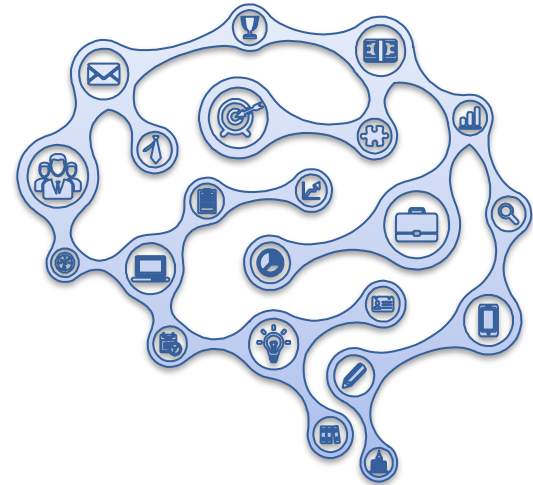


# 의료 Artificial Intelligence

기계 학습 (chap6)

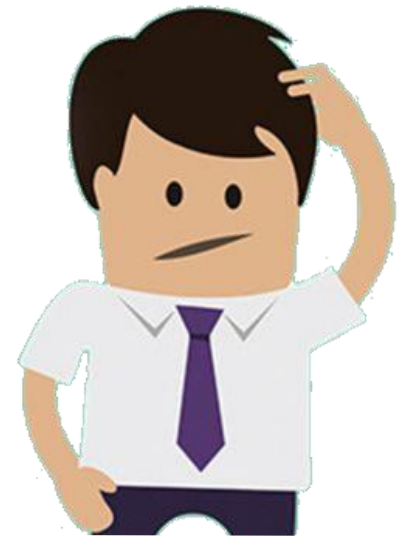
2022.05.19



# 오늘 배울 내용 ...

1. 기계 학습 (머신 러닝)
2. 인공지능 실습
3. mblock 실습

어렵지 않다  
쉬운 것도 아니다



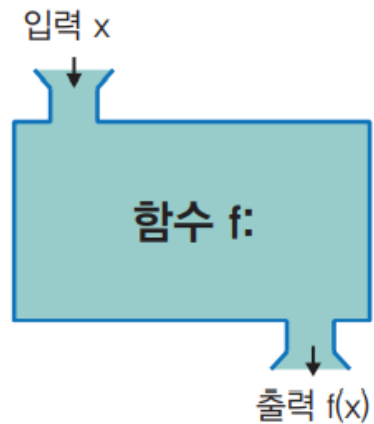
---

# 인공지능 이론

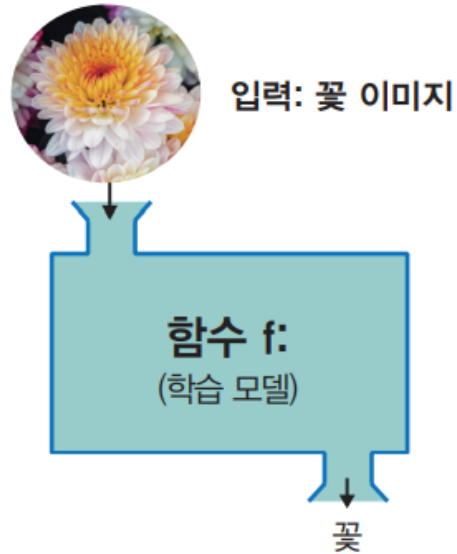
---

# 자료와 학습

- 머신러닝은 인간의 학습을 모방하여 예시(데이터)를 통해 학습
- 기계가 학습을 하기 위한 학습 모델(함수)이 필요



▲ 함수



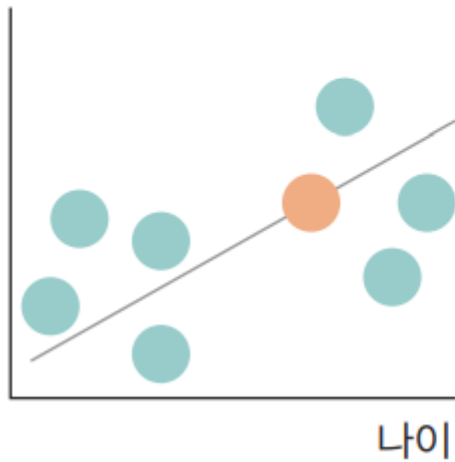
▲ 꽃을 판별하는 함수

# 머신러닝의 종류

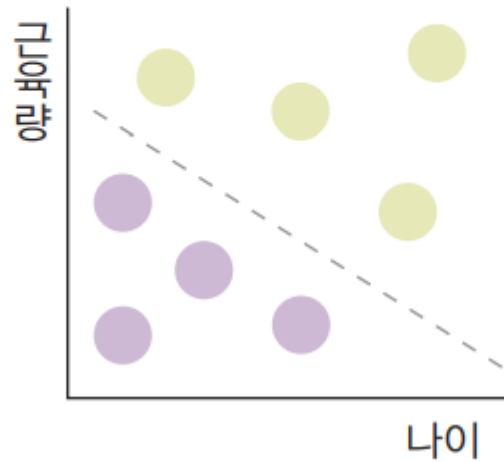
·머신러닝을 수학적으로 정의하면 회귀, 분류, 군집이라고 말할 수 있음

학습 = 회귀, 분류, 군집

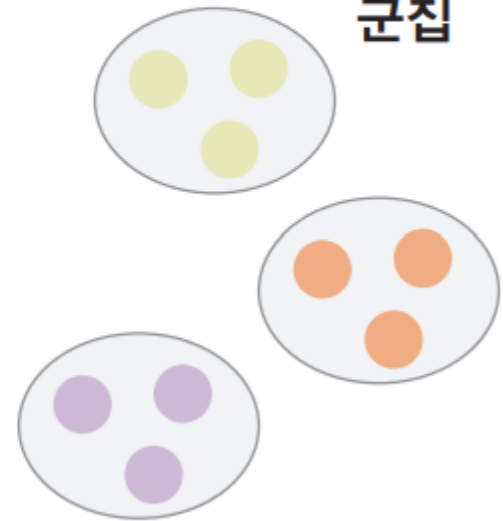
회귀



분류



군집



# 머신러닝과 데이터 마이닝

· 머신러닝과 통계적 분석(데이터 마이닝)의 차이점

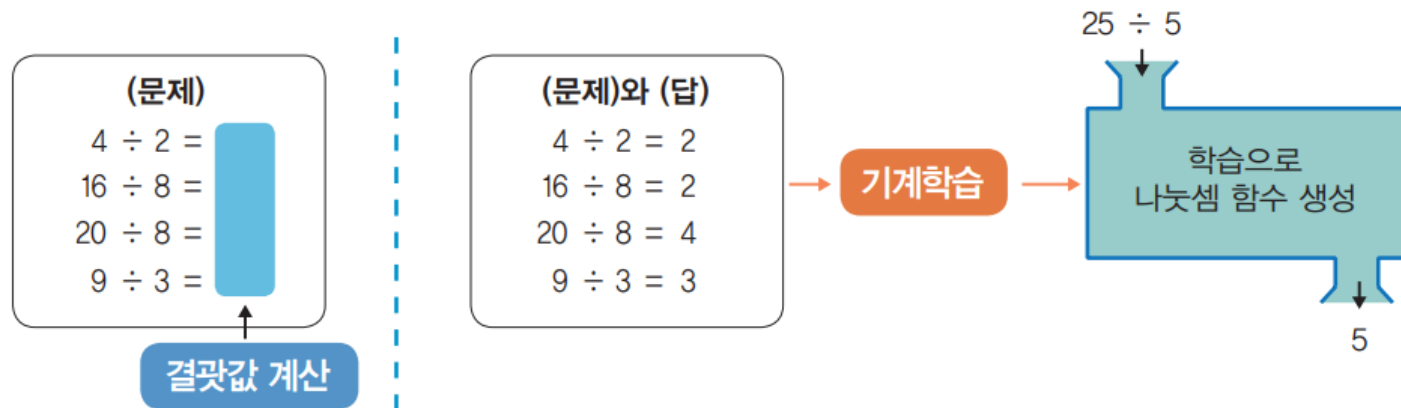
구분	머신러닝	통계학 (통계적 분석)
특징	정확한 예측, 분류에 집중	원인/이유 분석, 일반적인 통계정보
