

• 03/24 기출문제(2)

* $S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ ✓ ✓ *

S_{xy} 공분산

→ 코드를 이용해서 살펴보아요 ~

* 생각해볼 문제 정의 수학
 $(\text{점수}) \times (\text{점수})$

KOSPI 상상관계
 $(\text{주가}) \times (\text{주가})$

신장 수학
 $(\text{키}) \times (\text{몸무게})$

$\left[\frac{\text{cm}}{\text{m}} \right] \quad \frac{\text{kg}}{\text{mm}}$

(피어슨)
상관계수 (correlation Coefficient)

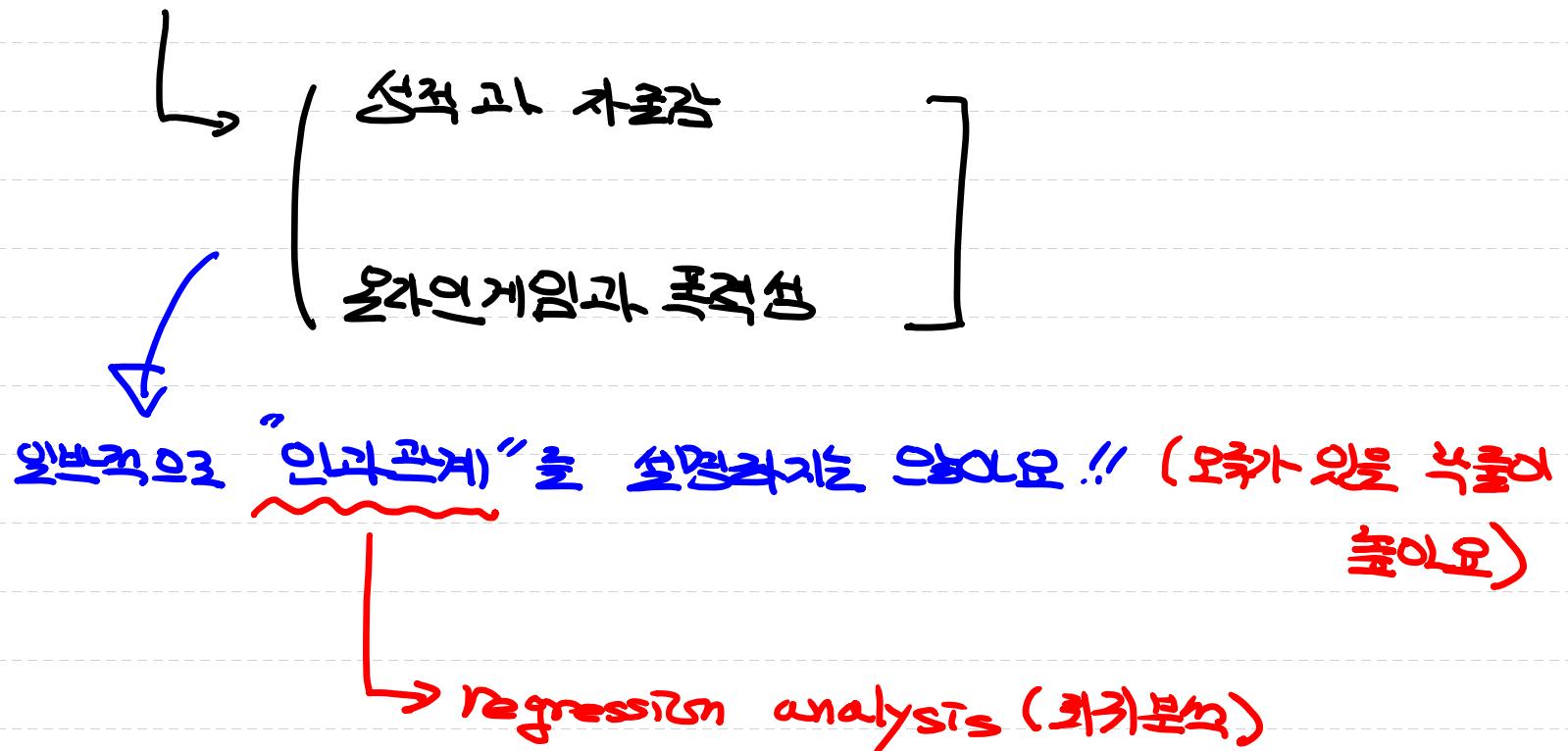
계산하기 $-1 \sim 1$ 사이
 $= -1 \sim 0 \sim 1$
= 음의 상관관계
= 무상관
= 양의 상관관계

* $r_{xy} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\text{std}(x)} \right) \times \left(\frac{y_i - \bar{y}}{\text{std}(y)} \right)$

r_{xy}

- 두 대상 (2번3, 2천원)이 서로 연관성이 있다고 추측되는 관계

⇒ 상관관계 (correlation)



→ 2차원 데이터의 시각화

⇒ Scatter (스ATTER)

"기초통계"
"추가통계"

① Machine Learning

AI (artificial Intelligence) → 인공지능

→ 사람의 사고능력을 구현한 프로그램(시뮬)

AI [Strong AI (사람과 구별할 수 있는 정도의 강한 AI)]
[Weak AI (특정 영역에서
작동하는 AI)]

이걸 만드는 데는
문제가 있어요!!

