

머신러닝 개요

Lecture 1: Concept of ML

College of Information and Electronic Engineering

Kyung Hee Univeristy

Prof. Wonha Kim

(wonha@khu.ac.kr)

김원하교수



- 교육

- 학사 : 연세대학교 전자공학 (1985)
- 석사 : Univ. of Wisconsin - Madison (1990, 전기공학)
- 박사 : Univ. of Wisconsin - Madison (1997, 전기공학)
논문 주제 : Wavelet 기반 영상 코딩 시스템

- 경력

- 1996. 01 ~ 1996. 12 : Motorola (USA), 인턴
- 1997. 08 ~ 2000. 02 : 미국 Los Alamos National Lab. 연구원 (USA MPEG delegate)
- 2000. 03 ~ 2003. 08 : 명지대학교 정보통신공학 조교수
- 2003. 09 ~ 현재 : 경희대학교 전자정보대학 교수
- 2009. 09 ~ 2013. 09 : California IT (CALIT2) 연구소 초빙 연구원

- 연구분야

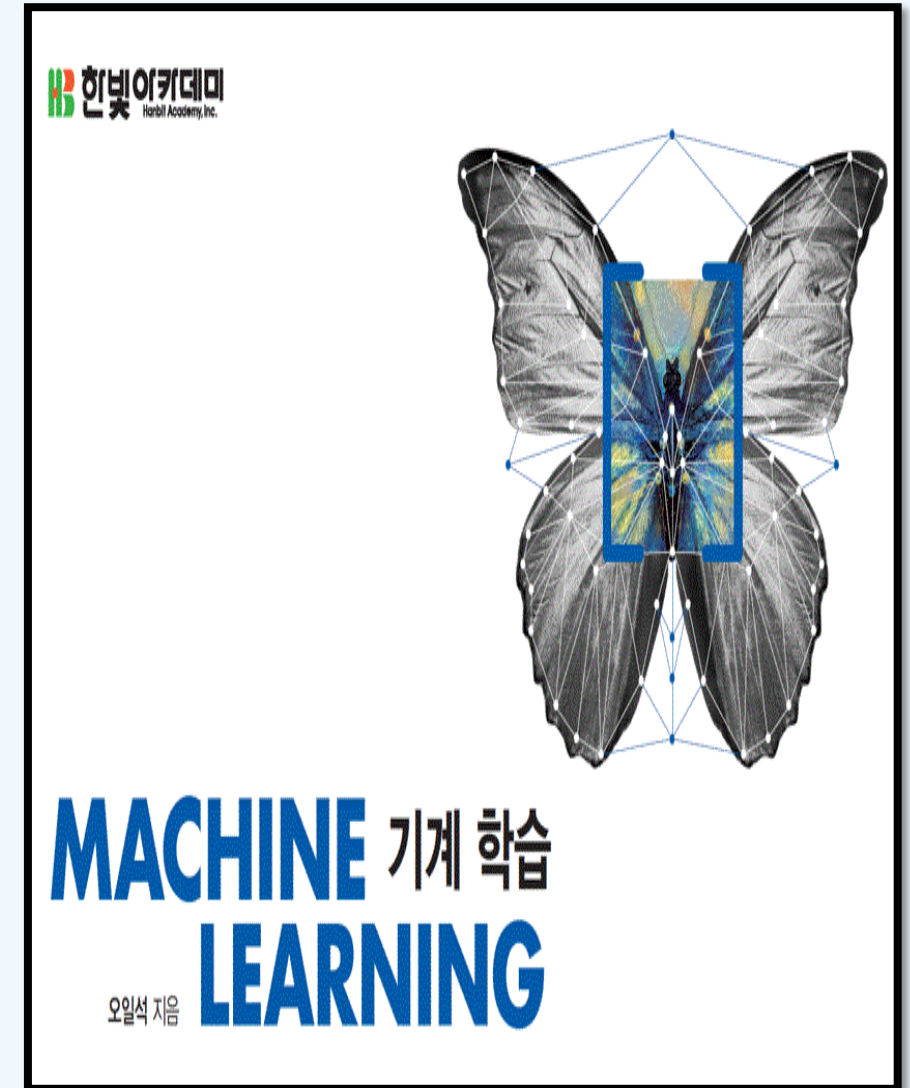
- 동영상 신호 처리, 영상 기반 자동 검측 시스템

- 연구활동 :

