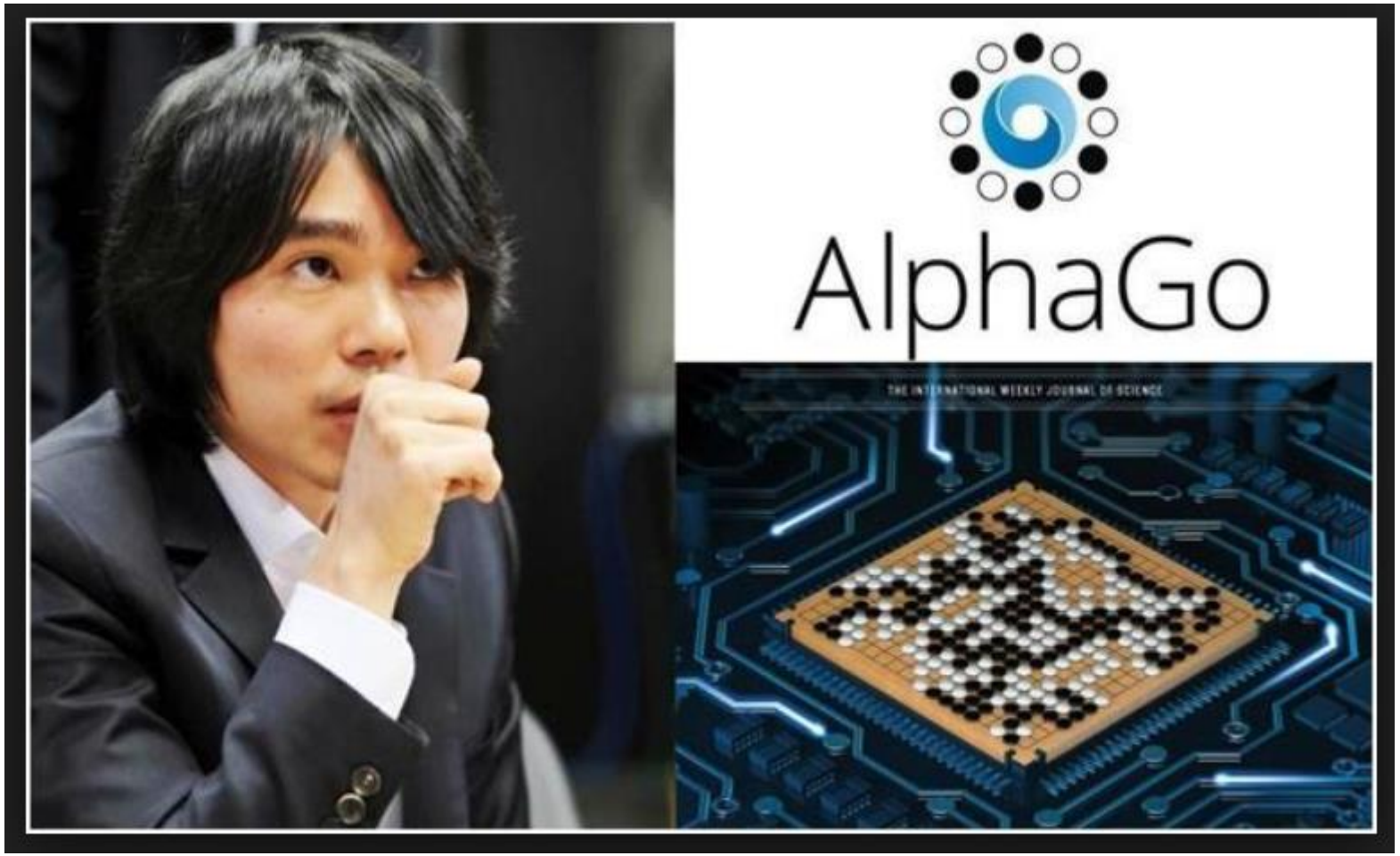
Lecture 01

Machine Learning Basics

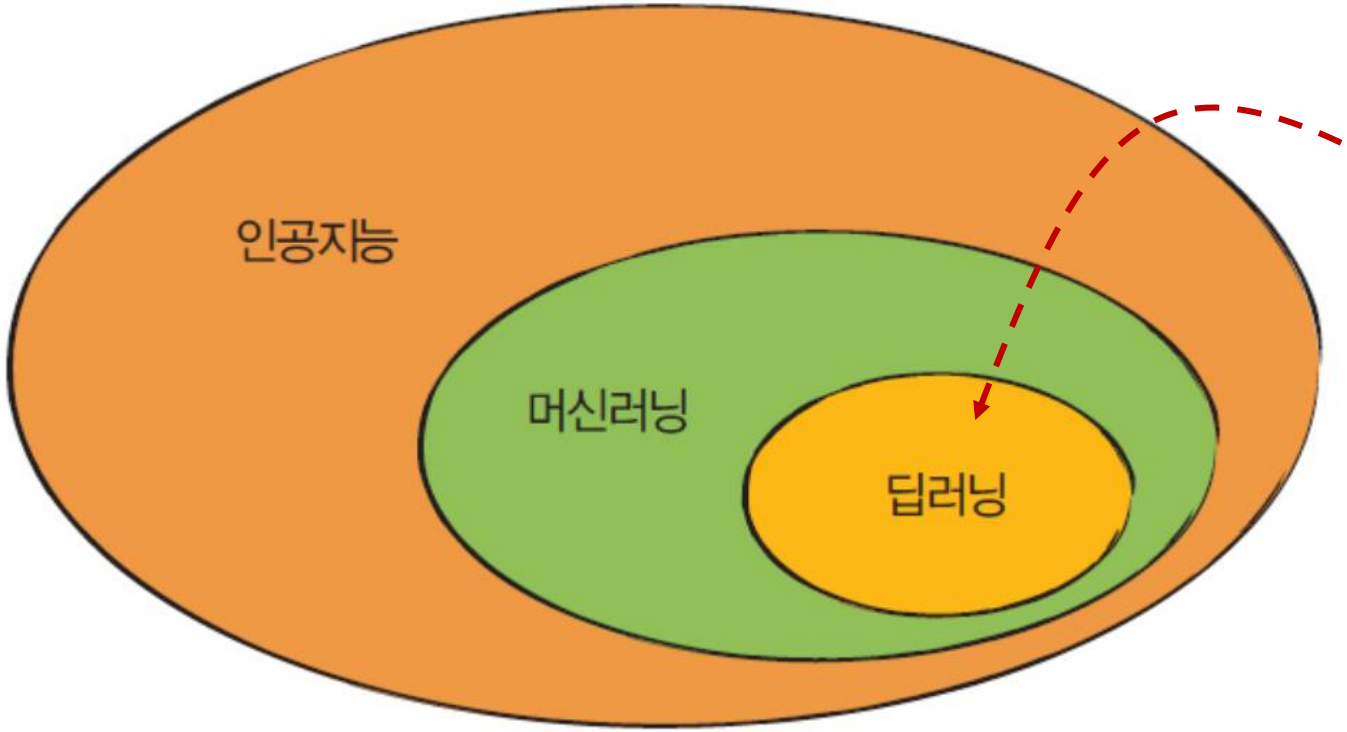
Basic concepts

- What is ML?
- What is learning?
  - supervised
  - unsupervised
