

D&A 2주차

# Index

(0) Review

(1) Regression

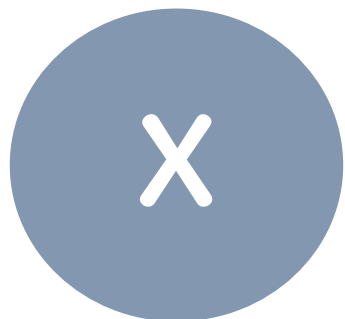
- Gradient decent

(2) Classification

- SVM (support vector machine)

(3) 과제 소개

# 기계학습 (지도학습)



정답지

X,Y가 주어졌을 때,  $F(X)$ 를 추정 하는 것.

이것은 나무다

Y



X

**이것은 나무가 아니다**



Y

X

X



Y

1

1

1

0

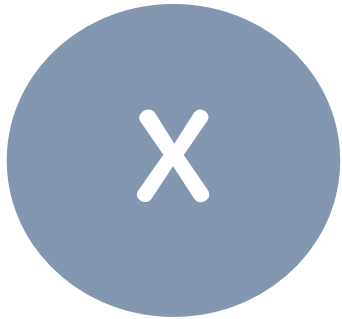
0

0

$F(x)$

수학적으로  $F(X)$ 를 구할 수 있을까?

$F(X)$ 를 구하는 여러 방법들이 있다.



정답지

X,Y가 주어졌을 때,  $F(X)$ 를 추정 하는 것.

그렇다면  $F(X)$ 를 왜 만들까?



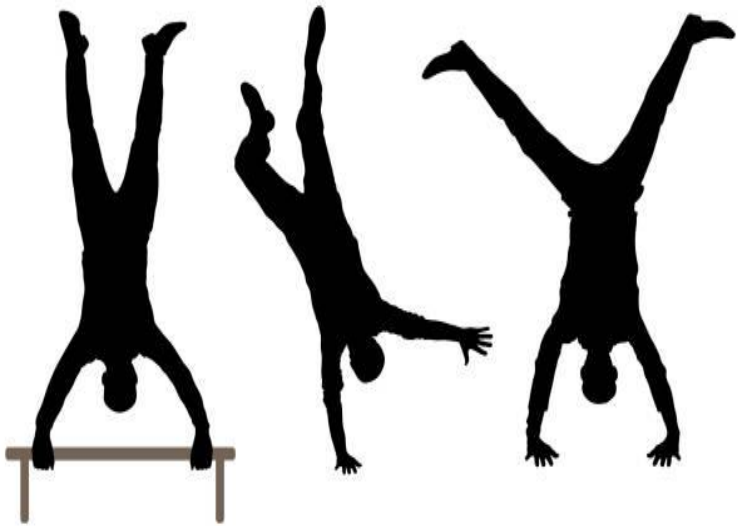
새로운 입력이 들어왔다.



우리가 만든  $F'(X)$ 는  
나무와 나무가 아닌 것을  
잘 구분할 수 있는  $F(X)$ 이다.

나무입니다  
 $\rightarrow 1$





새로운 입력이 들어왔다.

$F(x)$

우리가 만든  $F'(x)$ 는  
나무와 나무가 아닌 것을  
잘 구분할 수 있는  $F(x)$ 이다.

나무가 아닙니다  
 $\rightarrow 0$



기존 데이터에서 문제(X)와 정답(Y)을 학습하여  $F'(X)$ 를 추정합니다.



새로운 X가 들어왔을 때도 높은 성능을 갖는  $F'(X)$ 를 추정하는 것이 기계학습(지도학습)의 목표입니다.

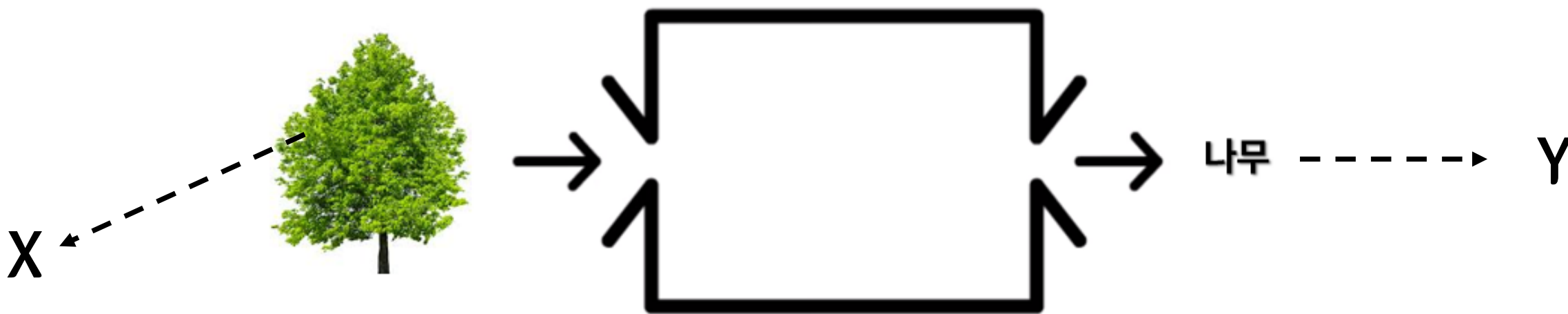

$$F(x)$$

**X, y을 활용하여 기계학습 알고리즘을 통해서  $F'(X)$ 를 만들 것입니다.**

**우리가 만든  $F'(X)$ 는 기존의 데이터의 학습을 통해 → 적합 되었다. (not 과소적합)**

**새로운 X가 등장했을 때도 잘 맞출 수 있는  $F'(X)$ 여야 합니다. → not 과적합**

**이 때의  $F'(X)$ 를 일반화된 모델이라고 합니다.**



기계학습: 문제와 정답을 학습하여 **블랙박스**를 만드는 것

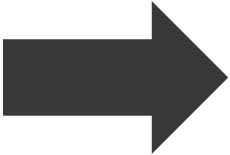
정교하게 학습된 블랙박스는 추후에 **새로운 문제를** 입력 받았을 때 맞는 정답 출력한다

CGV

이름	관람 영화 장르	함께 관람한 인원	관람 요일	팝콘 구매 여부
홍재성	공포	3	금요일	0
김보현	로맨스	1	금요일	1
김세홍	액션	2	토요일	1
한승수	공포	1	일요일	0
박새롬	공포	2	월요일	1
강지원	액션	2	화요일	0
...	...	...	...	...



고객 이름	관람 영화 장르	함께 관람한 인원	관람 요일
홍재성	공포	3	금요일
김보현	로맨스	1	금요일
김세홍	액션	2	토요일
한승수	공포	1	일요일
박새롬	공포	2	월요일
강지원	액션	2	화요일
...	...	...	...



Popcorn\_model

팝콘 구매 여부
0
1
1
0
1
0
...