- 대한전자공학회 이사, 방송공학회 이사, 영상이해 연구위원

Contents

1. Concept of Artificial Intelligence (AI)
 - Scope of AI.
 - Concept of Machine Learning

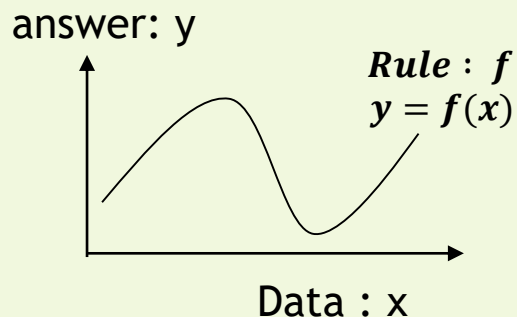
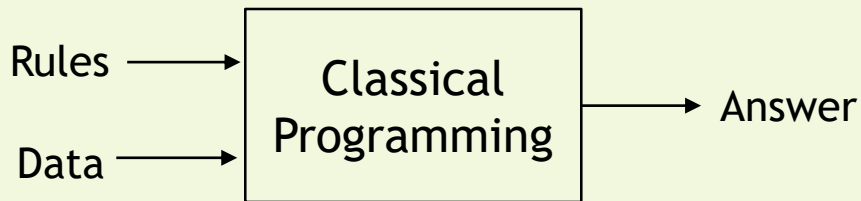


1.1 Scope of Artificial Intelligence (AI)

Artificial Intelligence :

Automation of intellectual tasks normally performed by humans.

- Without learning, Knowledge bases
- Humans specify rules (a program) and input data to be processed according to these rules, and out come answers



Machine Learning : with learning

- Logic bases
- study that lets systems be able to learn without being explicitly programmed.
- Random data
 - Needs training to decide a rule.
 - Needs training to cluster equivalent data
- States describing environments
 - Needs training to transit states to have more desirable environments

- KNN
- SVM
- EM
-

Deep Learning (DL) :
method representing
knowledge as hierarchy
of concepts



1.2 Concept of Machine Learning (ML) (1/4) : Example



- 직진 또는 좌회전의 방향에 맞도록 차선을 선택해야 함.
- 신호등의 순서는 직진(녹색:3분), 정지(빨강: 3분), 좌회전(1 분)
- 목적 : 교통 상황과 신호등의 상태에 따라서 신호 대기 시간을 최소화하는 직진 또는 좌회전 차선을 선택함.

• Method 1 : 교차로 도착 직전 신호등의 지시에 따라서 차선 선택

• Method 2 : 교차로의 신호등 상태와 기다리는 차량들의 숫자를 보고서 경험을 통하여 미리 차선 선택



- step 1 : Feature extraction : 신호등 상태 (feature 1), 대기 차량 수 (feature 2)
- step 2 : Data collection : Feature 1과 Feature 2에 따른 신호대기 시간에 대한 관찰 (data)
- step 3 : Model (F) 설정 : 신호등 지속 시간에 대한 확률 = $F(\text{신호등 상태, 대기차량 숫자})$
- step 4 : Training : Data로 부터 설정된 모델 F가 포함하고 있는 parameter를 추정

• Method 3 : Deep Neutral Network (DNN)