넷플릭스	영화 평가 <u>예측</u> 의 정확성	사람들이 영화를 좋아하는 <u>이유 분석</u>
의료	내년에 병원에 갈 사람들 숫자 <u>예측</u>	사람들이 병원에 가는 <u>이유 분석</u>
독감 예측	독감 발생 가능성 <u>예측</u>	독감 발생의 <u>요인 분석</u>

# 머신러닝 알고리즘

“어떤 문제Task T에 관련된 경험Experience E로 부터 성과 측정지표Performance Measure P를 가지고 학습Learn을 진행하는 컴퓨터 프로그램을 말한다.

이때 문제T에 대한 성과는 P로 측정되고, 경험E로 부터 개선을 진행한다.”

-토미첼 (1998)-

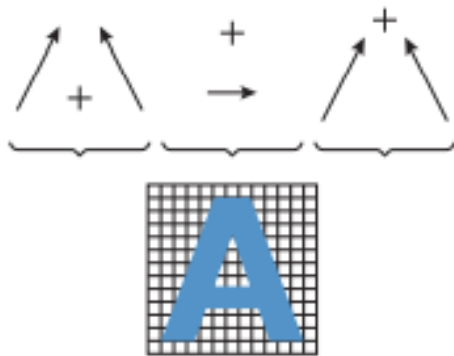


▲ 기존 계산 프로그램과 기계학습 프로그램

·작업 T는 나눗셈의 몫 구하기, 경험 E는 입력값, 성과지표 P는 수식이 정확한 확률.  
정답인 나눗셈의 몫 값을 찾으면 이 프로그램은 ‘학습한다’고 정의할 수 있음

# 머신러닝 알고리즘

- 머신러닝의 학습 결과는 매개변수 또는 가중값의 형태로 나타남
- 머신러닝이란 **입력값에 대응하는 결과값이 제대로 나오게 하는 최적의 가중값(매개변수)를 찾는 과정**



	특징			
문자	수직선	수평선	사선	곡선
L	1	1	0	0
P	1	0	0	1
T	1	1	0	0
E	1	3	0	0
A	0	1	2	0

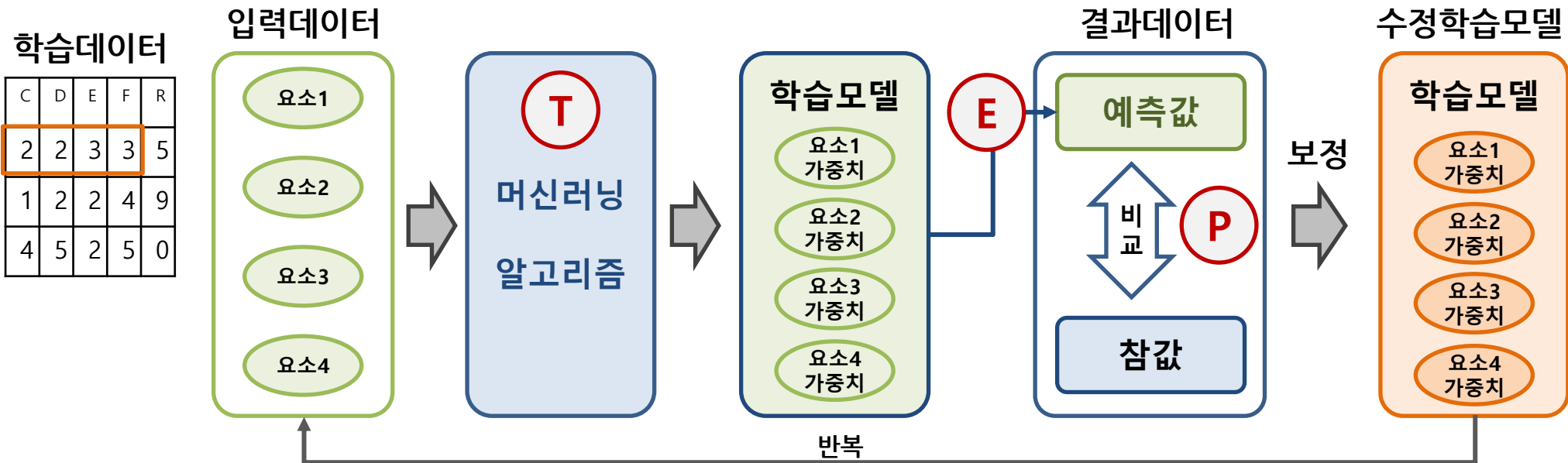
전통적인 영문자 인식 방법은 문자에 있는 사선, 수직선, 수평선, 곡선의 개수를 파악해 문자를 인식