새로운 고객

고객 이름	관람 영화 장르	함께 관람한 인원	관람 요일
권혁우	공포	3	금요일
민혜리	로맨스	1	금요일
서진영	액션	2	토요일
장성민	공포	1	일요일
조혁준	공포	2	월요일



Popcorn\_model

팝콘 구매 여부
0
0
1
1
0

문제 확인  
및 정의



가설 제시



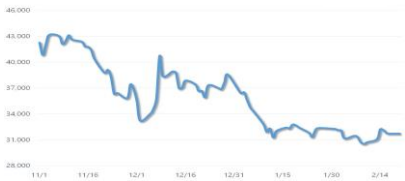
팝콘 구매 여부  
예측 모델 개발



프로모션 전략



검증



상반기 팝콘 매출 감소

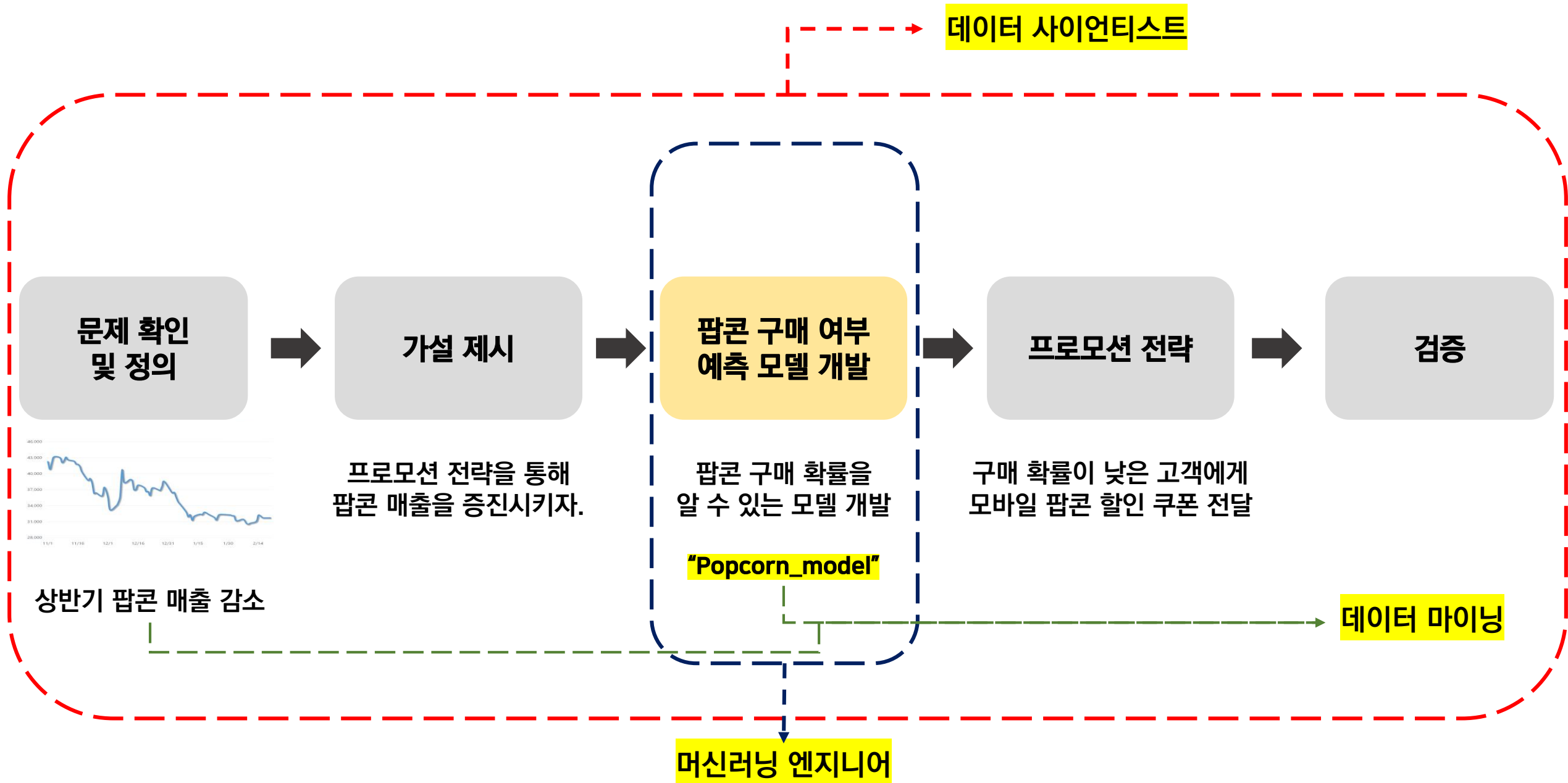
프로모션 전략을 통해  
팝콘 매출을 증진시키자.

팝콘 구매 확률을  
알 수 있는 모델 개발

“Popcorn\_model”

구매 확률이 낮은 고객에게  
모바일 팝콘 할인 쿠폰 전달







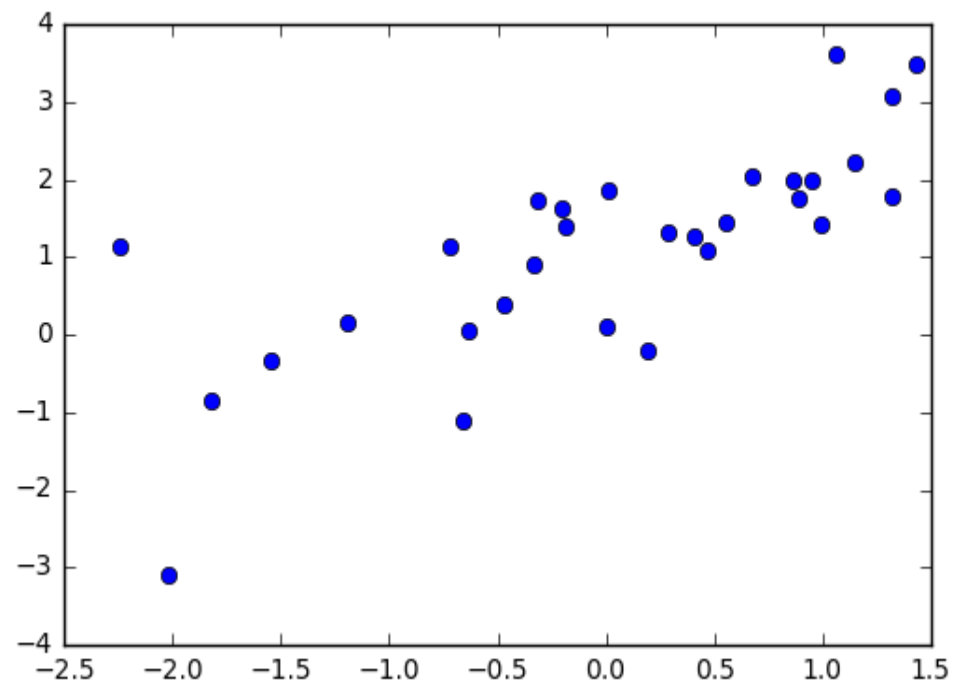
# Machine learning Algorithm



- ✓ 기계 학습은 예측 변수 (=종속변수) 존재 여부에 따라서 지도 학습과 비지도 학습으로 나뉜다.
- ✓ 종속 변수가 연속형 변수일 경우 회귀 문제, 종속 변수가 이산형 변수일 경우 분류 문제이다.
- ✓ 각 상황에 맞는 적절한 알고리즘을 사용할 수 있어야 한다.
- ✓ 지도 학습과 비지도 학습에서 사용되는 알고리즘은 구분되지만, 회귀와 분류에서 사용되는 알고리즘은 대부분 공통된다.