1.2 Concept of Machine Learning (ML) (2/4)

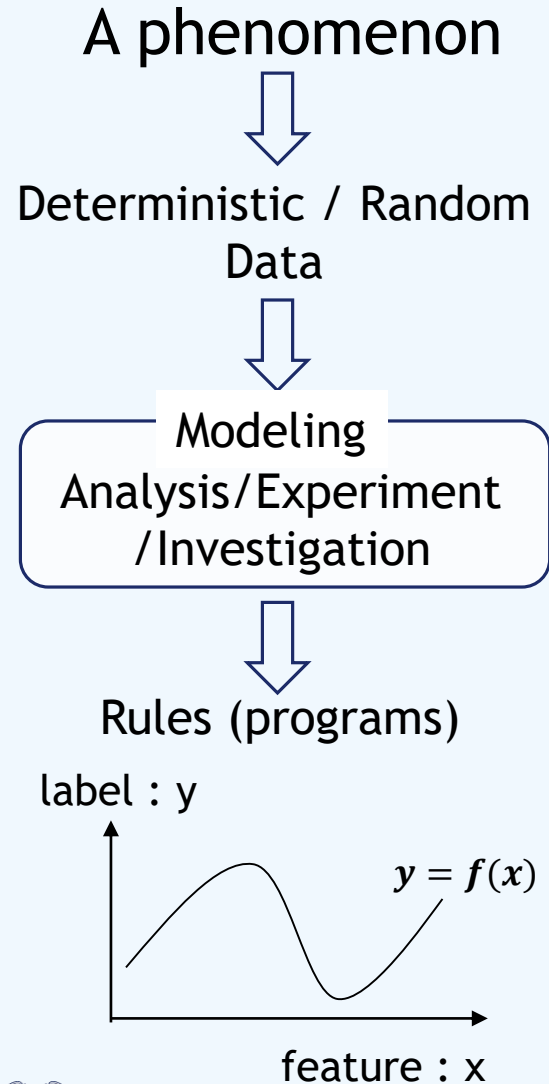
방법	Non Learning (지식 기반)	Machine Learning 기법 (데이터/경험 기반)	
특징	Learning이 없음. (method 1)	Modeling 기반 Learning (Method 2)	Deep Learning (DL) (Method 3)
장점	<ul style="list-style-type: none"> • 데이터가 거의 필요 없음 • 계산이 아주 간단함 	<ul style="list-style-type: none"> • 데이터가 상대적으로 적게 필요함 Ex: 2개월의 교차로 Data가 요구됨 • 계산이 상대적으로 간단함. 	<ul style="list-style-type: none"> • 복잡하고 난해한 문제도 용이하게 해결함. • 문제마다 모델을 크게 변화할 필요가 없음.
단점	<ul style="list-style-type: none"> • 문제가 복잡할 수록 성능이 떨어짐 	<ul style="list-style-type: none"> • 문제의 특성에 적절한 Feature을 설정해야 함 (Feature extraction). • 문제가 복잡할 수록 급격히 모델링이 어렵고 난해 함. 	<ul style="list-style-type: none"> • 다양한 정보가 포함된(즉, 복잡한) 데이터가 엄청나게 많이 필요함. Ex: 10년치 이상의 교차로 Data가 요구됨 • 엄청난 계산이 요구됨 Ex: ML보다 [십]만 배 이상의 계산량이 필요함.
응용	<ul style="list-style-type: none"> • Data 또는 문제의 해석이 간단한 경우 EX: 교통량이 적은 새벽 시간 	<ul style="list-style-type: none"> • 모델링이 용이하게 되는 응용에 적용 	<ul style="list-style-type: none"> • 다양한 정보가 포함된 데이터가 엄청나게 많이 확보된 경우 • 강력한 HW가 확보된 경우 • 문제의 모델링이 아주 난해한 경우

문제의 난이도, 확보된 데이터, 확보된 HW에 따라서 적절한 방법을 선택하거나 복합적으로 사용해야 함.

