

# Python 프로그래밍 - I

2023년도 1학기

조상구



# 과목명 : Python 프로그래밍- I

- Python의 기본적인 사용 방법을 익히고 당면 과제의 해결 방안을 찾아가는 방법을 실습해 본다.
- Python 기초 익히기
  - 코딩(Coding)이란 무엇인가?
  - 프로그래밍 기본
  - 다르게 생각하기
  - 문제해결 과정
- 데이터 다루기
  - 행렬 구조와 연산(Numpy)
  - 데이터 세트(Pandas)
  - 데이터 시각화(Data Visualization)
- 간단한 머신러닝(Machine Learning) 실행해 보기
  - 데이터 세트 준비
  - 머신러닝 모델 적용해 보기

# 강의 계획

주차	요일	주제	강의 내용
1	3.09(목)	Introduction, First Steps	인공지능 소개, 파이썬 기초 문법
2	3.16(목)	Basics, Operators and Expressions	
3	3.23(목)	Control Flow - if	
4	3.30(목)	Control Flow - while	
5	4.06(목)	Control Flow - for	
6	4.13(목)	Functions, Modules	
7	4.20(목)	Data Structures - lists, tuples, strings	
8	5.04(목)	[Mid-term Exam.]	
9	5.11(목)	Data Structures - dictionaries, sets, sequences	
10	5.18(목)	Input and Output, Standard Library	
11	5.25(목)	NumPy	
12	6.01(목)	NumPy	
13	6.08(목)	pandas	
14	6.15(목)	pandas	
15	6.22(목)	[Final Exam.]	

# 수업 내용

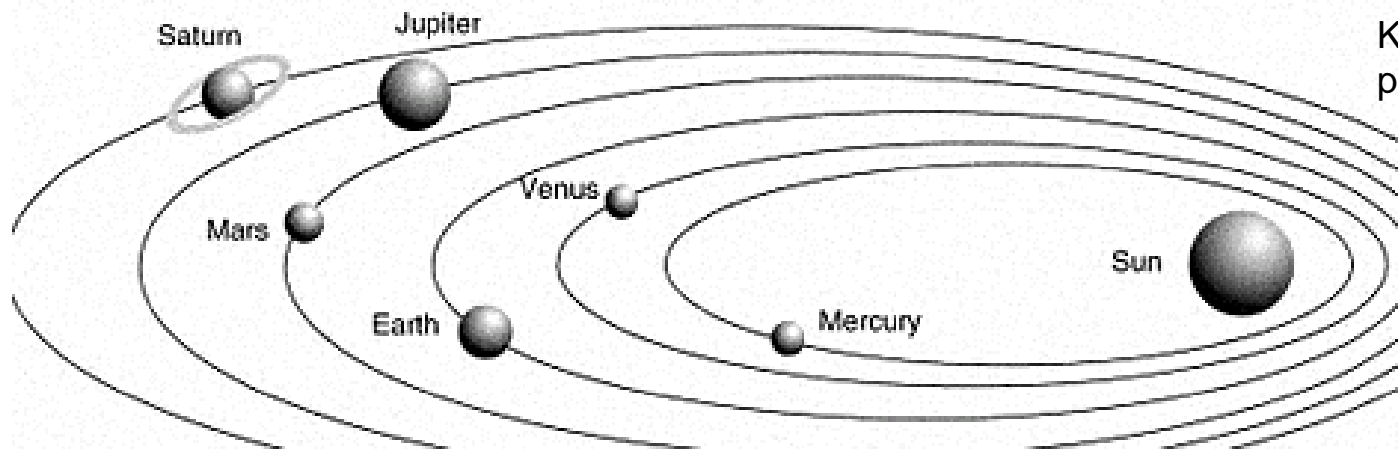


## • 주요 수업 내용

- MS Excel의 기본 기능을 익히고 데이터 처리와 간단한 시각화 및 매크로/VBA를 실습한다.
  - 데이터 분석에 활용하는 주요 기능인 엑셀표, 필터, 피벗, 차트 기능 학습
  - 엑셀의 다양한 함수를 이용하여 데이터 변환하는 방법을 실습
  - 파생변수를 만드는 방법을 실습
  - 자동화를 위한 매크로 작성과 실습
- 엑셀과 같이 제공되는 새로운 기술인 Power Query, Power BI Desktop을 활용 이해

# 패턴인식(Pattern recognition) 1

패턴인식(Pattern recognition)은 데이터에 존재하는 규칙성(regularities)을 발견하여 새로운(unseen) 데이터를 예측을 하는 것을 목적으로 한다.



