

동양미래대학교 인공지능소프트웨어학과

4차산업혁명 시대의 인공지능
“누구나 이해할 수 있는 인공지능”

Dongyang Mirae University
Dept. Of Artificial Intelligence

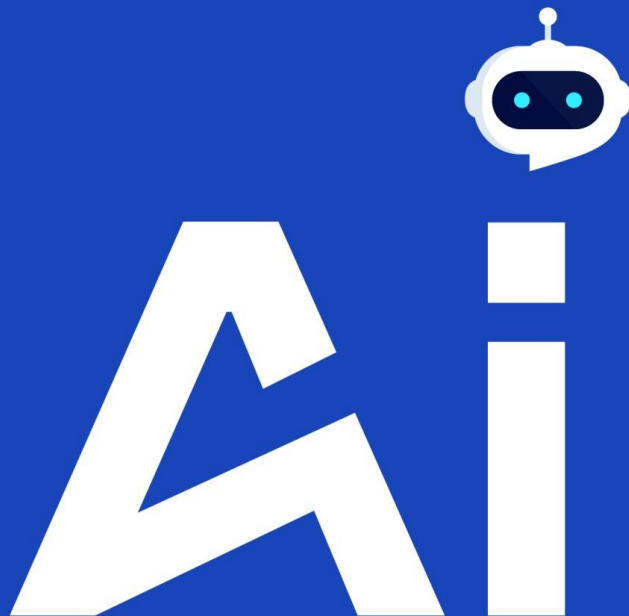


DONGYANG MIRAE UNIVERSITY
Dept. of Artificial Intelligence

누구나 이해하는 인공지능

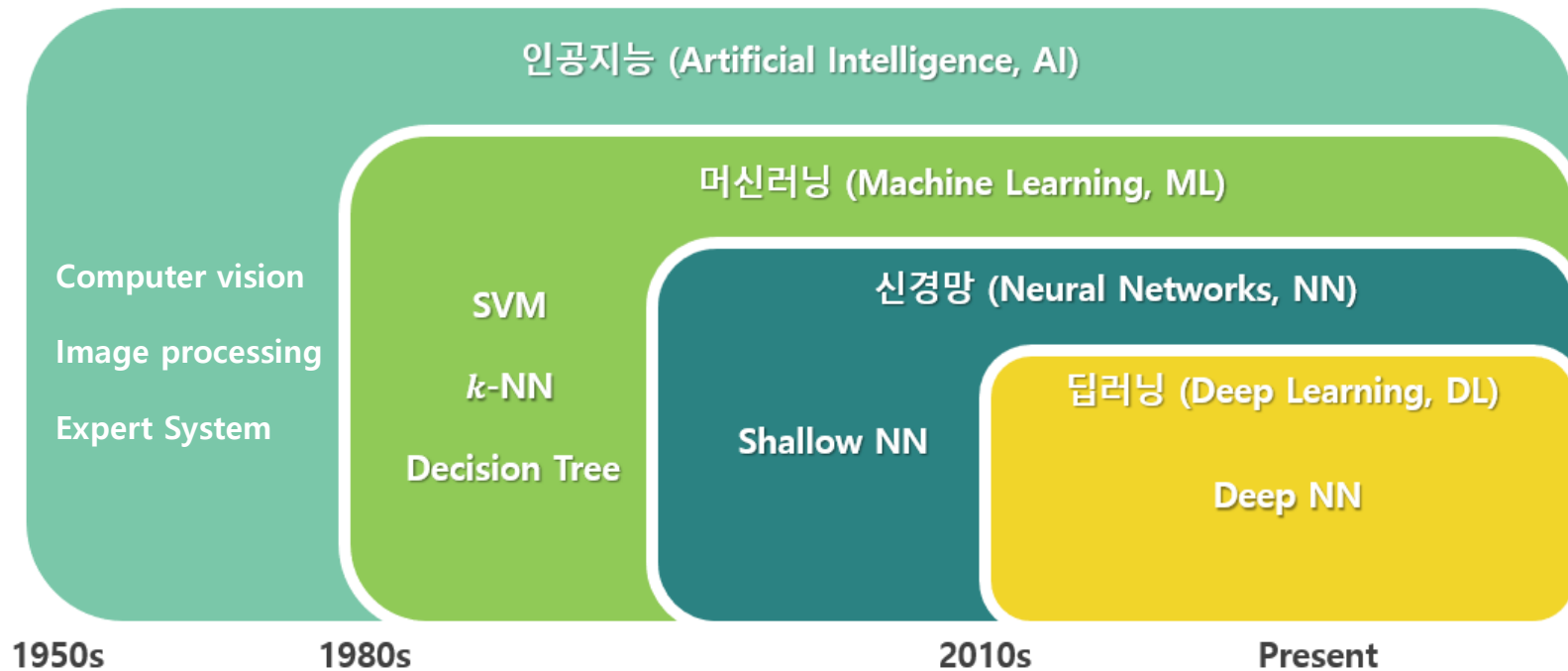
머신러닝 복습

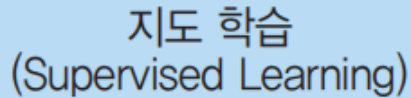
강환수 교수



DMUAi
동양미래대학교
인공지능소프트웨어학과

머신러닝의 다양성





문제와 정답을 모두 알려 주고
공부시키는 방법

비지도 학습 (Unsupervised Learning)



답을 가르쳐 주지 않고 공부시키는 방법

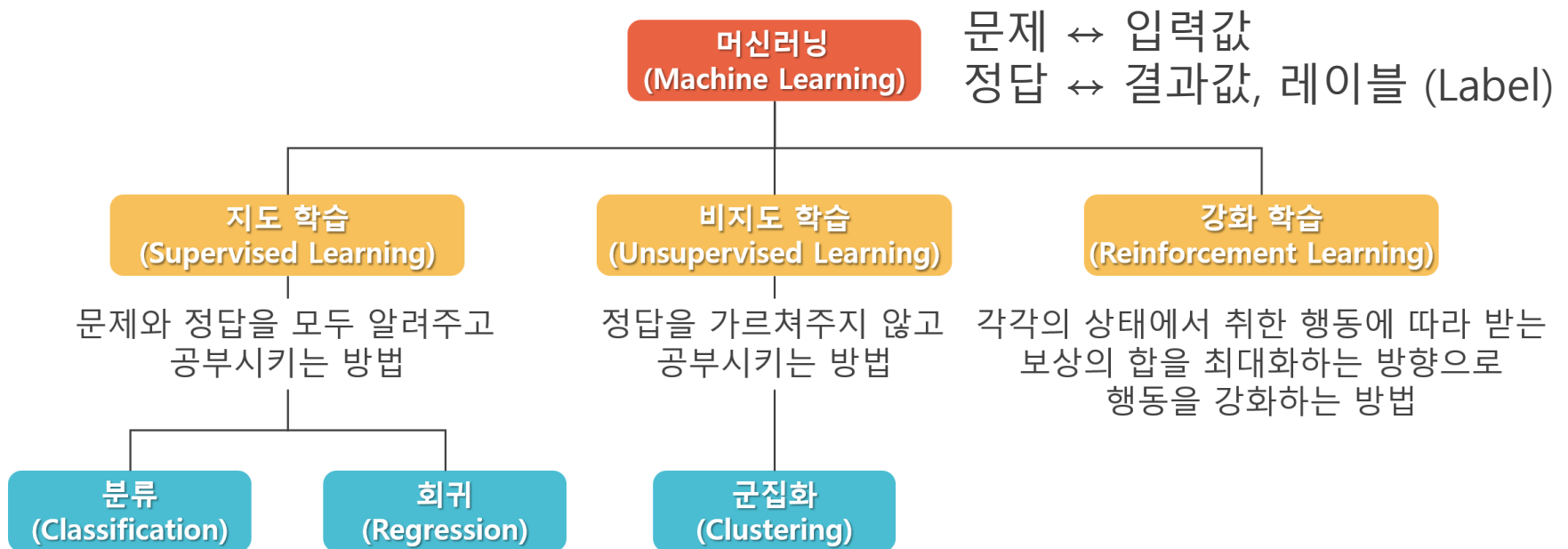
강화 학습 (Reinforcement Learning)



