# data 분석과 머신러닝기본 –with R


## data분석과 머신러닝 기본

제 1편. 분석을 위한
데이터 구조 및 전처리

With 정훈희


0.intro
데이터 분석, 머신러닝 개요

# 1. 머신러닝(machine learning)

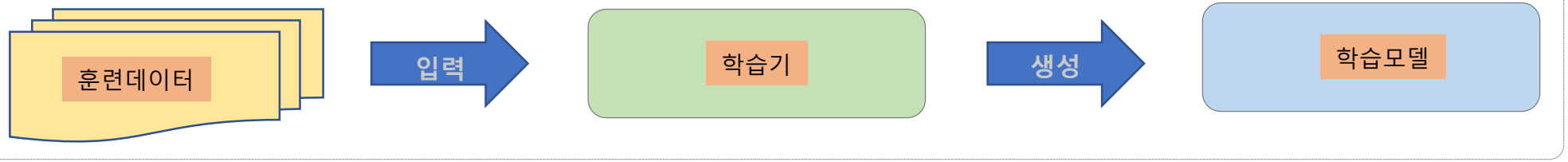

## 톰 미첼의 "머신러닝 " 정의

*"특정 작업 (T)는  성능 측정 방식( P)에 의해 측정되고, 경험 (E)에 의해 향상된다고 할 경우, 컴퓨터 프로그램은 작업( T)와 성능 측정 (P)에  관한 경험 (E)로부터 학습된다고 할 수 있다."*

시스템이 스스로 특정 작업의 수행을 학습할 수 있도록 가르치는 기법을 의미

학습 – 컴퓨터 또는 기계가 주어진 데이터로부터 어떤 패턴을 찾아낼 수 있도록 데이터 및 알고리즘에 근거한 지능을 갖게하는 것

# 머신러닝의 학습처리과정 및 종류



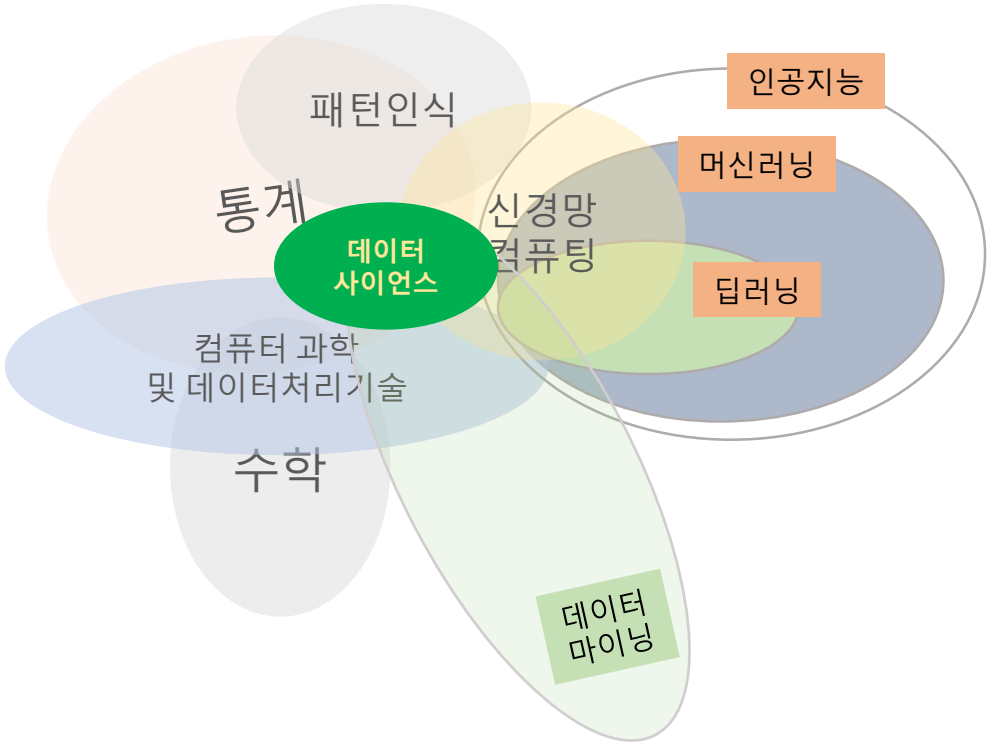

## 1. 지도 학습 알고리즘

입력데이터의 의미가 무엇인지 이미 밝혀진  데이터를 의미
훈련데이터에는 입력데이터와 지도학습데이터를 쌍으로 입력함.
사용 알고리즘 – 분류 알고리즘, 회귀 알고리즘

## 2. 비지도 학습 알고리즘

훈련데이터에 입력데이터만 입력함.
패턴인식, 서술형 모델링 작업에 주로 사용
사용 알고리즘 – 군집화  알고리즘, 연관성 규칙 알고리즘

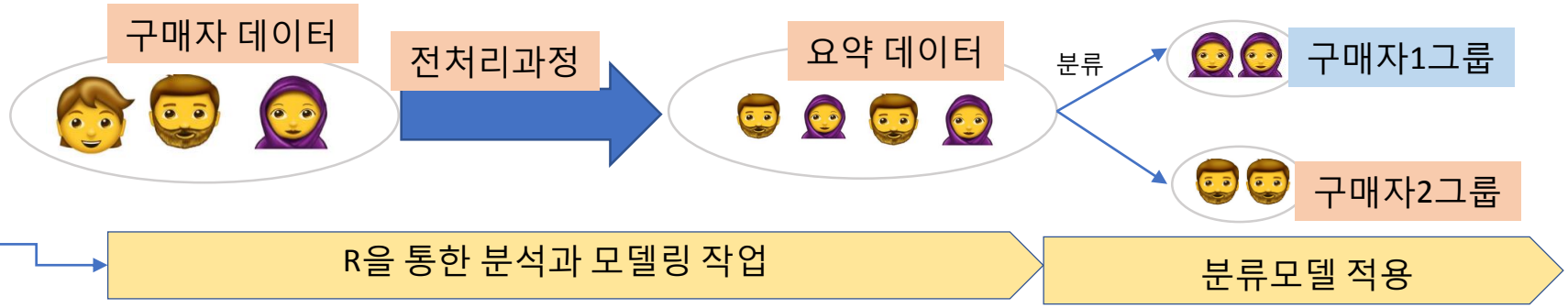
# 머신러닝 기본 베이스


패턴인식
통계
데이터
사이언스
신경망
컴퓨팅
인공지능
머신러닝
딥러닝
컴퓨터 과학
및 데이터처리기술
수학
데이터
마이닝

# R을 활용한 데이터 분석과 머신러닝

## R

- 통계용, 빅데이터 분석용 툴
- 인터렉티브형 언어
- 패키지 기반 언어
- 다양한 시각화 기법 제공
- 모든 프로그램 언어와 호환성




## R- 데이터 분석 및 머신러닝 관련 패키지

- e1071 –퓨리에 트랜스폼, 퍼지 클러스터링, 서포트 벡터 머신과 같은 알고리즘에 관련함
- randomForest – 분류회귀법에 쓰이는 랜덤 포레스트 구축에 관련된 패키지
- igraph – 통합 네트워크및 분석을 위한 패키지
- nnet – 신경망 예측모델 구축에 사용되는 패키지
- rpart – 회귀 구획 및 의사결정 트리에 중점을 둔 패키지
- arules – 연관규칙 학습 알고리즘에서 사용되는 패키지