

# 수치예측, 범주예측

김성범

## (다변량) 데이터

인자 (변수) 관측치	$X_1$	...	$X_i$	...	$X_p$	$Y$
$N_1$	$x_{11}$	...	$x_{1i}$	...	$x_{1p}$	20.5
$N_2$	$x_{21}$	...	$x_{2i}$	...	$x_{2p}$	22.2
...	...	...	...	...	...	...
$N_{n-1}$	$x_{n-11}$	...	$x_{n-1i}$	...	$x_{n-1p}$	72.3
$N_n$	$x_{n1}$	...	$x_{ni}$	...	$x_{np}$	82.8

- X (원인): 독립변수, 예측변수, 입력변수
- Y (결과): 종속변수, 반응변수, 출력변수

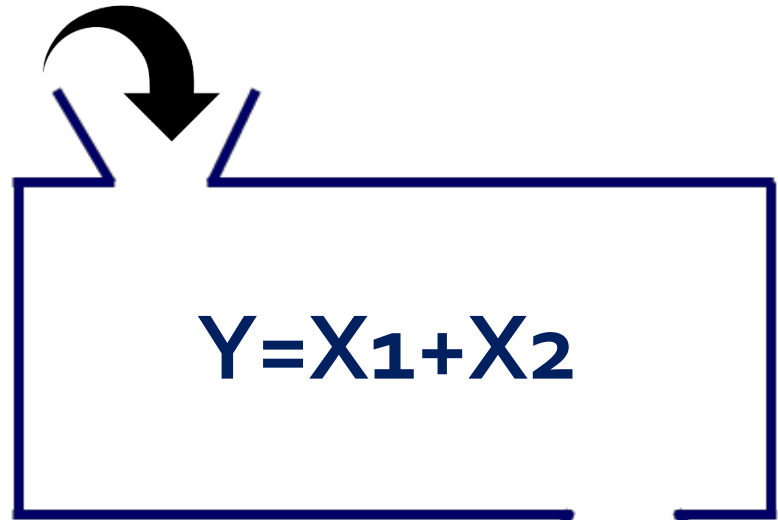
관측치 = 행 (row)  
변수 = 열 (column)