# 머신러닝 모델링 과정

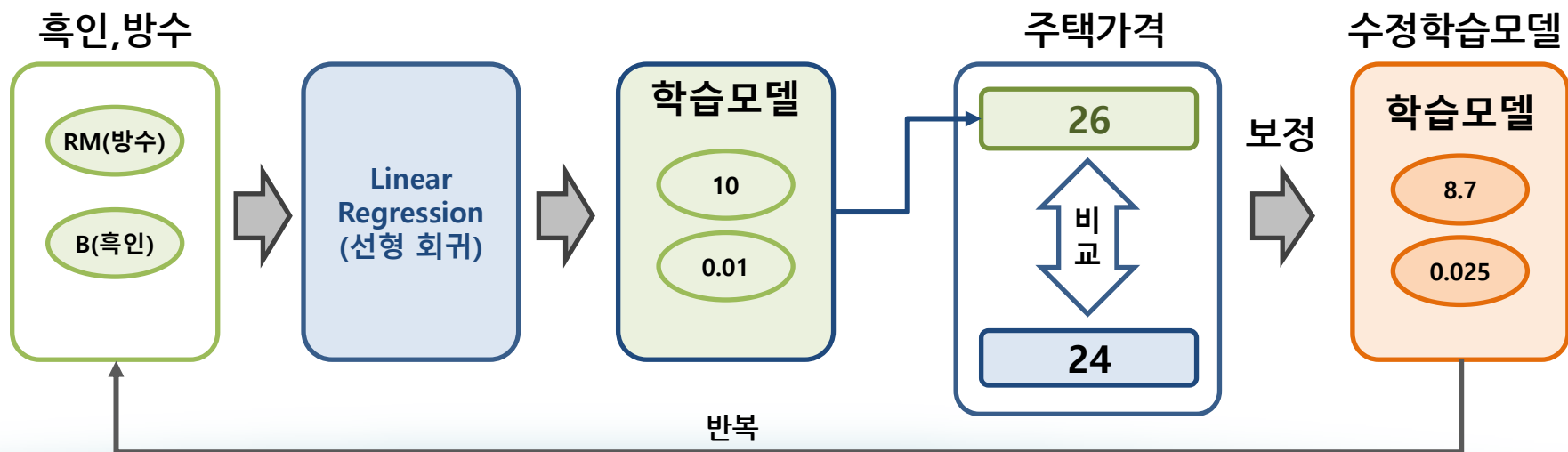


# 기계가 학습하는 원리 (회귀)

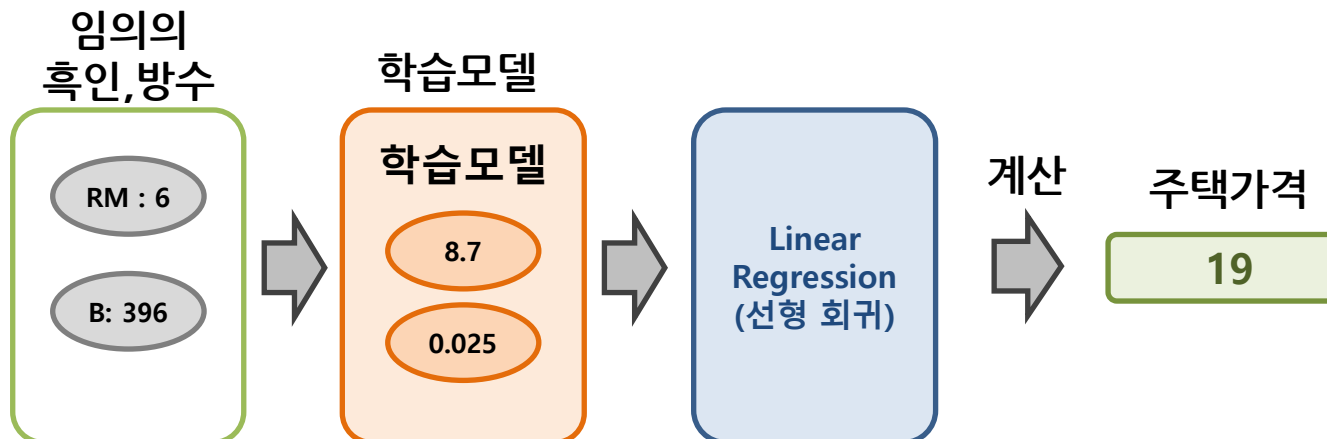
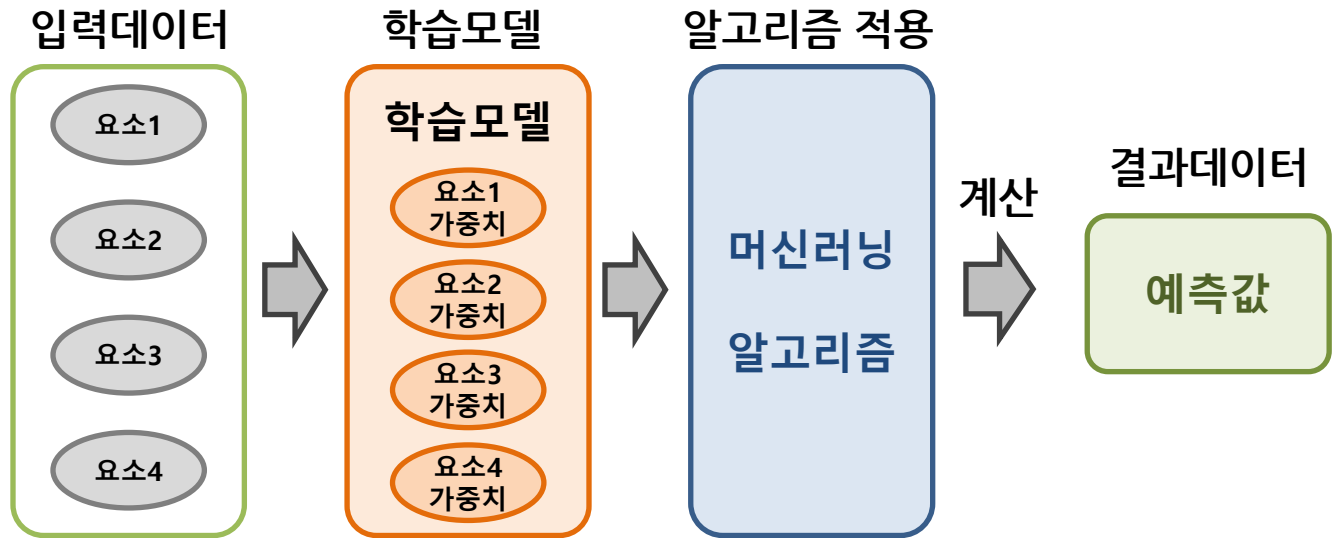


## Housing

MEDV	RM	B
24.1	6.575	396.90
21.6	6.421	396.90
34.7	7.185	392.83
33.4	6.998	394.63
36.2	7.147	396.90
28.7	6.430	394.12
22.9	6.012	395.60

