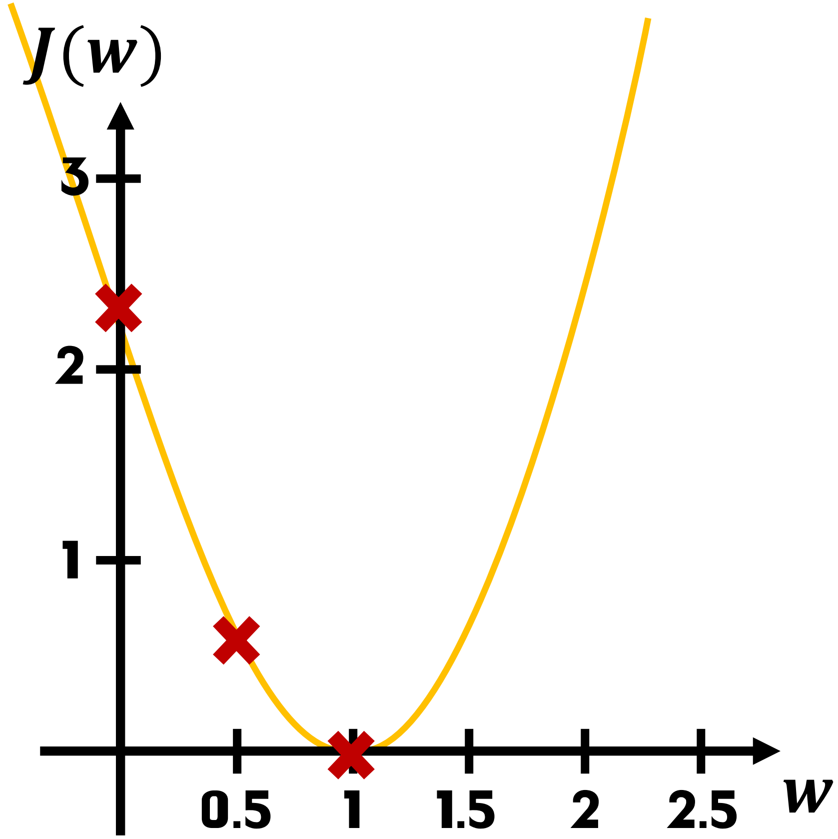
$f_w(x)$

For fixed w, function of x



$J(w)$

Function of w



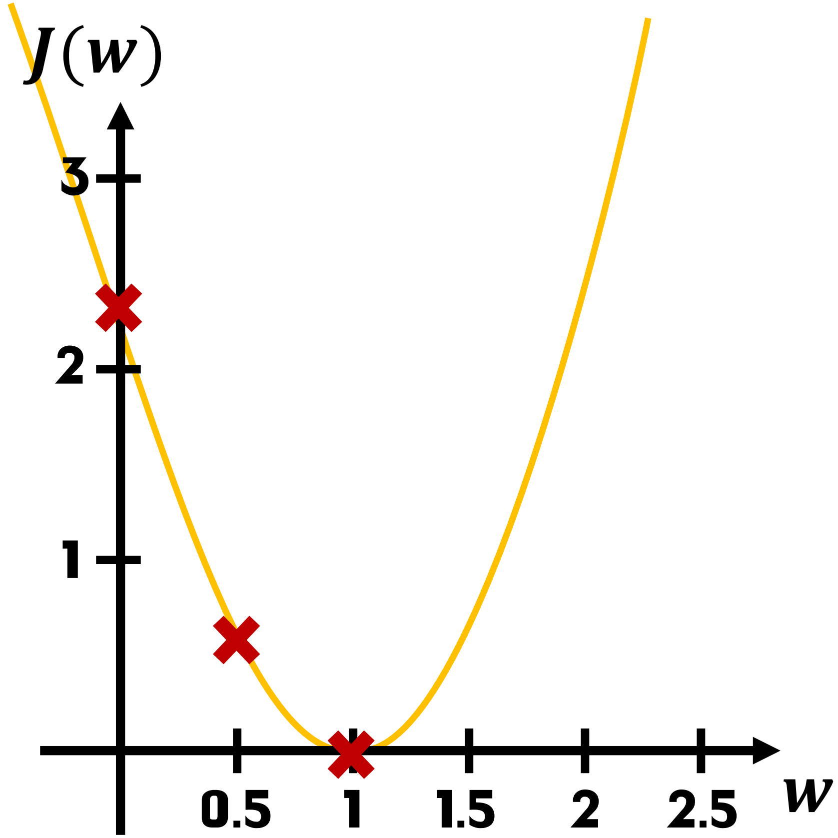
$$J(w) = \frac{1}{2m}\left((1)^2+(2)^2+(3)^2\right) = \frac{14}{6} = 2.3$$