## 예측 모델링

관측치 4개, 변수 3개

$(0, 2), (1, 3), (2, 4), (3, 5)$

$X_1$	$X_2$	$Y$
0	2	2
1	3	4
2	4	6
3	5	8

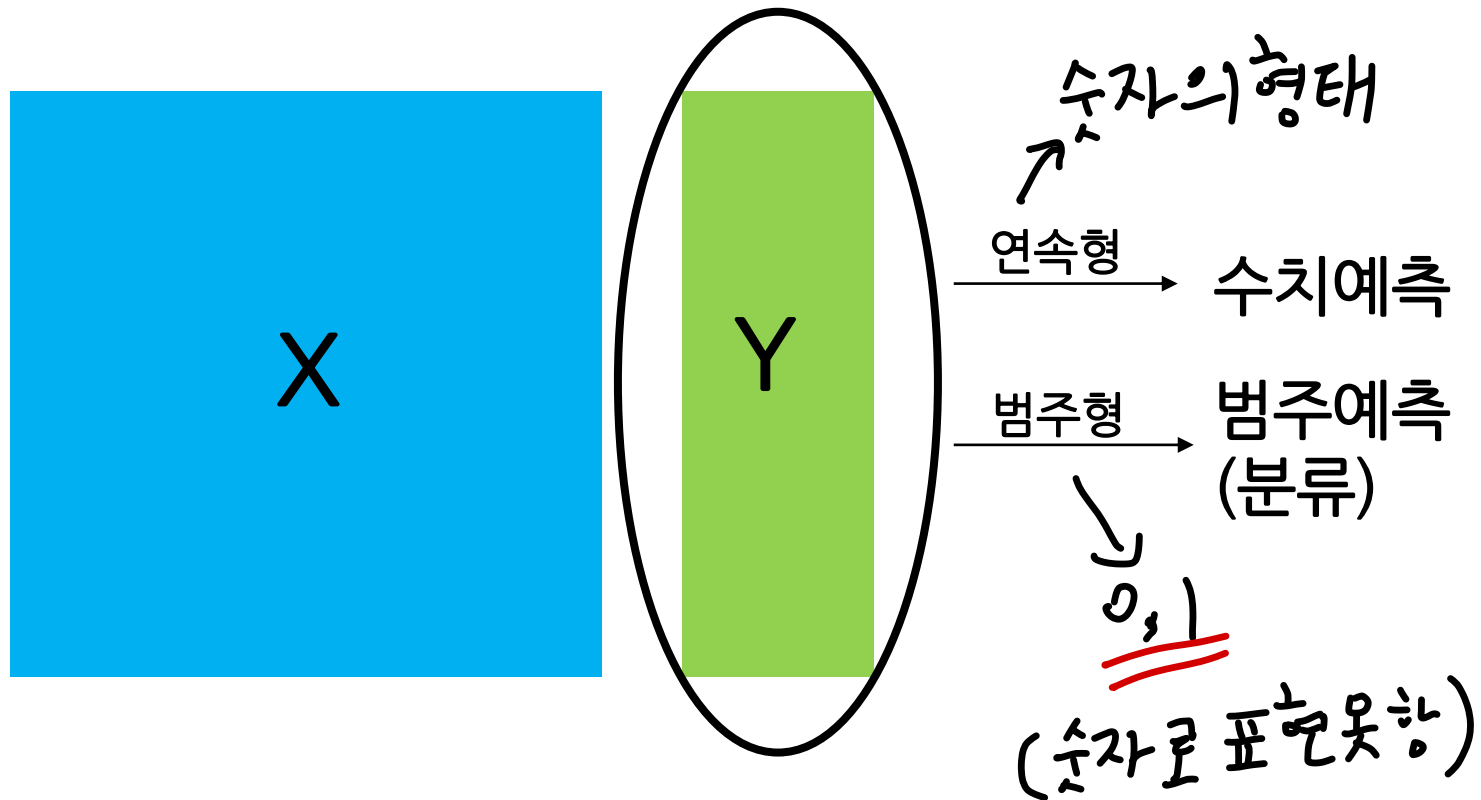

$$Y = X_1 + X_2$$

2, 4, 6, 8

→  $X$ 과  $Y$ 의 데이터를 이렇게 주어진 상황에서  
 $X$ 와  $Y$ 는 관계가 있기 때문에 이 관계를  
잘 설명하는 함수식을 찾는다.

(예측 모델링)을 한다.

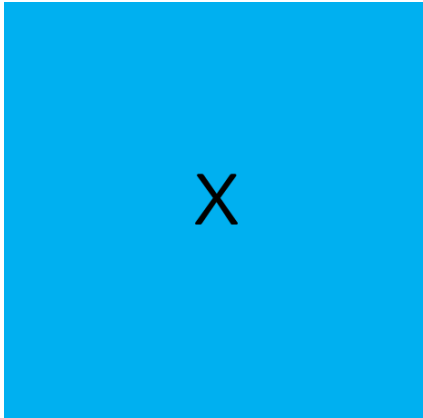
## 수치예측 / 범주예측 (분류)



- 연속형 데이터: 데이터 자체를 숫자로 표현  
예) 가격, 길이, 압력, 두께, ...
- 범주형 데이터: 원칙적으로 숫자로 표시할 수 없는 데이터  
예) 제품불량여부 (양품/불량), 보험사기여부(정상/비정상), ...

이는 예측하고자 하는 변수

# 수치예측 데이터

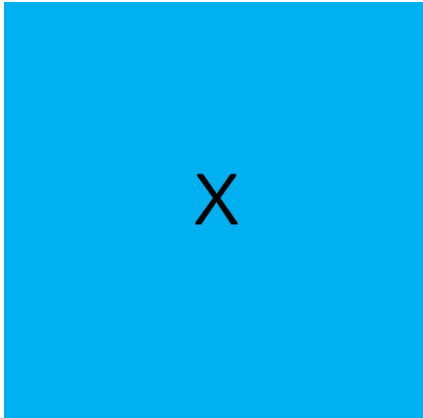


연속형 → 수치예측 (Regression)

연속형 데이터를 분석하는 것이  
Regression.