보상을 통해 상은 최대화, 벌은 최소화하는
방향으로 행위를 강화하는 학습

머신 러닝 분류 2/2

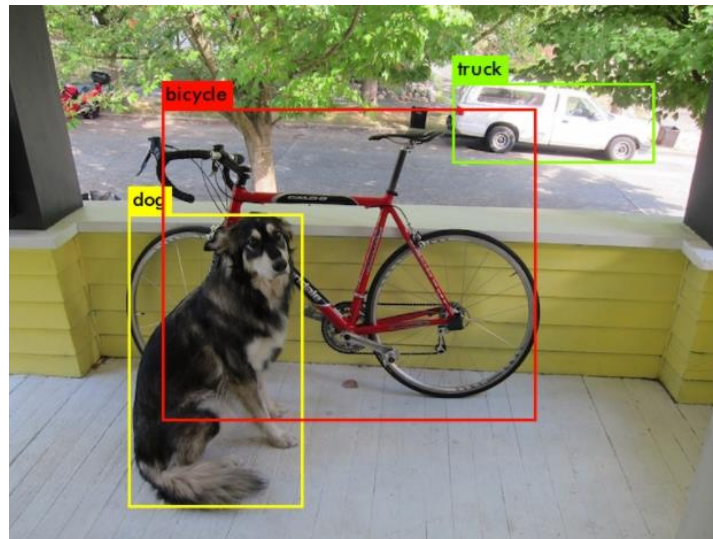
- 머신러닝 모델을 학습시킴에 있어
 - 입력값에 대한 결과값을 알고 있는지 여부에 따라
 - “지도학습”과 “비지도 학습”으로 나누어짐



머신러닝은 언제 주로 사용될까?

• 머신러닝의 응용 분야

- 주로, 복잡한 데이터들이 있고 이 데이터들에 기반하여 결정을 내려야 하는 분야
- 머신러닝 모델을 학습(Learning) 시키려면 많은 데이터가 필요하기 때문
 - 머신러닝은 빅데이터(Big Data)와 아주 밀접한 관계가 있음
- 예를 들어
 - 객체 검출 (Object Detection)
 - 음성 인식 (Voice Recognition)
 - 글자 인식 (Character Recognition)



머신러닝 모델의 생성 과정

머신러닝 모델을 생성

- 머신러닝으로 문제를 해결하기 위해서는 문제 적합한 머신러닝 모델을 생성해야 함

- 문제 속의 데이터를 해결에 잘 설명할 수 있는 머신러닝 모델 선정이 중요

학습과 적용

- 모델로부터 학습 데이터에 최적화된 구체적인 함수를 찾아내는 과정
 - 과정을 학습(Learning)이라고 함
- 학습된 모델(함수)를 실제 문제에 적용

시험 공부 시간으로 시험 성적을 예상하는 문제

- 머신러닝으로 해결
 - 머신러닝 모델 생성 과정을 개념적으로 표현하면 아래와 같음

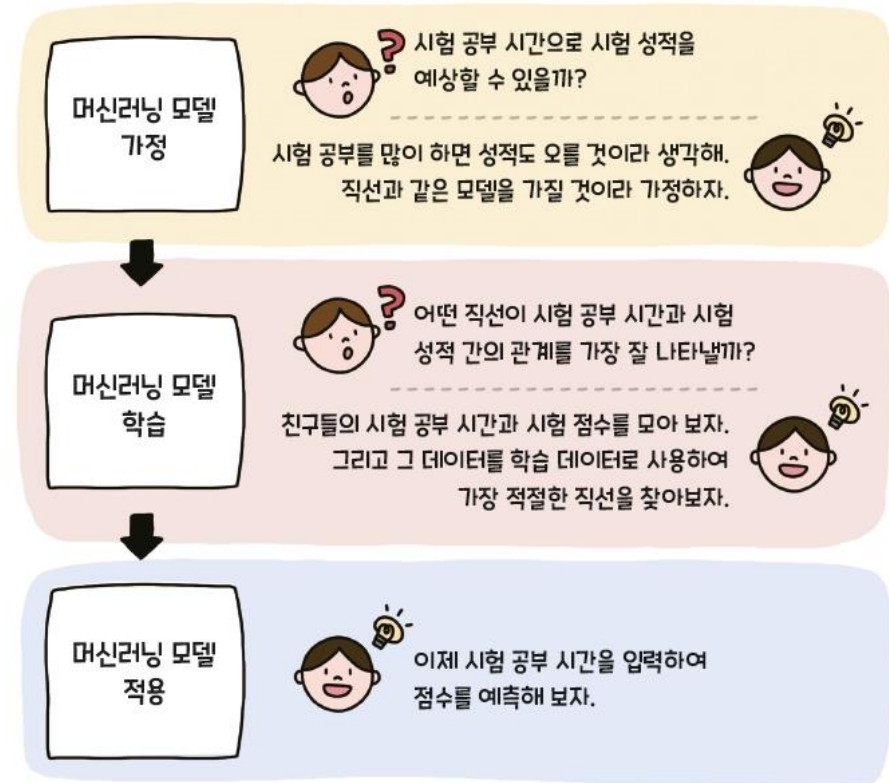


그림 2-4 머신러닝 모델 생성 과정의 예

머신러닝 구현 과정 예제 (1/4)

- 사람의 키를 입력했을 때 몸무게를 추측하는 작업을 머신러닝으로 구현하면?

- ① 일반적으로 키가 커지면 몸무게도 늘어날 것이라 가정해,
 $[몸무게 = a \times 키 + b]$ 와 같은 직선의 방정식을 만들어 머신러닝 모델로 가정

간단하지만 실제 사용하는 모델

$$몸무게 = a \times 키 + b$$

$$y = a \times x + b$$



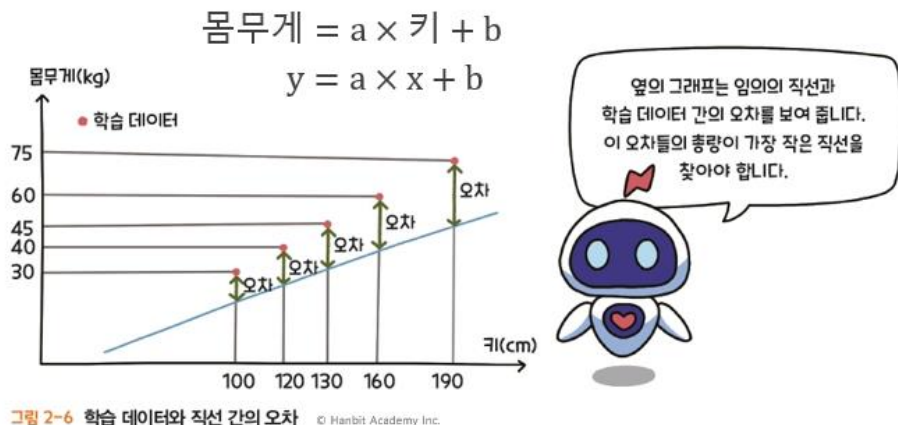
그림 2-5 머신러닝 구현 과정 예제의 학습 데이터 © Hanbit Academy Inc.