TheMAD

Cost function

$J(w)$

Function of w



Goal of linear regression:

minimize $J(w)$

Choose w to minimize $J(w)$

Logistic Regression
: classification
(로지스틱 회귀)

Classification

Question	Answer "y"	
Is this email spam ?	no	yes
Is the transaction fraudulent?	no	yes
Is the tumor malignant?	no	yes

y can only be one of two values

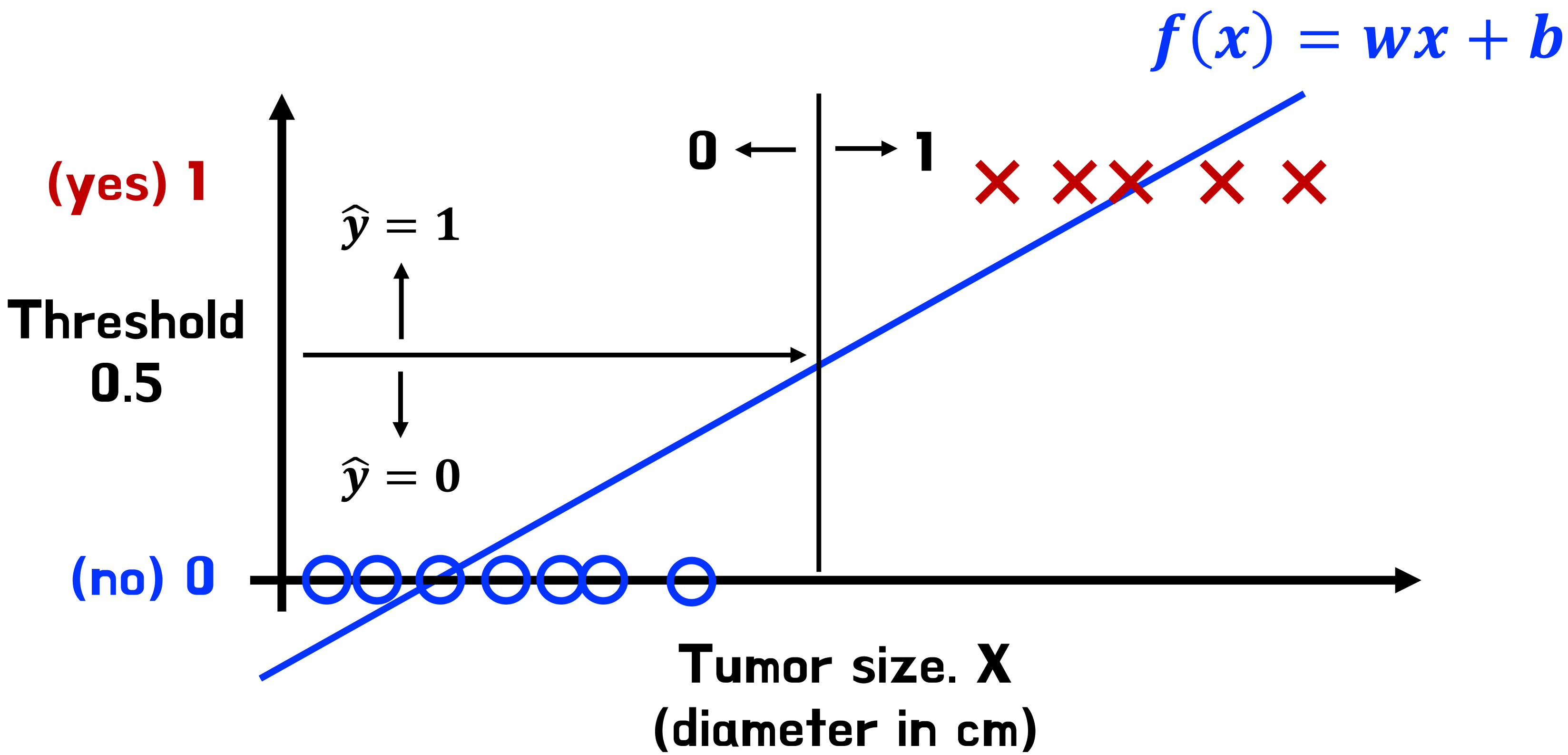
"binary classification"