↳ 자율주행, 첫봇, 감식 ✓

* 사람 → 생체학적 방식으로 사고
* Strong AI → 전자기적 방식으로 사고 (100만배 빨라요)

→ 얼마나 사고의 속도가
빠를거야??

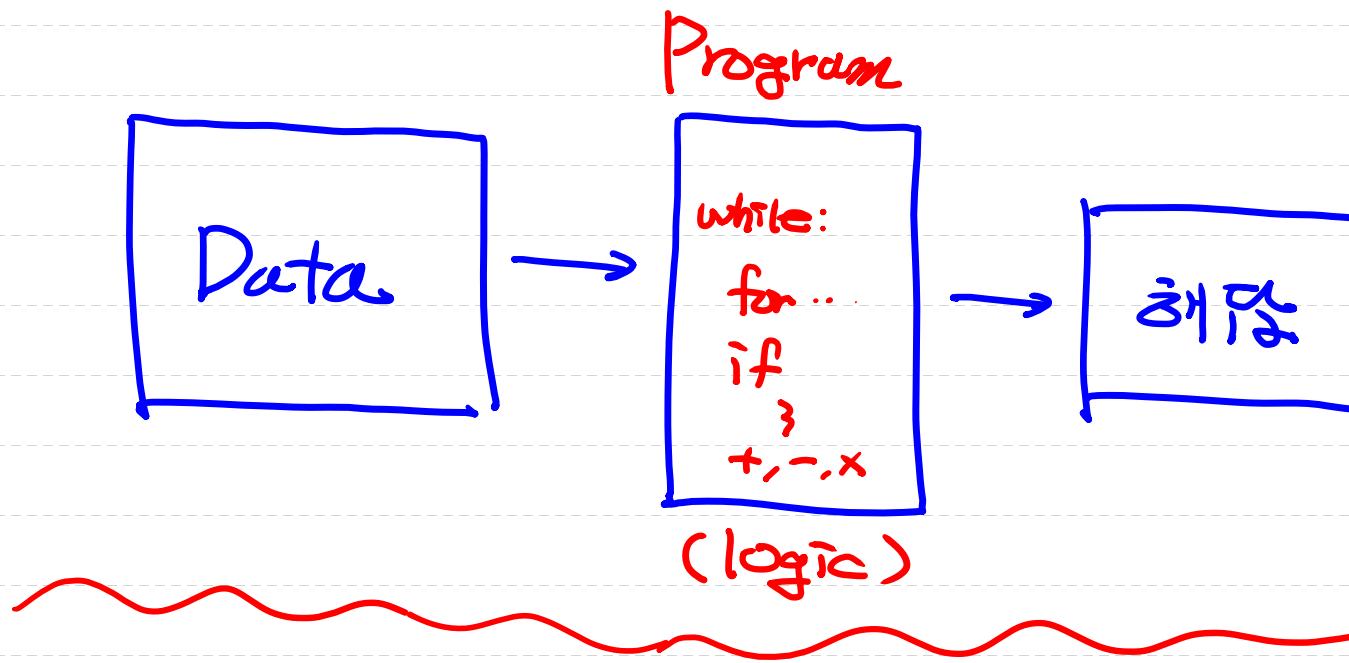
[MIT AI Lab] → 2만년 동안
할 일을
Strong AI → 1주일

- 기초 프로그램과 machine Learning을 비교.

→ 사진에 있는 품목이 개 몇개고 고양이 몇개요?