- What is regression?
- What is classification?

Artificial intelligence



Artificial intelligence / Machine learning / Deep learning



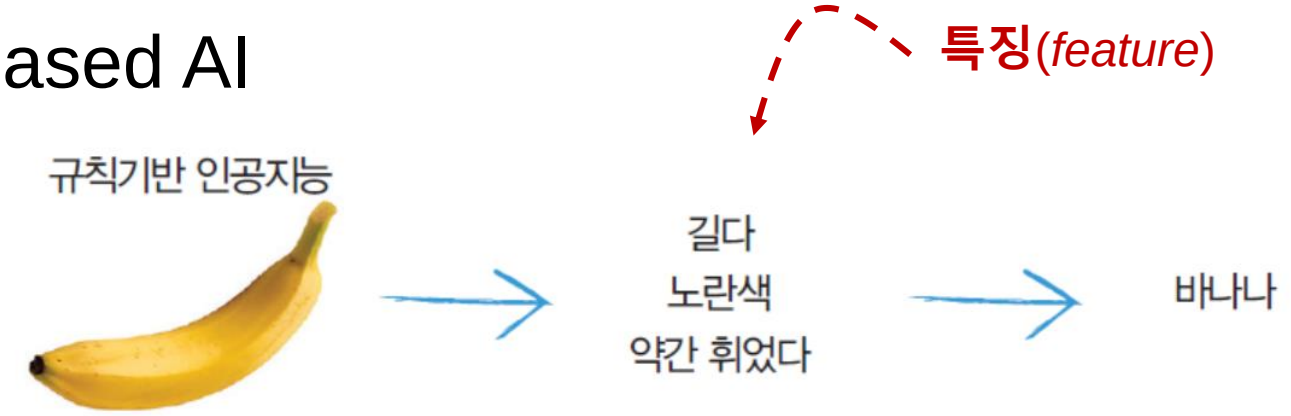
인공신경망(*artificial neural network*)