No.	이름	x	y
		키 (cm)	몸무게 (kg)
1	김민성	100	30
2	박다인	120	40
3	윤이안	130	45
4	최서연	160	60
5	문진승	190	75

머신러닝 구현 과정 예제 (2/4)

- 사람의 키를 입력했을 때 몸무게를 추측하는 작업을 머신러닝으로 구현 하면?
 - ② 머신러닝 모델로 가정한 직선의 방정식은 아직 기울기 a 와 y 절편 b 의 값이 결정 되지 않은 상태
 - 학습 데이터를 확보하여 최적화된 직선을 구함
 - 학습 데이터와 최적화된 직선을 구한다는 것
 - 학습 데이터와 오차가 가장 적은 직선의 기울기 ($= a$)와 y 절편 ($= b$)을 구한다는 의미

		x	y
No.	이름	키 (cm)	몸무게 (kg)
1	김민성	100	30
2	박다인	120	40
3	윤이안	130	45
4	최서연	160	60
5	문진승	190	75



머신러닝 구현 과정 예제 (3/4)

• 사람의 키를 입력했을 때 몸무게를 추측하는 작업을 ML로 구현하면?

② 직선의 방정식은 아직 기울기 a 와 y 절편 b 의 값이 결정되지 않은 상태로, 학습 데이터를 확보하여 최적화된 직선을 구함

- 학습 결과, 직선의 방정식 [몸무게 = $(0.5 \times \text{키}) - 20$]이 해당 학습 데이터에 최적화된 함수임

		x	y
No.	이름	키 (cm)	몸무게 (kg)
1	김민성	100	30
2	박다인	120	40
3	윤이안	130	45
4	최서연	160	60
5	문진승	190	75

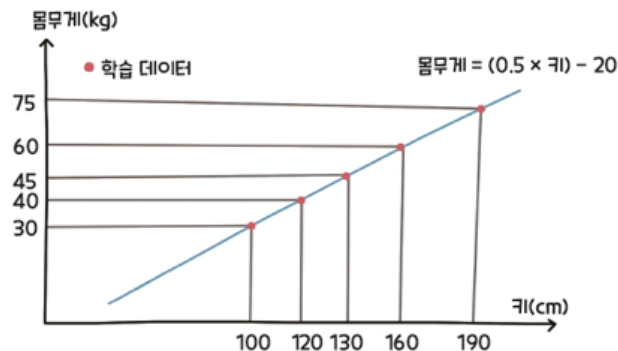


그림 2-7 몸무게와 키의 상관관계에 최적화된 직선의 방정식 © Hanbit Academy Inc.

$$\text{몸무게} = a \times \text{키} + b$$

$$y = a \times x + b$$



최적화 (Optimization)

$$a = 0.5$$

$$b = -20$$



$$y = 0.5 \times x - 20$$

머신러닝 구현 과정 예제 (4/4)

• 사람의 키를 입력했을 때 몸무게를 추측하는 작업을 ML로 구현하면?

- ③ 최적화된 머신러닝 모델을 실제 문제에 적용해 확인
 - 학습 데이터에는 없던 키 180cm를 방정식에 대입하면 예측되는 몸무게는 $[70\text{kg} = (0.5 \times 180) - 20]$ 이 나옴

Q) 키가 180cm 인 사람은 몸무게가 몇 kg일까?

No.	이름	x	y
		키 (cm)	몸무게 (kg)
1	김민성	100	30
2	박다인	120	40
3	윤이안	130	45
4	최서연	160	60
5	문진승	190	75

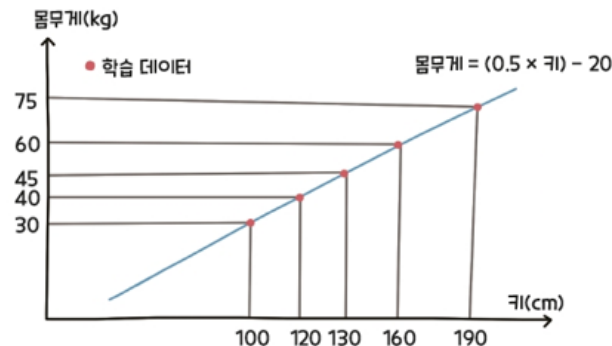


그림 2-7 몸무게와 키의 상관관계에 최적화된 직선의 방정식 © Hanbit Academy Inc.

$$\begin{aligned}
 y &= 0.5 \times x - 20 \\
 &\downarrow \\
 x &= 180 \text{ 대입} \\
 &\downarrow \\
 y &= 0.5 \times 180 - 20 \\
 &= 90 - 20 \\
 &= 70
 \end{aligned}$$

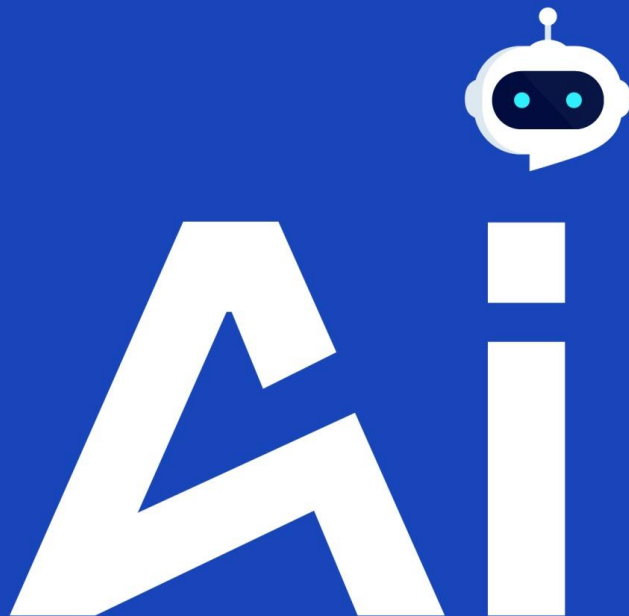
회귀와 분류 (regression and classification)

DONGYANG MIRAE UNIVERSITY
Dept. of Artificial Intelligence

누구나 이해하는 인공지능

회귀와 분류

강환수 교수



Ebs 수학

<https://www.ebssw.kr/info/intro/infoTchmtrHeaderView.do?tabType=AI>

DMUAi
동양미래대학교
인공지능소프트웨어학과

회귀(regression)와 분류(classification)

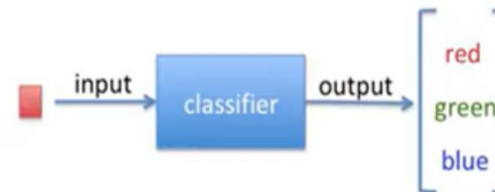
• 회귀 모델

- 연속적인 값(실수)을 예측
 - 캘리포니아의 주택 가격이 얼마인가요?
 - 사용자가 이 광고를 클릭할 확률이 얼마인가요?
 - 키(Height) 정보가 주어졌을 때, 몸무게를 예측
 - 공부한 시간 정보가 주어졌을 때, 시험 성적 예측
 - 커피를 몇 잔 마셨는지에 대한 정보가 주어졌을 때, 수면 시간 예측
 - 사과와 전년도 수확량과 날씨, 고용 인원 수 등으로 올해 수확량 예측

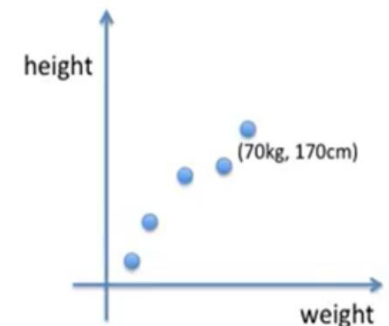
Classification VS Regression

• 분류 모델

- 불연속적인 값(유한 개 이산 값)을 예측
 - 주어진 이메일 메시지가 스팸인가요, 스팸이 아닌가요?
 - 이 이미지가 강아지, 고양이 또는 햄스터의 이미지인가요?



