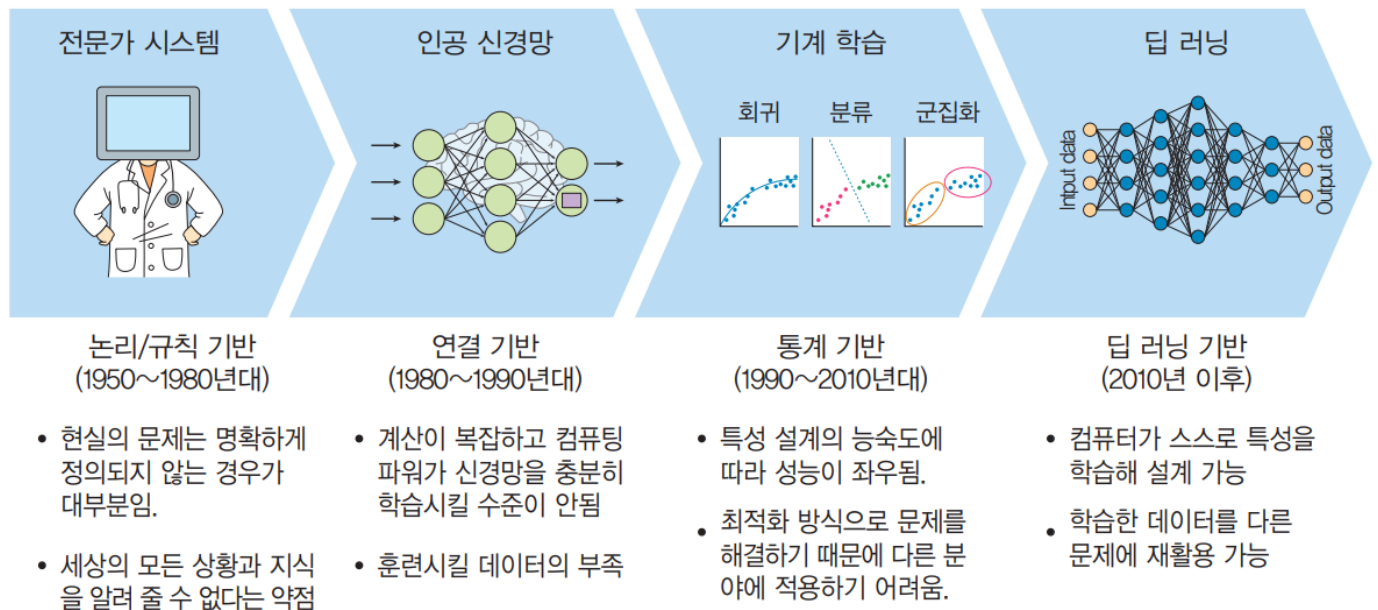


머신러닝 개요

김정현 강사



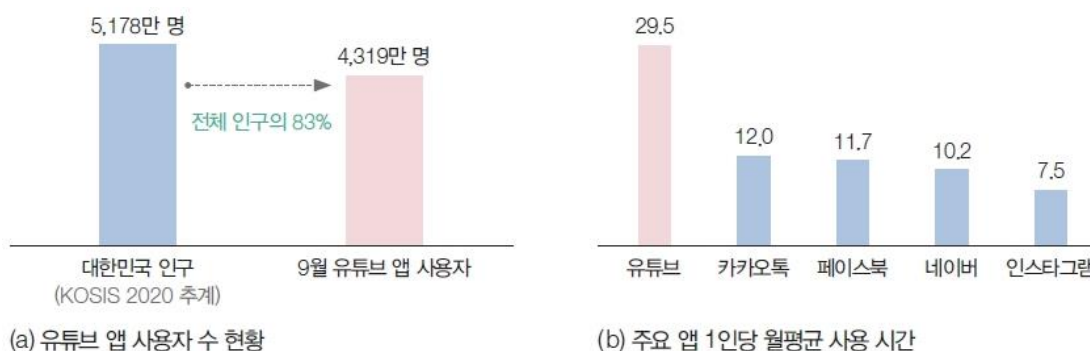
▲ 인공지능의 발전 과정

머신러닝이란?

알고리즘 - 어떠한 문제를 해결하기 위한 일련의 절차나 방법

머신러닝 - 기계가 패턴을 학습하여 자동화하는 알고리즘

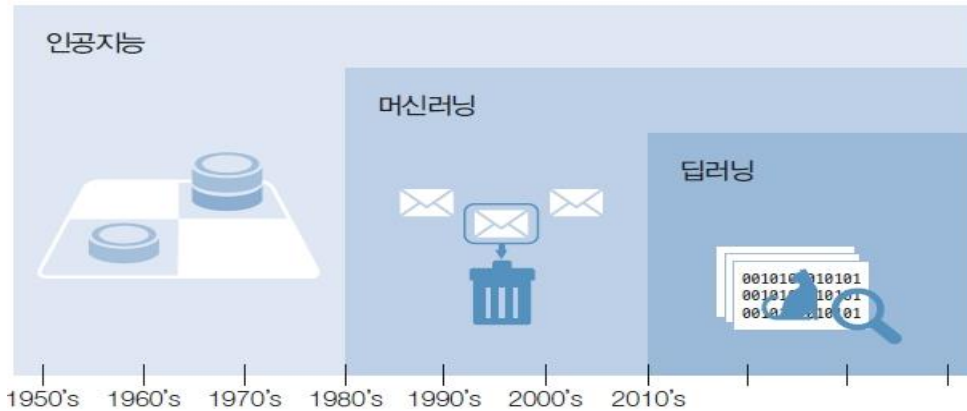
예를들어 유튜브는 개인이 유튜브 영상을 보는 패턴에 대해 학습하는 프로그램을 머신러닝으로 만든 다음 그 패턴(알고리즘)에 맞게 다음 영상을 계속 추천해 줌으로써 사용자의 사이트 유지율을 올려 지금과 같이 대중화되었다.



[머신러닝의 실제 사용 사례]

- 구매추천
- 번역
- 자율주행차
- 챗봇

인공지능,머신러닝,딥러닝



[인공지능(Artificial Intelligence, AI)]

- 컴퓨터가 학습하고 생각하여 스스로 판단할 수 있도록 만드는 기술
- 1956년에 열린 '다트머스 컨퍼런스'에서 '인공지능'이라는 표현을 처음 사용
- 여기에 참석했던 '인공지능의 아버지들'이라고 부르는 10명의 과학자들이 인공지능의 미래에 대해 논의하면서 본격적인 학문으로서 인공지능이 발전하기 시작
- 인공지능의 다양한 시도중의 하나가 머신러닝

[머신러닝(Machine Learning)]

- 데이터를 컴퓨터에 학습시켜 그 패턴과 규칙을 컴퓨터가 스스로 학습하도록 만드는 기술
- 데이터를 분류할 수 있는 수학적 모델을 프로그래밍하여 데이터만 입력하면 이미 만들어진 수학 모델이
- 규칙으로 적용되어 여러 문제를 풀 수 있도록 시도한 것
- 통계학에서 시작된 여러가지 수학 모델들이 컴퓨터 프로그램과 만나 큰 데이터들을 빠르게 처리할 수 있게 됨

[딥러닝(Deep Learning)]

- 사물이나 데이터를 군집화하거나 분류하는데 사용하는 기술
- 머신러닝 기법 중 '신경망(neural network)'을 기반으로 한 기술
- 머신러닝의 한 종류이지만 2012년을 기점으로 그 성능의 탁월성이 증명되어
- 오늘날 대부분의 인공지능 시스템에는 이 딥러닝 기법이 들어가 있음
- 딥러닝의 근간이 되는 신경망 알고리즘은 완전히 새로운 것이 아니라 1980년에 제안된 개념
- 엄청나게 많이 쌓인 데이터와 고속화 컴퓨터 성능으로 인해서 이전에 불가능했던 패턴 인식을 가능하게 한 것