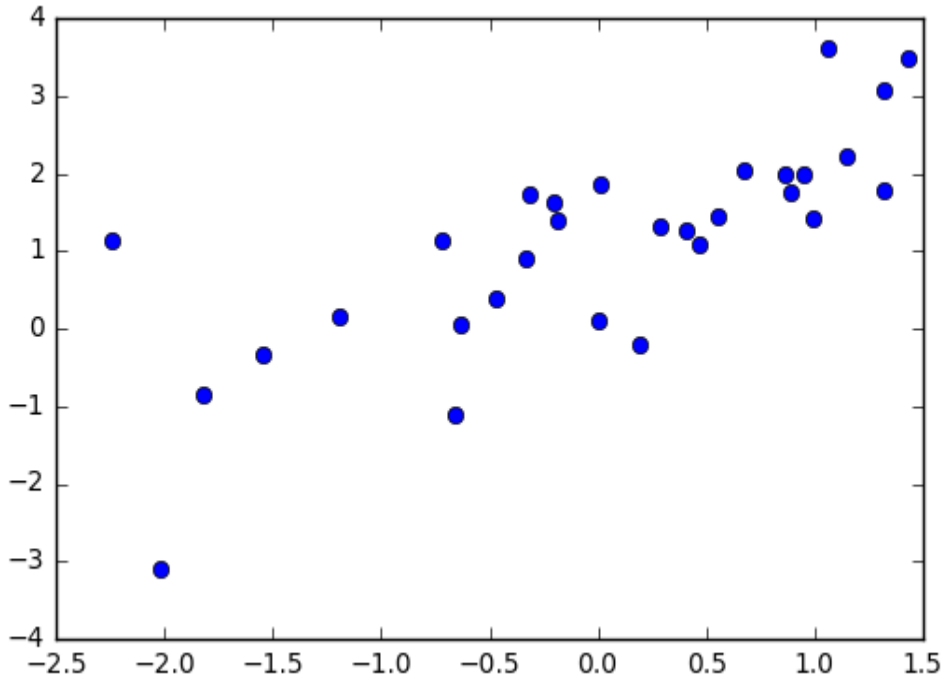


지도학습(회귀) - Gradient decent



X와 Y값은 모두 존재한다.  
X에서 Y로 가기 위한  $F(x)$ 는 무엇일까?

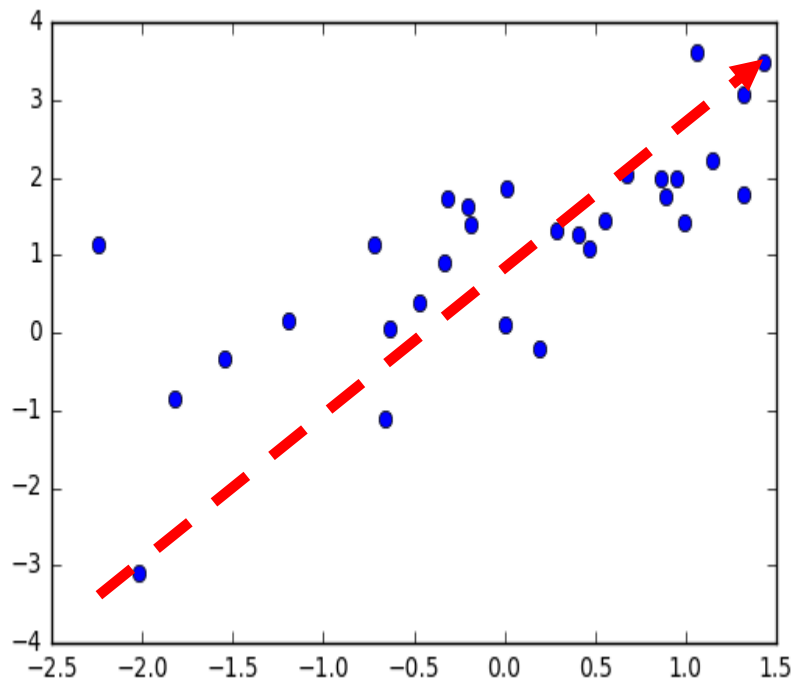
## 지도학습(회귀) - Gradient decent



$F(x)$  = 모델 = 모형 = 블랙박스

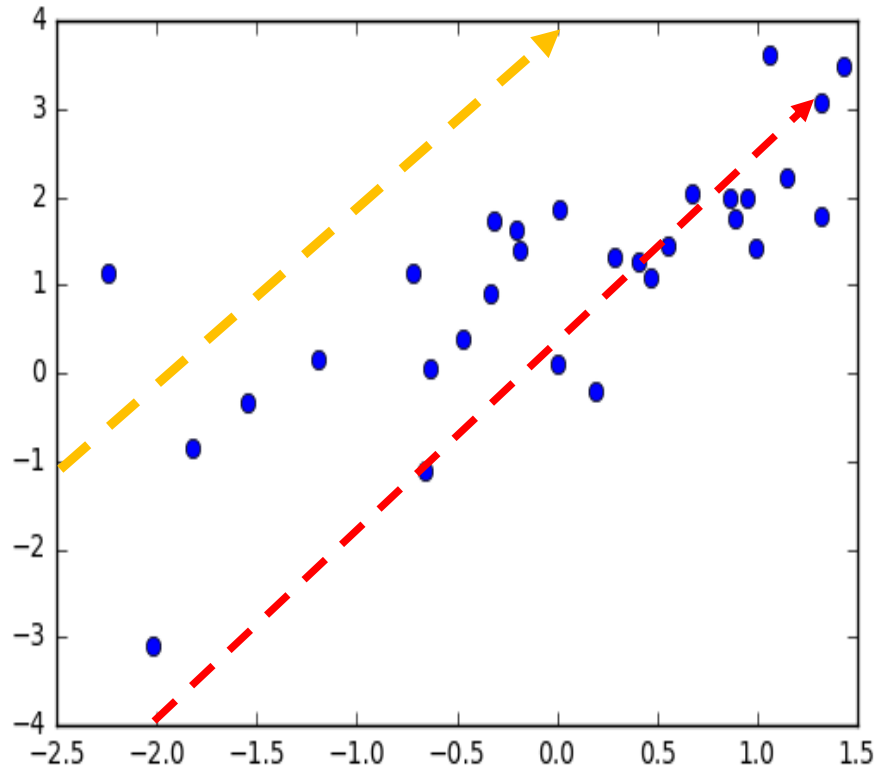
- ✓ X가 증가할수록 y가 증가하는 선형적인 모습을 띤다.
- ✓  $F(x) = wx + b$  의 형태로 표현할 수 있다.
- ✓ 모형은( =  $F(x)$ ) 주어진 데이터의 패턴을 가장 잘 표현할 수 있어야 한다. 앞의 예시에서 X와 Y를 가장 잘 표현할 수 있는 모형은  $Y = 2X$  였다.
- ✓ 즉, 모형은 데이터의 패턴을 정의한 것을 의미한다.
- ✓ 데이터의 패턴을 가장 잘 정의할 수 있는 모형을 찾아야 한다
- ✓  $Y = wx + b$ 에서 모형의 형태를 결정짓는 변수는 무엇인가?
- ✓  $w$ (기울기)와  $b$ (절편)이다. 여기서 모형의 형태를 결정짓는 변수를 모수라고 하며, 파라미터라고도 부른다.
- ✓ 기계학습은 주어진 데이터의 패턴을 가장 표현할 수 있는 파라미터를 찾는 것이다.
- ✓ 선형 회귀에서는 주어진 데이터의 패턴을 가장 잘 표현할 수 있는 기울기와 절편을 찾는 것이 기계학습의 목표이다.
- ✓ 그렇다면, 위의 식에서 최적의 파라미터 ( $W, b$ )를 어떻게 구할 수 있을까?