- 사람들은 이 문제를 아주 쉽게 해결 가능!
- 하지만 기초 프로그램 방식으로는 이 문제를 해결하는게 쉽지 않아요!
(Rule based programming)

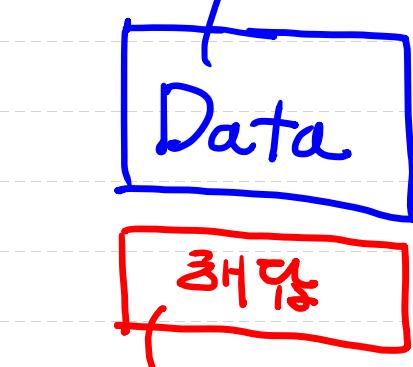
② Rule based Programming



Machine Learning (지도학습)

Machine Learning (지도학습)

사진 Data



1번사진 고양이
2번사진 명령이

Program

1번사진 : 고양이
2번사진 : 명령이
3번사진 : 고양이
4번사진 : 명령이
⋮

(machine
Learning)

만들어야
해요!!!
(model)
수식!!



개(65%)
고양이(35%)

새로운 사진



구르침의 흐름에 만족하지 않아요
점점 모델이 (구식이, 구르침) 외설이 되가요!

"Learning" (학습)

- 사용도 Learning 과정을 거쳐요!

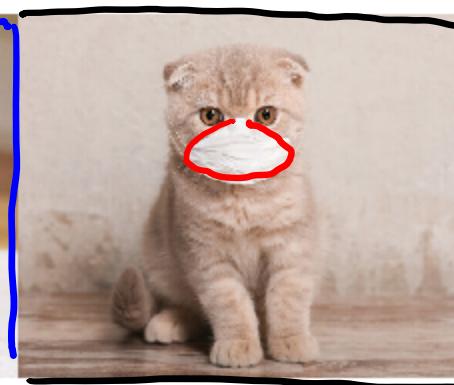
*



✓ 고양이 1



고양이 2



*

Predict

(예측)



Learning이 뭘 뜻하면?

- ★ [① 데이터가 많이야 해요 !
② 데이터가 양질의 데이터어야 해요 !]



• AI : 가장 표준적인 개념. 인지이 가지는 학습능력, 추론능력을 컴퓨터로 구현.

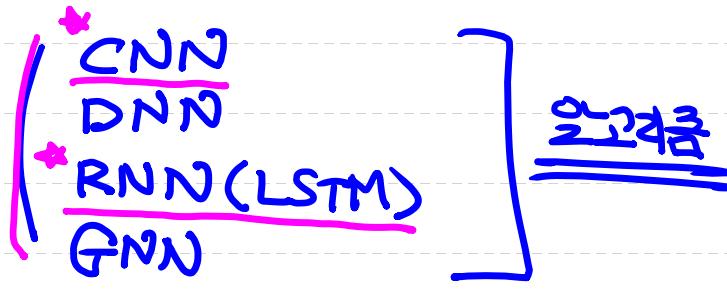
① machine Learning : AI (weak AI)를 구현하기 위한 하나의 분야로.
데이터를 기반으로 특성과 패턴을 파악한 후 그 결과(model)를
이용해 디지털 데이터에 대한 추천과 예측하는 프로그램

기법(手法)

- * ● Regression (회귀)
- SVM (Support Vector Machine)
- Decision Tree, Random Forest
- Naive Bayes
- KNN (k-Nearest Neighbor)

- * ● Neural Network (신경망)
- Clustering (k-Means, DBScan)
- Reinforcement Learning
(강화 학습)

- Deep Learning: Machine Learning의 한 분야인 "Neural Network" (신경망)을 이용해 학습하는 구조의 알고리즘을 지칭



* Deep Learning (Neural Network)을
제외한 Machine Learning 알고리즘
⇒ 정밀도가 높아 처리가 조금 더複雜, 시간이 더 걸려요

Deep Learning (Neural Network)

⇒ 비정형데이터 처리가 좋았고, 시간이 오래 걸려요!
(이미지, 소리, 대용량 Text, etc)

↳ "으쌰"가 고전된 데이터 처리가 좋았고.

- Machine Learning을 왜 필요할까요?

→ Rule based programming이 기본
↓

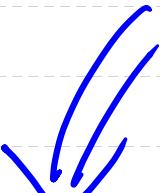
Explicit programming



* 조건에 따라
~~통계학적~~ 프로그램

그러나 우리가 해결해야 하는 문제들이 조건이 너무 많아서
그 조건을 일일히 다 구현하기 못하는 경우가 있어요!

→ 자율주행, spam filter, 바둑
대~한
대~한



이 문제를 해결하기 위해 나온 개념이 Machine Learning
"1959"에서 사무엘

① Machine Learning Type.