Machine Learning

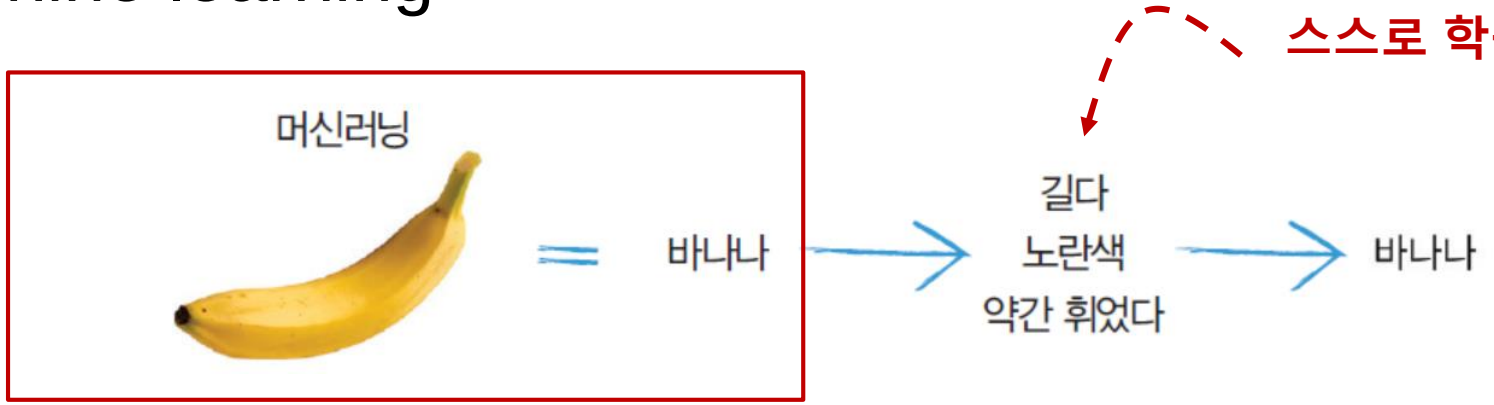
- Limitations of explicit programming
  - **Spam filter** : many rules
  - **Automatic driving** : too many rules
- Machine learning : "Field of study that gives computers the ability to learn without being explicitly programmed" Arthur Samuel (1959)