Kepler's empirical laws of planetary motion



## Rule 1, Rule 2, Rule 3

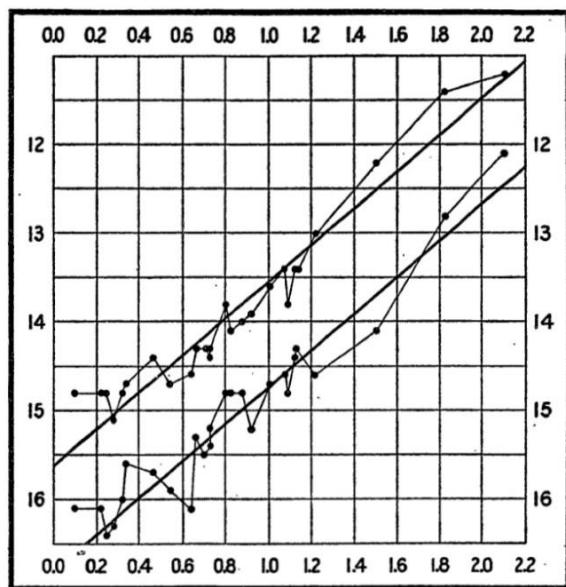
- 케플러 제1법칙(타원궤도 법칙),
- 케플러 제2법칙(면적속도 일정의 법칙),
- 케플러 제3법칙(조화의 법칙)

## 패턴인식(Pattern recognition) 2

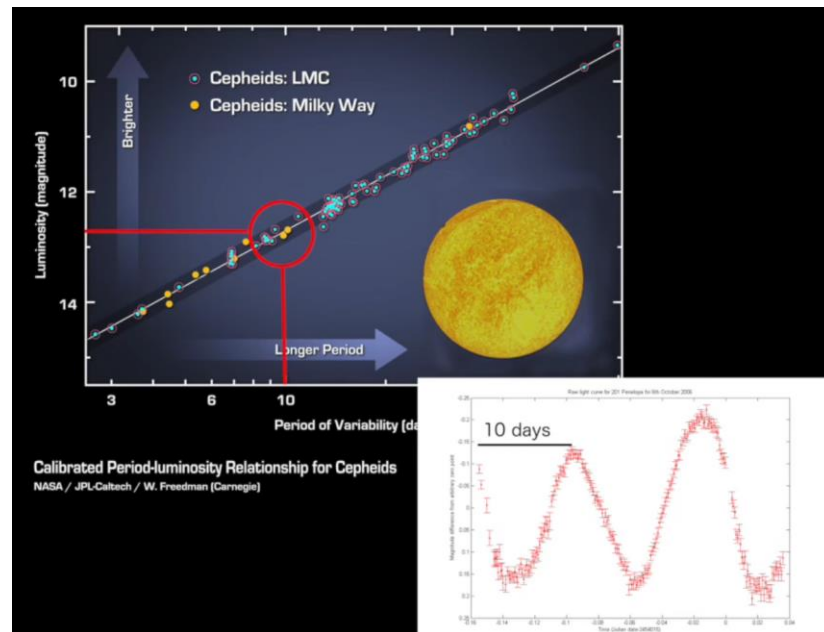
우주공간의 은하계의 특정한 세페이드 변광성을 기준으로 빛의 파동주기와 거리 간의 패턴, 상관 관계를 통해 우주공간의 모든 별까지의 거리를 발견

\* 세페이드 변광성은 직경과 온도가 모두 변하고 안정적인 주기와 진폭으로 밝기 변화를 생성하는 방사형으로 맥동하는 별

Brightness(광도)



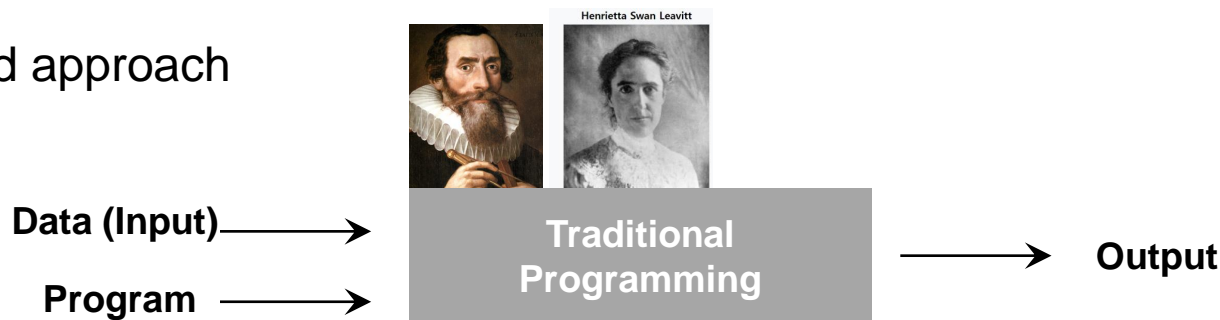
Period(주기)