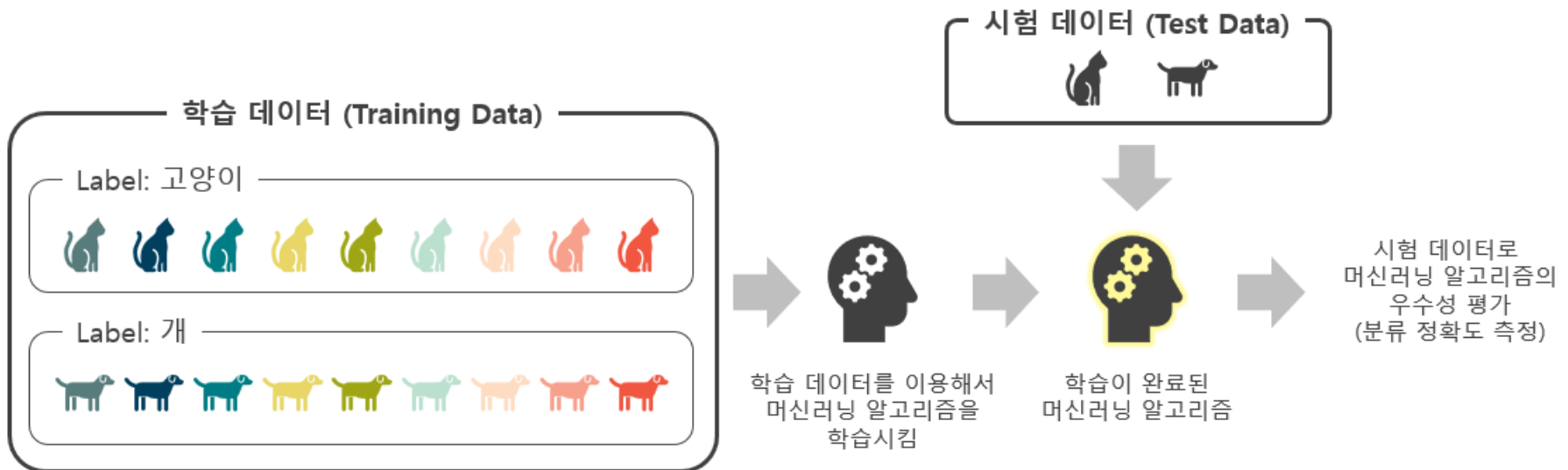
classify input into categorical output



how tall is he if his weight is 80kg?

분류 (Classification)

- **정답(레이블)이 포함된 데이터를 학습하여**
 - 유사한 성질을 갖는 데이터끼리 분류한 후
 - 새로 입력된 데이터가 어느 그룹에 속하는지를 찾아내는 기법
 - 어떤 입력 데이터가 들어오더라도
 - 학습에 사용한 레이블(Label) 중 하나로 결과값을 결정
 - 레이블(Label)이 이산적인(Discrete) 경우 (즉, [0, 1, 2, 3, ...]와 같이 유한한 경우)



이진분류와 다중분류

이진 분류 (Binary Classification)

- 데이터를 2개의 그룹 (Class)으로 분류

Label: 고양이



Label: 개



"고양이" or "개"
둘 중에 하나로 분류

다중 분류 (Multiclass Classification)

- 데이터를 3개의 그룹 (Class) 이상으로 분류

Label: 고양이



Label: 개



Label: 토끼



"고양이" or "개" or "토끼"
셋 중에 하나로 분류

분류의 이해

• 대표적인 예: 필기체 (손글씨) 인식

- 0부터 9까지의 손글씨 숫자 이미지와 레이블 (Label) 정보를 학습 데이터로 사용
- 어떤 이미지라도 (심지어는 숫자 이미지가 아니더라도) 0부터 9까지의 레이블 중 하나로 결과값을 결정함

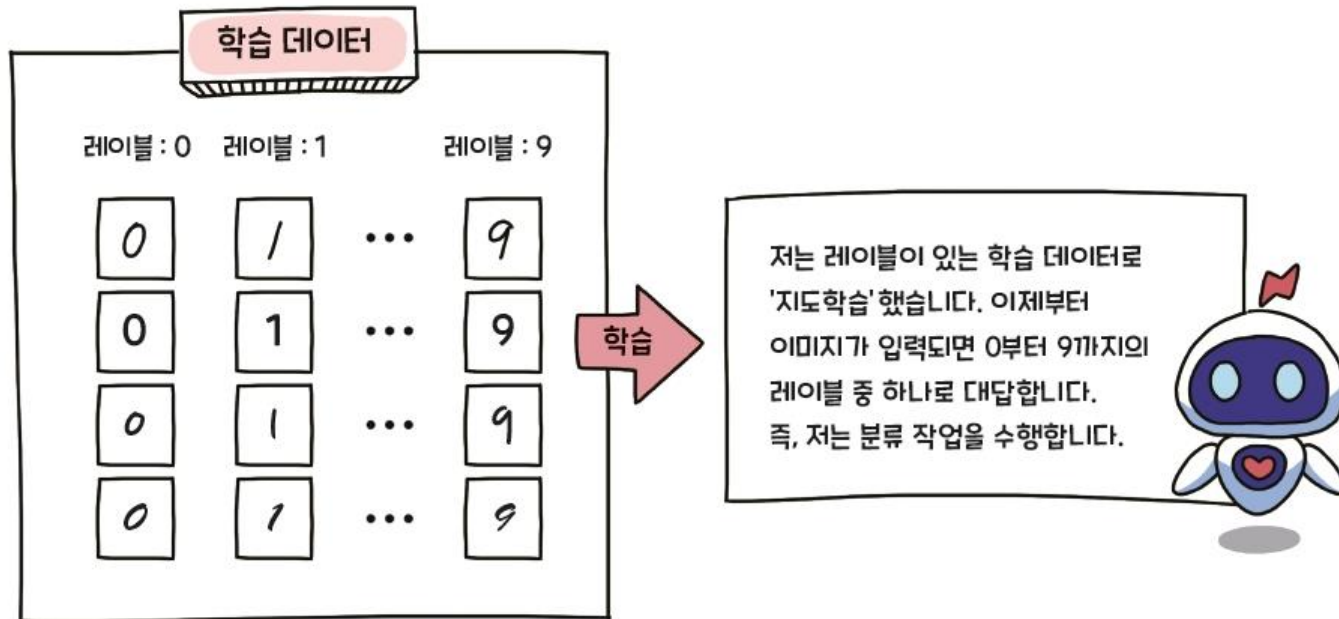


그림 2-9 머신러닝의 분류 작업 예: 숫자 필기체 인식하기

숫자 필기체 인식 실험

- <https://digit-recog-torch.uc.r.appspot.com/>

CNN Digit Recognition WebApp using PyTorch, Flask

<https://github.com/nai-kon/CNN-Digit-Recognition>
[Donation to the author \(via PayPal\)](#)

Clear

Probability : 100.00%

Overview

Pre-Processing

input → detect → centering

Neural Network

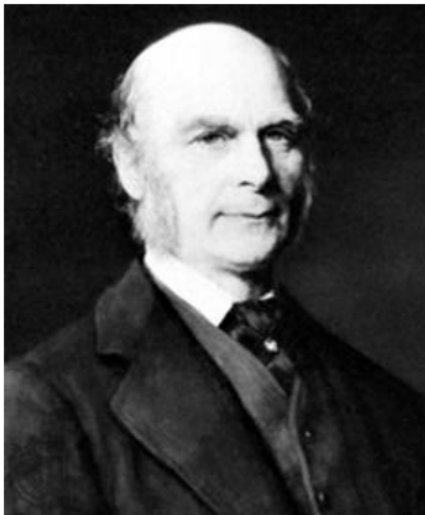
- 2xCNN layers
- Batch Normalization
- Adam Optimizer
- 99.3% of accuracy at MNIST test images