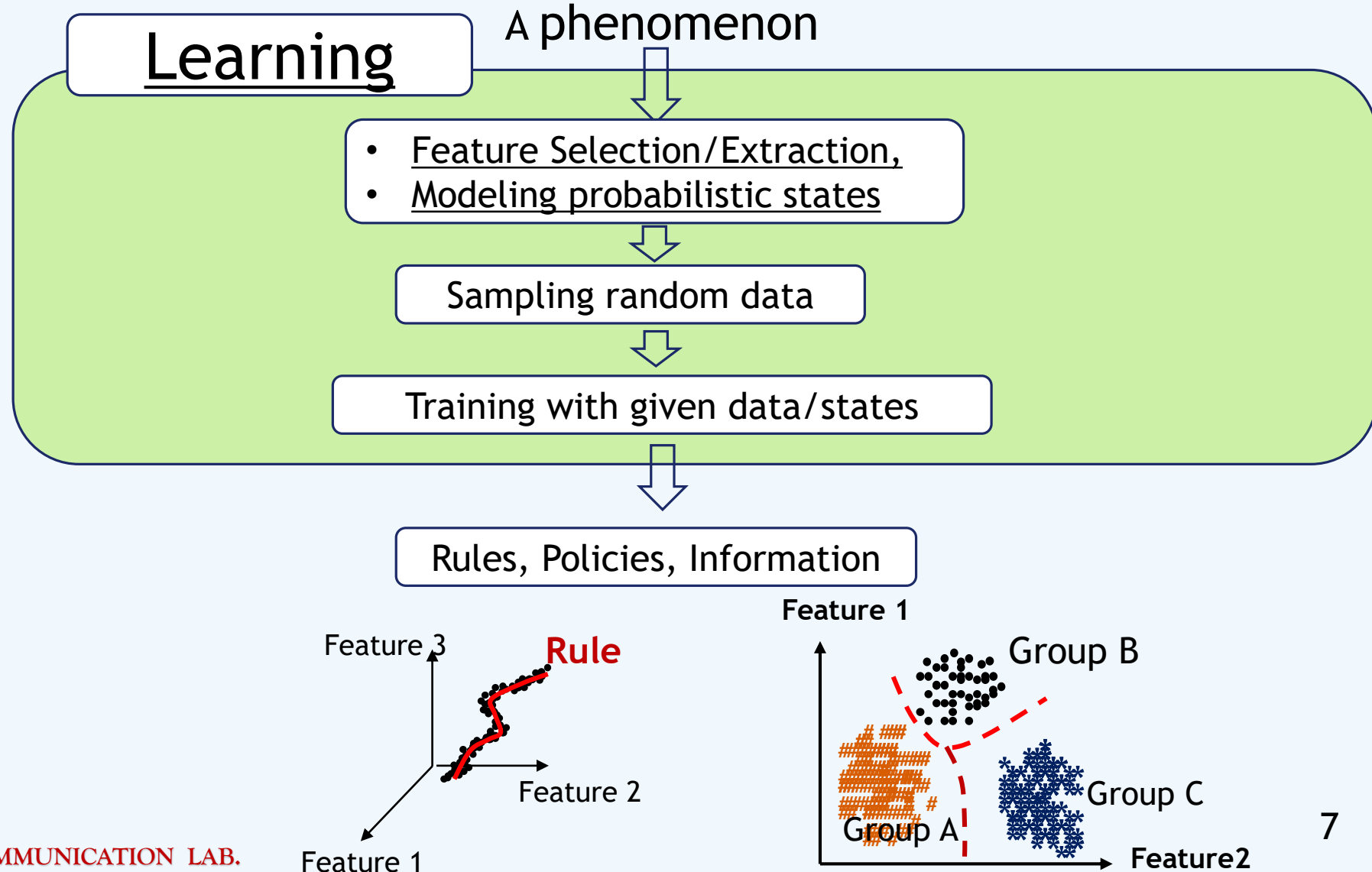


1.2 Concept of Machining Learning (ML) (3/4) : Procedure

- Non-ML problems



- ML problems



1.2 Concept of Machine Learning (4/4) : Definition

- 인공 지능 초장기 (사무엘의 정의)

“Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort. 컴퓨터가 경험을 통해 학습할 수 있도록 프로그래밍할 수 있다면, 세세하게 프로그래밍해야 하는 번거로움에서 벗어날 수 있다[Samuel1959].”

- 현대적 정의

“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 . 어떤 컴퓨터 프로그램이 T 라는 작업을 수행한다. 이 프로그램의 성능을 P 라는 척도로 평가했을 때 경험 E 를 통해 성능이 개선된다면 이 프로그램은 학습을 한다고 말할 수 있다[Mitchell1997(2쪽)].”

“Programming computers to optimize a performance criterion using example data or past experience 사례 데이터, 즉 과거 경험을 이용하여 성능 기준을 최적화하도록 프로그래밍하는 작업[Alpaydin2010]”

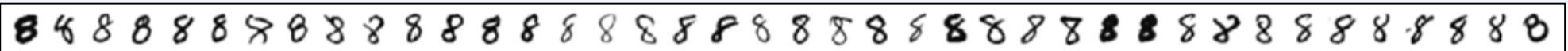
“Computational methods using experience to improve performance or to make accurate predictions 성능을 개선하거나 정확하게 예측하기 위해 경험을 이용하는 계산학 방법들[Mohri2012]”



1.2 Concept of ML : 지식에서 경험 / 데이터로의 전환

- 8의 예

- 지식 : “구멍이 2개이고 중간 부분이 홀쭉하며, 맨 위와 아래가 둥근 모양이라면 8이다”

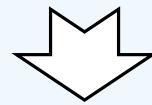
- Data : 

- 단추의 예 :

- Data :



- 지식 : ????