Rule-based AI / Machine learning

- Rule-based AI

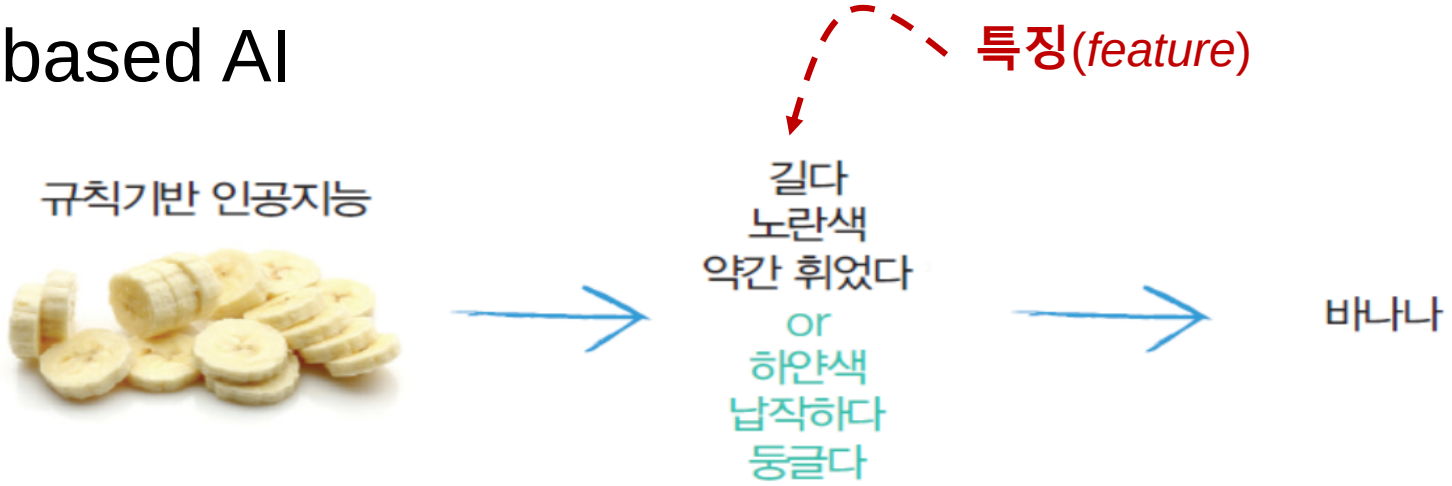


- Machine learning

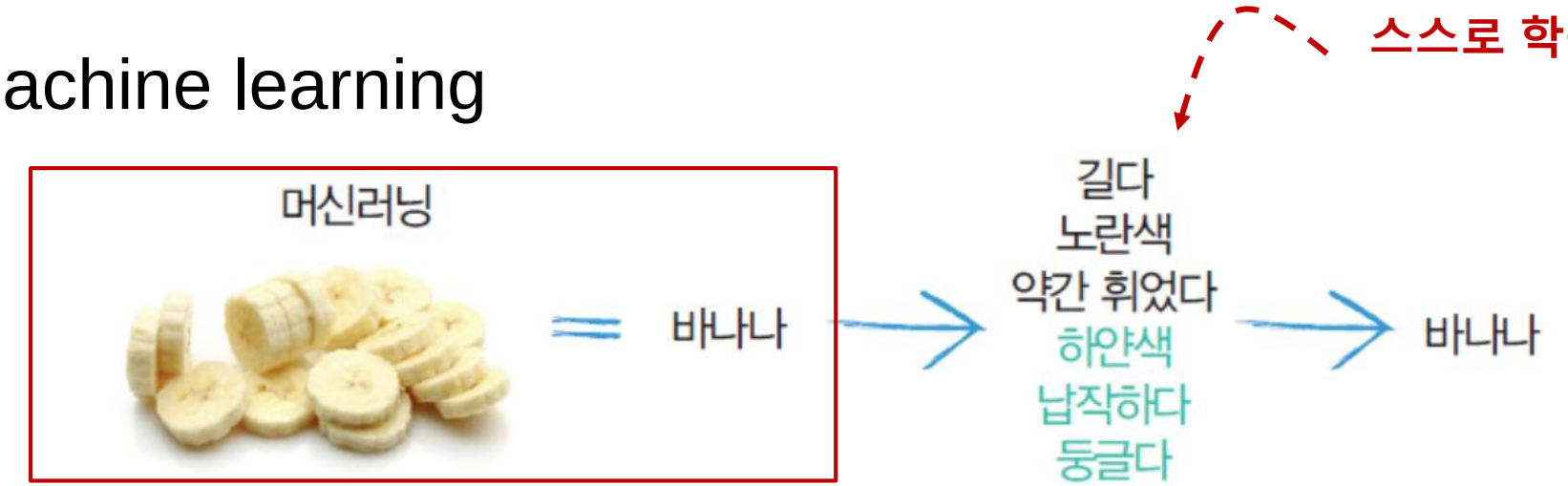


Rule-based AI / Machine learning

- Rule-based AI

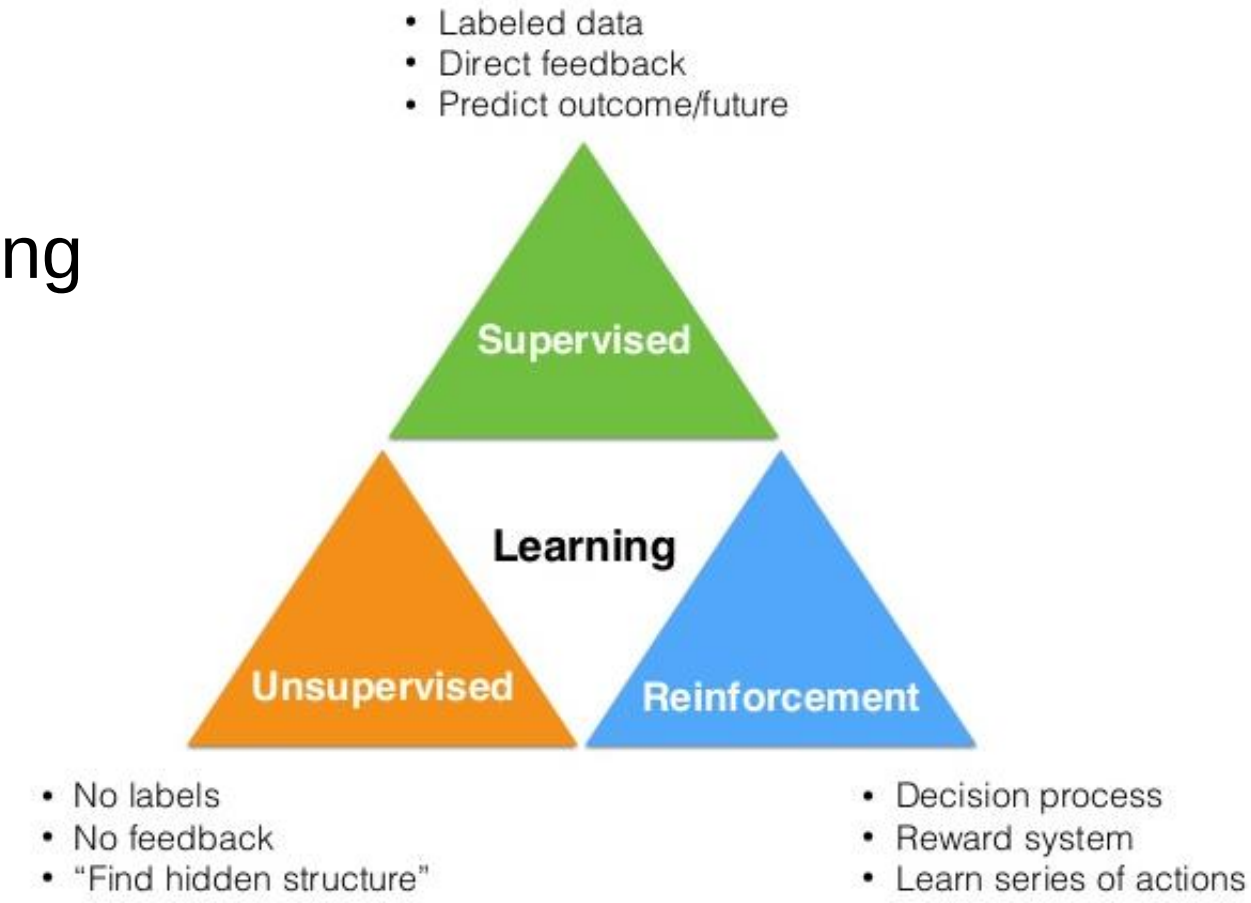


- Machine learning



Machine learning

- Supervised learning
- Unsupervised learning



Supervised learning

- Supervised learning (지도 학습)
  - learning with labeled examples

data , label

1

learning

training set

2

label이 지정되지 않은 **test set**을 이용해서 학습된 알고리즘이 얼마나 정확히 prediction하는지를 측정

Supervised learning

- Supervised learning (지도 학습)
  - 예측하는 결과값이 continuous value(연속값)이면 regression(회귀) 문제

    ex) 3개월뒤 이 아파트 가격은 2억1천만원 일 것인가? 2억2천만원 일 것인가?
  - 예측하는 결과값이 discrete value(이산값)이면 classification(분류) 문제

    ex) 예측값이 1인가? 2인가?
  - Methodology
    - ✓ Convolutional Neural Network(CNNs)
    - ✓ Recurrent Neural Networks(RNNs)