# 머신러닝

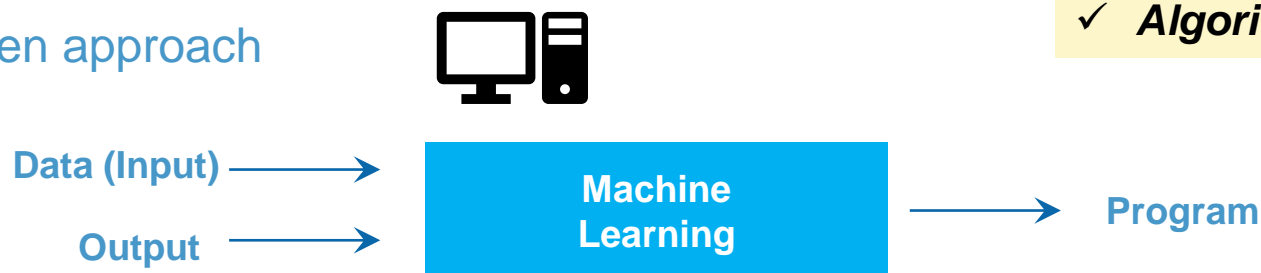
컴퓨터에게 데이터에 존재하는 규칙성(regularities)을 발견하게 하여 새로운 데이터의 예측을 컴퓨터가 스스로 하게 하는 것 (사람이 규칙을 찾아서 문제를 해결하는 방법은 이제는 너무 어려움)

## ■ Rule based approach



- ✓ **Big data**
- ✓ **High dimensionality**
- ✓ **Computing power**
- ✓ **Algorithms**

## ■ Data driven approach



# AI와 빅데이터 핵심역량

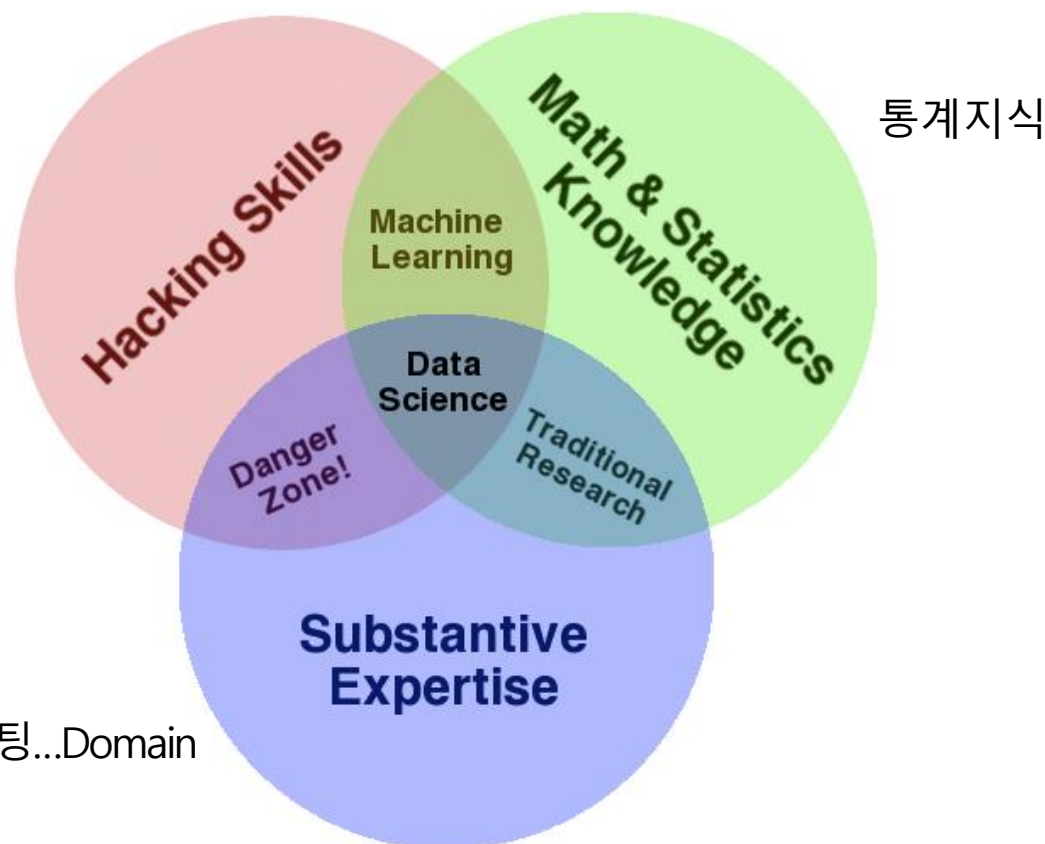
프로그래밍 코딩능력(Hacking Skills), 통계 추론 및 기초 수학 능력(Math & Statistics), 산업분야 Domain 전문 지식 등을 모두 갖추어야 데이터 사이언스가 완성