사람은 변화가 심한 장면을 아주 쉽게 인식하지만, 왜 그렇게 인식하는지 서술하지는 못함
현상에는 반드시 Random 성분이 있음



지식 기반으로 현상을 기술하거나 표현하는데 한계가 있음
지식 기반에서 데이터 중심의 기계학습으로 전환이 요구됨

1.3 Procedures and Method of ML (1/4)

• 기계학습 방법

- Step 1 : Data Collection
- Step 2 : Data Modeling
- Step 3 : Training
- Step 4 : Verification

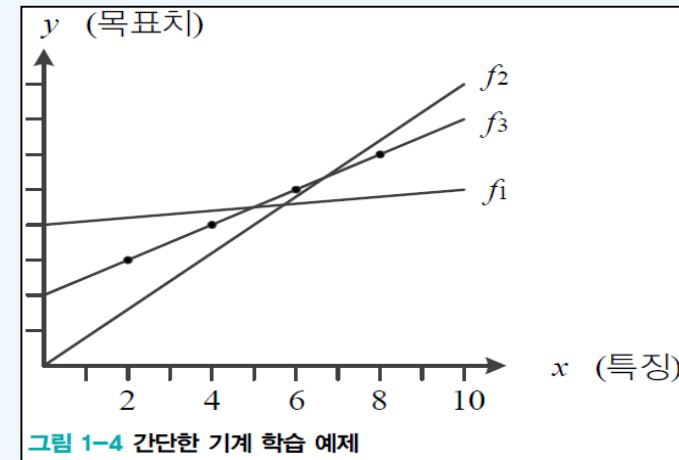
• Training의 목적

- Training에 사용되지 않은 data sample (Test data)에 대하여 오류를 최소화
- Test data에 대한 높은 성능을 generalization능력이라 함.

- 학습 후 모델로 입력된 feature data에 대하여 output 예측

• 간단한 기계 학습 예제

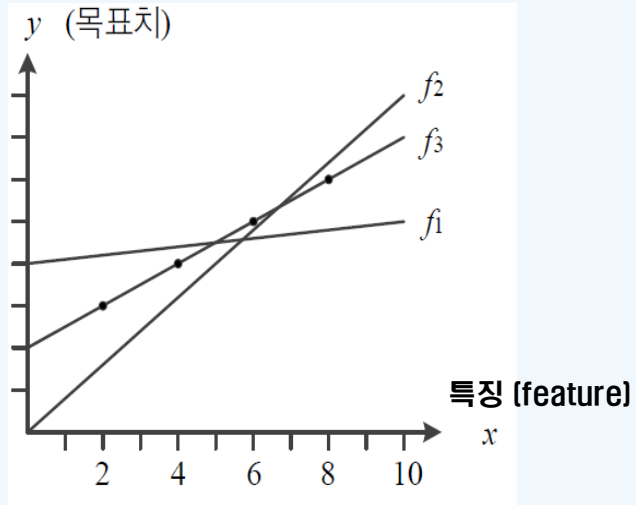
- 가로축은 시간, 세로축은 이동체의 위치
- 관측한 4개의 점이 데이터



- Task : Prediction
임의의 시간에 따라서 이동체의 위치를 추정

1.3 Procedures and Method of ML (2/4)

• Step 1 : Data Collection



- 가로축은 **특징**, 세로축은 **목표치**
- 관측한 4개의 점이 **Training Data set**
- Training data set:

$$\mathbb{X} = \{x_1, x_2, \dots, x_n\}, \quad \mathbb{Y} = \{y_1, y_2, \dots, y_n\}$$

$$\mathbb{X} = \{x_1 = (2.0), x_2 = (4.0), x_3 = (6.0), x_4 = (8.0)\}$$
$$\mathbb{Y} = \{y_1 = 3.0, y_2 = 4.0, y_3 = 5.0, y_4 = 6.0\}$$

• Step 2 : Data Modelling

- Feature와 value 관계를 (경험적으로) 직선으로 modelling
- '2개의 매개변수 w 와 b 를 도입하여 : $y = wx + b$

• Step 3 : Training

- 가장 정확한 예측 값을 산출하는 최적의 매개변수를 찾는 작업
- 처음에는 최적의 w 와 b 를 알 수 없기 때문에 임의의 값에서 시작하고, 점점 성능을 개선하여 최적에 도달 (다양한 방법이 존재)
- f_1 에서 시작하여 $f_1 \rightarrow f_2 \rightarrow f_3 \Rightarrow$ 최적인 f_3 은 $w=0.5$ 와 $b=2.0$
- Training 후에는 예측에 사용 : $x=10.0$ 이면 $f_3(10.0)=0.5*10.0+2.0=7.0$

• Step 4 : 검증 (Verification)