인자 (변수) 관측치	$X_1$	...	$X_i$	...	$X_p$	Y
$N_1$	$x_{11}$	...	$x_{1i}$	...	$x_{1p}$	20.5
$N_2$	$x_{21}$	...	$x_{2i}$	...	$x_{2p}$	22.2
...	...	...	...	...	...	...
$N_{n-1}$	$x_{n-11}$	...	$x_{n-1i}$	...	$x_{n-1p}$	72.3
$N_n$	$x_{n1}$	...	$x_{ni}$	...	$x_{np}$	82.8

# 범주 예측 데이터



범주형  $\longrightarrow$  범주 예측, 분류 (Classification)

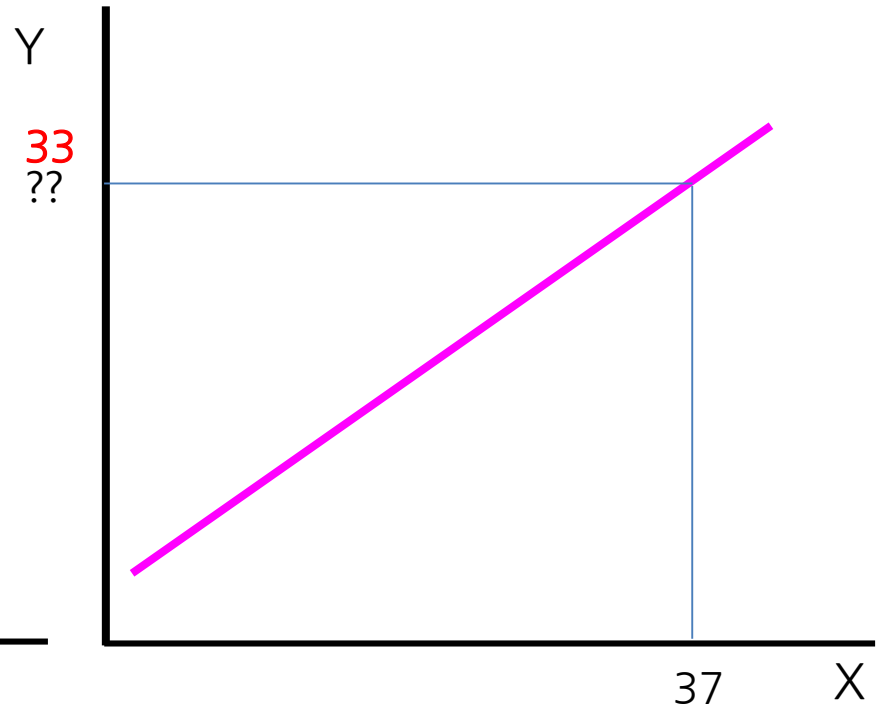
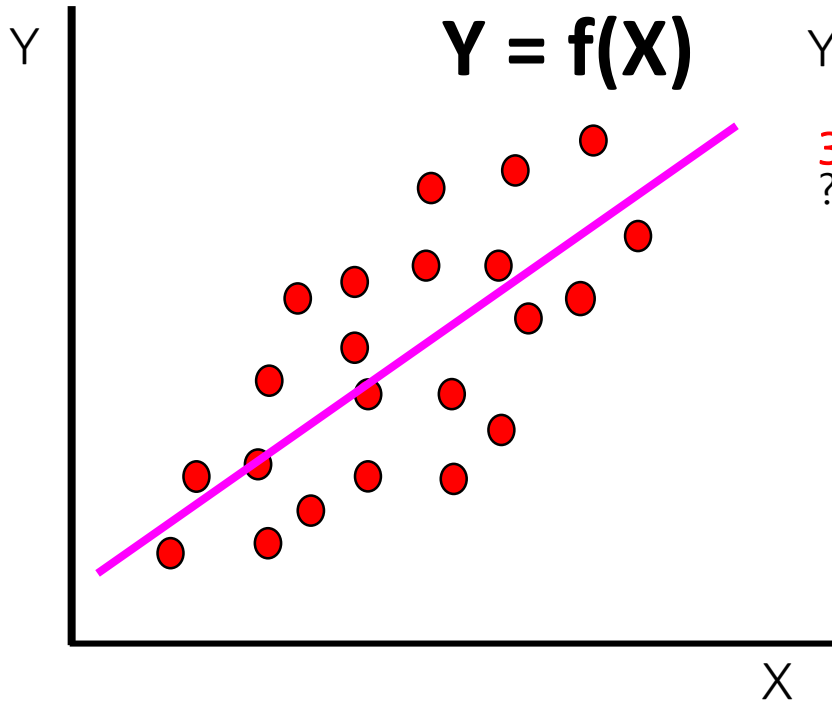
Y가 범주형인 경우  
분류 문제로 해결

이진 분류

인자 (변수)	$X_1$	...	$X_i$	...	$X_p$	$Y$
관측치						
$N_1$	$x_{11}$	...	$x_{1i}$	...	$x_{1p}$	0 (정상)
$N_2$	$x_{21}$	...	$x_{2i}$	...	$x_{2p}$	0 (정상)
...	...	...	...	...	...	...
$N_{n-1}$	$x_{n-11}$	...	$x_{n-1i}$	...	$x_{n-1p}$	1 (불량)
$N_n$	$x_{n1}$	...	$x_{ni}$	...	$x_{np}$	1 (불량)

## 수치예측 모델링 개요

→ 선형적인 관계가 있다.