## 지도학습(회귀) - Gradient decent



- ✓ X, Y의 형태 (데이터의 패턴)을 표현할 수 있는 모형을 만드는 것이 기계학습의 목표이다.
- ✓ 그렇다면 100% 정확하게 X와 Y를 표현할 수 있는 모형을 만들 수 있을까?
- ✓ 거의 불가능하다.
- ✓ 따라서 오차를 최소화 할 수 있는 모형을 만들어야 하며, 오차를 최소화할 수 있는  $w$ 와  $b$ 값을 찾아야 한다.
- ✓ 여기서 오차에 대한 식이 **손실함수**이며, 오차를 최소화 할 수 있는 방법 중 하나가 gradient decent이다.
- ✓ 즉, 손실함수를 최소화 할 수 있는 직선을 찾아야 하는데 그 방법이 Gradient decent 이다.

## 지도학습(회귀) - Gradient decent



$$MSE = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$$

$Y = 2x$ 라는 식이 위의 데이터를 가장 잘 표현할 수 있는 모형인 것 같아서 만들어봤다. 여기서 파라미터는 무엇인가?

$MSE = 3.4$

$Y = 2x + 8$  라는 식을 만들어 보았다.

$MSE = 6.7$

둘 중에 어떻게 좋은 모형인가?

아! MSE를 최소화할 수 있는 모형을 찾으면 되겠구나!

아! MSE를 최소화할 수 있는 파라미터를 찾으면 되겠구나!

아! MSE를 최소화할 수 있는 기울기와 절편을 찾으면 되겠구나!  
Gradient decent 를 통해서 MSE를 최소화 할 수 있는 파라미터를 찾을 수 있다.



## 지도학습(회귀) – Gradient decent

MSE = 파라미터에 대한 식 =

주어진 데이터에서 (X,Y) 오차를 최소화 할 수 있는 모수를 찾는 것

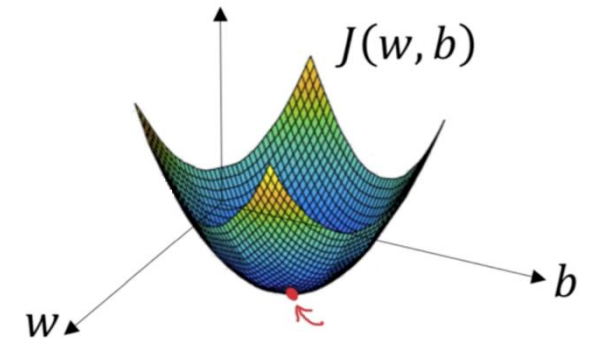
주어진 데이터에서 (X,Y) 오차를 최소화 할 수 있는 파라미터를 찾는 것

주어진 데이터에서 (X,Y) 오차를 최소화 할 수 있는 기울기와 절편을 찾는 것

주어진 데이터에서 (X,Y) MSE를 최소화 할 수 있는 기울기와 절편을 찾는 것

주어진 데이터에서 (X,Y) 손실함수를 최소화 할 수 있는 기울기와 절편을 찾는 것

주어진 데이터에서 (X,Y)  $MSE = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$  최소화 할 수 있는 기울기와 절편을 찾는 것

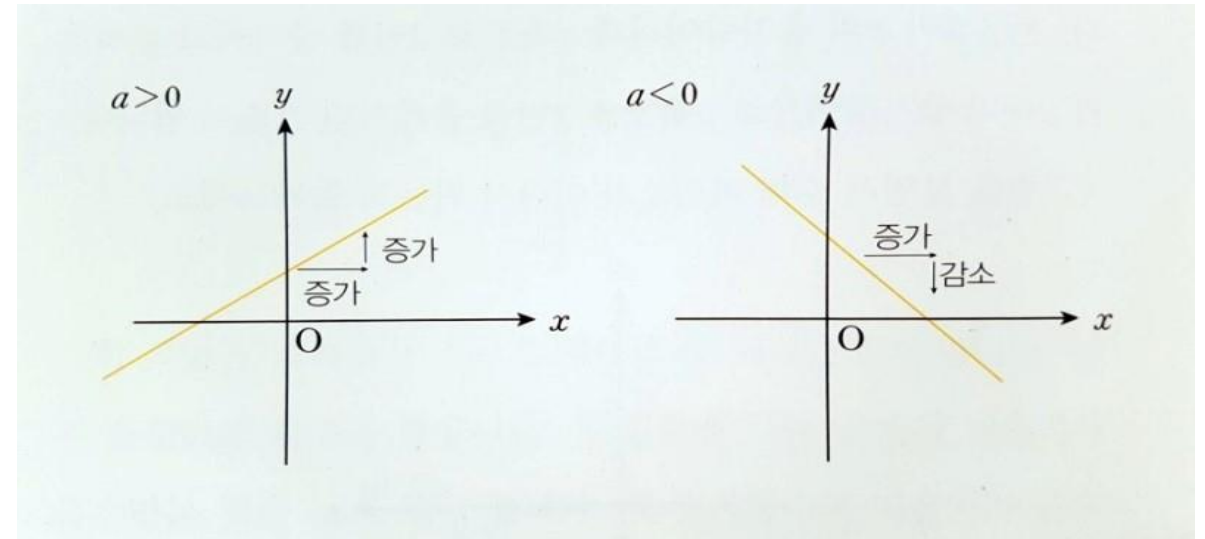
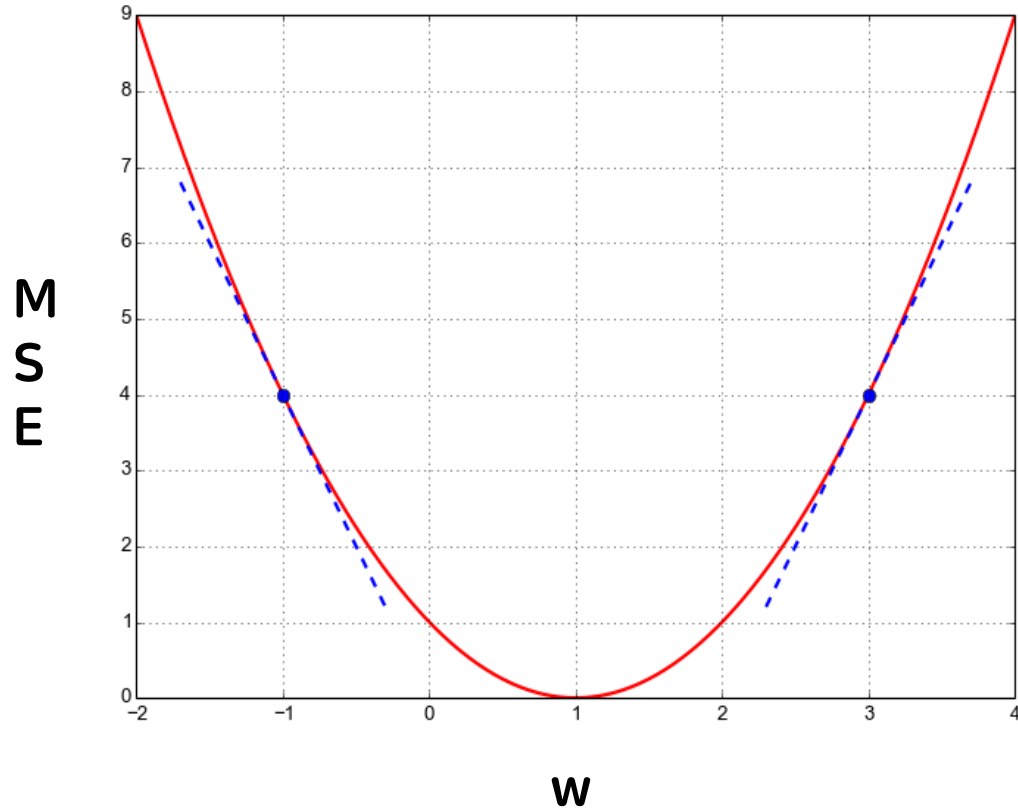


어떻게 할 수 있을까? → gradient decent를 통해서

MSE는 W와 b에 따라 달라지는데 미분을 통해 해결할 수 있을 것 같은 감이 오지 않은가?