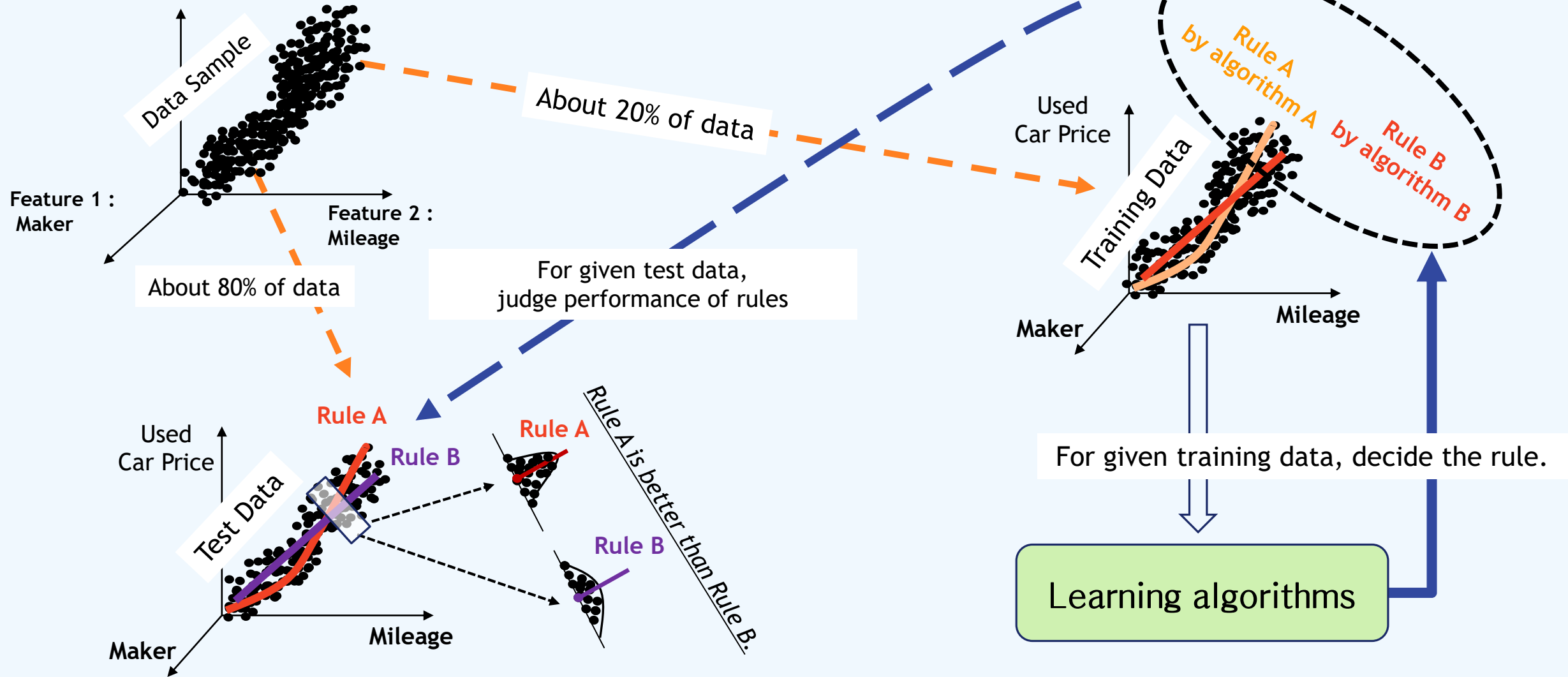
- Training data에 없는 새로운 sample data(test data)로 정확성 검증
- Test data에 대하여 높은 성능이 나면 사용 (즉, generalization 함.)



1.3 Procedures and Method of ML (3/4)

Used Car Price Estimation

Label : Used Car Price



1.3. Procedures and Method of ML (4/4)

표 1-1 사람의 학습과 기계 학습의 비교

기준	사람의 학습	기계 학습
학습 과정	능동적	수동적
데이터 형식	자연에 존재하는 그대로	일정한 형식에 맞추어 사람이 준비함
동시에 학습 가능한 과업 수	자연스럽게 여러 과업을 학습	하나의 과업만 가능
학습 원리에 대한 지식	매우 제한적으로 알려져 있음	모든 과정이 밝혀져 있음
수학 의존도	매우 낮음	매우 높음
성능 평가	경우에 따라 객관적이거나 주관적	객관적(수치로 평가, 예를 들어 정확률 99.8%)
역사	수백만 년	60년 가량