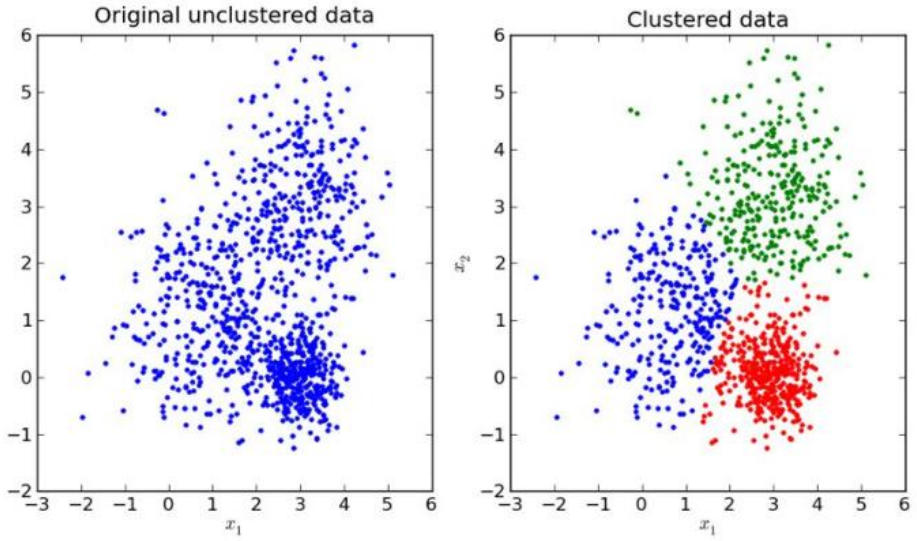
Supervised learning

- An example training set for four visual categories.

http://cs231n.github.io/classification/

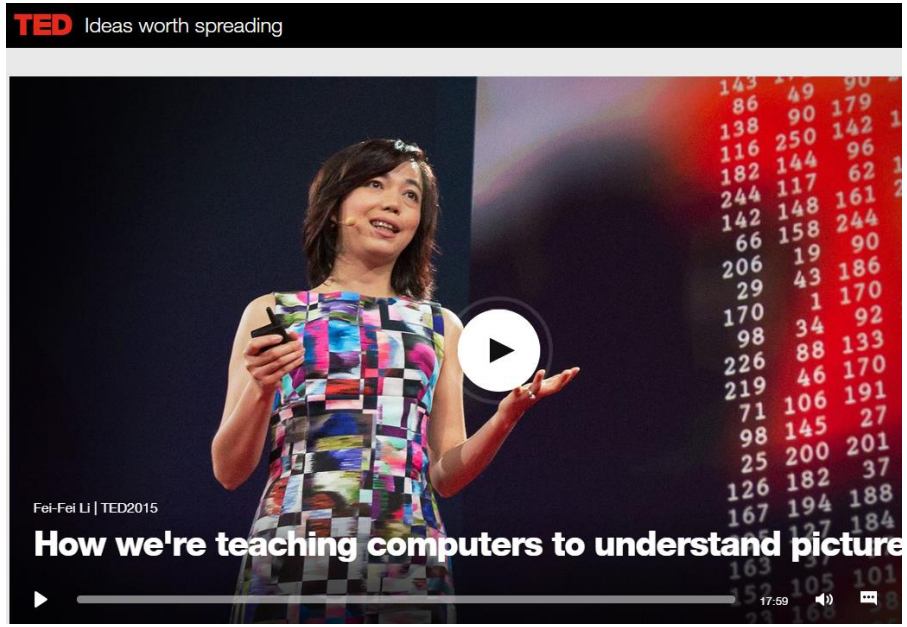
Unsupervised learning

- Unsupervised learning : un-labeled data (비지도 학습)
  - Data 형태로 학습을 진행하는 방법
  - 데이터가 무작위로 분포되어 있을 때 비슷한 특성을 가진 부류로 Clustering
  - Google news grouping
  - Word clustering
  - Methodology
    - ✓ Autoencoders