통계학적 머신러닝

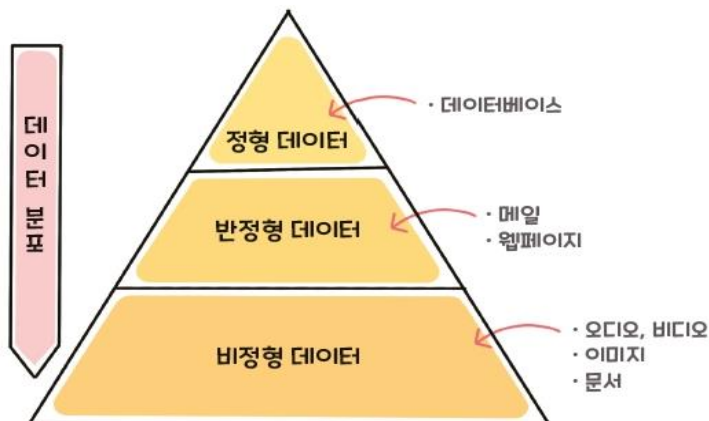
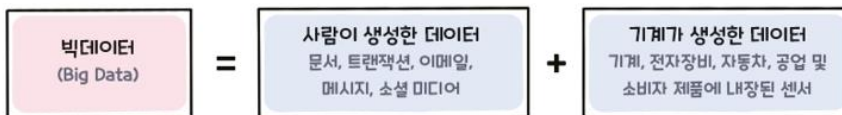
IT 기반 딥러닝

빅데이터

- 기존의 데이터베이스로는 수집, 저장, 분석 등을 수행하기 어려울 만큼 방대한 양의 데이터
- 빅데이터의 기원은 데이터베이스임

• 빅데이터(Big Data)

- 디지털 환경에서 생성되는 데이터
- 과거 아날로그 환경에서 생산되던 데이터에 비해 그 규모가 방대하고 생성 주기도 짧으며 수치 데이터뿐만 아니라 문자와 영상 데이터를 포함하는 대규모 데이터



• 정형 데이터(Structured Data)

- 일정한 형식이나 규칙(Rule)에 맞게 저장된 데이터
- 가장 쉽게 접할 수 있는 유형은 스프레드시트
- 정형 데이터의 예 : 스프레드시트, 관계형 데이터베이스, CSV 등

ID	Name	Age	Degree
1	John	18	B.Sc.
2	Jason	32	Ph.D.
3	Robert	52	Ph.D.
4	Ricky	35	M.Sc.
5	Gibb	26	B.Sc.

• 비정형 데이터(Unstructured Data)

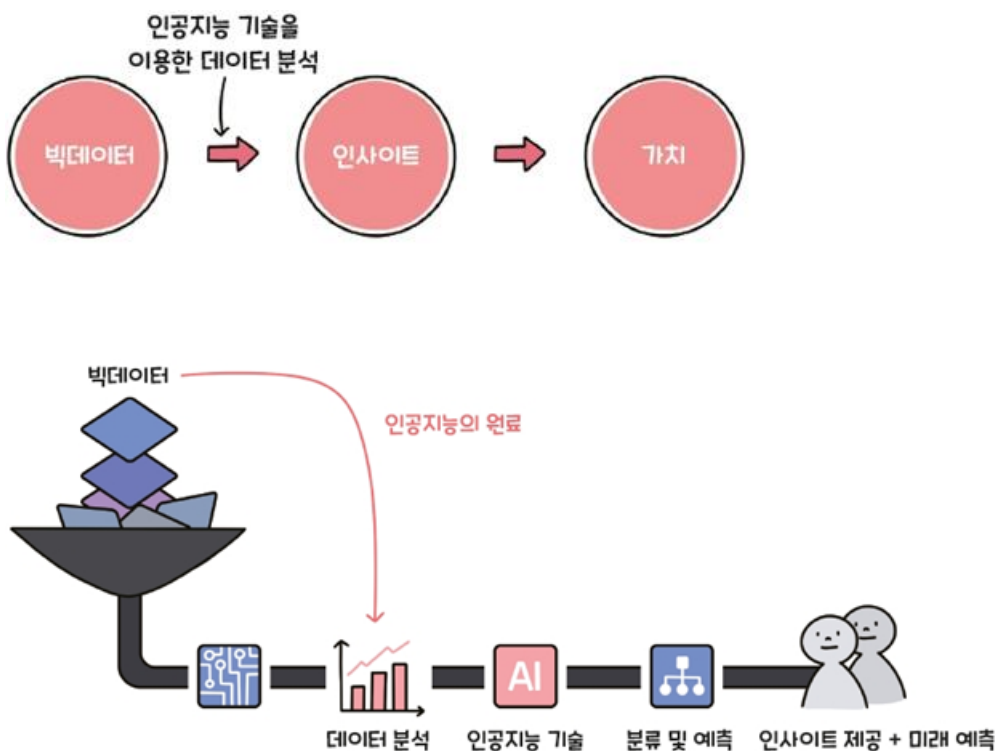
- 정형 데이터와 반대되는 개념으로, 정해진 규칙이 없어서 값의 의미를 쉽게 파악하기 힘든 경우가 많음
- 인공지능 기술의 발전과 함께 비정형 데이터로부터 인사이트(Insight)를 얻는 사례가 많아지면서 비정형 데이터에 대한 중요성이 부각되고 있음
- 비정형 데이터의 예 : SNS, 영상, 이미지, 음성, 텍스트 등

Artificial intelligence(AI), is intelligence demonstrated by machines, unlike the natural intelligence displayed by humans and animals, which involves consciousness and emotionality. The distinction between the former and the latter categories is often revealed by the acronym chosen. 'Strong' AI is usually labelled as AGI(Artificial General Intelligence) while attempts to emulate 'natural' intelligence have been called ABI(Artificial Biological Intelligence).

• 반정형 데이터(Semi-structured Data)

- 스키마(Schema)와 메타데이터(Metadata)의 특성을 갖는 데이터
- 반정형 데이터가 주목을 받는 이유는 인터넷에 존재하는 수많은 데이터를 인공지능에 사용할 수 있기 때문
- 반정형 데이터의 예 : XML, JSON, NoSQL, 로그 등

```
<University>
<Student ID="1">
  <Name> John </Name>
  <Age> 18 </Age>
  <Degree> B.Sc. </Degree>
</Student>
<Student ID="2">
  <Name> Jason </Name>
  <Age> 32 </Age>
  <Degree> Ph.D. </Degree>
</Student>
...
</University>
```





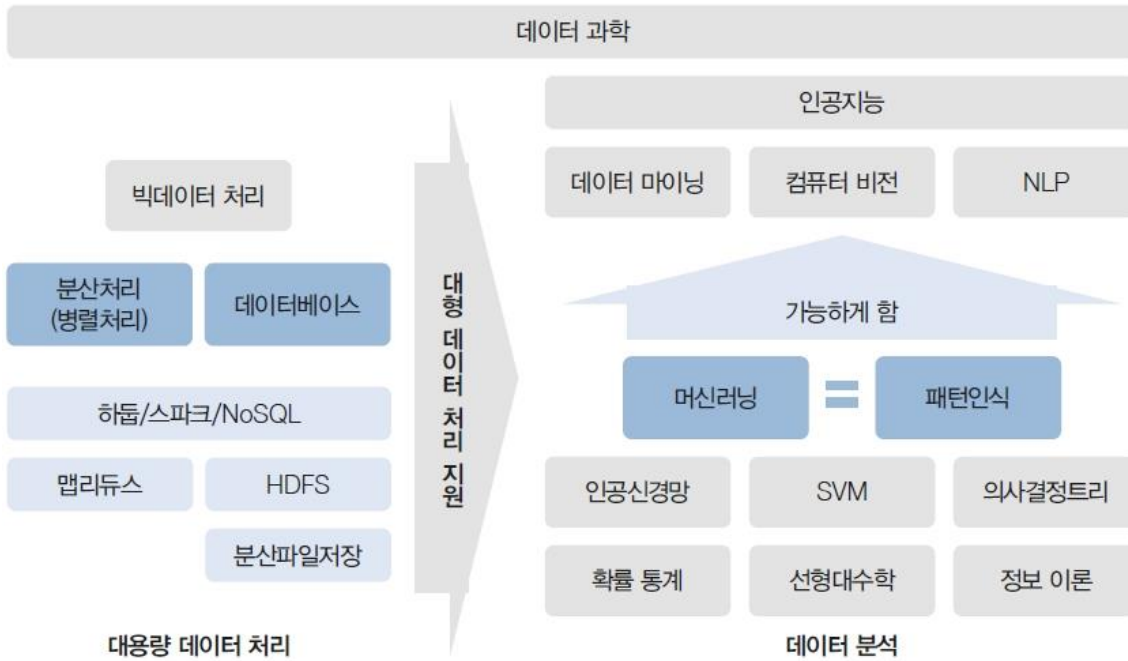
- 머신러닝 시대를 열었던 많은 기술 중 대표적인 것이 구글의 GFS와 맵리듀스임
- GFS(Google File System)은 구글의 거대한 검색 시스템을 여러 대의 컴퓨터에 분산해서 저장 및 관리하기 위해 설계된 시스템
- 맵리듀스(Map Reduce)는 이렇게 분산되어 저장되어 있는 데이터를 어떻게 여러 대의 컴퓨터가 동시에 협력해서 처리할 것인가에 대한 방법을 알려준 시스템
- 구글의 분산처리 시스템 개념이 확장되어 이후에 **하둡(Hadoop)**이라는 시스템으로 개발되었고 이 시스템이 발전하여 오늘날의 대용량 데이터를 처리하는 데 사용되고 있다.