## 지도학습(회귀) - Gradient decent

미분 : 한 지점에서의 변화율



- ✓ 미분했을 때 기울기가 양수이다.  $Y$ 값을 감소시키려면  $X$ 값을 어떻게 해야 할까?
- ✓ 미분했을 때 기울기가 음수이다.  $Y$ 값을 증가시키려면  $X$ 값을 어떻게 해야 할까?

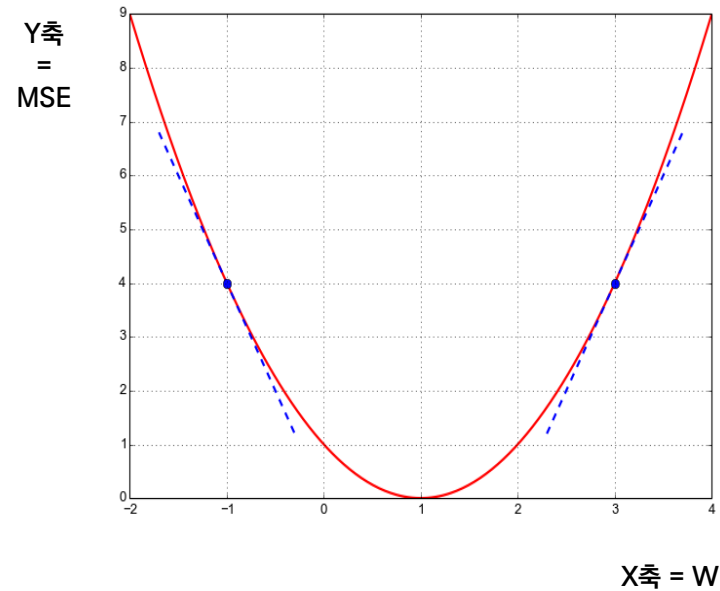
## 지도학습(회귀) – Gradient decent

### 손실함수의 기하학적 표현

아이디어의 핵심 : 미분 !

$$MSE = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$$

손실함수(MSE)는 모수(파라미터)에 대한 2차식이다. 기하학적으로 표현해보자



우리의 목표는 MSE를 최소화 할 수 있는  $W$ 를 찾는 것이다. 어떻게 하면 좋을까?

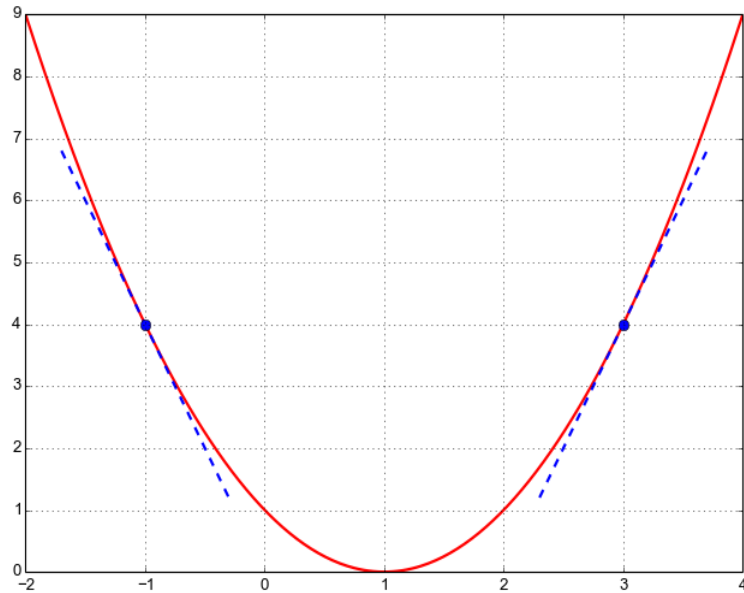
최적화 = 손실함수의 값을 최소화 하는 파라미터를 찾는 것

## 지도학습(회귀) - Gradient decent

$$w_{t+1} = w_t - \eta \frac{dE}{dw}$$

갱신된 가중치      초기의 가중치      학습율

Gradient = 손실함수를 가중치  $w$ 에 대해 편미분 = 가중치  $w$ 에 의한 손실함수의 변화율



1. 임의의 파라미터를 설정한다

2. 손실함수를 파라미터에 대해서 미분한다. 이를 통해 파라미터에 의한 손실함수의 변화율을 구할 수 있다.

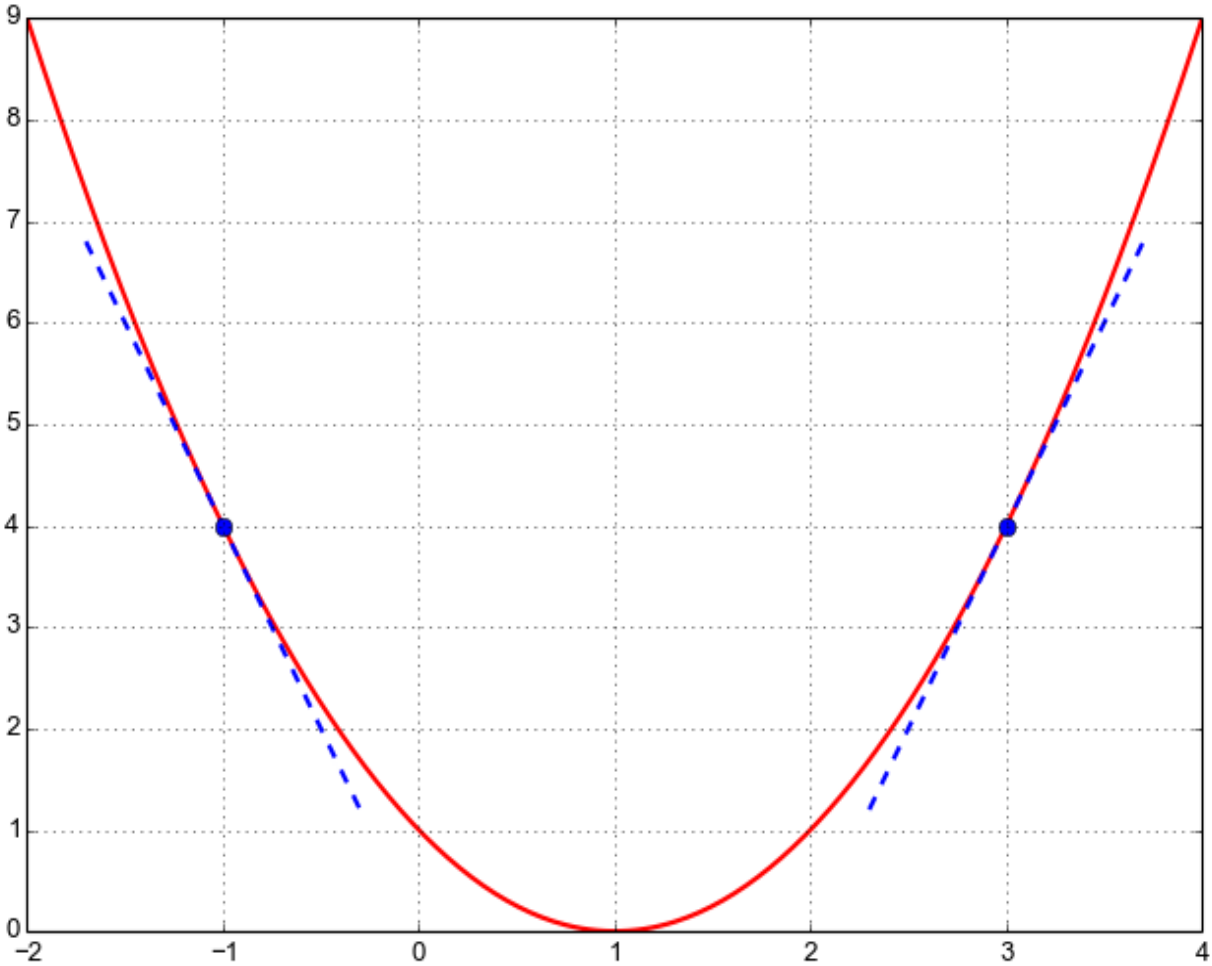
3. 파라미터에 대한 오차의 미분이 (+) 인 경우 손실함수를 감소시키려면 파라미터를 어떻게 조정해야 할까?