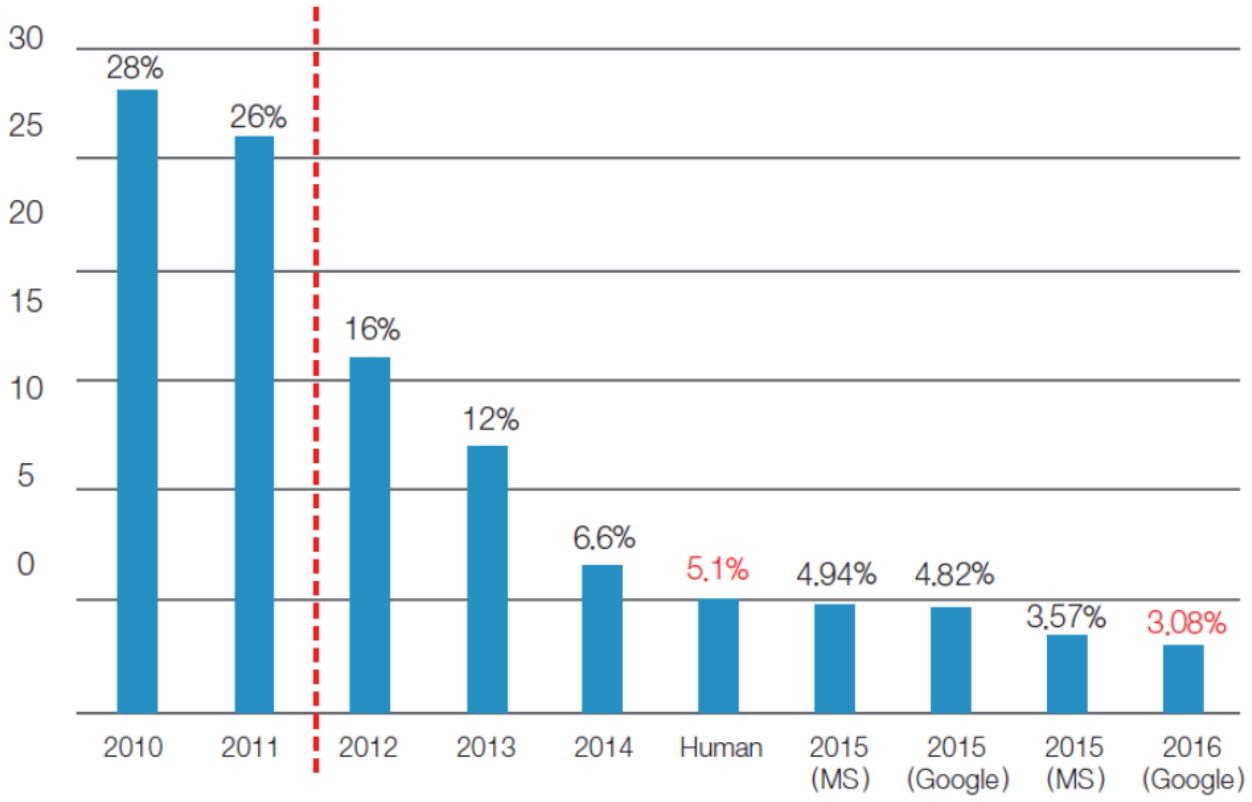


Fei-Fei Li (Computer scientist)

- https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?language=ko



ILSVRC ImageNet Large Scale Visual Recognition Competition



ILSVRC recognition error rate

Why is machine learning evolving?

- Big data
- GPU evolution
- Invention of Deep Learning Algorithm

Supervised learning

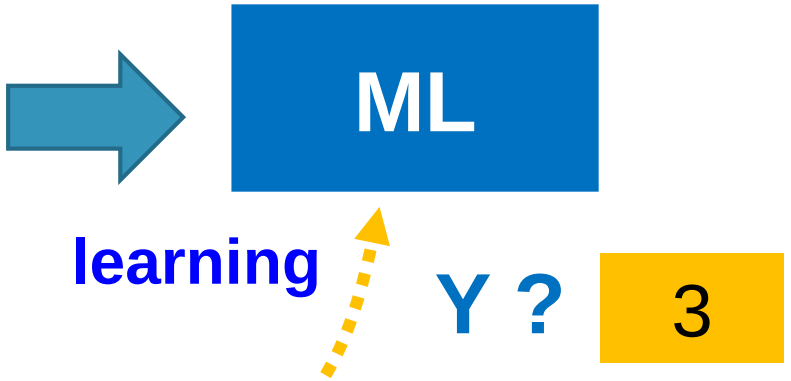
- Most common problem type in ML
  - **Image labeling: learning** from **tagged images**
  - **Email spam filter**: **learning** from **labeled (spam or ham) email**
  - **Predicting exam score**: **learning** from **previous exam score and time spent**