# 수치예측 예제 - 중고차 가격 예측

원인 결과

X

Y

모델

TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors
TOYOTA Corolla 1800 T SPORT VVT I 2/3-Doors
TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors
TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 2/3-Doors
TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT BNS 2/3-Doors
TOYOTA Corolla 1.6 VVTI Linea Terra Comfort 2/3-Doors
TOYOTA Corolla 1.6 16v L.SOL 2/3-Doors
TOYOTA Corolla 1.6 16V VVT I 3DR TERRA 2/3-Doors
TOYOTA Corolla 1.6 16V VVT I 3DR SOL AUT4 2/3-Doors
TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-Doors
TOYOTA Corolla 1.4 VVTI Linea Terra 2/3-Doors
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors
TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors
TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors
TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors
TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-Doors
TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-Doors

TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-Doors

주행거리	마력	용량 (CC)
46986	90	2000
19700	192	1800
71138	69	1900
31461	192	1800
43610	192	1800
21716	110	1600
25563	110	1600
64359	110	1600
43905	110	1600
56349	110	1600
9750	97	1400
27500	97	1400
49059	97	1400
44068	97	1400
46961	97	1400
110404	90	2000
100250	90	2000
84000	90	2000
79375	90	2000
75048	97	1400

가격
13500
21500
12950
20950
19950
17950
16750
16950
16950
15950
12950
14750
13950
16750
13950
16950
16950
19000
17950
15800