Created by Katz Sasaki
Hosting on Google App Engine

회귀의 어원

• 회귀 분석(regression analysis)

- 관찰된 연속형 변수들에 대해 두 변수 사이의 모형을 구한 뒤 적합도를 측정해 내는 분석 방법
- 회귀분석은 시간에 따라 변화하는 데이터나 어떤 영향, 가설적 실험, 인과 관계의 모델링 등의 통계적 예측에 이용



[사진출처] https://en.wikipedia.org/wiki/Francis_Galton

프랜시스 골턴

- 아버지와 자식의 키를 분석함
- 사람의 키 (Height)는 세대를 거듭할 수록 평균에 가까워지는 경향이 있다는 것을 발견
- 키가 큰 아버지의 자식은 아버지보다 키가 작고, 키가 작은 아버지의 자식은 아버지보다 키가 크다
- 즉, 세대를 거듭할 수록 큰 키는 작아지고, 작은 키는 커지고 평균에 수렴한다
- 이를 프랜시스 골턴은 "평균으로 돌아간다 (=회귀)"라고 표현함

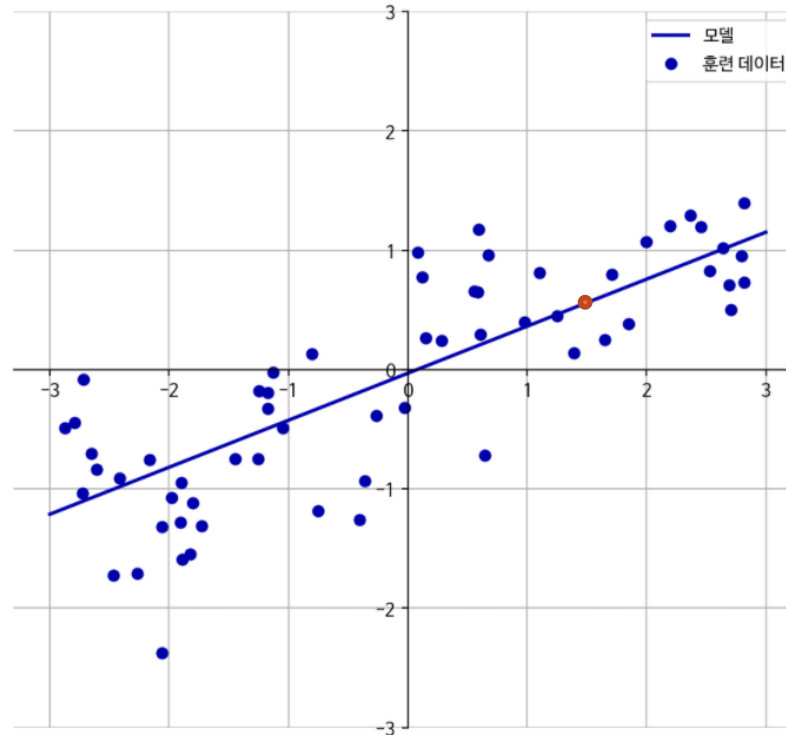
회귀(영어: regress 리그레스[*])의 원래 의미

- 변수 사이의 관계를 분석하는 방법을 역사적인 이유 때문에
 - “회귀 (Regression)” 라고 부름
- 19세기, 통계학자이자 인류학자인 프랜시스 골턴 (Francis Galton)이 처음 사용
 - 옛날 상태로 돌아가는 것을 의미
 - 프랜시스 골턴은 “평균으로의 회귀(regression to the mean)”
 - 부모의 키와 아이들의 키 사이의 연관 관계를 연구
 - 부모와 자녀의 키 사이에는 선형적인 관계가 있고 키가 커지거나 작아지는 것보다는
 - 전체 키 평균으로 돌아가려는 경향이 있다는 가설을 세웠으며
 - 이를 분석하는 방법을 “회귀분석”이라고 함
 - 이러한 경험적 연구 이후, 칼 피어슨은 아버지와 아들의 키를 조사한 결과를 바탕으로 함수 관계를 도출하여 회귀분석 이론을 수학적으로 정립

선형 회귀 (Linear Regression) (1/4)

- 데이터로 부터 예측
 - x 로 y 를 예측, 특징 수: 1
 - 직선으로 예측
- 테스트 데이터, 붉은 점의 y (타깃)?

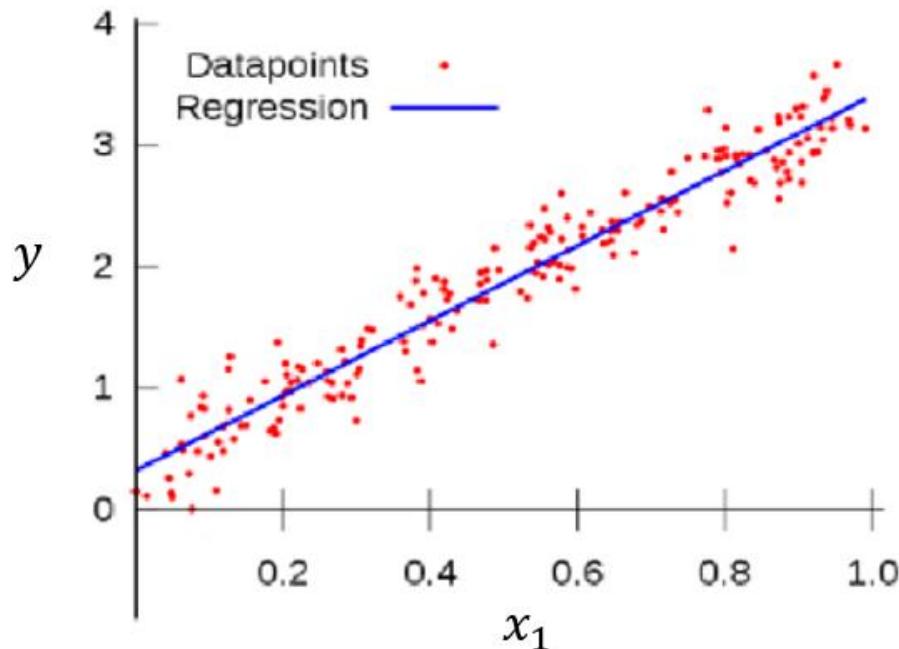
w[0]: 0.393906 b: -0.031804



선형 회귀 (Linear Regression) (2/4)

• 학습 데이터 (Training Data)로부터

- 종속 변수 y 와 한 개 이상의 독립 변수 x 와의 선형 상관 관계를 모델링하는 방법
 - 쉽게 이해하면, 학습 데이터를 잘 표현하는 직선 하나를 찾아내겠다는 의미
 - 입력이 하나이면 “직선”이 됨



[사진출처] <https://bangu4.tistory.com/100>

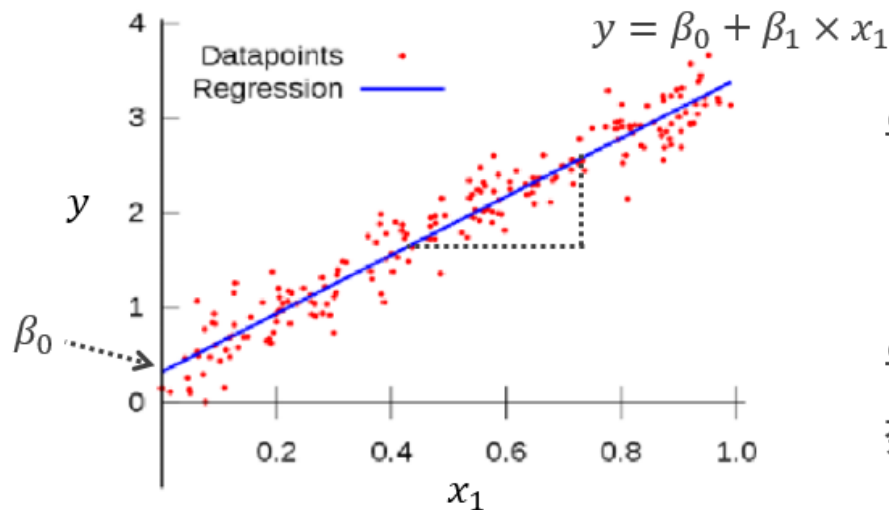
빨간색 점들이 학습 데이터이고,

이 데이터들을 잘 표현하는
파란색 직선을

찾아내는 것이 “선형 회귀”가 하는
역할임

선형 회귀 (Linear Regression) (3/4)