= category

false — 0 → Negative class ≠ "bad" absence

true — 1 → Positive class ≠ "good" presence

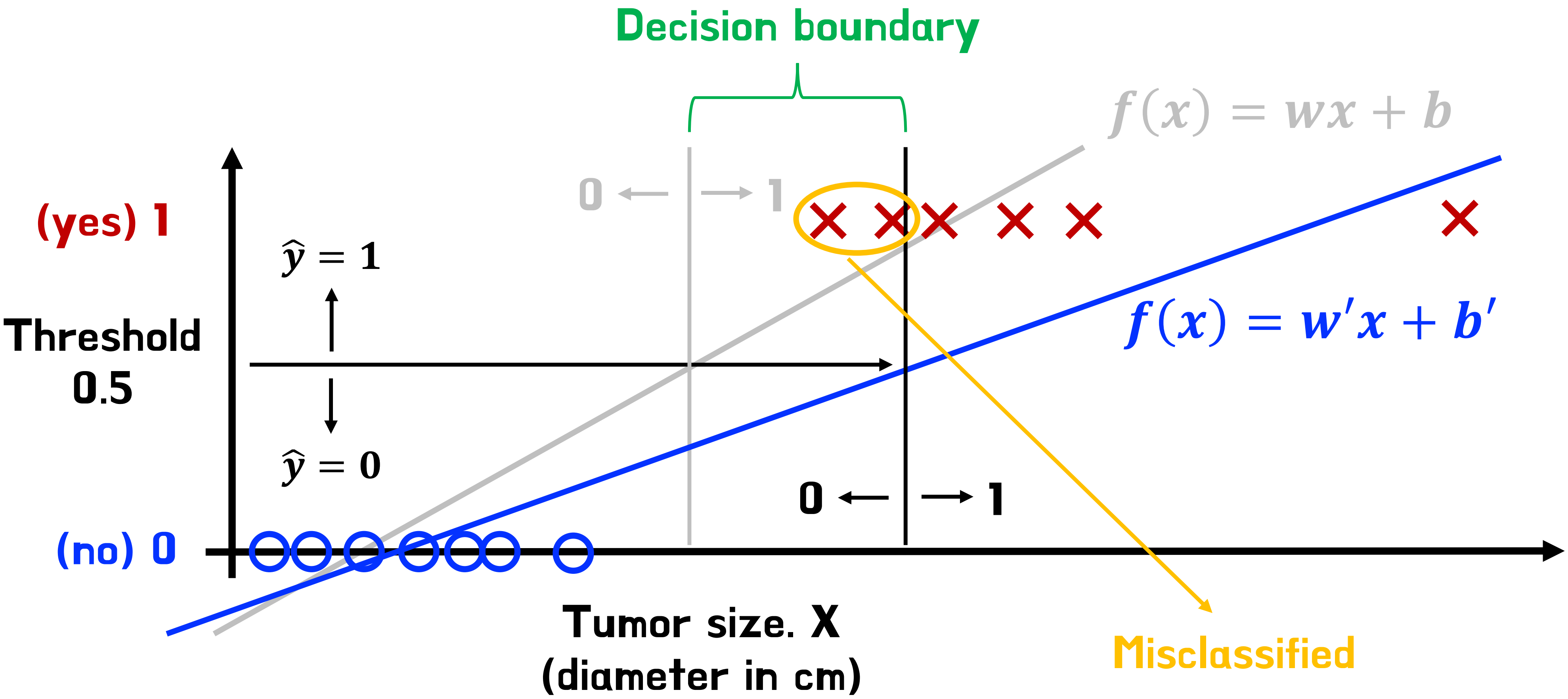
Linear regression



If $f(x) < 0.5 \rightarrow \hat{y} = 0$

If $f(x) \geq 0.5 \rightarrow \hat{y} = 1$

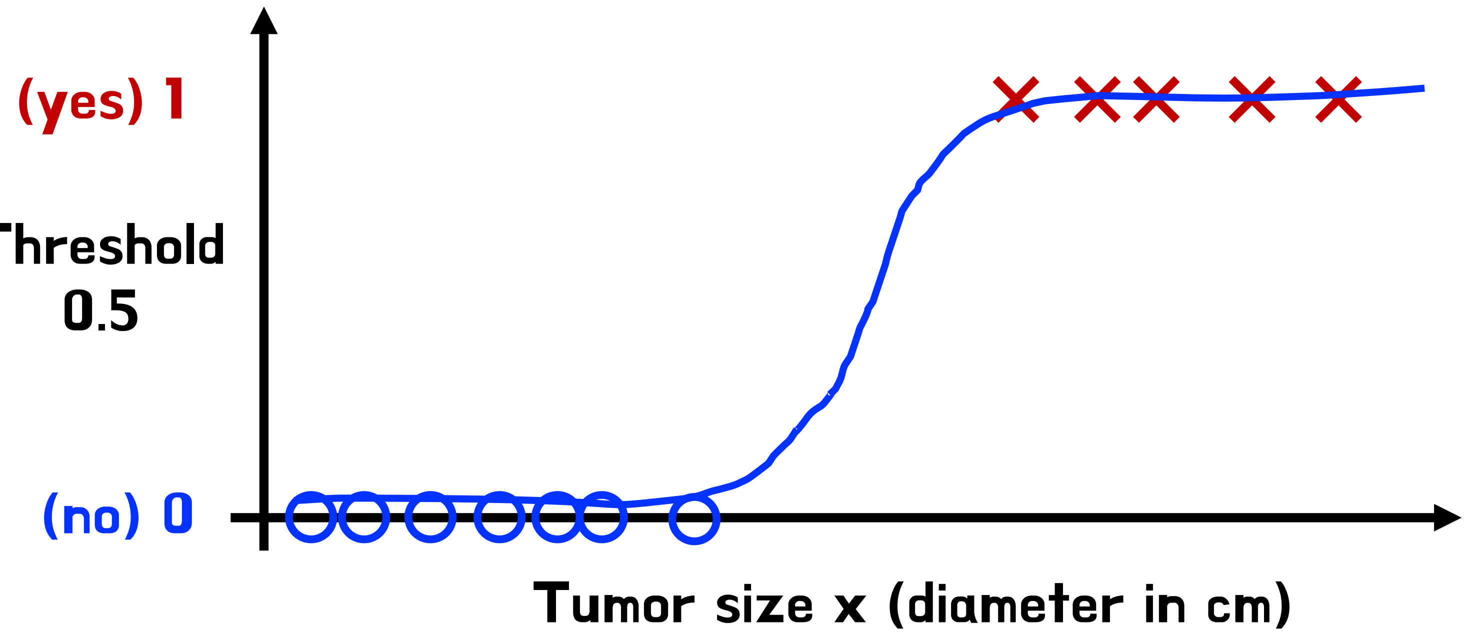
Linear regression

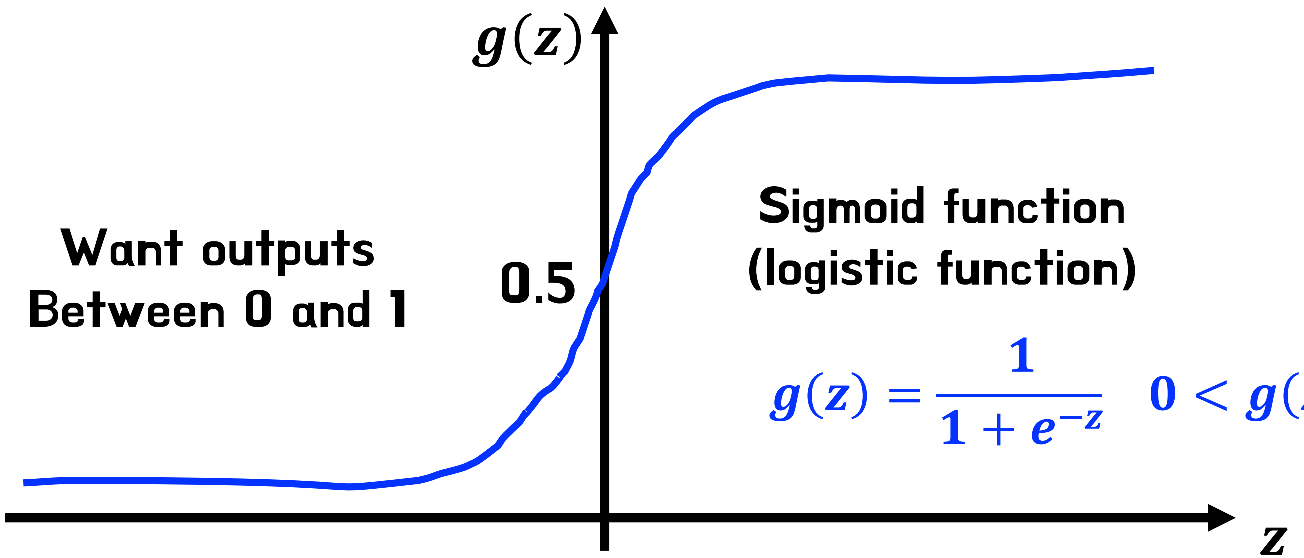


If $f(x) < 0.5 \rightarrow \hat{y} = 0$

If $f(x) \geq 0.5 \rightarrow \hat{y} = 1$

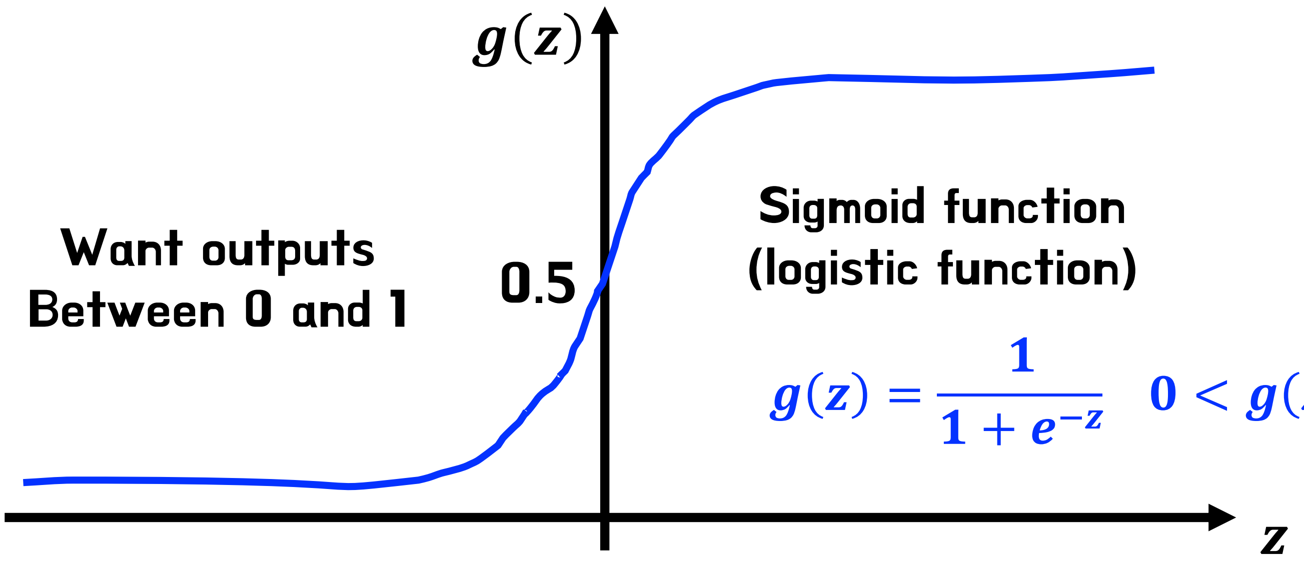