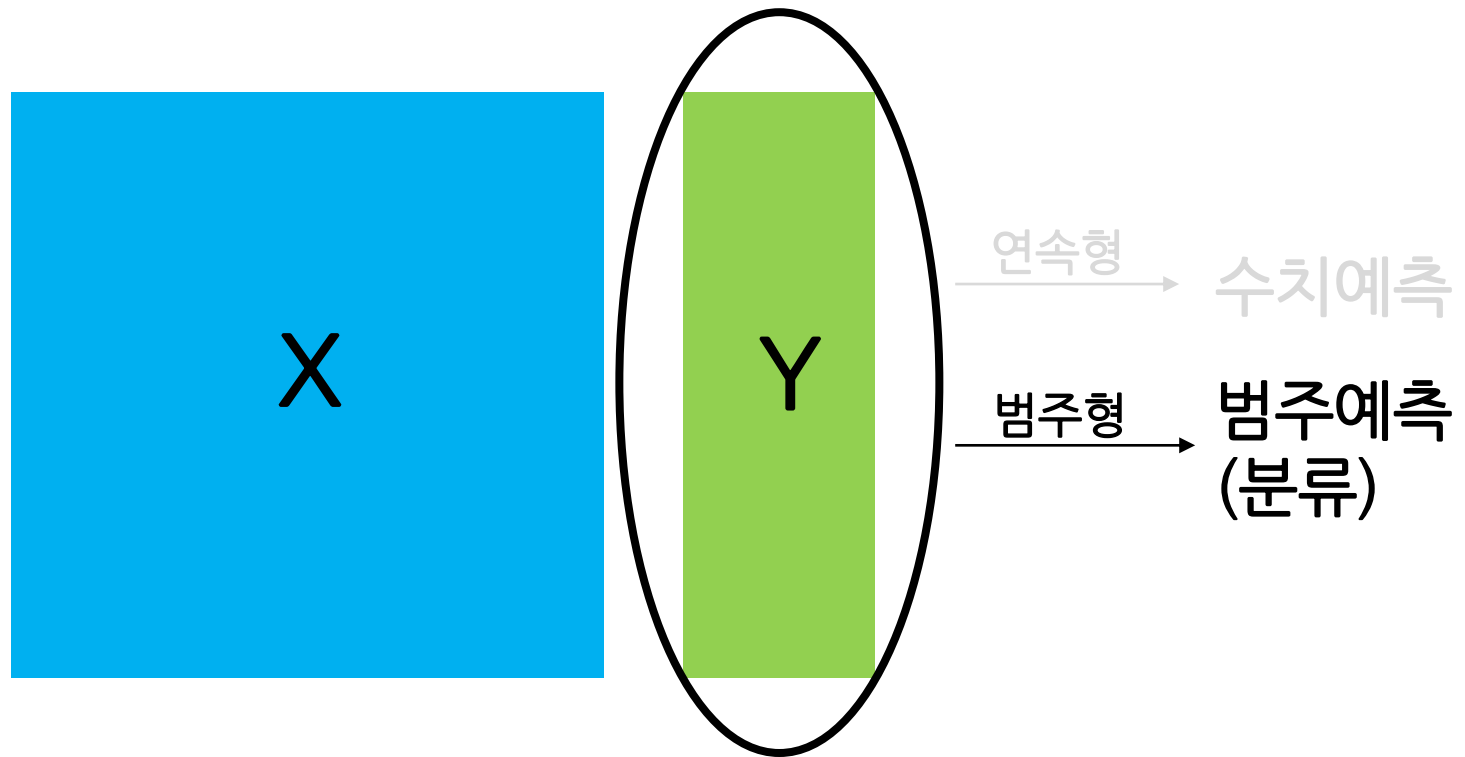
132151 110 1600

??????

18200



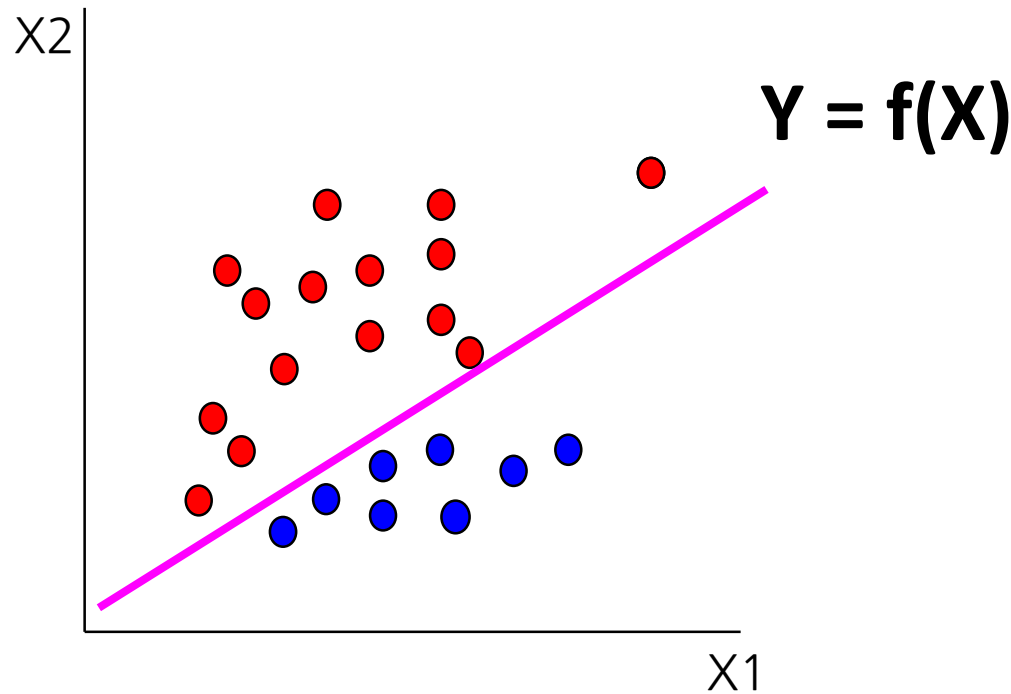
## 수치예측 / 범주예측 (분류)