빅데이터와 머신러닝은 기본적으로 서로 별개의 것이지만 머신러닝을 가능하게 한 것이 빅데이터이고 빅데이터 기술이 가장 빛나는 순간이 바로 머신러닝에 활용될 때임

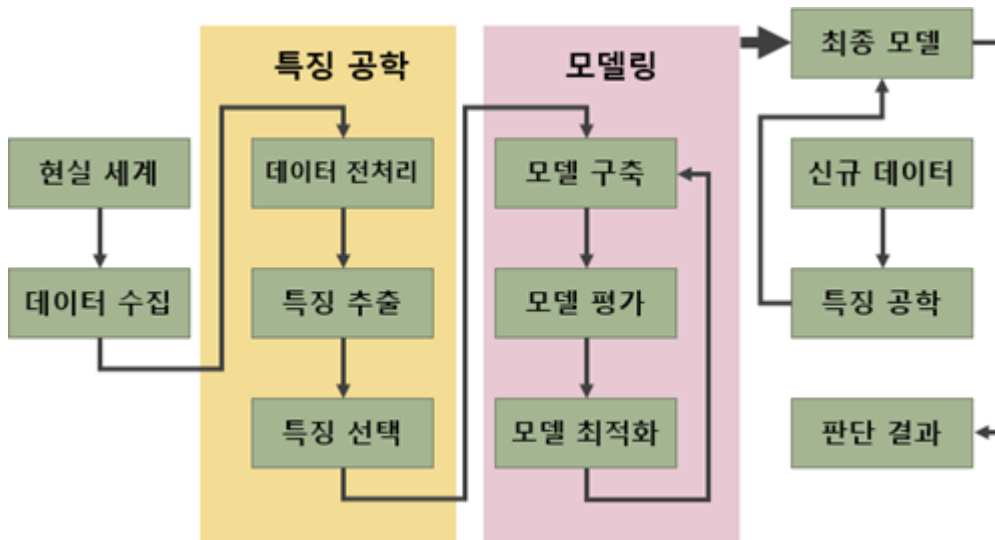
얼굴인식 기술 - 지금만큼의 데이터가 모이지 않았다면 개발되기 어려운 것임

많은 데이터를 전처리하여 머신러닝 기술에 사용하기 위해서는 거대한 데이터를 처리하는 빅데이터 기술이 필수적이다. 두 기술은 각자의 생태계에서 발전했지만 결국은 오늘날의 인공지능 생태계를 완성한 핵심기술 영역들이다.



머신러닝 처리 프로세스

[특징 == 특성 == 데이터프레임이나 매트릭스의 컬럼]



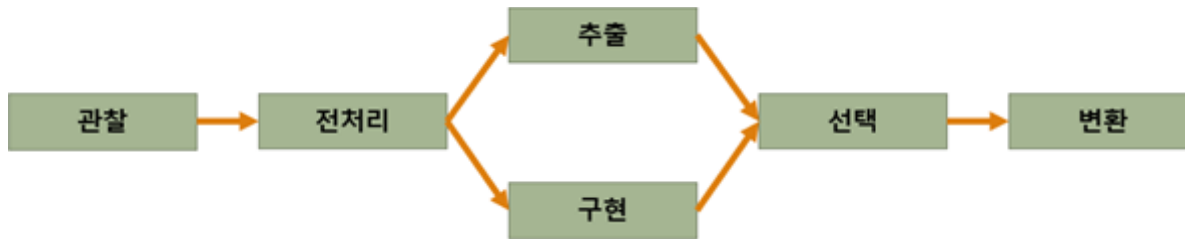
1) 데이터 수집 단계

모든 데이터는 '컴퓨터가 이해할 수 있는' 형식을 갖추어야 한다.

데이터 수집은 다양한 채널을 통해 이루어질 수 있다. 하드웨어로 수집 한다거나 웹 서버의 정보를 크롤링하거나, 이미 운영 중인 솔루션 또는 운영체제 자체에서 자동으로 기록하는 로그 정보도 좋은 데이터가 될 수 있다.

데이터가 어느 정도로 필요한 지 미리 규정짓는 것은 의미가 없다. 가능한 많은 데이터 확보가 우선 시 된다.

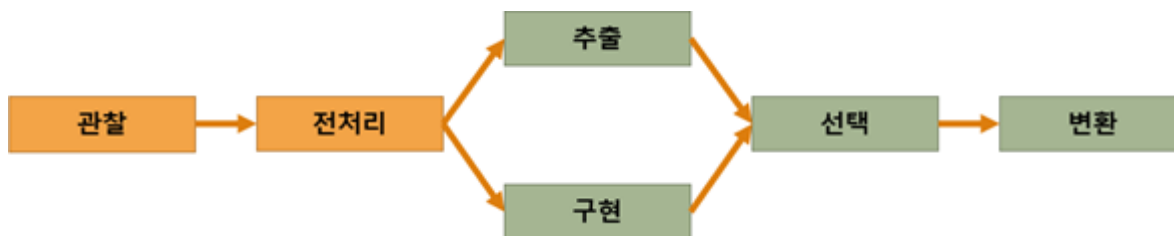
2) 특징(특성) 공학(Feature Engineering)



특징 공학 : 모델의 성능을 좌우하는 핵심 요소로, 머신 러닝 프로세스에서 가장 중요한 역할을 담당한다.

특징 공학은 데이터의 의미를 이해하고, 조작하고, 특징을 추출하고, 변환하는 모든 과정을 포함한다.

[관찰과 전처리]



특징 공학은 데이터 관찰에서 시작한다.

- 데이터 관찰:

문제 해결에 도움을 주는 정보를 찾기 위함이 아니라 데이터 자체를 이해하기 위함이다.

데이터가 어떠한 형식으로 저장되어 있는지, 누락된 값은 없는지, 데이터 분포는 어떠한지 등을 종합적으로 살펴본다.

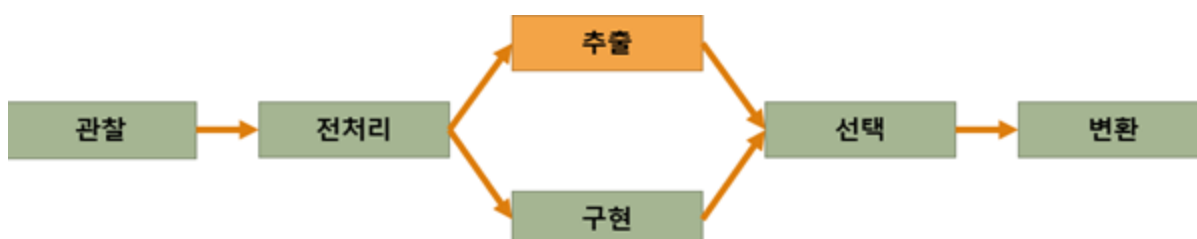
- 전처리:

누락된 데이터나 오류 값이 있으면 수정 또는 제거해주고, 단순 텍스트 형식의 로그 파일일 경우 가독성이 높은 테이블 형태로 파싱한 후 데이터베이스 또는 파일로 저장해주는 과정.