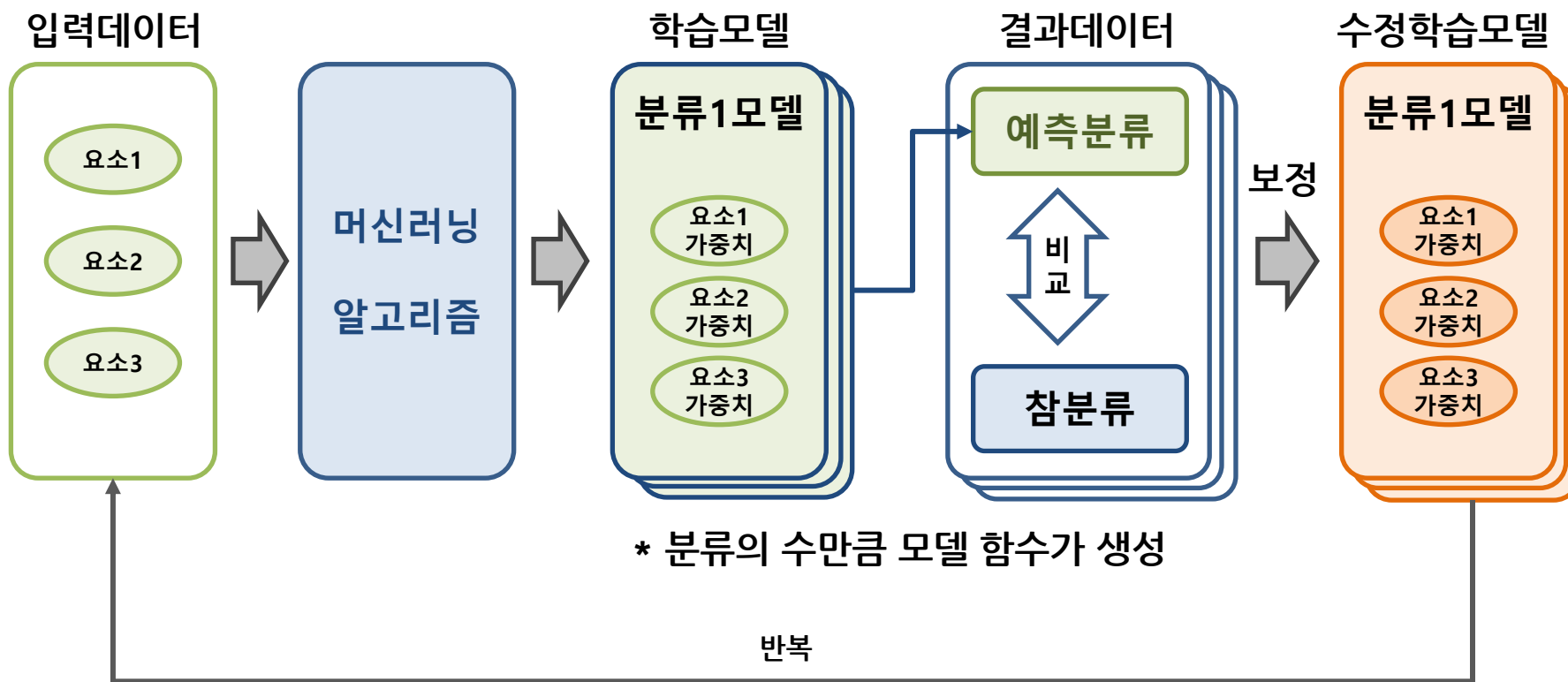


# 기계가 예측하는 원리 (회귀)

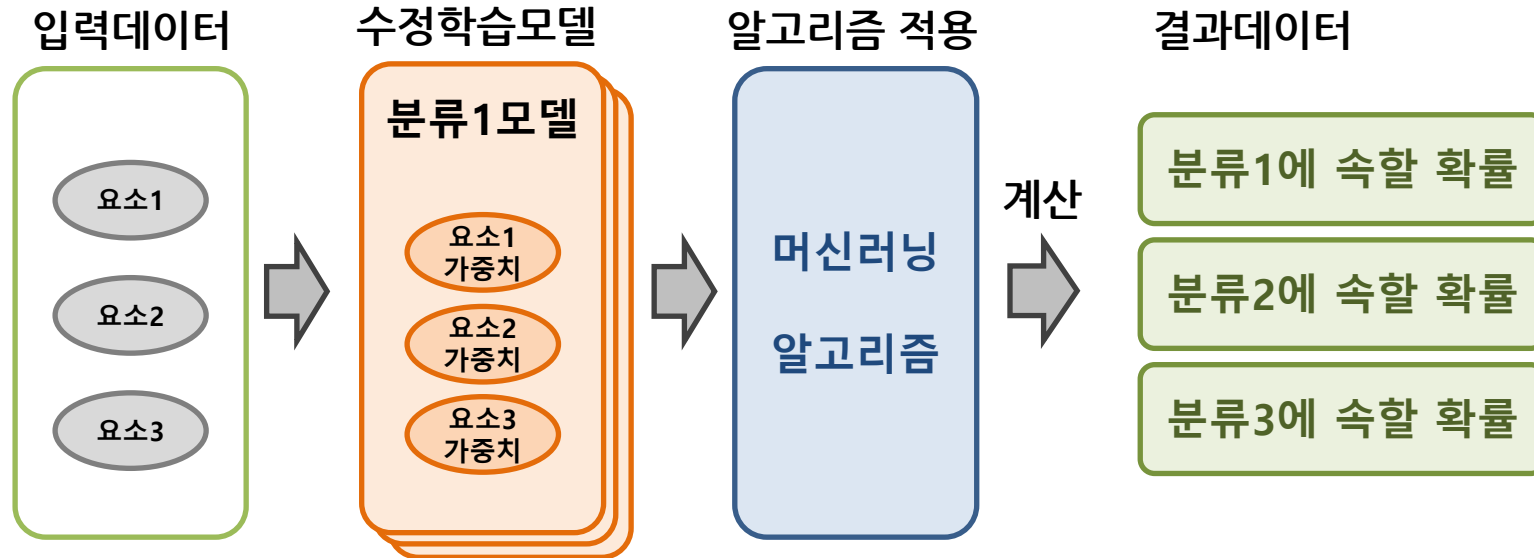


# 기계가 학습하는 원리 (분류)

\* 3가지로 분류하는 머신 구축



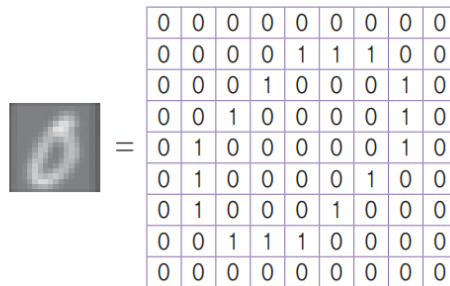
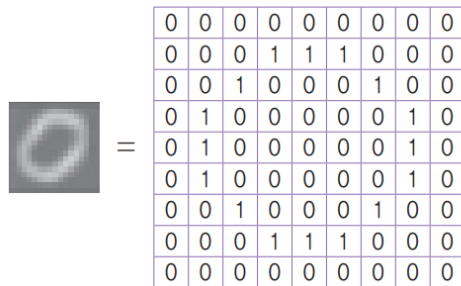
# 기계가 예측하는 원리 (분류)



# 손글씨 인식원리

## ① 입력 자료 정하기

- 28\*28 픽셀(총 784개 셀)로 이뤄진 이미지
- 기존의 선의 특징을 추출하지 않고  
각 픽셀의 입력값을 그대로 입력



# 손글씨 인식원리

## ② 정답 결정하기

- 정답(출력값)  $y$ 는 각 이미지에 쓰인 숫자값
- 벡터값으로 표현

---

$$\begin{array}{cccccc} y_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & y_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & y_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \cdots & y_9 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} & y_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ 1 & 2 & 3 & \cdots & 9 & 0 \end{array}$$

---

# 손글씨 인식원리

## ③ 학습 모델 만들기

- 컴퓨터가 0부터 9의 숫자를 분류하기 위한 최적의 값(가중값)을 찾도록 적당한 모델 알고리즘이나 수식을 정의해줘야 함
- 수식에서 매개변수  $\theta$  값(가중치)  
→ 가중치의 조합이 학습모델임

$$h_i(X) = \theta_1^i x_1 + \theta_2^i x_2 + \dots + \theta_{399}^i x_{398} + \theta_{400}^i x_{400}$$



# 손글씨 인식원리

## ④ 머신러닝 시작하기

- 처음에는 각 데이터 요소별로 임의의 값으로  $\theta$  값(가중치)을 정해서 시작
- 머신러닝 알고리즘이나 수식에 의해 계산된 결과값(예측값)과 정답을 비교하여 정답에 가까워질 때까지  $\theta$  값(가중치)을 조절해 주는 과정을 반복

## ⑤ 학습 모델 구축

- 학습이 끝나면 10개 숫자에 대한 입력데이터의  $\theta$ 값(4,000개 = 400픽셀 x 10개 숫자)이 결정되고 이를 이용한 10개의 함수가 준비됨  
→ 학습 모델

# 손글씨 인식원리

## ⑥ 학습 결과 출력

- 학습 결과로 얻은 필기체 숫자의 예측 함수(학습 모델 + 모델 알고리즘)의 적용 결과는 해당 숫자에 해당하는 위치만 1이고 나머지는 0인 벡터를 출력
- 10개의 함수를 이용해 숫자를 인식시킴.