- 독립 변수 x 의 개수에 따라 선형 회귀는 아래와 같이 분류됨
 - 단순 선형 회귀 (Simple Linear Regression)
 - 독립 변수 x 의 개수가 1개
 - 샘플 문제 제외하곤 실제 문제에서는 거의 없는 경우
 - 다중 선형 회귀 (Multiple Linear Regression)
 - 독립 변수 x 의 개수가 2개 이상



[사진출처] <https://bangu4.tistory.com/100>

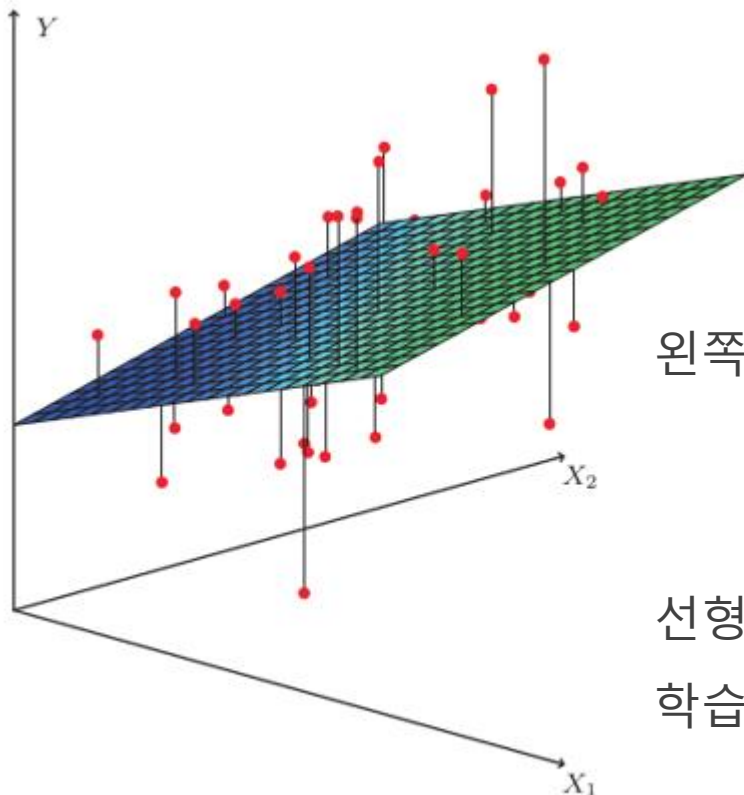
왼쪽 예제는 독립 변수의 개수가 1개인 경우임

$$y = \beta_0 + \beta_1 \times x_1$$

위 직선에서 기울기 ($= \beta_1$) 값과 y절편 ($= \beta_0$) 값을 찾는 것이 선형 회귀 알고리즘에서 수행하는 동작

선형 회귀 (Linear Regression) (4/4)

- 독립 변수 x 의 개수에 따라 선형 회귀는 아래와 같이 분류됨
 - 단순 선형 회귀 (Simple Linear Regression): 독립 변수 x 의 개수가 1개
 - 다중 선형 회귀 (Multiple Linear Regression): 독립 변수 x 의 개수가 2개 이상



독립 변수(입력의 차수)가
여러 개이면 그만큼 복잡

왼쪽 예제는 독립 변수의 개수가 2개 (x_1, x_2)인 경우임

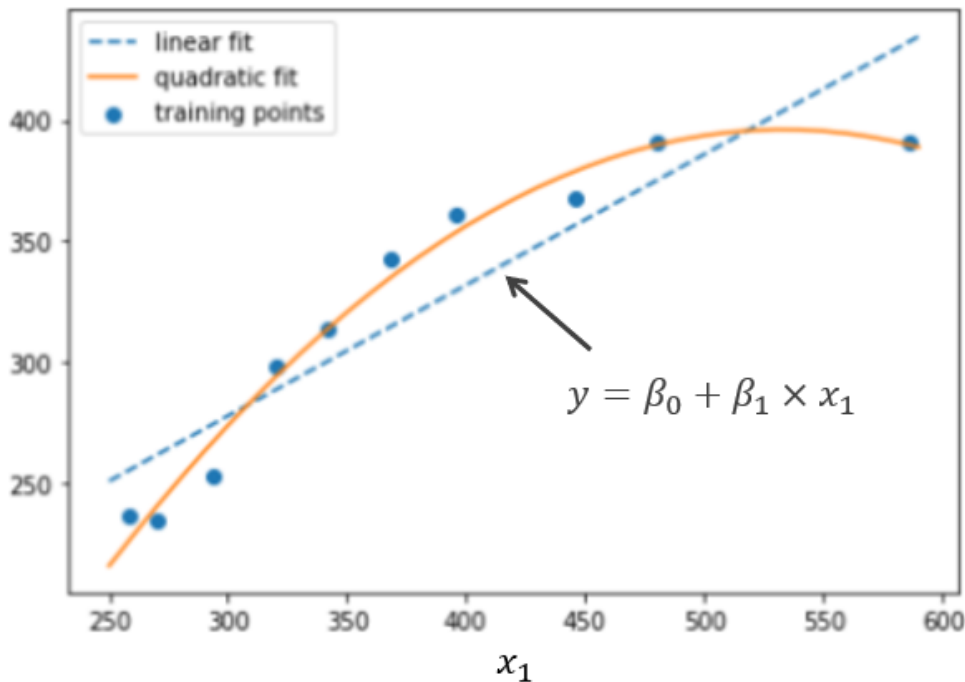
$$y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2$$

선형 회귀 알고리즘 수행을 통해서,
학습 데이터를 가장 잘 표현 하는 $\beta_0, \beta_1, \beta_2$ 값을 찾아냄

다항 회귀 (Polynomial Regression) (1/2)

• 선형 회귀의 단점

- 학습 데이터 내, 종속 변수 y 와 독립 변수 x 사이의 상관 관계
 - 선형이 아닐 수 있음!



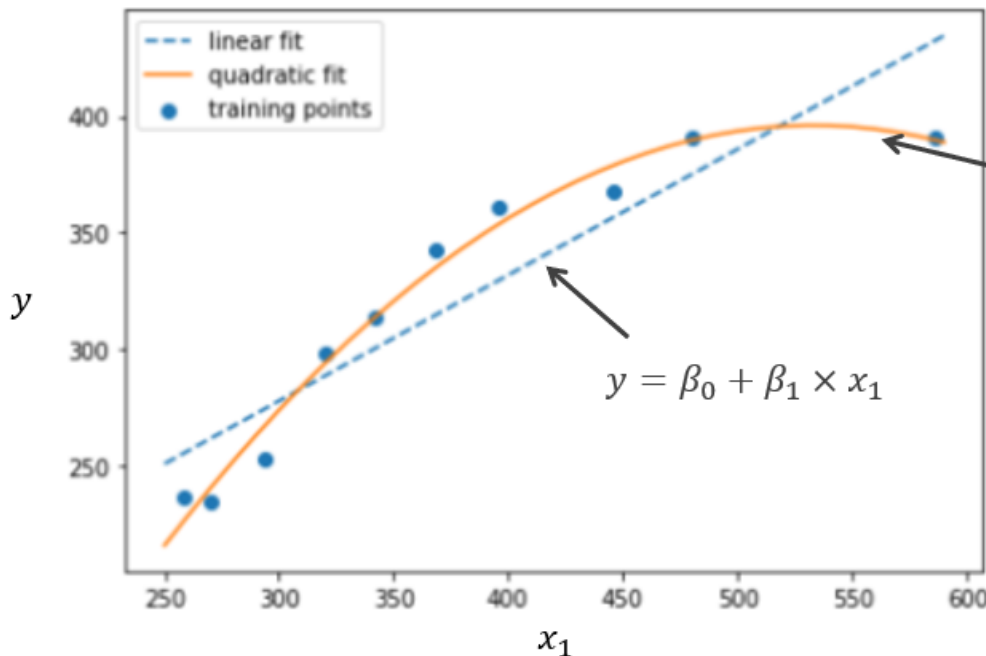
[사진출처] <https://hongl.tistory.com/128>

학습 데이터 ●가 선형 회귀 모형 (점선)으로 잘 표현되지 않고 있다 (오차가 크다)