데이터 관찰과 전처리 과정은 시각화와 함께 진행해야 한다.

그러나 시각화로 본 내용이 그대로 결과에 반영될 것이라고 가정하면 안된다. 그저 보조 도구일 뿐이다.

[특징 추출]



데이터 전처리 완료 후에는 모델링에 사용될 특징을 추출해야 한다. 특징 추출에는 크게 3가지 방법이 존재한다.

특징 추출의 3가지 방법:

유형	장점	단점	예시
데이터를 그대로 특징으로 사용	- 추출이 간단함	- 대부분 데이터는 그대로 사용이 불가능 - 좋은 특징을 보장하지 않음	실행 파일 또는 패킷의 헤더 값
통계 수치 활용	- 추출이 간단함 - 데이터 특성 반영	- 통계 수치 해석 능력 필요 - 적용 가능 영역이 제한됨	10초 동안 서버로 들어온 프로토콜별 평균 요청 횟수
도메인 특성을 토대로 추출	- 좋은 특징 추출 가능 - 최소 특징 최대 효율	- 도메인 지식이 필요 - 해당 분야 전문가의 도움이 필요	프로그램 데이터 영역의 엔트로피 값

특징 추출 방법 비교

통계 수치 : 통계학에서 데이터의 분포와 경향을 표현할 때 사용하는 수치 값들을 특징에 활용하는 방법이다. 주로 시계열 데이터에 적용하는 특징이다.

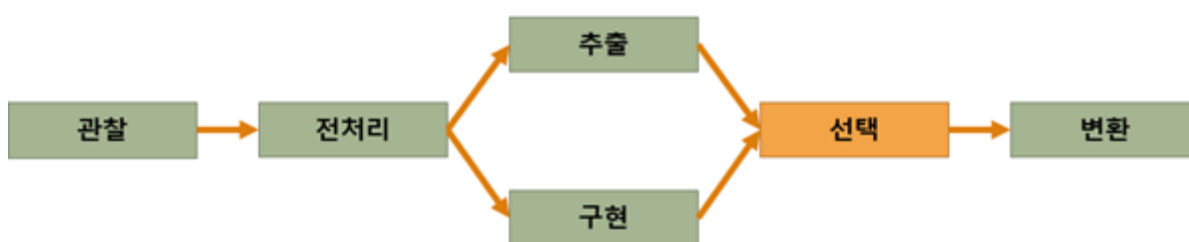
하나의 특징 = 하나의 속성(정보) = 1차원

특징 추출 단계에서 해당 특징이 모델링 정확도에 주는 영향도를 고려할 필요 없이 가능한 모든 특징을 뽑아 두는 것이 좋다.

n개의 특징을 가진다는 것은 n차원으로 표현한다고 말할 수 있다.

즉, 특징이 많아질수록 복잡도는 굉장히 높아진다. 그러나 컴퓨터는 100차원을 이해할 수 있다. 그렇다고 해도 특징이 너무 많으면 모델 성능에 독이 될 수 있다. ('차원의 저주' 문제라고 한다.)

[특징 선택]



- 특징 선택 방법: 최적의 조합을 찾을 때까지 특징 개수를 하나씩 늘리거나 줄이는 방법.

→ 간단하지만 특징이 많은 경우 선택에 엄청나게 많은 시간이 소요될 수 있다.

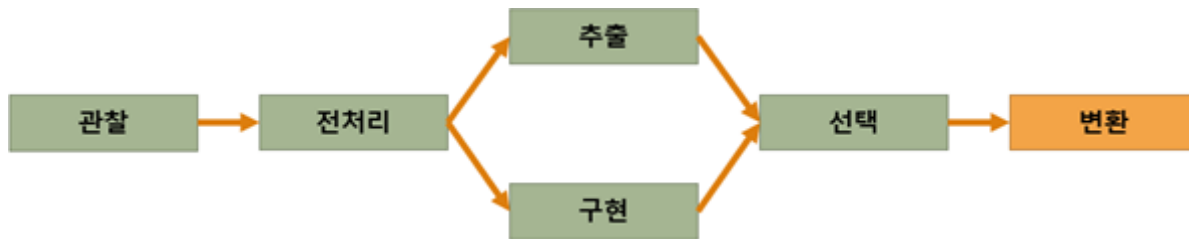
특징 선택을 위한 모델링은 해결하고자 하는 문제에 적용 가능한 모델을 3~5가지 정한 다음 여기에 선택한 특징 조합을 대입해보는 방식을 말한다. 예를 들어 특징 추출 단계에서 뽑아낸 10개의 특징으로 악성코드와 정상 프로그램을 '분류'하는 것이 목표일 때 대표적인 알고리즘인 SVM, 랜덤 포레스트, 나이브 베이스 분류기, 딥 뉴럴 네트워크를 테스트 모델로 선정한다.

다음으로 특징 조합을 하나 선택한 후 네 개의 모델에 각각 넣고 정확도를 계산한다. 가능한 모든 조합에 대한 정확도를 계산한 후에 네 개 모델 정확도 평균값이 가장 높게 나온 특징 조합을 선택하면 된다.

- 특징 선택 방법2: 도메인 지식을 바탕으로 휴리스틱(경험에 기반한 문제 해결)에 의존해 특징을 선택하는 방법

분석가의 입장에서 볼 때 의미가 없고 아무런 정보도 제공하지 않는 특징이라고 해도 머신 러닝 모델에는 정확도에 크게 기여하는 좋은 특징이 될 수 있기 때문에 무작정 골라내는 것보다 후보군을 선정한 후 전진선택 또는 후진제거 방식을 적용하는 편이 좋다.

[변환 단계]

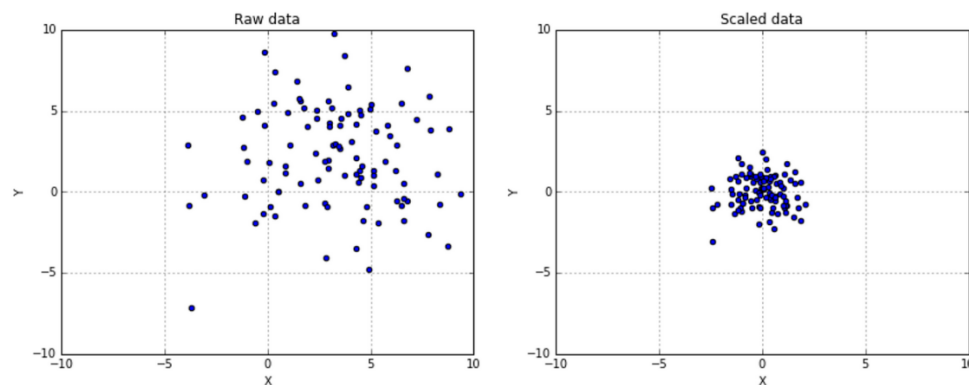


변환 단계:

특징 값을 모델링에 적합한 형태로 변환하는 과정이다. 정규화, 스케일링, 범주형 데이터 처리 등이 여기에 속한다.

정규화 : 모든 특징의 값 범위를 동일한 범위로 맞춰주는 과정이다.

스케일링 : 특정 데이터를 정규 분포 형태로 변환해주는 과정이다. 가장 일반적인 방식은 데이터를 로그 변환하는(자연 로그) 것으로, 큰 수를 같은 비율의 작은 수로 바꿔주며(정규화), 복잡한 계산을 쉽게 만들고, 왜도(skewness)와 첨도(kurtosis)를 줄여 데이터 분석 시 의미 있는 결과를 도출할 수 있도록 해준다



특징 값이 수치형이 아닌 **범주형** 데이터일 경우 수치형 특징으로 변환해야 한다.

가장 대표적인 것은 **One-Hot 인코딩**이다. One-Hot 인코딩은 개별 특징 내에 포함된 유일한 문자의 개수만큼 특징을 추가한 뒤 각 특징을 0 또는 1로 표현하는 방식이다. 범주형 특징이 모델 판단에 중요한 정보를 포함하고 있을 경우 유용하게 쓰일 수 있으나, 특징의 개수가 범주의 개수만큼 증가한다는 단점이 있다.