숫자 0의 이미지를 입력했을 때  $h_0$ 의 함수의 값이 1에 가깝게 계산되며, 나머지는 0이라는 결론을 출력. 즉, 0일 확률이 가장 높으므로 이 모델에서는 입력 이미지를 0이라고 인식

숫자 0의 이미지(0)를  $x$ 값에 입력했을 때

$$h(x) = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \\ h_0 \end{bmatrix} \quad h(x) = \begin{bmatrix} 0.000000 \\ 0.000005 \\ 0.000023 \\ 0.000000 \\ 0.000560 \\ 0.000000 \\ 0.000000 \\ 0.000000 \\ 0.000028 \\ 0.000000 \\ 0.999531 \\ 0.999528 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Leftarrow \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 0 \end{bmatrix}$$

# 손글씨 인식원리

실제 프로그램에서는 숫자의 모양에 따라 확률이 달라짐

	$h_0=0.000003, h_1=0.000000, h_2=0.002263, h_3=0.000110, \underline{h_4=0.001144}$ $h_5=0.000423, \underline{h_6=0.984752}, h_7=0.000001, h_8=0.024688, h_9=0.002832$
	$h_0=0.000000, h_1=0.000000, h_2=0.001421, h_3=0.000210, \underline{h_4=0.236924}$ $h_5=0.001327, \underline{h_6=0.897365}, h_7=0.000001, h_8=0.000487, h_9=0.001499$

같은 숫자 6이라도 첫번째는 거의 1에 가까운 확률로 6이라고 인식한 반면,  
두번째는 4일 가능성도 다른 값들에 비해 높게 나옴

---

# 인공지능 실습


---

# 인공지능 실습

<https://teachablemachine.withgoogle.com/train>

## 새 프로젝트

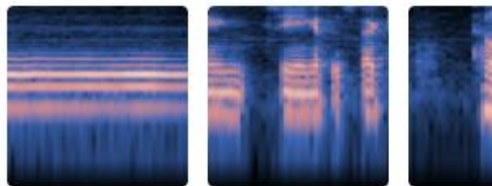
 Drive에 있는 기존 프로젝트를 엽니다.

 파일에서 기존 프로젝트를 엽니다.



### 이미지 프로젝트

파일 또는 웹캠에서 가져온 이미지를 기반으로 학습시키세요.



### 오디오 프로젝트

파일 또는 마이크에서 가져온 1초 분량의 사운드를 기반으로 학습시키세요.



### 포즈 프로젝트

파일 또는 웹캠에서 가져온 이미지를 기반으로 학습시키세요.

# 인공지능 실습

## 이미지 셋 추가

The interface is divided into three main sections: Class Management, Training, and Preview/Output.

**Class Management (Left):**

- 바위 (Rock):** 9 이미지 샘플. Includes buttons for '촬영' (Capture) and '업로드' (Upload). A horizontal bar shows 9 hand gesture samples.
- 보 (Paper):** 10 이미지 샘플. Includes buttons for '촬영' (Capture) and '업로드' (Upload). A horizontal bar shows 10 hand gesture samples.
- 가위 (Scissors):** 11 이미지 샘플. Includes buttons for '촬영' (Capture) and '업로드' (Upload). A horizontal bar shows 11 hand gesture samples.
- 클래스 추가:** A dashed box at the bottom with the text '클래스 추가' (Add Class).

**Training (Center):**

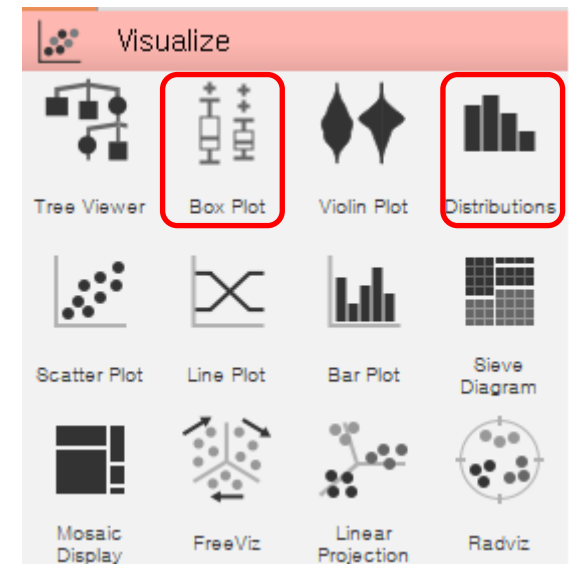
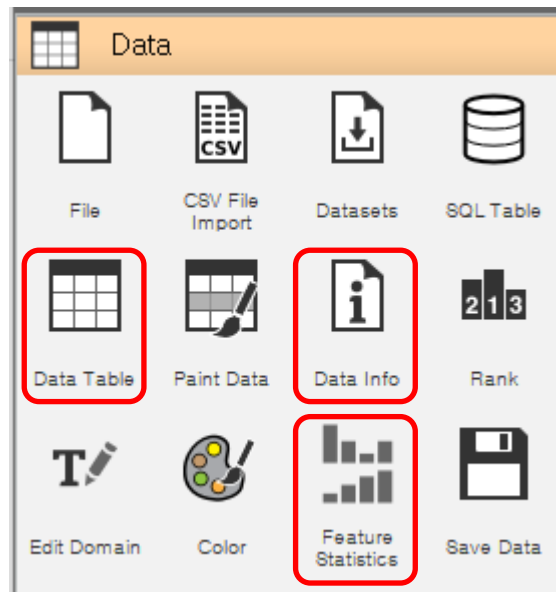
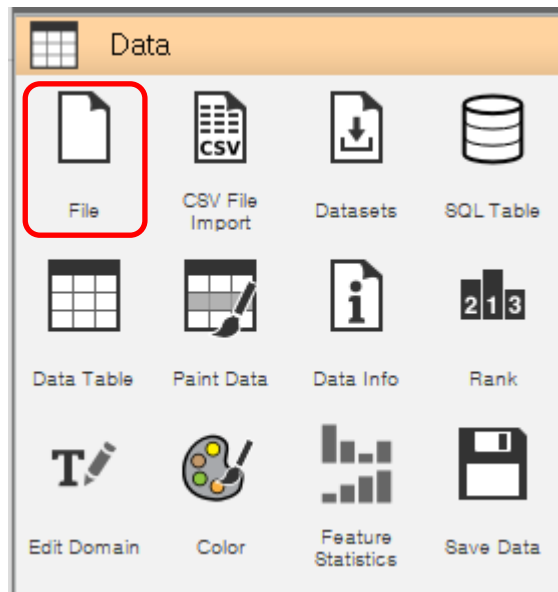
- 학습 (Training):** 모델 학습 완료됨 (Model training completed).
- 고급 (Advanced):** Includes settings for '에포크' (Epochs) set to 50, '배치 크기' (Batch size) set to 16, and '학습률' (Learning rate) set to 0.001. There are also buttons for '기본값 초기화' (Reset to default) and '고급 설정' (Advanced settings).

**Preview/Output (Right):**

- 미리 보기 (Preview):** Includes a '모델 내보내기' (Export model) button and a 'Webcam' dropdown menu. Below is a live video feed of a desk with a lamp and bookshelves.
- 출력 (Output):** Shows the predicted class for each frame. The '바위' (Rock) class is highlighted in orange, '보' (Paper) in red (99% confidence), and '가위' (Scissors) in purple.