Logistic regression





$$g(z) = \frac{1}{1+e^{-z}} \quad 0 < g(z) < 1$$

Logistic regression



Want outputs
Between 0 and 1

Sigmoid function
(logistic function)

$$g(z) = \frac{1}{1+e^{-z}} \quad 0 < g(z) < 1$$

$$f_{w,b}(x) = wx + b$$

in linear regression

$$z = wx + b$$

$$f_{w,b}(x) = g(\underbrace{wx+b}_{=z}) = \frac{1}{1+e^{-(wx+b)}}$$

$$g(z) = \frac{1}{1+e^{-z}}$$

"Logistic regression"

Logistic regression

$$f_{w,b}(x) = \frac{1}{1+e^{-(wx+b)}}$$

"probability" that class is 1

Example:

x is "tumor size"

y is 0 (양성)

or 1 (악성)

$$f_{w,b}(x) = 0.7$$

70% chance that y is 1

$$f_{w,b}(x) = P(y=1|\, x; w\,, b)$$

Probability that y is 1,

given input x, parameters $w\,, b$

$$P(y=0) + P(y=1) = 1|$$

Logistic regression

$$f_{w,b}(x) = g(wx + b) = \frac{1}{1 + e^{-(wx+b)}}$$

$$= P(y = 1 \mid x; w, b)$$

Is $f_{w,b}(x) \geq 0.5$? **Yes:** $\hat{y} = 1$ **No:** $\hat{y} = 0$

When is $f_{w,b}(x) \geq 0.5$?