특징		A	B	C	D	E
A	One-Hot 인코딩 적용 후	1	0	0	0	0
E		0	0	0	0	1
A		1	0	0	0	0
B		0	1	0	0	0
C		0	0	1	0	0
A		1	0	0	0	0
D		0	0	0	1	0

머신러닝은 모델을 생성하는 과정

머신러닝을 간단히 설명하면 다음과 같은 함수이다.

$$\hat{y} = f(x)$$

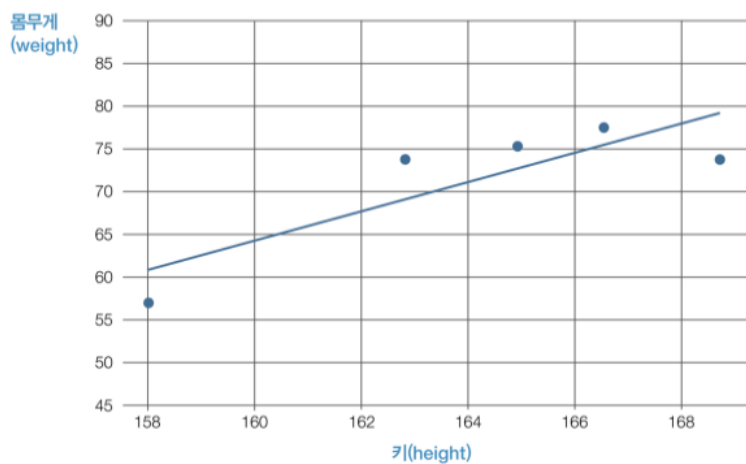
\hat{y} - 머신러닝 알고리즘에 의해서 산출되는 예측치

x - 가지고 있는 데이터

$f(\cdot)$ - 일종의 함수로 머신러닝에서는 모델이라고 함

[머신러닝 모델을 찾는 과정]

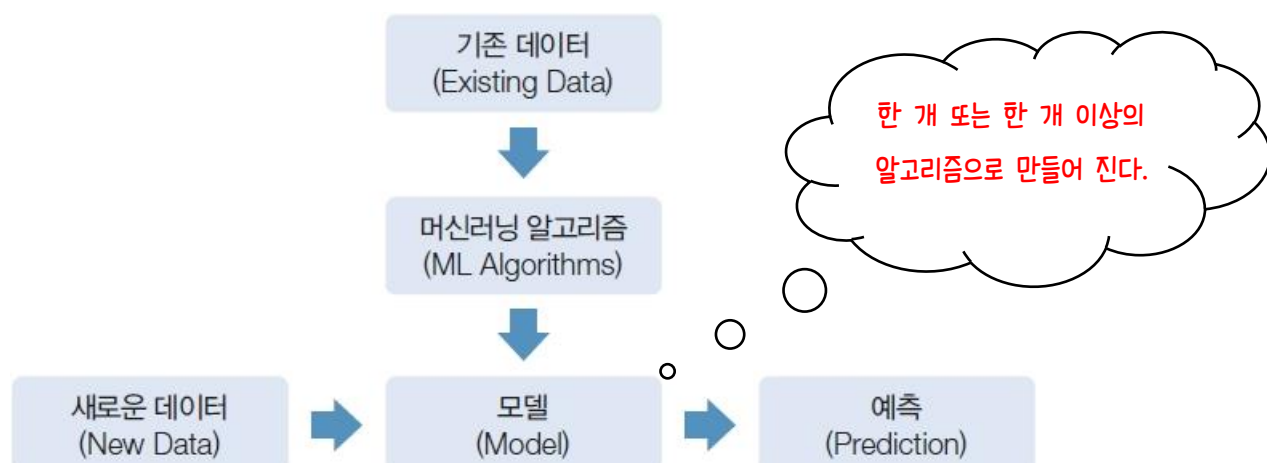
키	158	168.7	162.8	166.5	164.9
몸무게	57.1	74.1	74.1	77.9	75.5



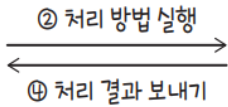
키와 몸무게 간의 상관관계를 다음과 같이 1차 방정식으로 표현할 수 있다.

$$y = \alpha x + \beta$$

바로 이와 같은 식을 모델(model)이라고 부른다. 다음은 머신러닝의 프로세스이다.



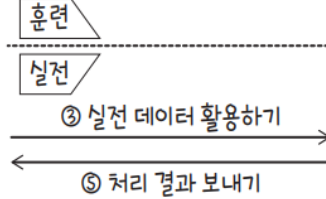
① 처리 방법 만들기



컴퓨터

▲ 일반적인 프로그래밍

① 원하는 결과 샘플 준비하기 (데이터)



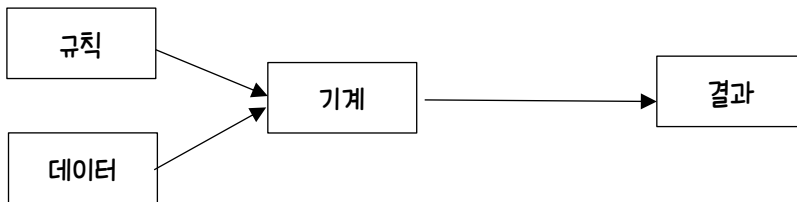
② 처리 방법 만들기



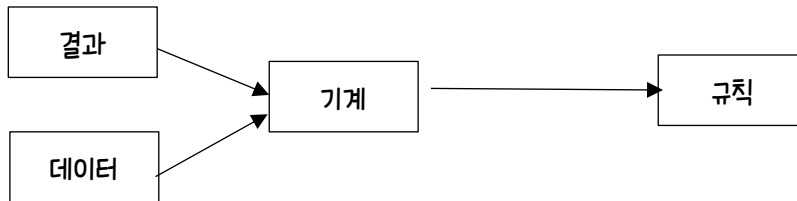
컴퓨터
④ 처리하기

▲ 기계 학습 프로그래밍

[전통적인 프로그래밍 접근]



[머신러닝 기반 프로그래밍 접근]



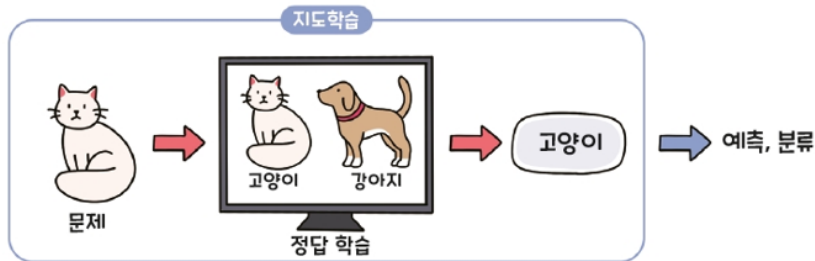
머신러닝의 종류

머신러닝을 기본적으로 분류하면 지도학습과 비지도 학습으로 나눌 수 있다. 지도학습(supervised learning)이란 문제와 답을 함께 학습함으로써 미지의 문제에 대한 올바른 답을 예측하는 학습이다. 비지도학습(unsupervised learning)이란 지도학습과 다르게 조력자의 도움없이 컴퓨터 스스로 학습하는 형태로 컴퓨터가 훈련데이터를 이용하여 데이터들 간의 규칙성을 찾아내는 학습이다.