# 인공지능 실습

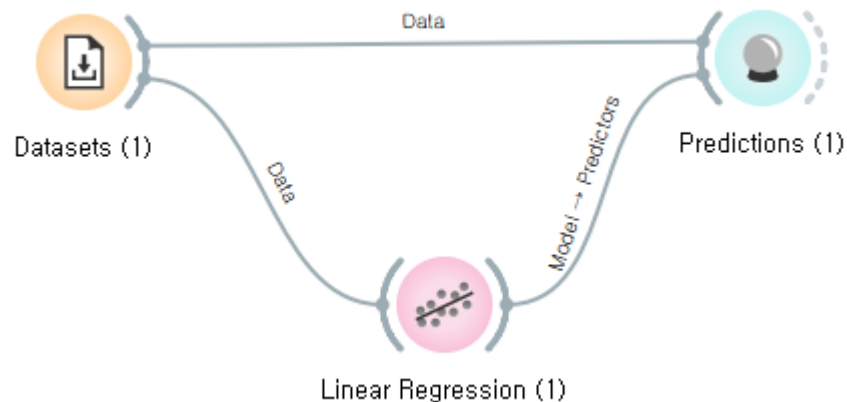
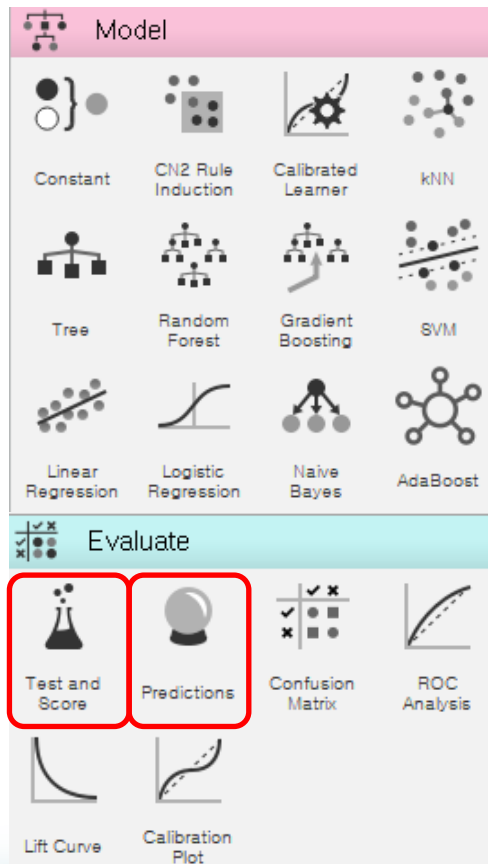
- 데이터 불러오기 : File, CSV File Import, Datasets
- 데이터 보기 : Data Table, Data Info, Feature Statistics
- 데이터 차트 : Box plot, Distributions



# 인공지능 실습

## 데이터 분석 : 모델

- Model 위젯
- Evaluate 의 'Predictions'나 'Test and Score' 와 연결
- Data는 따로 연결해 주어야 함

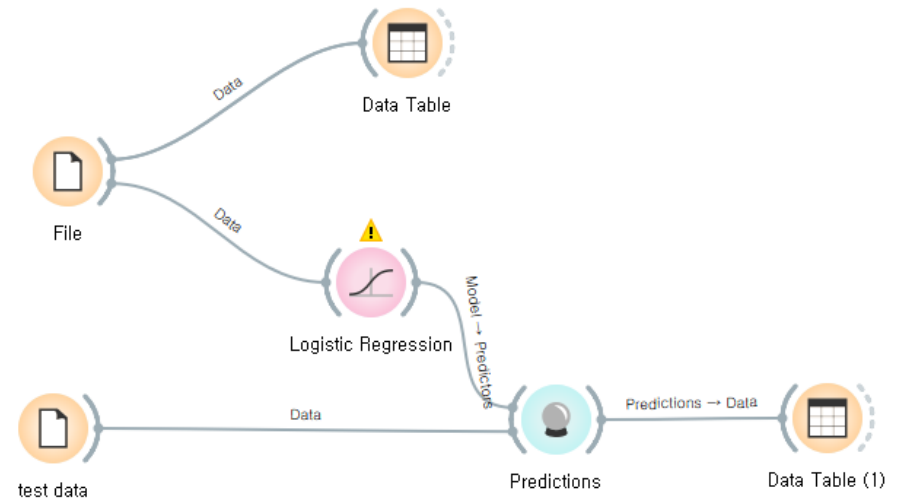
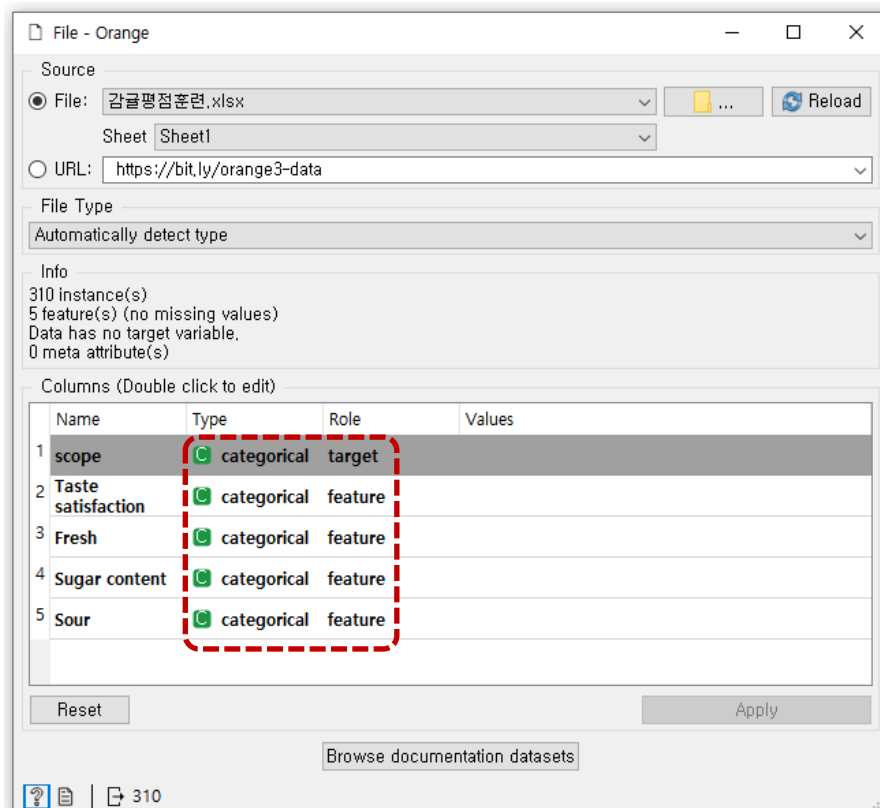




# 인공지능 실습

## 실습 1

- 데이터 불러오기 : '감귤평점훈련', '감귤평점테스트'
- 데이터 보기 및 평가 : Data Table, Predictions
- 데이터 모델 : Logistic Regression