$$g(z) \geq 0.5$$

$$z \geq 0$$

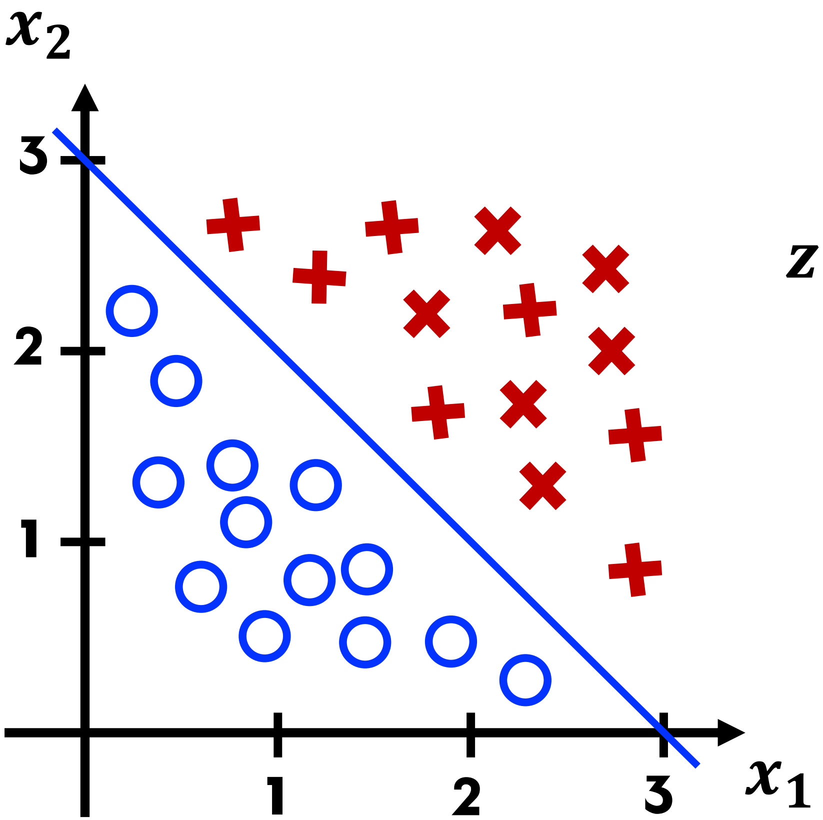
$$wx + b \geq 0 \qquad wx + b < 0$$

$$\hat{y} = 1 \qquad \hat{y} = 0$$

Decision boundary

$$f_{w,b}(x) = g(z) = g(w_1 x_1 + w_2 x_2 + b)$$

$w_1 = 1$, $w_2 = 1$, $b = -3$



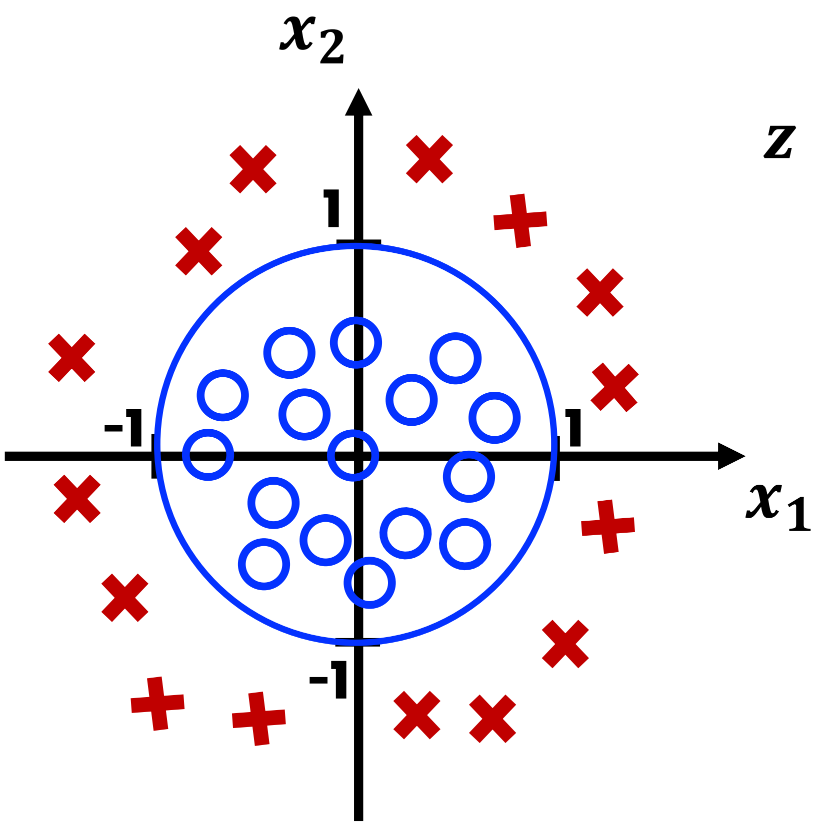
$$z = wx + b = 0$$
$$= x_1 + x_2 - 3 = 0$$

$$x_1 + x_2 = 3$$

Non-linear decision boundaries

$$f_{w,b}(x) = g(z) = g(w_1 x_1{}^2 + w_2 x_2{}^2 + b)$$

$w_1 = 1$, $w_2 = 1$, $b = -1$



$$z = x_1{}^2 + x_2{}^2 - 1 = 0$$

$$x_1{}^2 + x_2{}^2 = 1$$

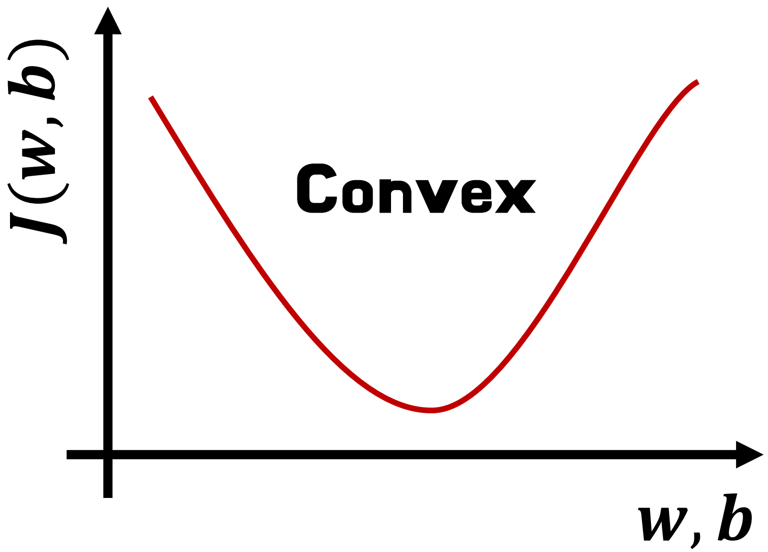
Squared error cost

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2}\left(f(x^{(i)}) - y^{(i)}\right)^2$$