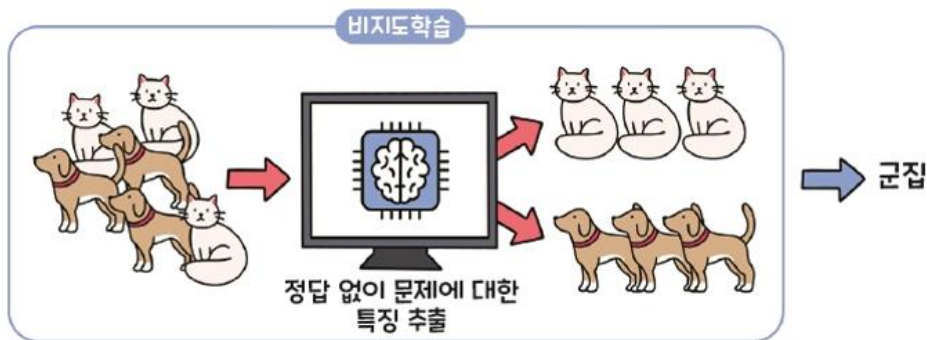
• 지도학습(Supervised Learning)

- 문제와 답을 함께 학습함으로써 미지의 문제에 대한 올바른 답을 예측하는 학습
- 지도학습에서 사용하는 모델로는 크게 예측과 분류가 있음



• 비지도학습(Unsupervised Learning)

- 지도학습과 다르게 조력자의 도움 없이 컴퓨터 스스로 학습하는 형태
- 컴퓨터가 훈련 데이터를 이용하여 데이터들 간의 규칙성을 찾음



지도학습과 비지도학습 시 필요한 데이터

구분	지도학습	비지도학습
필요한 데이터 종류	x(학습 데이터), y(레이블)	x(학습 데이터)

지도학습	회귀	연속형 값인 y의 특징을 찾아 데이터 x를 사용하여 y 값을 예측하는 기법
	분류	이산형 값인 y의 특징을 찾아 데이터 x를 사용하여 y 값을 예측하는 기법
비지도학습	군집	y값이 주어지지 않고 데이터의 특징이 유사한 값들의 모임을 군집으로 표현하는 기법

내용	연속형 데이터	이산형 데이터
개념	<ul style="list-style-type: none"> 값이 끊어지지 않고 연결되어 표시되는 형태 일반적인 실수나 정수 값들 	<ul style="list-style-type: none"> 값이 연속적이지 않고 끊어서 표현되는 값들 라벨로 구분되는 값들 숫자일 경우에도 그 값들 간에 스케일이 존재하지 않음
예시	온도, 시험 평균, 자동차 속도 등	성별, 주소, 설문조사 척도 등

회귀

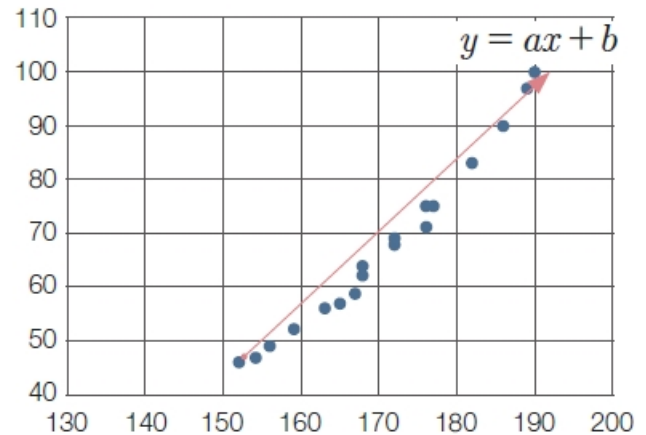
회귀는 '독립변수 x 와 종속변수 y 의 관계를 함수식으로 설명한 것

(독립변수는 영향을 주는 변수, 종속변수는 영향을 받는 변수)

두 값의 관계를 하나의 선(추세선)으로 나타내는 것으로 이 선을 표현하는 수학적 모델을 만드는 기법을 회귀라고 부름

키	몸무게	키	몸무게
152	46	172	69
154	47	172	68
156	49	176	71
159	58	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

(a) 키와 몸무게 데이터



(b) 회귀 표현

- 회귀는 연속형 변수를 예측하는 데 사용되는데, 즉 연속적인 숫자나 실수를 예측하는 것(예 : 주식 및 부동산 가격 예측 등)
- 회귀는 종속변수와 독립변수 간의 관계를 살펴볼 때 유용하게 사용

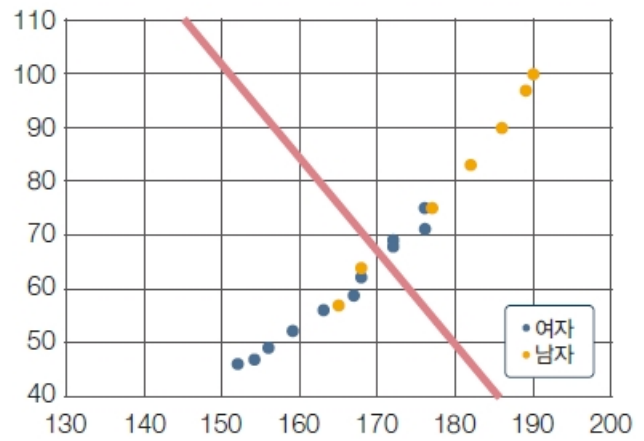


분류

데이터를 어떠한 기준에 따라서 나누는 기법이다. 이미지를 입력했을 때 그 이미지가 개인지 고양이인지, 뉴스 텍스트를 입력했을 때 그 뉴스가 스포츠 뉴스인지 연예뉴스인지 분류하는 것이 이에 해당된다. 이 때 y 는 이산값(discrete value)을 가지는 것이 일반적이는데 2개의 값중에서 1개는 분류하는 것을 이진분류(binary classification)이라 하며 3개 이상의 분류를 실행하는 것을 다중분류(multi-class classification)이라 한다.

성별	키	몸무게	성별	키	몸무게
여자	152	46	남자	172	69
여자	154	47	여자	172	68
여자	156	49	여자	176	71
여자	159	58	여자	176	75
여자	163	56	남자	177	75
남자	165	57	남자	182	83
남자	167	59	남자	186	90
남자	168	64	남자	189	97
여자	168	62	남자	190	100

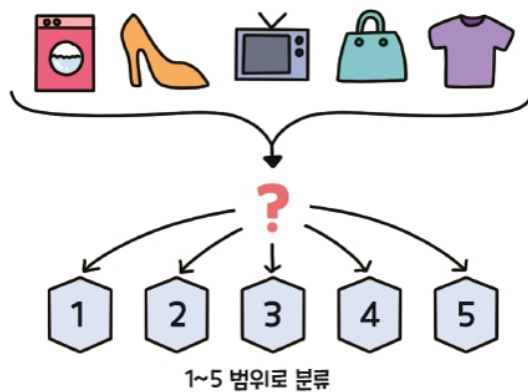
(a) 성별, 키, 몸무게 데이터



(b) 분류 표현

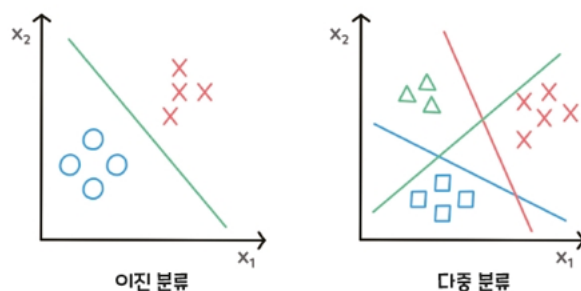
• 분류(Classification)