오히려 주황색 실선이 학습 데이터 ●를 잘 표현하고 있다 (오차가 작다)

다항 회귀 (Polynomial Regression) (2/2)

- 각 독립 변수 x 에 대한 고차원의 다항식을 이용
 - 종속 변수 y 의 관계를 비선형적(Non-linear)으로 모델링하는 방법



[사진출처] <https://hongl.tistory.com/128>

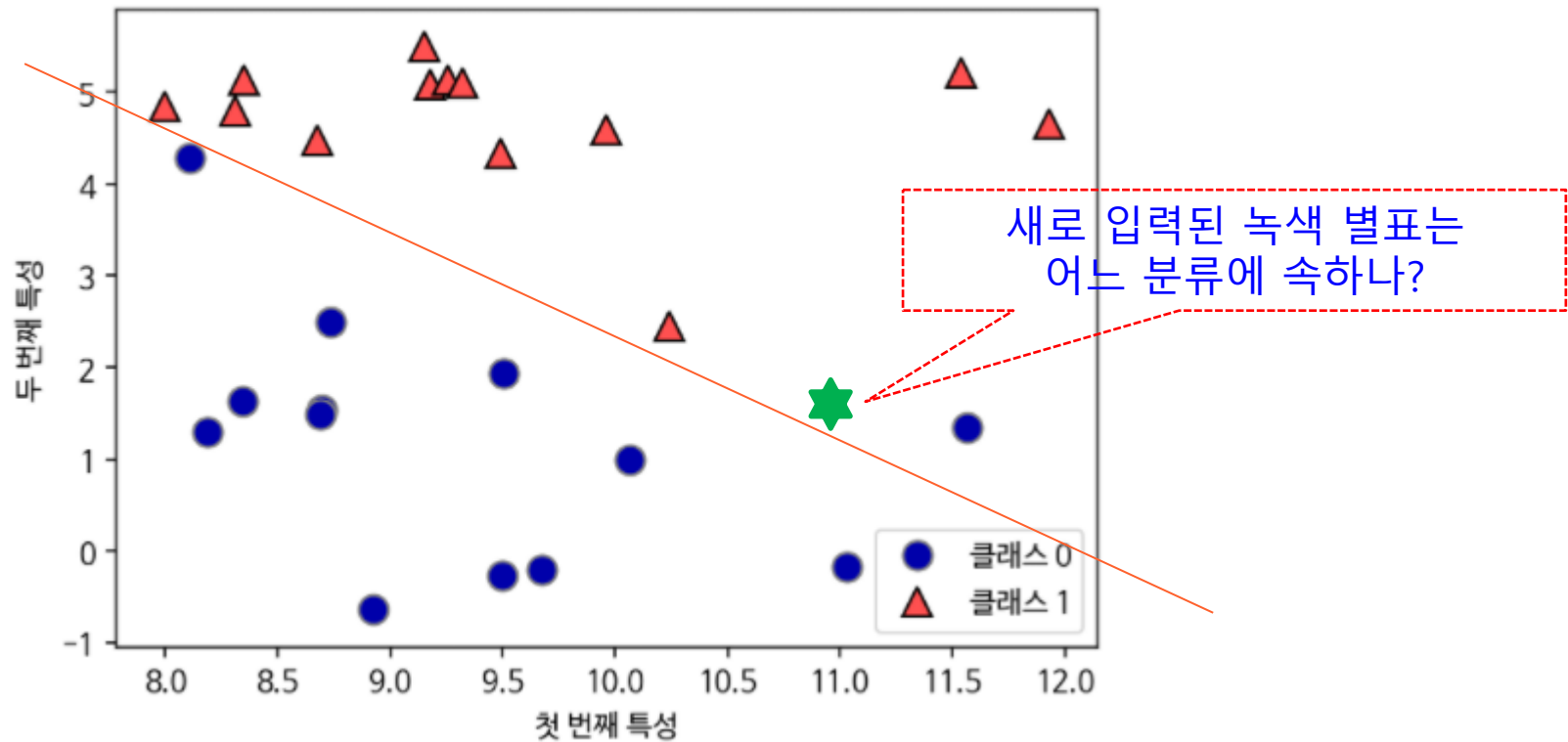
왼쪽 예제는 독립 변수의 개수가 1개인 경우임

$$y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_1^2 + \dots + \beta_n \times x_1^n$$

다항 회귀 알고리즘 수행을 통해서,
학습 데이터를 가장 잘 표현 하는
 $\beta_0, \beta_1, \dots, \beta_n$ 값을 찾아냄