개발지식

- Python, Java
- R, Julia

전문지식

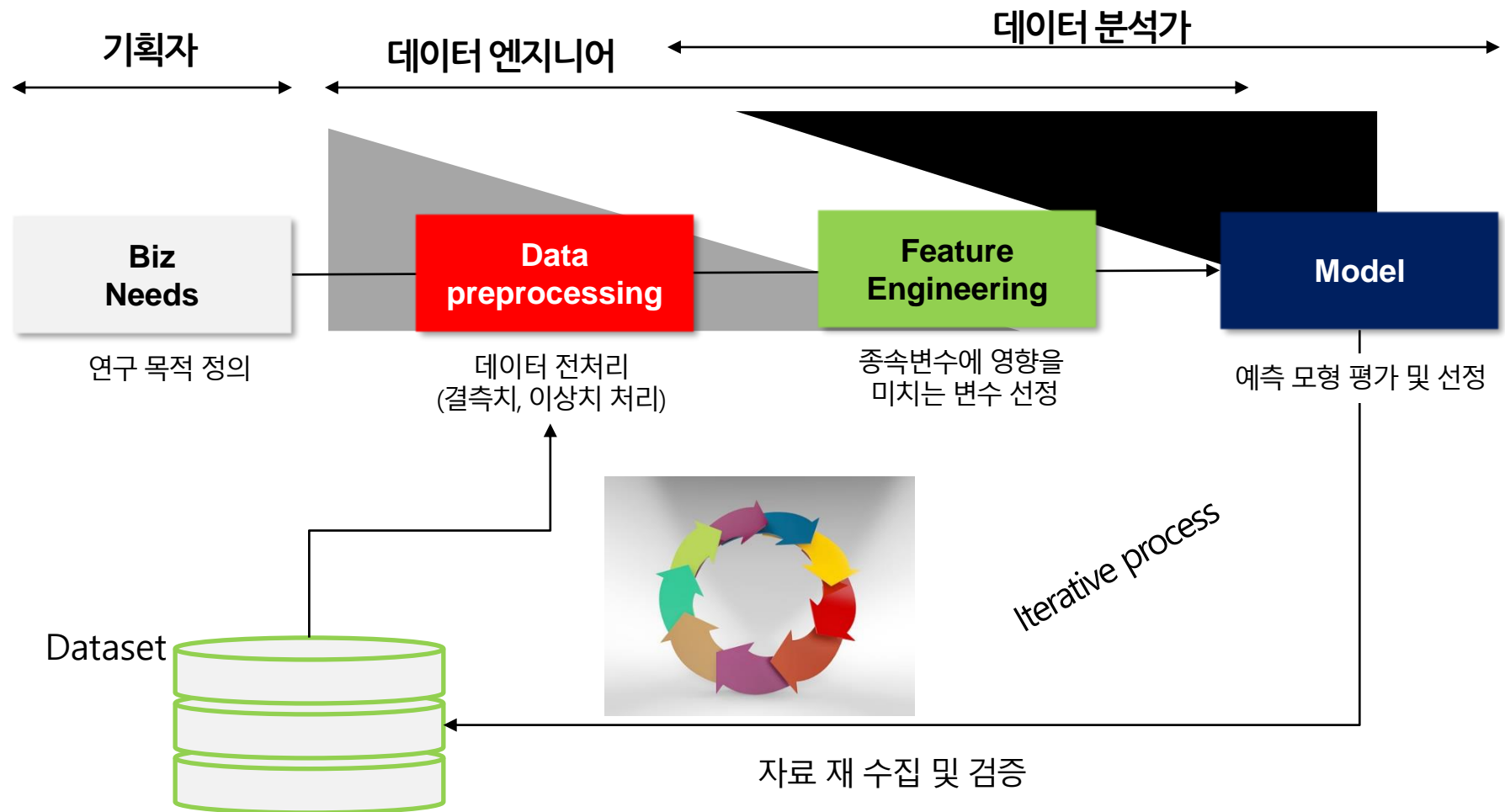
- 제조, 마케팅...Domain



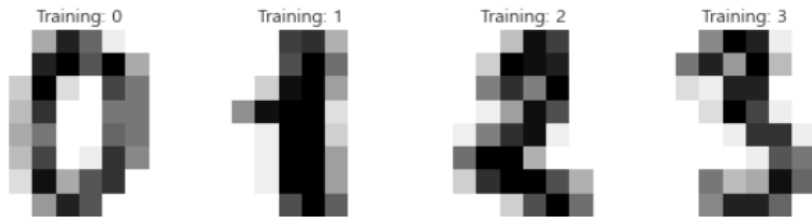
통계지식



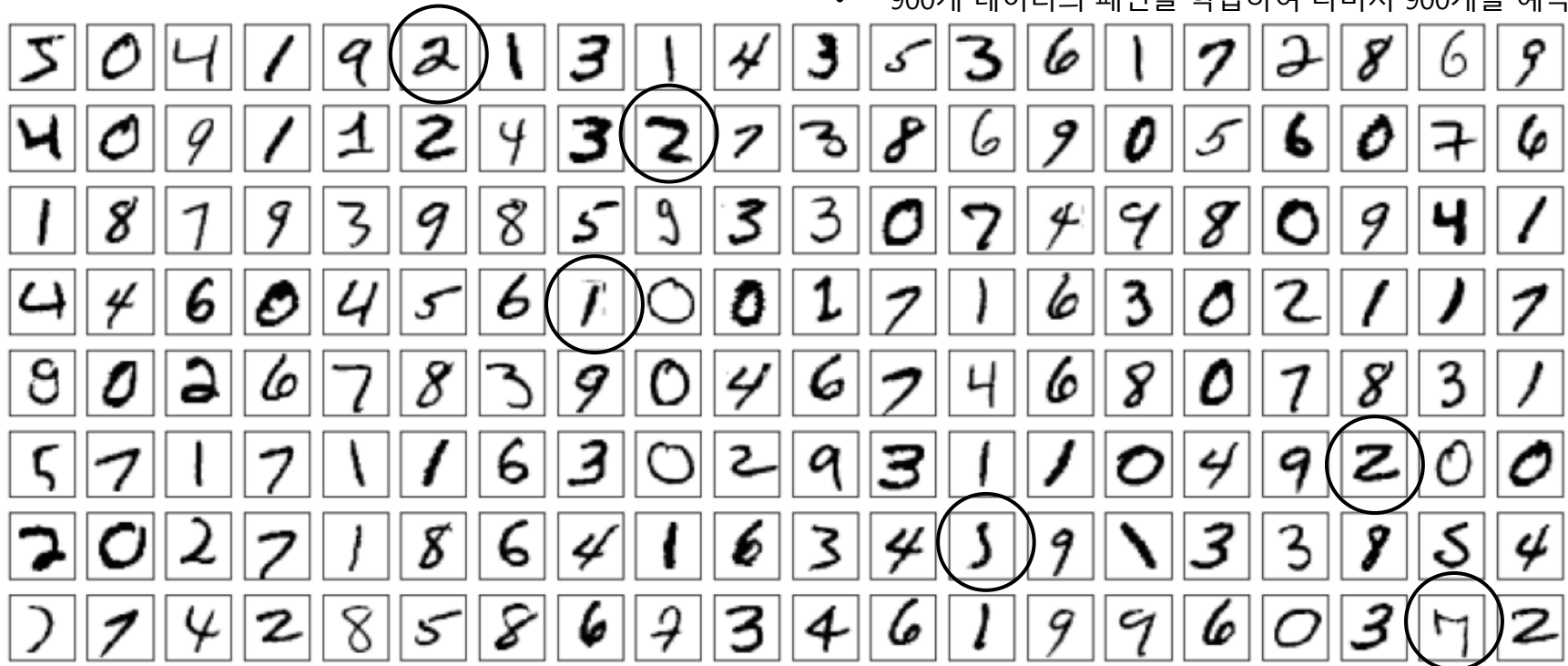
# 머신러닝 수행 절차



# 패턴인식\_MNIST

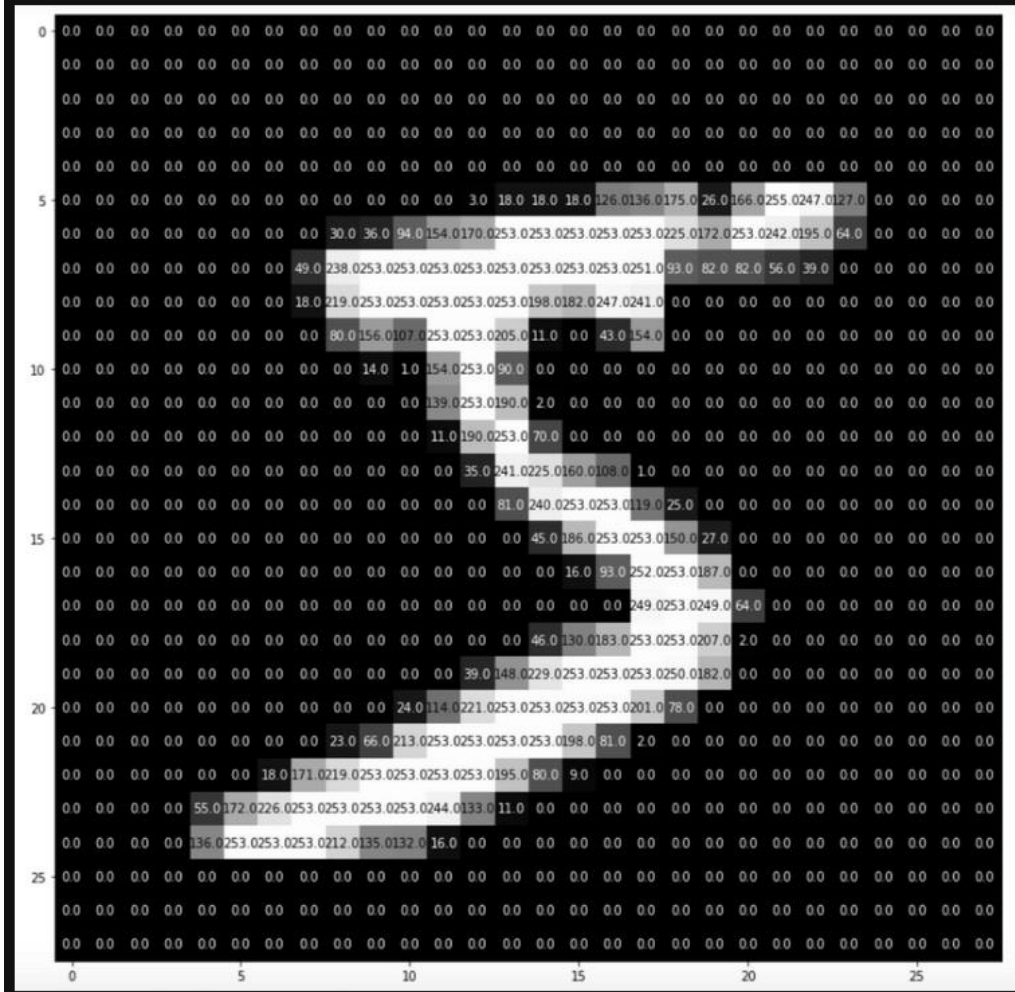


- 1,797개의 다양한 필기체 아라비아 숫자
- 사람이 Rule을 찾기가 거의 불가능함