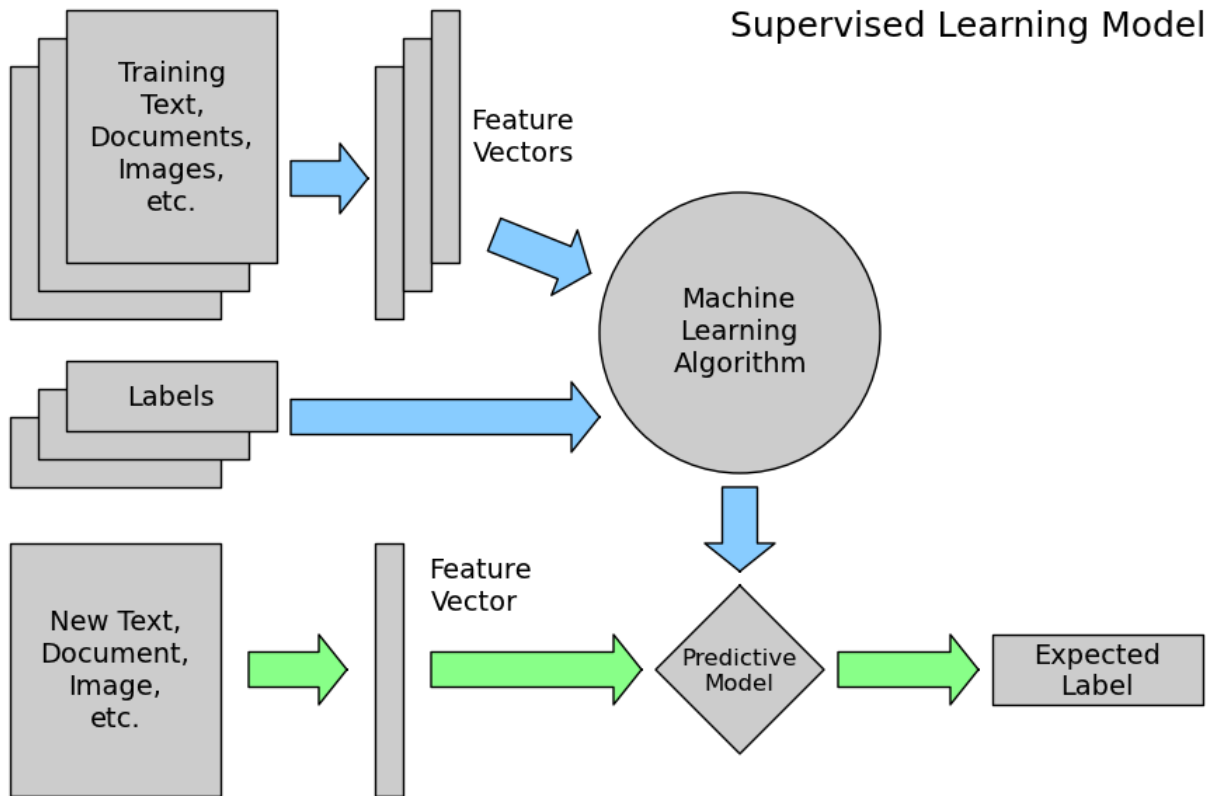
- 레이블이 포함된 데이터를 학습하고 유사한 성질을 갖는 데이터끼리 분류한 후, 새로 입력된 데이터가 어느 그룹에 속하는지를 찾아내는 기법



• 분류의 종류

- 이진 분류(Binary Classification) : 데이터를 2개의 그룹으로 분류
- 다중 분류(Multiclass Classification) : 데이터를 3개의 그룹 이상으로 분류





군집

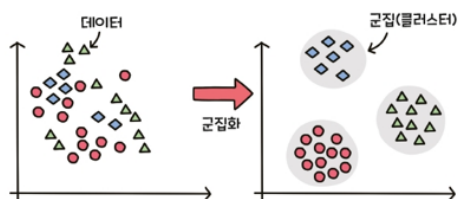
군집(clustering)은 기존에 모여 있던 데이터에 대해 따로 분류 기준을 주지 않고 모델이 스스로 분류 기준을 찾아 집단을 모으는 기법이다. 분류의 기준이 되는 y 값이 존재하지 않고 알고리즘이 스스로 데이터의 특성을 고려하여 군집을 나눈다.

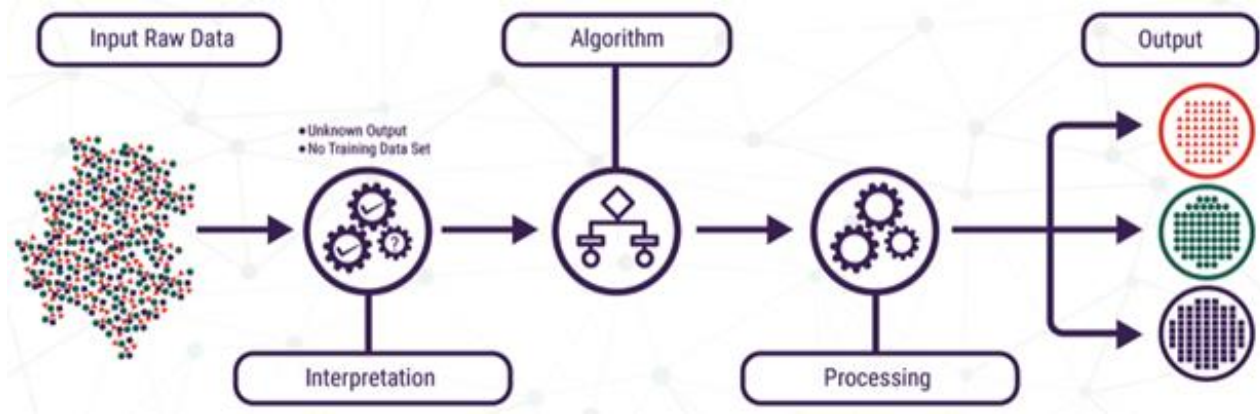
- **군집(Cluster, 클러스터)**

- 비슷한 특징을 가진 데이터들의 집단

- **군집화(Clustering, 클러스터링)**

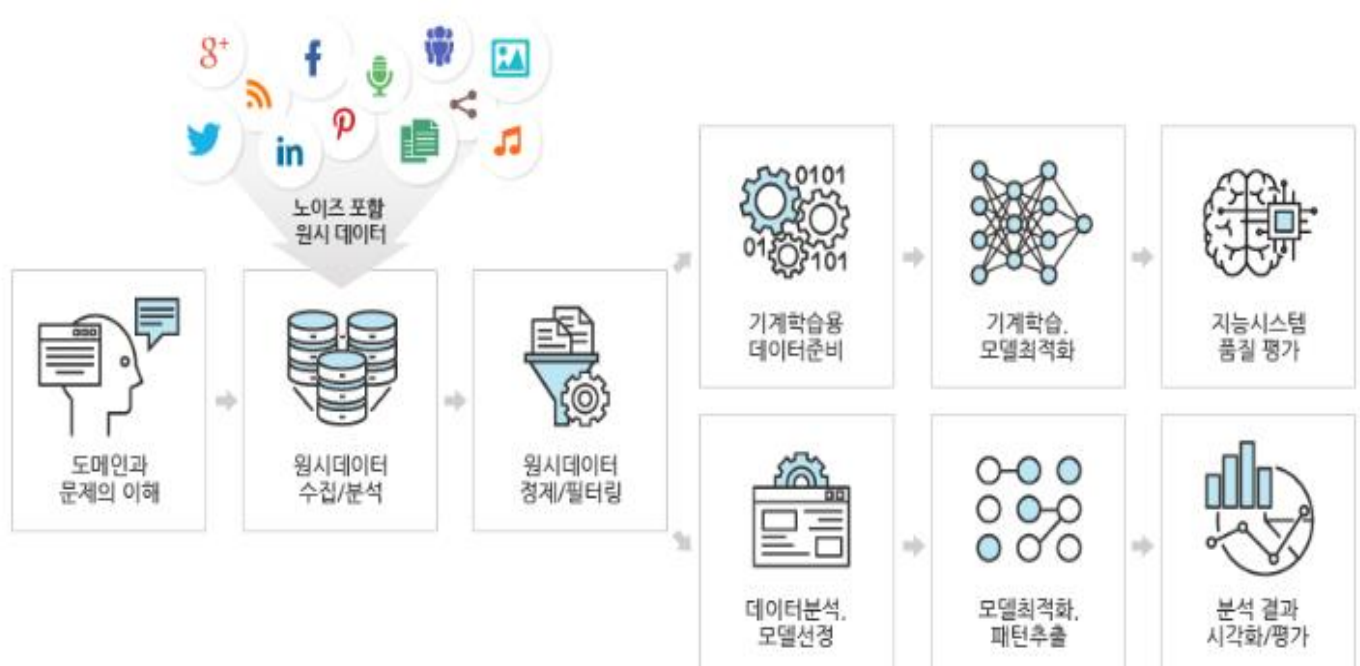
- 데이터가 주어졌을 때 그 데이터들을 유사한 정도에 따라 군집으로 분류하는 것
- 왼쪽 그래프를 보면 다양한 데이터들이 서로 섞여 있지만, 군집화 과정을 진행하면 오른쪽 그래프와 같이 비슷한 데이터끼리 군집으로 묶임





[파이썬 머신러닝의 도구들]