$$= L\left(f\left(x^{(i)}\right), y^{(i)}\right)$$ **Loss**

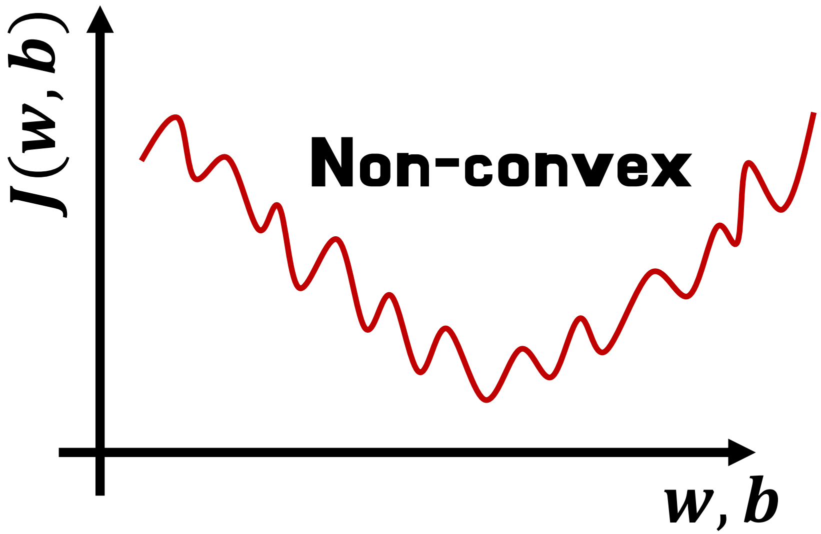
Linear regression

$$f_{w,b}(x) = wx + b$$



Logistic regression

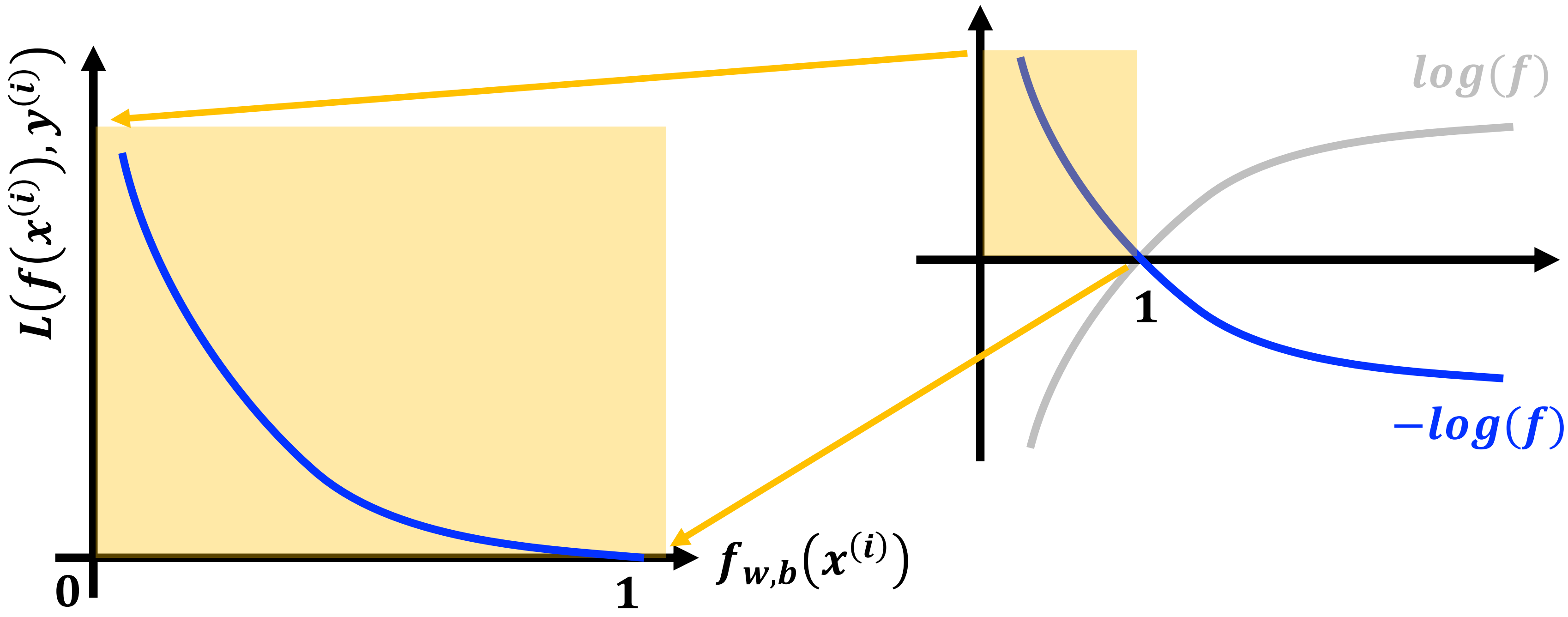
$$f_{w,b}(x) = \frac{1}{1 + e^{-(wx+b)}}$$