- 64개 특성변수(8 \* 8 pixel)
- 밝기는 각 픽셀에 0~255 실수 부여
- 900개 데이터의 패턴을 학습하여 나머지 900개를 예측



# 패턴인식\_MNIST

Images are numbers !



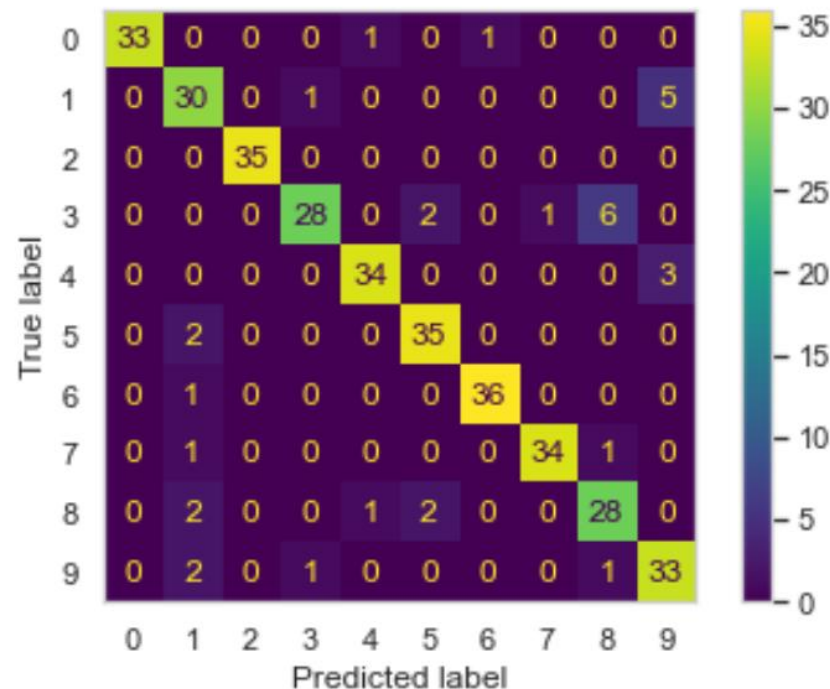
# 패턴인식\_MNIST

간단한 프로그램으로 데이터의 패턴을 인식하여 숫자를 0 ~ 9 구분하는 정확도는 90.55%

- '8', '9'번을 구분 못함
- '8', '9'번 판정의 정답률이 낮음

	pixel_0_0	pixel_0_1	pixel_0_2	pixel_0_3	pixel_7_5	pixel_7_6	pixel_7_7
0	0.0	0.0	5.0	13.0	0.0	0.0	0.0
1	0.0	0.0	0.0	12.0	10.0	0.0	0.0
2	0.0	0.0	0.0	4.0	16.0	9.0	0.0
3	0.0	0.0	7.0	15.0	9.0	0.0	0.0
4	0.0	0.0	0.0	1.0	4.0	0.0	0.0
5	0.0	0.0	12.0	10.0	10.0	0.0	0.0
6	0.0	0.0	0.0	12.0	11.0	3.0	0.0
7	0.0	0.0	7.0	8.0	0.0	0.0	0.0
8	0.0	0.0	9.0	14.0	11.0	1.0	0.0
9	0.0	0.0	11.0	12.0	3.0	0.0	0.0