## 범주예측 모델링 개요

- 불량범주
- 양품범주

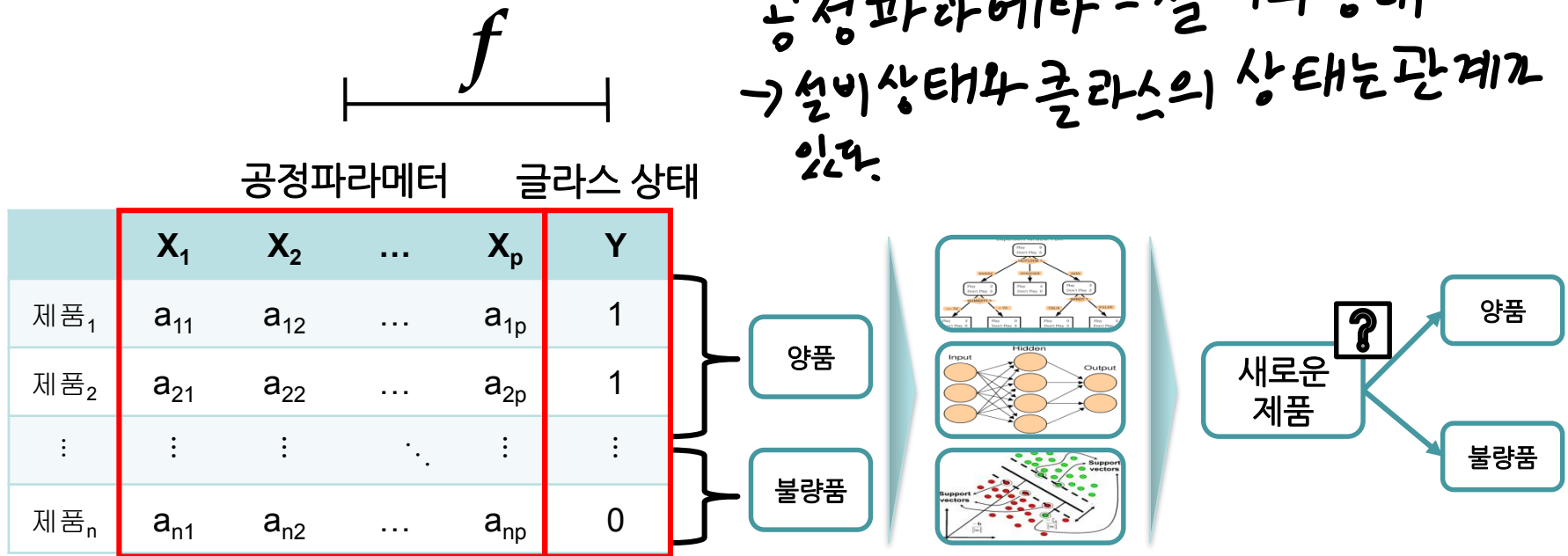
→ 새로운데이터가 들어왔을 때 이 데이터가  
어디에 속하는지 범주예측을  
하는데 사용



# 범주예측 예제 - 불량 예측

- 디스플레이 공정에서 공정 파라미터의 측정값들을 이용하여,  
해당 글라스 가 양품인지 불량품인지의 여부를 예측

공정파라메타 = 설비의 상태  
→ 설비상태와 클래스의 상태는 관계가 있다.