 TheMAD

Loss function

$$L\left(f\left(x^{(i)}\right), y^{(i)}\right) = \begin{cases} -log\left(f_{w,b}\left(x^{(i)}\right)\right) & if \;\; y^{(i)} = 1 \\ -log\left(1 - f_{w,b}\left(x^{(i)}\right)\right) & if \;\; y^{(i)} = 0 \end{cases}$$

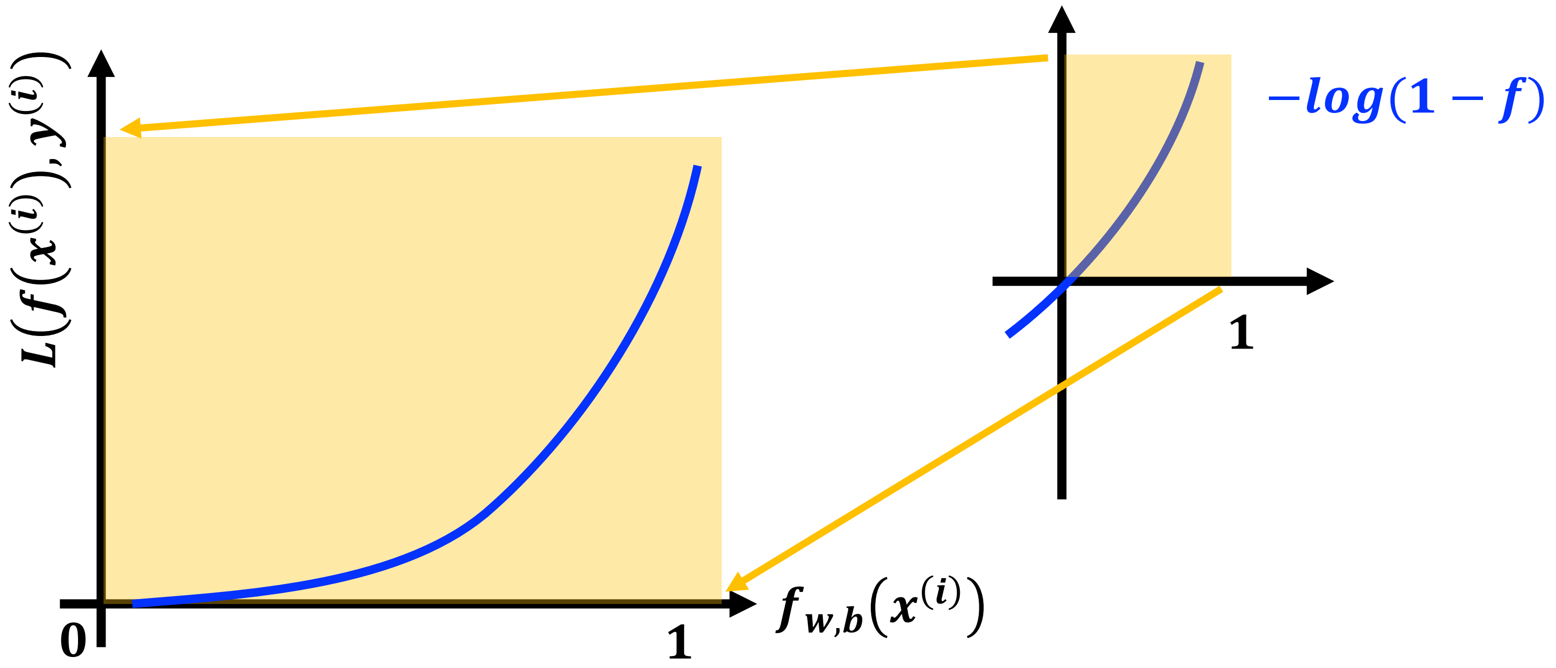


As $f_{w,b}\left(x^{(i)}\right)$ → 1 then loss → 0

As $f_{w,b}\left(x^{(i)}\right)$ → 0 then loss → infinite

Loss function

$$L\left(f\left(x^{(i)}\right), y^{(i)}\right) = \begin{cases} -log\left(f_{w,b}\left(x^{(i)}\right)\right) & if \;\; y^{(i)} = 1 \\ -log\left(1 - f_{w,b}\left(x^{(i)}\right)\right) & if \;\; y^{(i)} = 0 \end{cases}$$



As $f_{w,b}\left(x^{(i)}\right)$ → 1 then loss → infinite

As $f_{w,b}\left(x^{(i)}\right)$ → 0 then loss → 0