Training data set

Ex

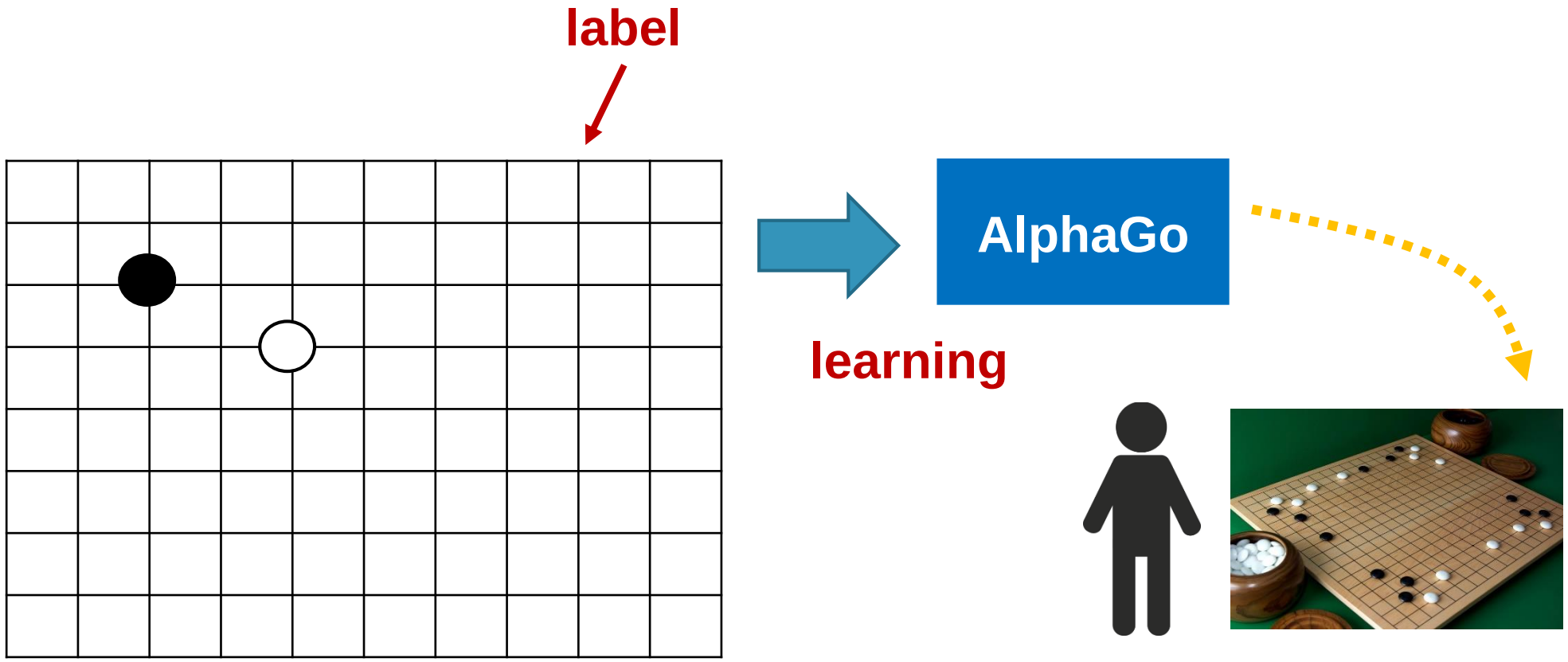
label

X	Y
3, 6, 9	3
2, 5, 7	2
2, 3, 5	1



$X_{test} = [9,3,6]$

Training data set



Types of supervised learning

- Predicting final exam score based on time spent
  - **regression**
- Pass/non-pass based on time spent
  - **binary classification**
- Letter grade (A, B, C, D and F) based on time spent
  - **multi-label classification**

Predicting final exam score based on time spent

x (hours)	y (score)
10	90
9	80
3	50
2	30

regression

if) x =7

y ? 75

Pass/non-pass based on time spent

x (hours)	y (pass/fail)
10	P
9	P
3	F
2	F

binary classification

Letter grade (A, B, …) based on time spent

x (hours)	y (grade)
10	A
9	B
3	D
2	F

multi-label classification