4. 파라미터에 대한 오차의 미분이 (-) 인 경우 손실함수를 감소시키려면 파라미터를 어떻게 조정해야 할까?

5. 파라미터를 줄이고 키우는 과정을 더 이상 오차가 줄어들지 않을 때 학습을 중단한다.

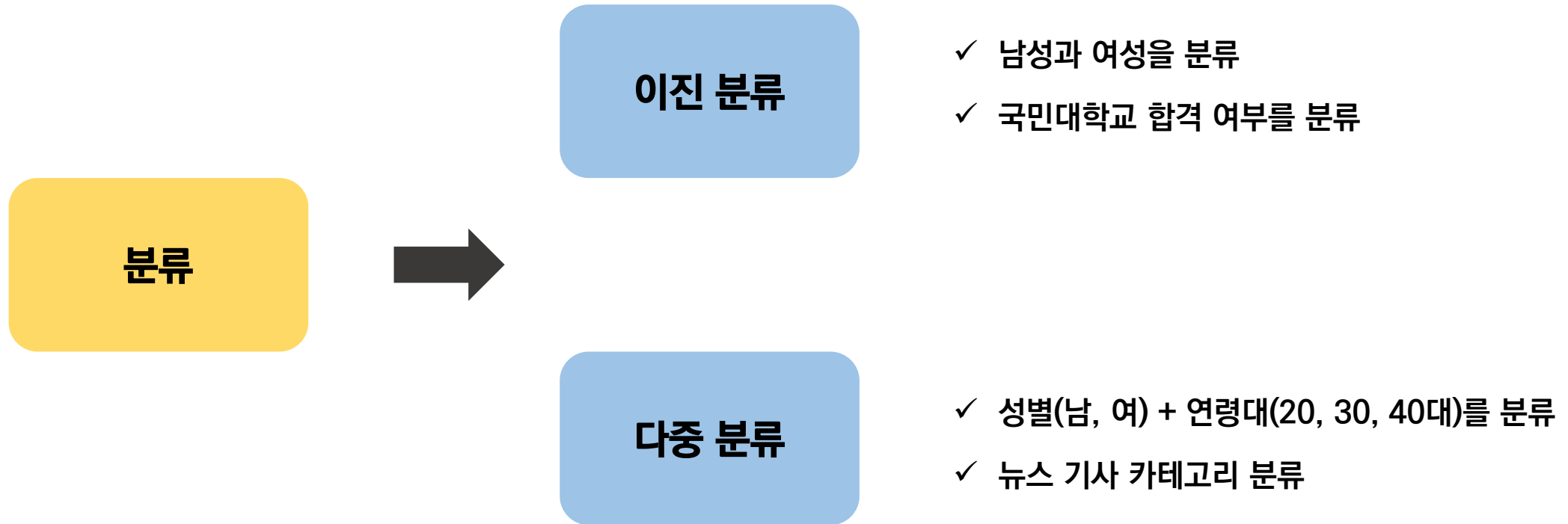
지도학습(회귀) - Gradient decent

$$w_{t+1} = w_t - \eta \frac{dE}{dw}$$





# Machine learning Algorithm - 분류

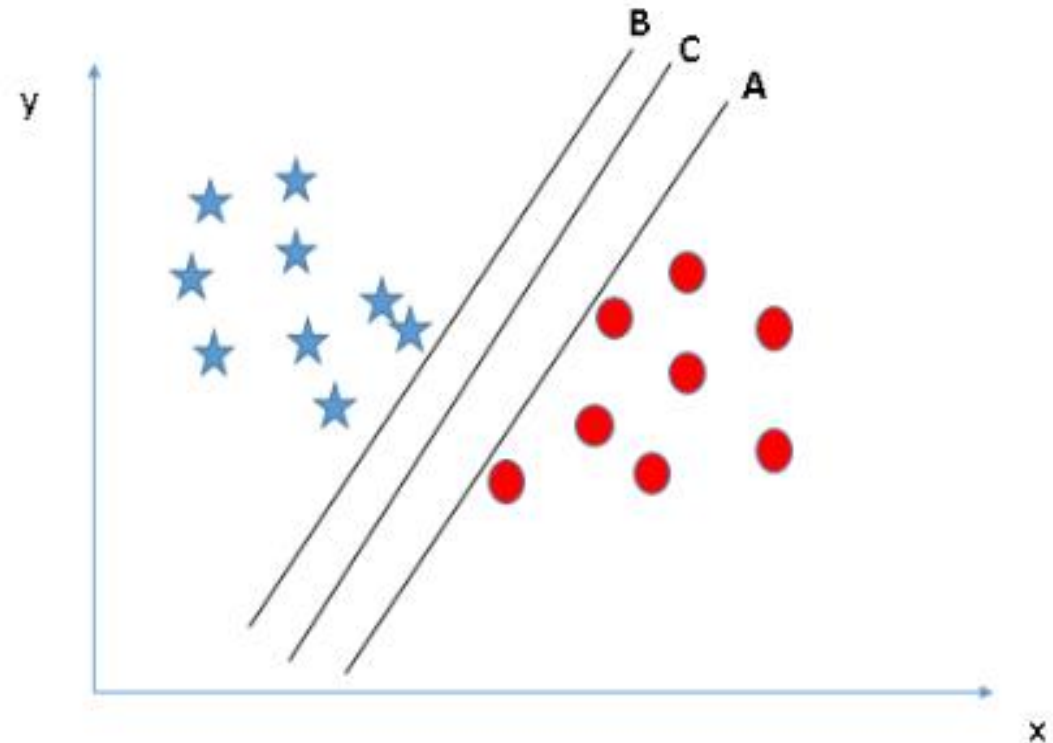


- ✓ 분류 문제는 이진 분류 (binary classification)와 다중 분류 (multi classification)으로 나뉜다.

# SVM (Support Vector Machine)



- ✓ 별 모양과 동그라미 모양을 구분하려고 한다. 이는 이진 분류 문제로 취급할 수 있다.
- ✓ 주어진 데이터에서 별 모양과 동그라미 모양을 어떻게 구분할 수 있을까?
- ✓ 별 모양과 동그라미 모양의 데이터 사이에 **적절한** 선을 긋는 방법을 생각해 볼 수 있다.