디스플레이공정 데이터

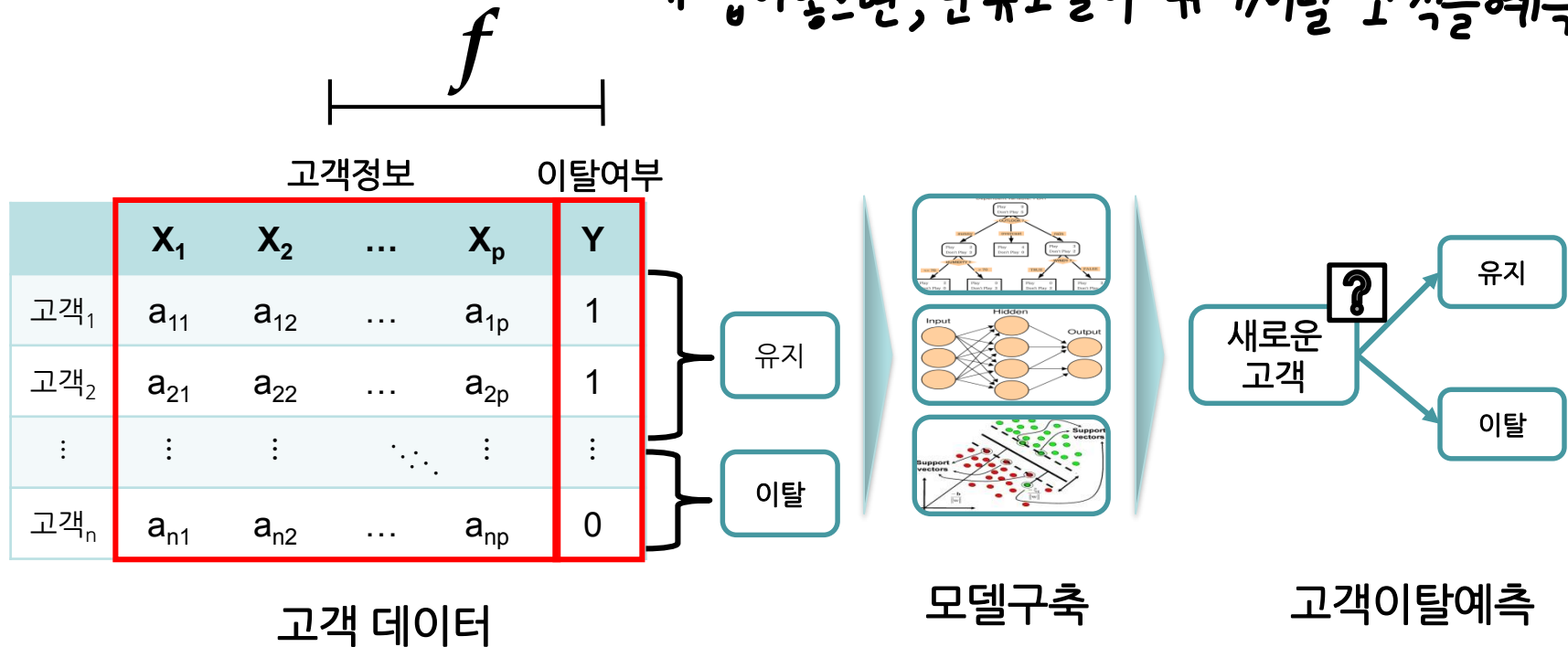
모델구축

불량 글라스 예측

→  $f(\text{함수})$ 를 찾는 것이 예측모델링이다.