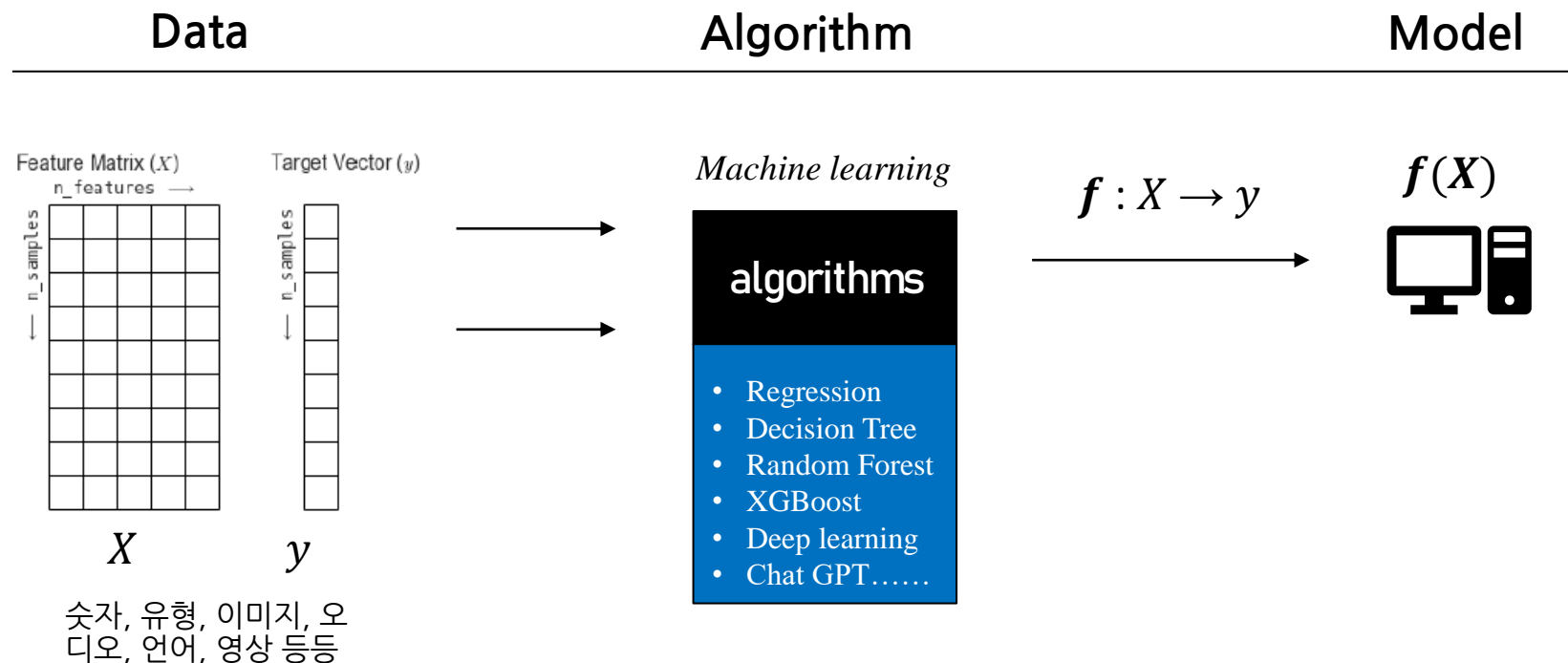
컴퓨터가 보는 세상



array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

# 데이터, 모델과 알고리즘(Data, Model & Algorithms)

- 모델이란 구체적인 과거의 데이터를 학습한 특정 표현으로 생각하는 것 그리고 그것을 학습하는 과정에서 필요한 알고리즘
- $\text{Model} = \text{Algorithm}(\text{Data})$



# 머신러닝 분야 <https://scikit-learn.org/stable/>

## scikit-learn

Machine Learning in Python

Getting Started

Release Highlights for 1.2

GitHub

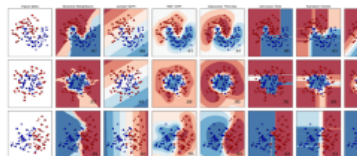
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

### Classification

Identifying which category an object belongs to.

**Applications:** Spam detection, image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, and more...



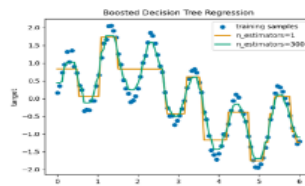
Examples

### Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, nearest neighbors, random forest, and more...



Examples

### Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, and more...



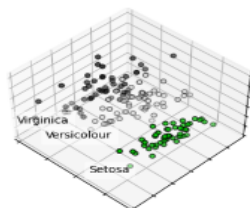
Examples

### Dimensionality reduction

Reducing the number of random variables to consider.

**Applications:** Visualization, Increased efficiency

**Algorithms:** PCA, feature selection, non-negative matrix factorization, and more...



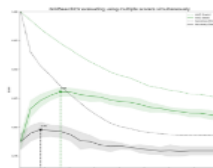
Examples

### Model selection

Comparing, validating and choosing parameters and models.

**Applications:** Improved accuracy via parameter tuning

**Algorithms:** grid search, cross validation, metrics, and more...



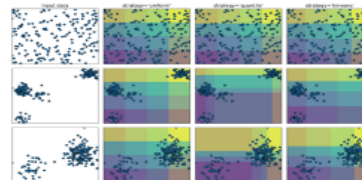
Examples

### Preprocessing

Feature extraction and normalization.