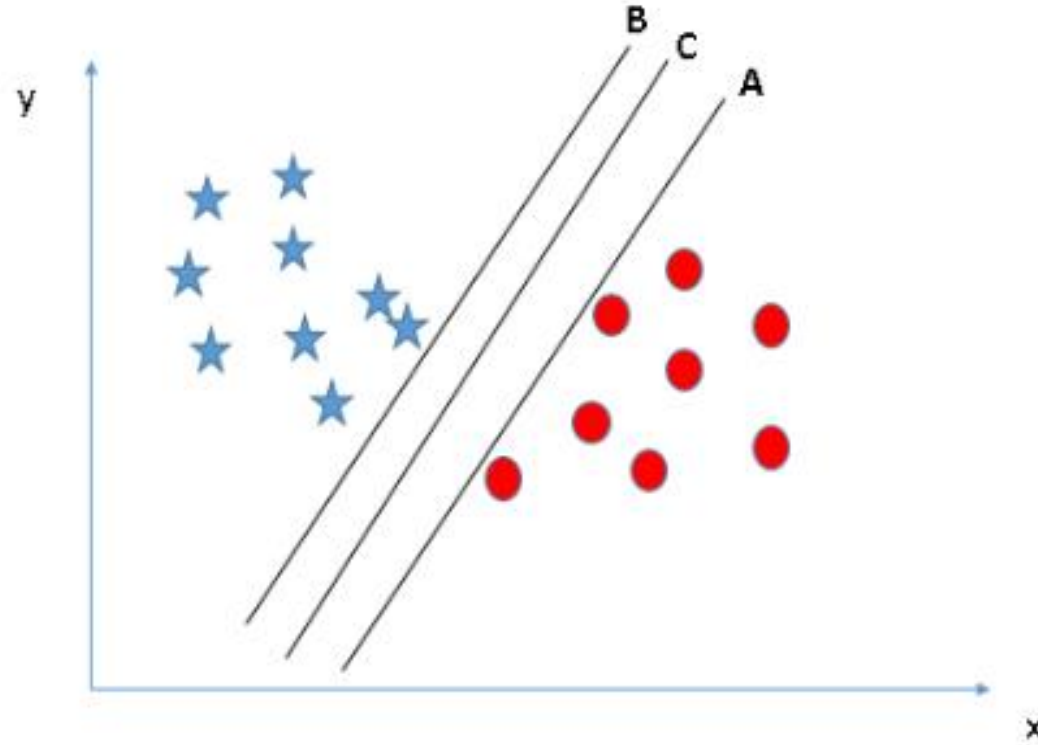
- ✓ A, B, C 중에 어떤 선이 **적절한** 선일까?
- ✓ 직감적으로 C라는 선이 적절한 선임을 추측해볼 수 있다.
- ✓ 기계 학습의 본질에 대해 고민해 봄으로서 왜 C가 적절한 선인지 알아보자.



### 일 반 화

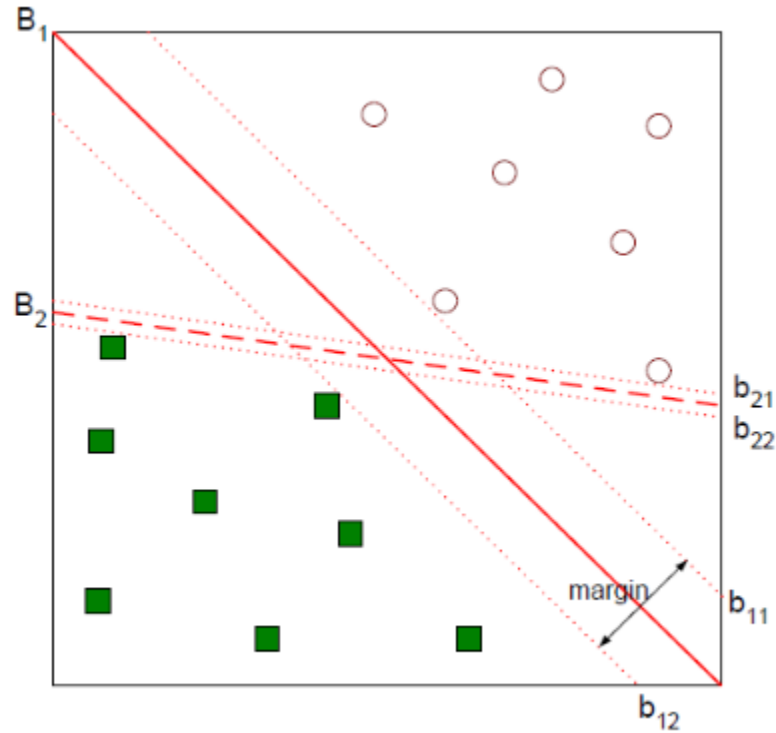
- ✓ 학습 데이터를 통해  $F(X)$ 를 만들었다. 우리는 Unseen 데이터에 대해서도 좋은 성능을 갖는 모델을 만들어야 한다.
- ✓ 학습 데이터만 잘 맞추는 것이 아니라 새로 등장할 데이터 또한 잘 맞춰야 한다.
- ✓ 학습 데이터와 Input 데이터가 달라져도 성능 차이가 나지 않게 하는 것을 일반화 라고 한다.

# SVM (Support Vector Machine)



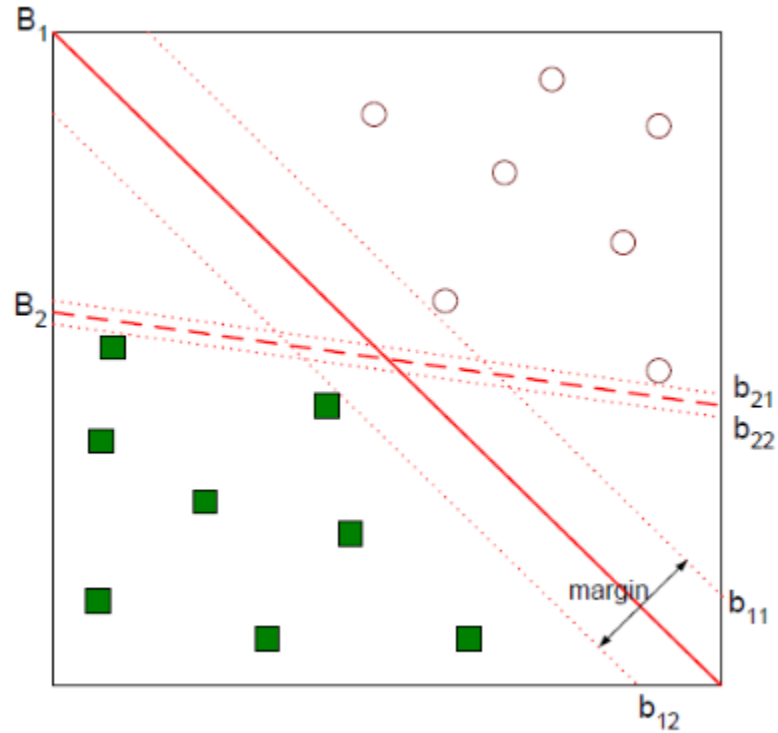
- ✓ A, B, C 중에 어떤 선이 **적절한** 선일까? → C
- ✓ C가 적절한 선인 이유는 두 범주를 **여유 있게** 구분하고 있기 때문이다.
- ✓ A, B를 사용해도 범주를 구분할 수 있다. 하지만 새로운 데이터에 대한 **일반화**된 성능을 기대하기 힘들 수 있다.
- ✓ 그렇다면 우리는 어떻게 복잡한 데이터에서 C와 같은 적절한 선을 찾을 수 있을까?