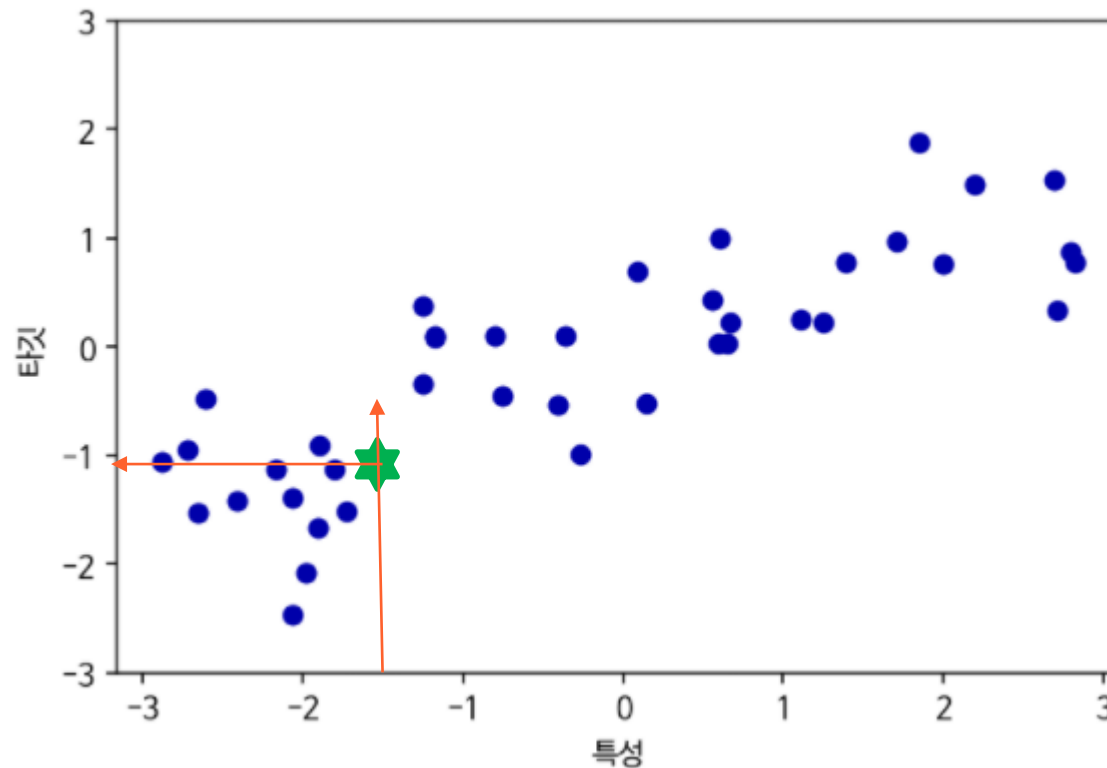
분류 문제

- 특성이 2개이며, # of classes=2인 자료



회귀 문제

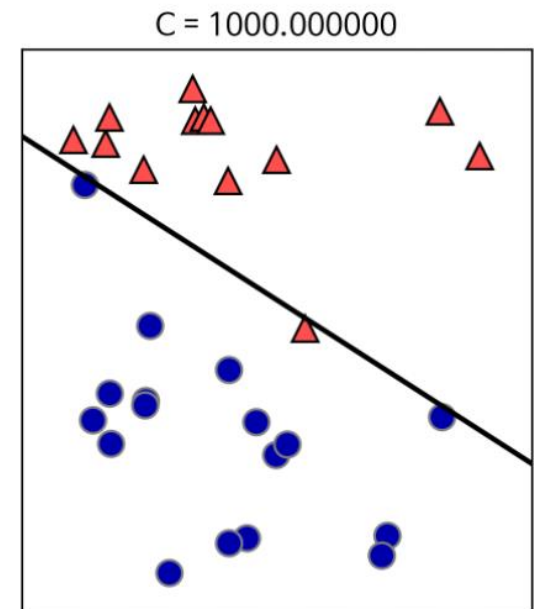
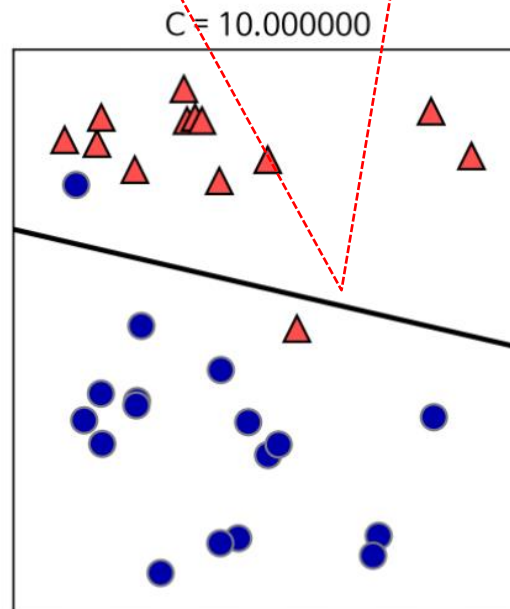
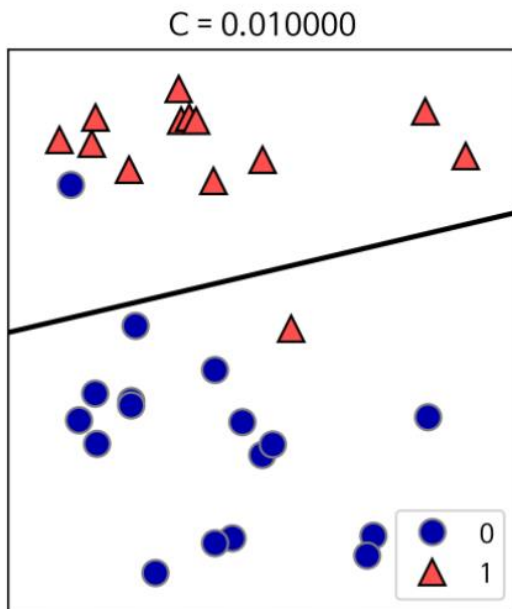
- 특성이 1개(X축)이며, 타겟(label)이 연속된 값으로 Y축으로 표시



이진 분류

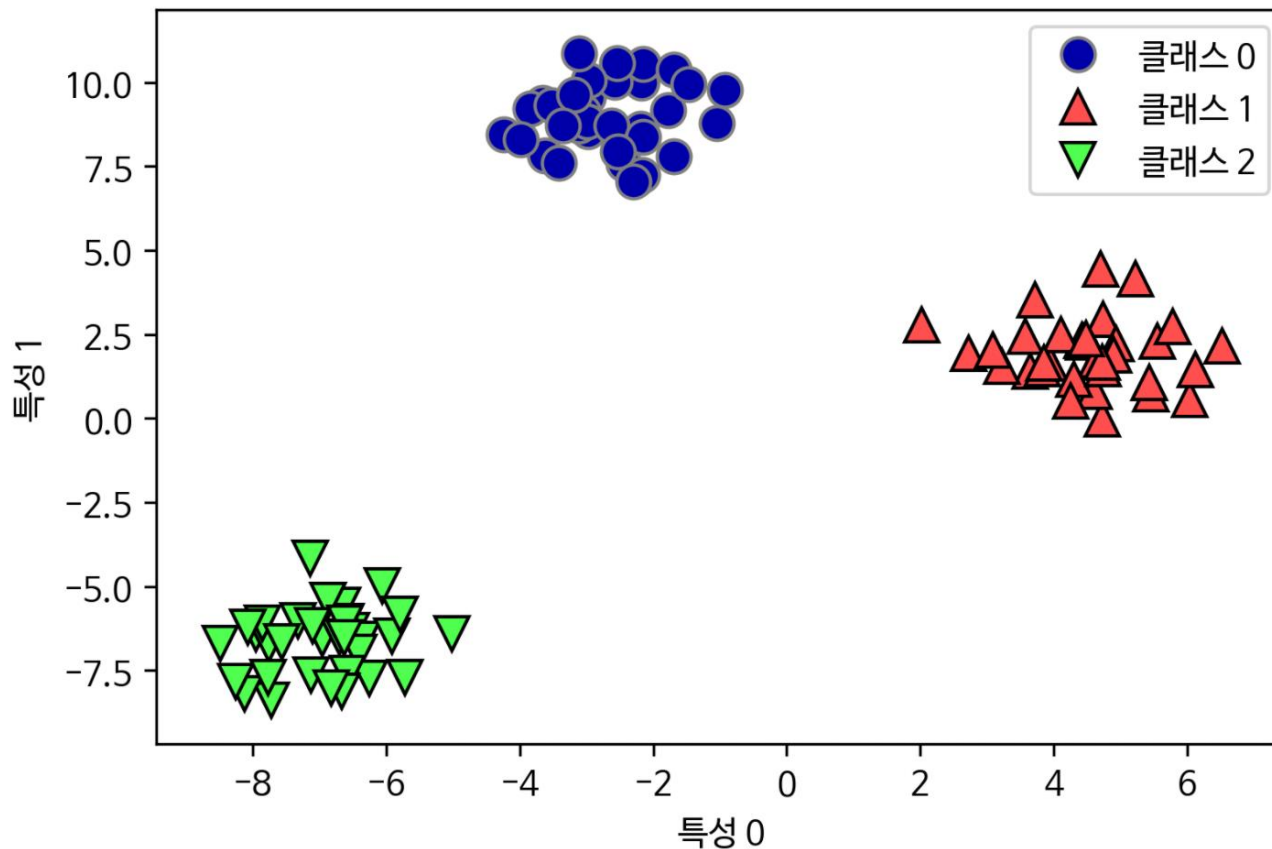
- 직선으로 분류하는 방법

테스트 데이터를 위해 이것이 더 좋을 듯



다중 클래스 분류

- 삼중 클래스 분류
 - 이진 분류를 3 번 이용



삼중 클래스 분류

- 직선 3개로 분류
 - 영역으로 구분