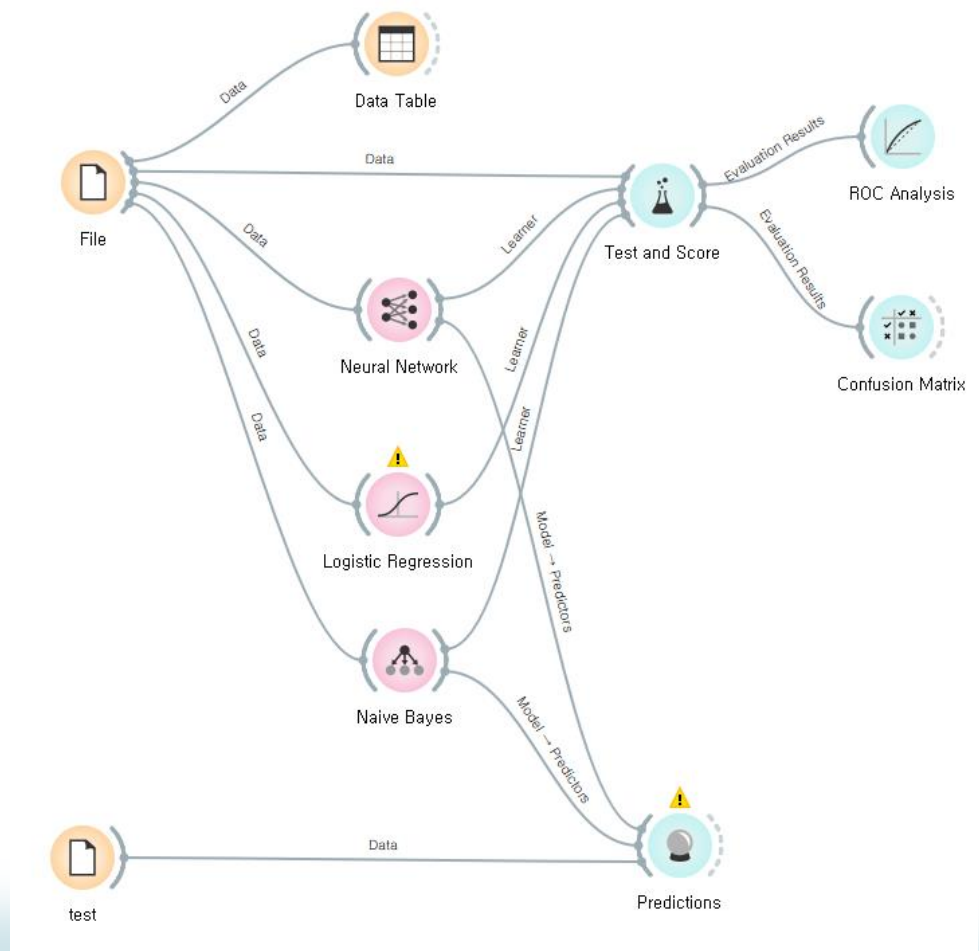


# 인공지능 실습

## 실습 1 : 데이터 모델 비교

- 데이터 불러오기 : 'heart\_train', 'heart\_test'
- 데이터 보기 및 평가 : Data Table, Predictions
- 데이터 모델 : Logistic Regression, Neural Network, Naïve Bayes

Columns (Double click to edit)				
	Name	Type	Role	Values
1	id	numeric	meta	
2	age	numeric	feature	
3	sex	categorical	feature	0, 1
4	cp	categorical	feature	
5	trtbps	numeric	feature	
6	chol	numeric	feature	
7	fbs	categorical	feature	0, 1
8	restecg	categorical	feature	
9	thalachh	numeric	feature	
10	exng	categorical	feature	0, 1
11	oldpeak	numeric	feature	
12	slp	categorical	feature	
13	caa	numeric	feature	
14	thall	categorical	feature	
15	output	categorical	target	0, 1



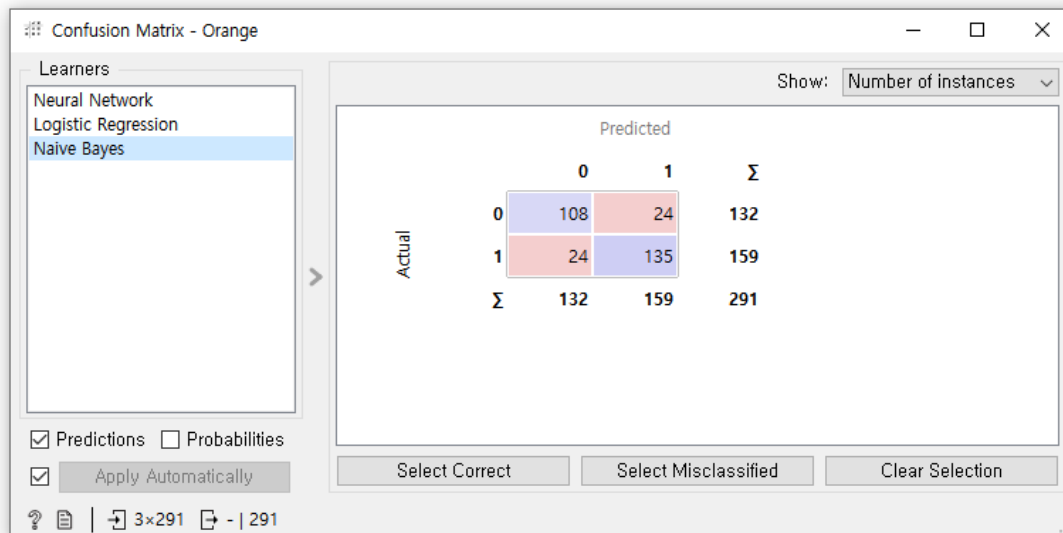
# 인공지능 실습

## Confusion Matrix 해석

- 예측결과와 참값을 참/거짓 매트릭스로 구분



Confusion Matrix



	Positive	Negative
True	<b>TP</b> 참인 것을 참으로 예측	<b>FN</b> 참인 것을 거짓으로 예측
False	<b>FP</b> 거짓을 참으로 예측	<b>TN</b> 거짓을 거짓으로 예측

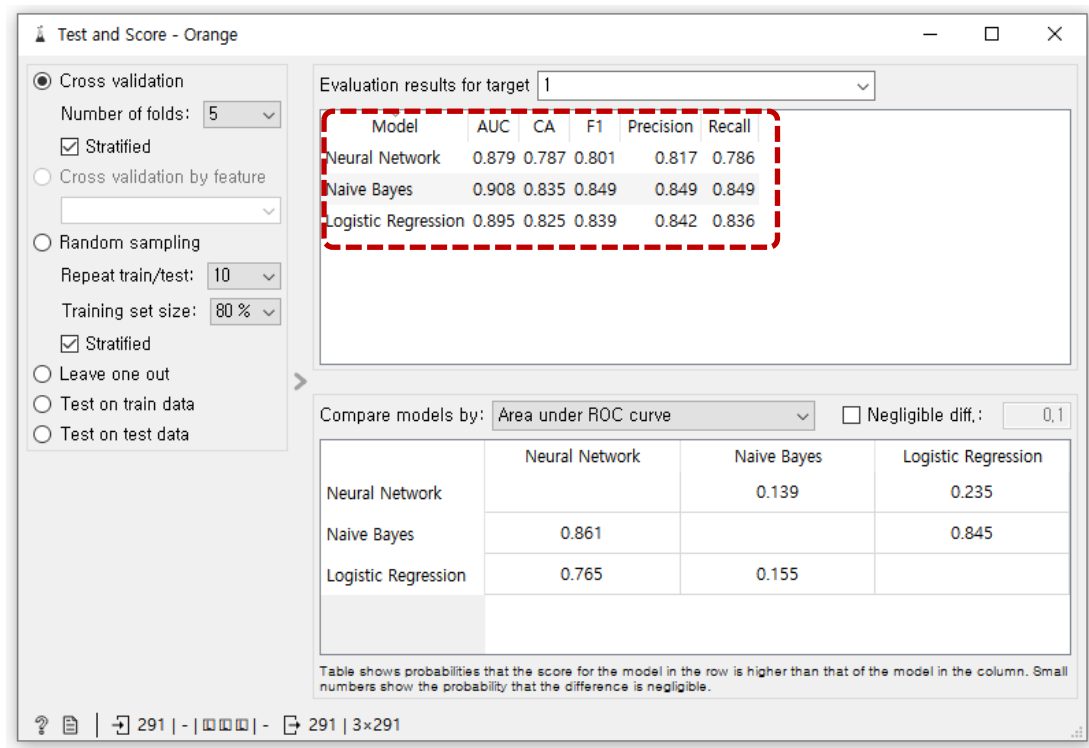
# 인공지능 실습

## Test and Score 해석



Test and Score

- AUC : 재현율 (Recall : 참을 참으로 분류)과 위양성률(Fall-out : 거짓을 참으로 분류)의 비율관계
- CA : 정확하게 분류한 비율. 가장 일반적 성과지표  
 $(TP + TN) / (TP + FP + TN + FN)$
- Precision : 예측한 참에서 실제 참의 분류 비율  
 $TP / (TP + FP)$
- Recall : 실제 참 중에 예측한 참의 분류 비율  
 $TP / (TP + FN)$
- F1 : Precision와 Recall에 대한 평균