**Applications:** Transforming input data such as text for use with machine learning algorithms.

**Algorithms:** preprocessing, feature extraction, and more...



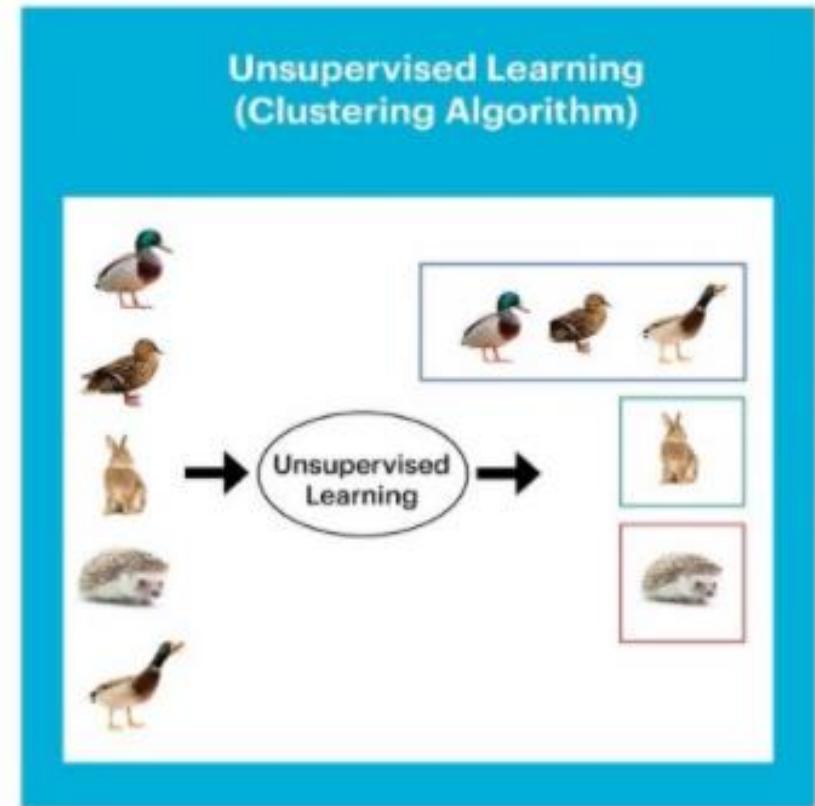
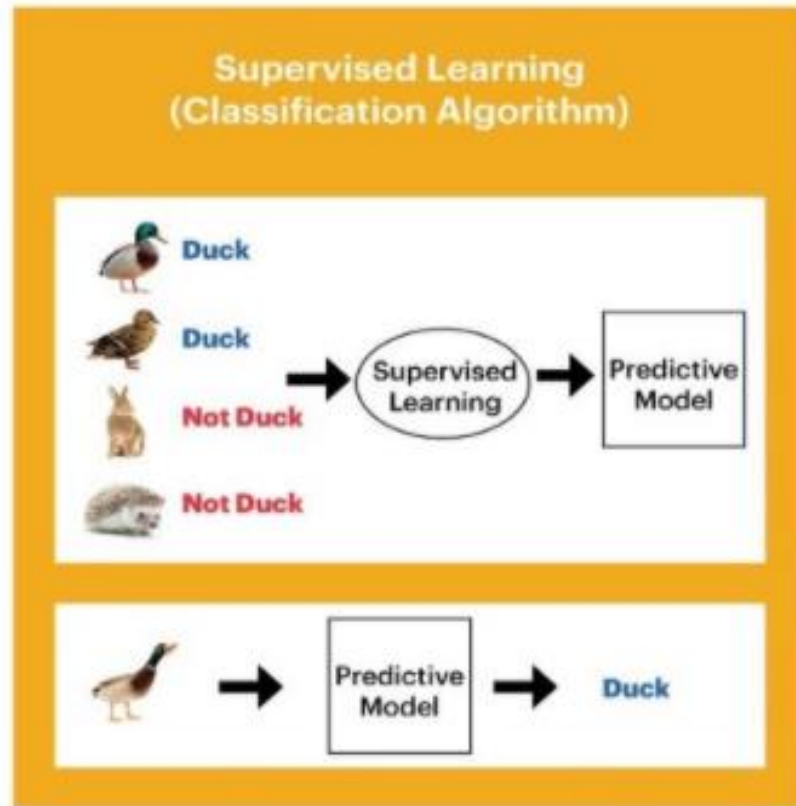
Examples



# 지도학습과 비지도학습

지도학습(Supervised learning)과 비지도학습(Unsupervised learning)

- 지도학습(Supervised learning) : Input과 Output 이 정해져 있음
- 비지도학습(Unsupervised learning) : Input의 특성만을 가지고 분류, Output 이 없음



# 분류와 회귀생성

분류(Classification)와 회귀생성(Regression)

- 분류(Classification) : 명목형 변수(Categorical variables) 결과를 예측
- 회귀생성(Regression) : 수치형 변수결과를 예측



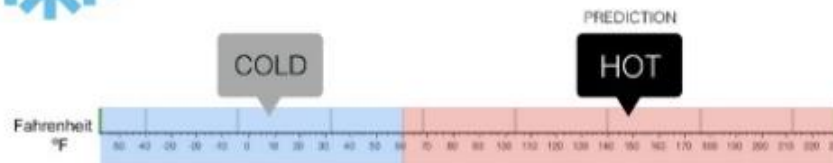
## Regression

What is the temperature going to be tomorrow?



## Classification

Will it be Cold or Hot tomorrow?



[https://medium.com/@ali\\_88273/regression-vs-classification-87c224350d69](https://medium.com/@ali_88273/regression-vs-classification-87c224350d69)