도구 종류	설명	사용 도구
파이썬 인터프리터	파이썬 코드를 해석하고 실행시키기 위한 실행 프로그램	아나콘다(Anaconda)
코드 편집기	파이썬 코드를 수정할 때 사용하는 프로그램	주피터(Jupyter), VS코드(VSCode)
통계 분석 및 전처리 도구	데이터를 로드하고 전처리하기 위한 도구들	넘파이(Numpy), 판다스(Pandas), 사이파이(SciPy)
시각화 도구	데이터의 상태를 파악하기 위해 시각화를 지원하는 도구	맷플롯립(Matplotlib), 시본(Seaborn), 플롯리(Plotly)
머신러닝 프레임워크	실제 머신러닝 모델을 생성하고 데이터에 적용할 수 있도록 도와주는 도구	사이킷런(Scikit-learn)

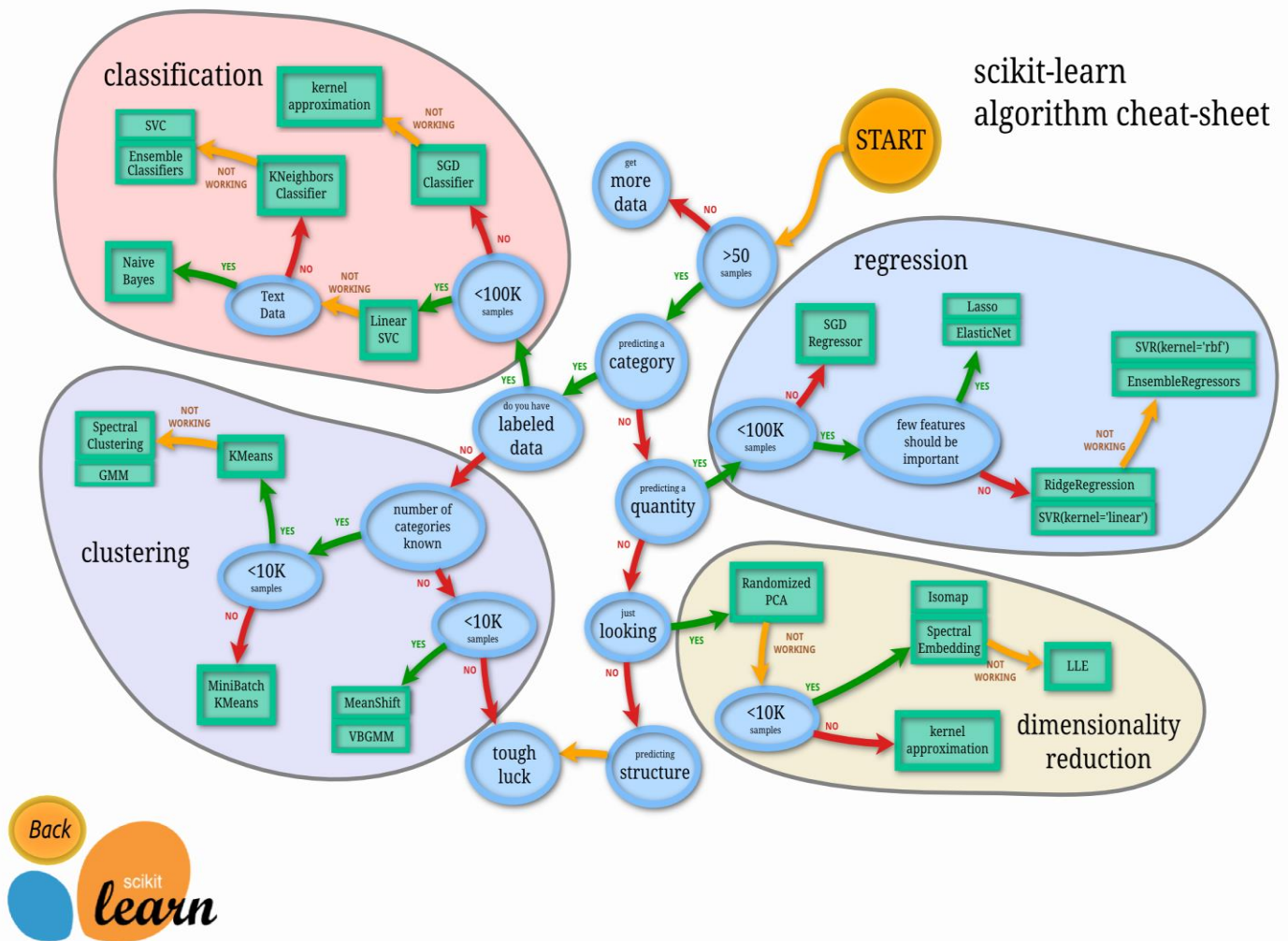


Step		목적	
1	문제 정의		<p>문제의 목표는 무엇인가? 문제의 범위는? 문제 해결의 성공 및 실패의 척도는? 문제 해결에 있어서의 제약 조건은?</p> <ul style="list-style-type: none"> - 머신러닝 유형파악 (지도 학습 vs 비지도 학습) - 회귀 문제 : 연속적 숫자 예측(ex. 교육, 나이, 조거지 연간 소득 예측) - 분류 문제 : 클래스 레이블 중 하나 예측 (ex. 암환자 예측)
2	데이터 준비(수집)		<ul style="list-style-type: none"> - 데이터 수집 - 데이터 조합 (정형데이터 + 비정형데이터) - 웹크롤링, 웹스크래핑 - 반드시 포함 되어야할 속성?
3	데이터 전처리 (Pre-Processing)	Transforming Munging Wrangling Preprocessing	<ul style="list-style-type: none"> - 활용 목적에 따른 필터링 - 분석시간 단축 - 저장공간 효율적 활용 - 오류와 중복 제거 --> 품질개선 - Missing Data, Value 처리 - Encoding - Noise 제거 - Rescaling - Normalization - 불완전(imcomplete) - 잡음(Noisy) - 모순된 (Inconsistent) <p>학습데이터와 평가데이터 나누기 --> 일반적으로 Train Set 70%, Test Set 30% --> 별도의 Test 데이터 필요 (학습과정 확인, 목적에 맞는 성능 검증)</p>
4	탐색적 분석 (Exploratory Data Analysis)		<ul style="list-style-type: none"> - 분석을 위한 데이터 특성 파악 - 상관관계 등 기본 통계적 분석 - 데이터 값/분포 오류 확인

	=EDA		<ul style="list-style-type: none"> - 데이터의 특징/패턴 발견 - 데이터의 문제점이나 한계 확인 - 데이터 재수집 또는 추가 수집 - 다양한 관점에서 집계, 시각화 - 기계 학습에 넣을 속성 선택 <p>데이터 탐색 -> 패턴발견 -> 가설 수립 단계</p> <p>목적 : 데이터를 잘 이해하고 검토, 탐구할 만한 가티 있는 가설 설정</p> <p>머신러닝 알고리즘에 넣을 중요한 변수</p> <p>* 시각화 인포메틱스(INFORMATIX) 기술 -> 중요</p>
5	모델 생성		<ul style="list-style-type: none"> - 예측 알고리즘 선택 - 예측 모델 구축 시, 기계학습 기법 이용 - 머신러닝 학습 알고리즘 선택 - 문제의 특성에 따른 구분 : 분류/회귀 - 데이터의 특성 : 데이터의 양, 속성의 수 등 - 모델의 용도 : 결과 해석이 중요한 경우
6	학습/예측	Machine Learning	<p>-학습시킬 변수 선택 -> 데이터 중심</p> <p>-하이퍼 파라미터 조정 -> 모델 중심</p>
7	평가	Reporting	<ul style="list-style-type: none"> - 예측 결과의 정확성 검증(Score) - 문제 유형에 따라 다양한 평가지표 존재 - 회귀모델 : RMSE - 분류기법 : 정확도(Accuracy)
8	커뮤니케이션		<ul style="list-style-type: none"> - 예측 결과를 이해 당사자와 효과적으로 소통 - 결과 -> 소통 -> 파급효과

[사이킷런에서 제공하는 각 상황별 알고리즘 선택 로드맵]

scikit-learn
algorithm cheat-sheet



Machine Learning Algorithms Cheat-sheet