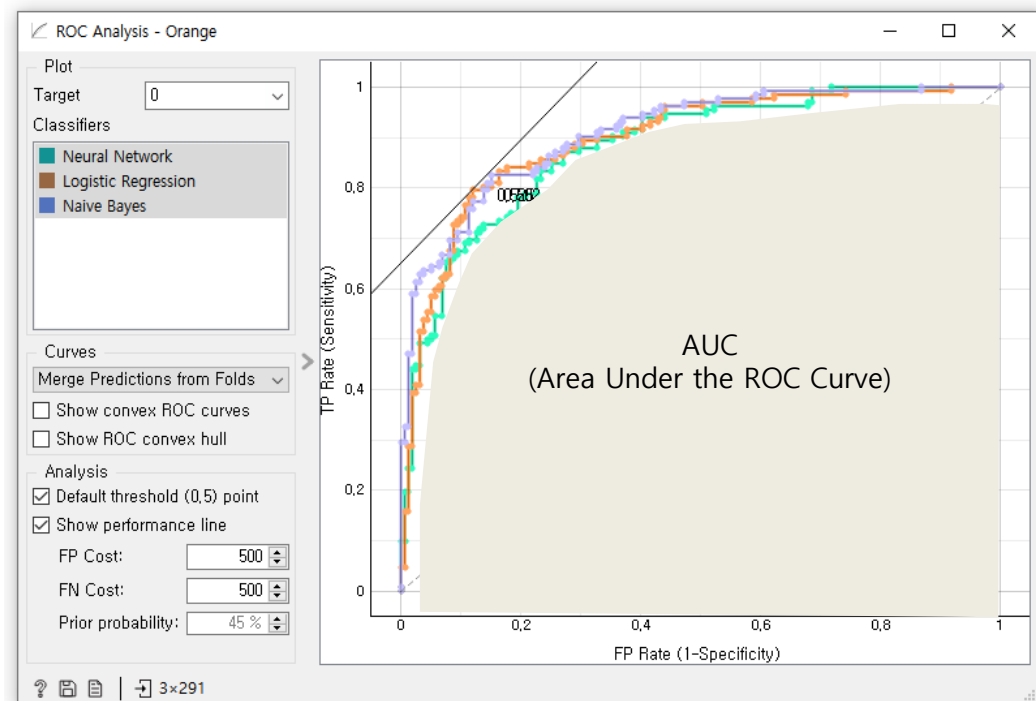


# 인공지능 실습

## ROC(Receiver Operating Characteristic Curve) Analysis 해석



- AUC는 그래프의 아랫 쪽 면적
- 1에 가까울수록, 면적이 클 수록 모델의 성능이 우수함

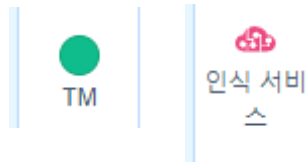
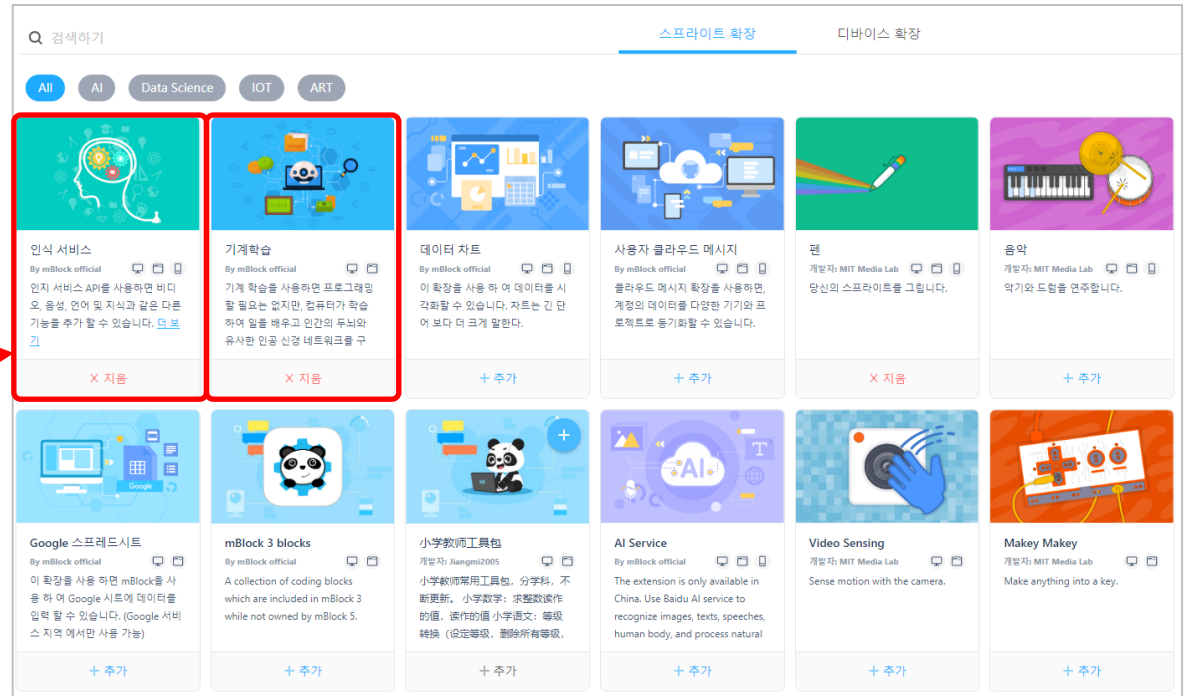
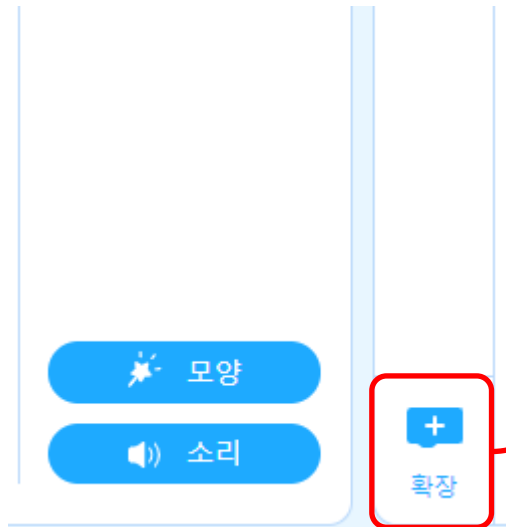


---

# MBlock 실습

---

# 인공지능 MBlock 실습



# 나이 인식 프로그래밍





# 감정인식 프로그래밍



팀 활동