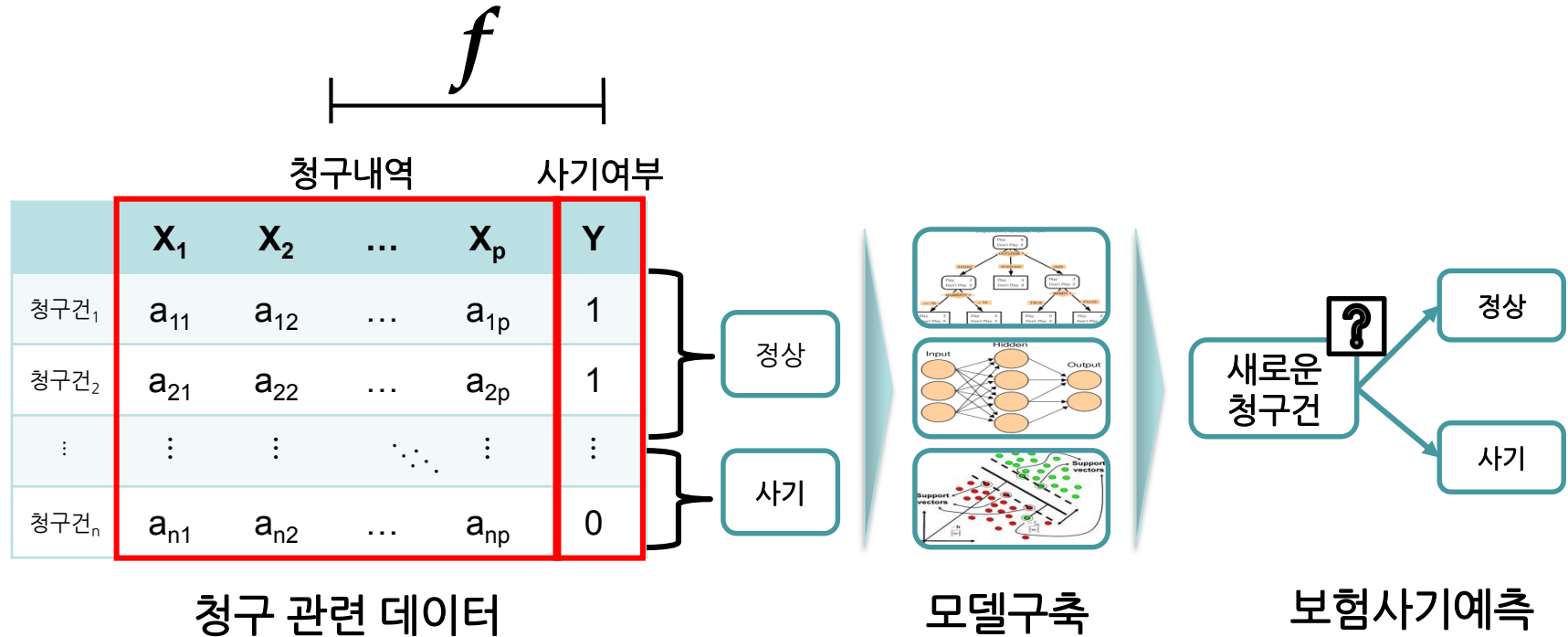
## 범주예측 예제 - 고객 이탈 예측

- 고객의 정보(성별, 연령, 직업, 연봉 등)를 이용하여,  
고객 이탈 여부를 예측  
→ 관계를 잘 설명하는 범주모델(분류모델)을 찾고 이 모델은 새로운 고객이 왔을 때 새로운 고객의 정보를 우리가 만든 모델에 집어넣으면, 분류모델이 유지/이탈 고객을 예측



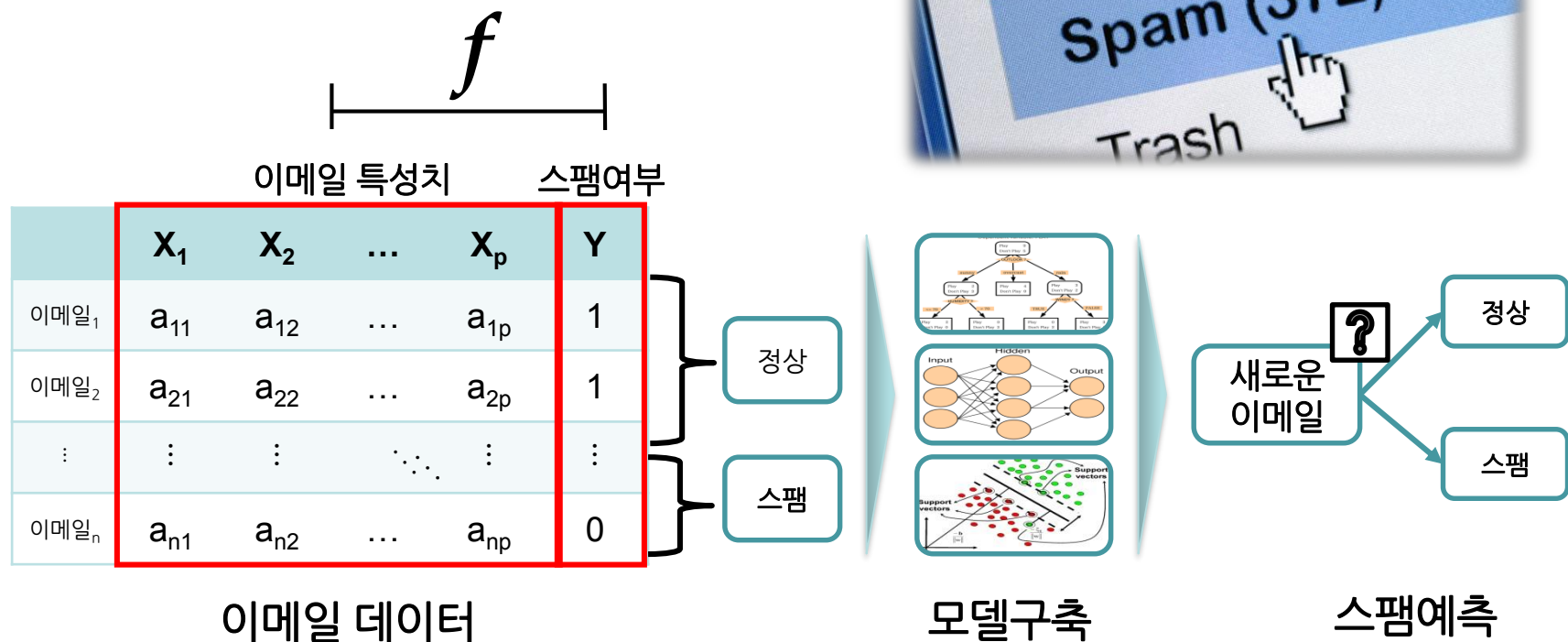
## 범주예측 예제 - 보험 사기 여부 예측

- 각 청구 건에 대한 내역 분석을 통해 청구 건에 대한 사기 여부 예측



## 범주예측 예제 - 이메일/스팸 분류

- 이메일이 정상메일인지 스팸 인지 분류



---

**EOD**