# SVM (Support Vector Machine)



- ✓ margin: 결정 경계와 가장 가까이에 있는 학습 데이터까지의 거리
- ✓ Support vector: 결정 경계로부터 가장 가까이에 있는 학습 데이터들
- ✓ SVM: 분류 오차를 줄이면서 동시에 margin을 최대로 하는 결정 경계를 찾는 모델
- ✓ margin을 최대화 시킨다는 것의 의미 → 두 범주를 여유 있게 가르는 결정 경계를 찾는다는 뜻

# SVM (Support Vector Machine)



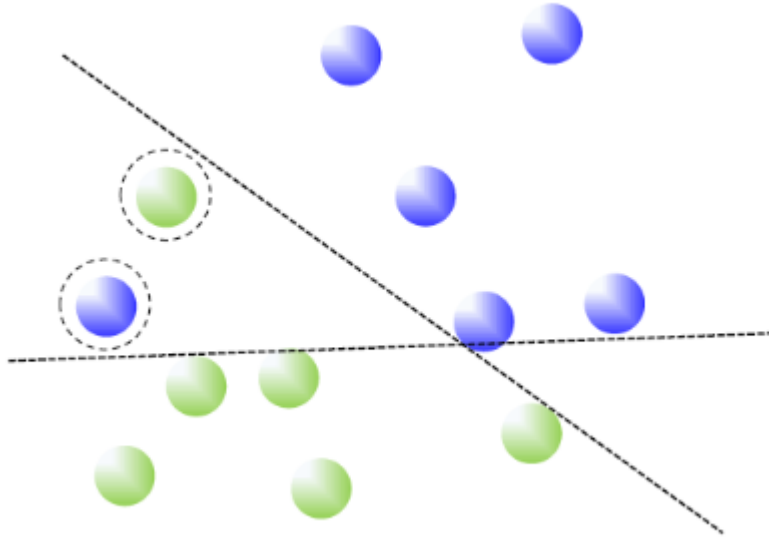
- ✓ margin을 최대화 시킨다는 것의 의미 → 두 범주를 여유 있게 가르는 결정 경계를 찾는다는 뜻
- ✓ 목적함수(object function) ?

## 지도학습(분류) - SVM

C-SVM

Kernel-SVM

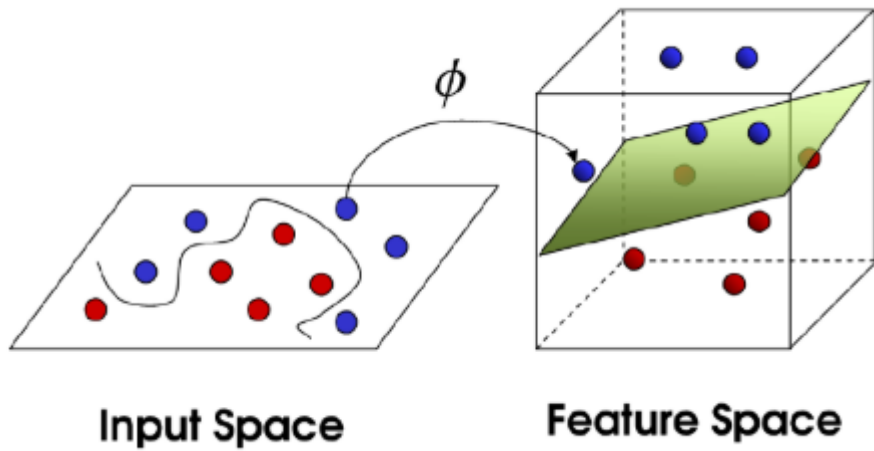
# C-SVM



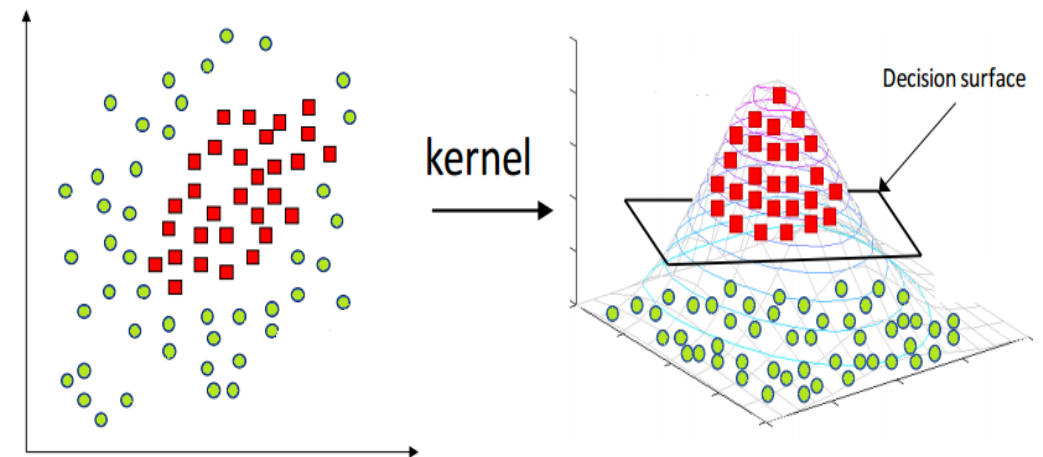
- ✓ 이 그림처럼 어떤 직선을 그어도 두 범주를 완벽하게 분류하기 어려운 경우가 많다.
- ✓ 기존의 SVM은 margin안에 관측치가 들어오는 것이 허용 안된다.
- ✓ 하지만 C-SVM은 margin안의 관측치의 존재를 허용한다.
- ✓ 이를 soft-margin 이라고 한다.

- ✓ C 값이 커질 수록 마진의 폭이 줄어든다.
- ✓ C 값이 커진다 → 마진 안의 관측치를 많이 허용한다.
- ✓ C 값이 작아진다 → 마진 안의 관측치를 많이 허용하지 않는다.
- ✓ C 값은 파라미터? 하이퍼 파라미터?

# Kernel - SVM



- ✓ 원 공간 (Input Space) → 고차원 공간 (Feature Space) 매핑
- ✓ 두 범주를 분류하는 초평면을 찾는다.



- ✓ 기존 데이터 공간에서는 SVM의 높은 성능을 기대하기 어렵다.
- ✓ Kernel trick을 통해 기존 데이터를 고차원 공간으로 매핑 시킨다.
- ✓ 그 후 초평면을 이용해 쉽게 구분한다.

## 참고 자료

경사하강법 코드 [http://doc.mindscale.kr/km/unstructured/gradient\\_descent.html](http://doc.mindscale.kr/km/unstructured/gradient_descent.html)

경사하강법 참고자료 [http://doc.mindscale.kr/km/data\\_mining/dm02.html](http://doc.mindscale.kr/km/data_mining/dm02.html)

SVM [https://drive.google.com/open?id=1QTVWfRh3eXCSBwHI0AY\\_y5dchF5e8n-j](https://drive.google.com/open?id=1QTVWfRh3eXCSBwHI0AY_y5dchF5e8n-j)

C-SVM <https://ratsgo.github.io/machine%20learning/2017/05/29/SVM2/>

Kernel SVM <https://ratsgo.github.io/machine%20learning/2017/05/30/SVM3/>



# 과제 및 시험 공지

## Test

- ✓ 객관식 10 ~ 15 문제 Test 예정
- ✓ 서술형 - gradient decent에 대해서 서술 하시오.