[보류]



① 지도학습 (Supervised Learning)

우리가 알고 있는 것에
초점이 맞춰져 있어요~

② 비지도학습 (Unsupervised Learning)

③ 준지도학습 (Semi-supervised Learning)

별도 ④ ④ 강화학습 (Reinforcement Learning)

① 지도학습 (Supervised Learning)

정답으로 들어가는 값이

Data (입력값) : x

✓ Training
Data Set
(학습 데이터셋)

예) 공부시간에 따른 시험점수
Regression

3 5	10 15	8 9	80 95
7시간	→	75	

Data의 Label (해답, 정답) : t

- 궁부시그니 대회
총적예측

시각(x)	총적예부(t)
1	F
3	F
4	F
6	F
8	T
9	T
11	T

↑
입력값(x) Label(결과)
↑
 t

지도학습을

어떤 종류의 미래값을 예측하나?

[Regression (회귀) → 예측값이 (continuous value) 연속적인
수치값으로 나오는 경우!
Classification (분류) → 예측값이 어느 분류에 속하는가로 나오는 경우!]

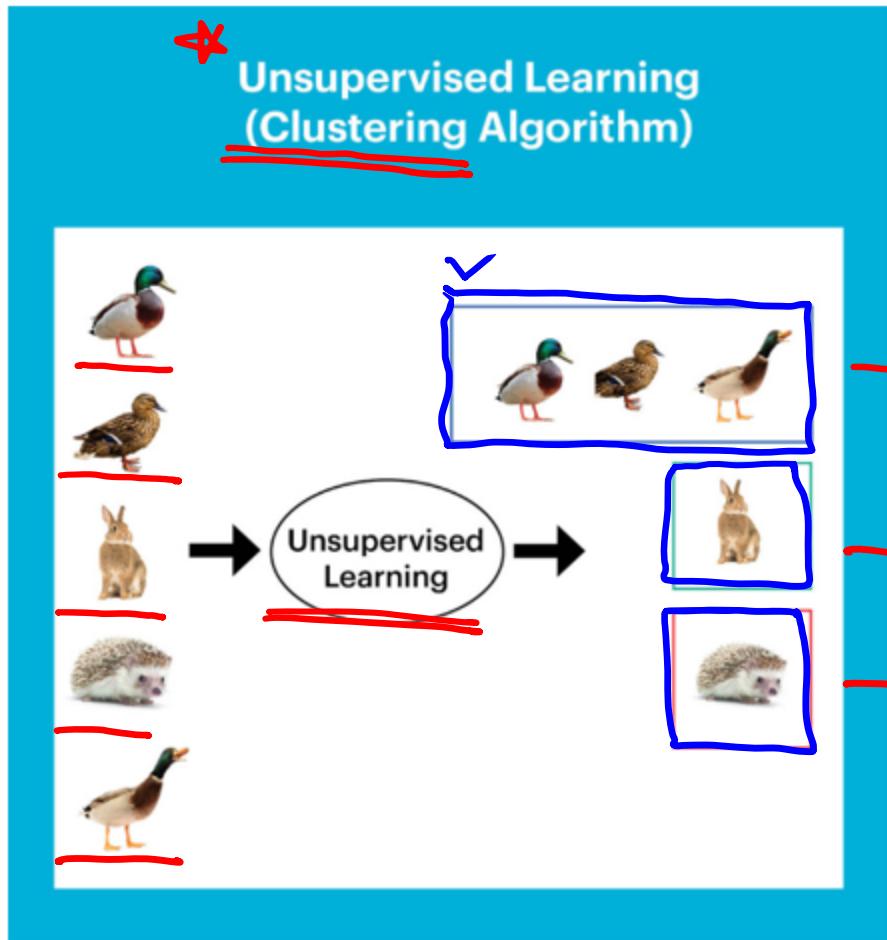
Classification

Model

기준선

T, F
(45%) (55%)

• 비지도 학습 (Unsupervised Learning) → 입력에 Label이 돌아가지
않아요 !!



"Classification" 표시하지 않아요

지도학습

Clustering (군집학)



이 쪽은 허가하는 대표적인
알고리즘

⇒ k-Means, DBScan

- 조지도 학습 (Semi supervised Learning)

[비지도학습 + 지도학습]

양적(x)	Label(t)
"	O
"	O
"	X
"	O
"	X

⇒ "google photos"

↳ cluster



- 강화학습 (Reinforcement Learning) → Action, Reward 개념을 이용해 최상의 policy(정책) 찾기

◦ 머신러닝 process

*

① 문제파악 : Domain 보고

*

② EDA (탐색적 데이터 분석) : 데이터 구조, 내부 데이터 및 외부 데이터 활용

여부

결측치, 이상치 찾으!

보호자료, 빠진값의 속성 분석

→ *

③ 모델링의 pre-processing

④ 모델 학습 →

- ① Python 구현 (주로 python)
- ② Tensorflow 구현
- ③ scikit-learn

⑤ Prediction