

# Machine Learning

# 머신러닝이란 무엇인가?

- **Machine learning** is a field of computer science that uses statistical techniques to progressively improve the performance of the computer program in detecting patterns in data
- 인공지능 연구 과제 중의 하나로, 인간의 노가 자연스럽게 수행하는 학습이라는 능력을 컴퓨터로 구현하는 방법
- 예를 들면, 손글씨 문자 인식을 생각해보면 일단 수많은 손글씨 문자 데이터를 입력하고, 이러한 데이터를 기반으로 일정한 규칙이 있는지 찾아냄
- 활용범위
  - 문자인식, 음성 인식, 바둑, 장기, 게임전략, 의료진단, 로봇 개발에 이용됨

# 어떻게 규칙을 찾아갈까?

- 머신러닝은 벡터 공간이라는 것이 가장중요함
- 벡터 공간(vector)이란 크기와 방향성을 가진것을 의미함
- 특징추출
  - 머신러닝은 다양한 텍스트(이미지, 텍스트, 주가등)를 분석할 수 있는데. 분석하려면 벡터형태로 만들어야함
  - 메일을 학습기라고 입력한다면, 텍스트가 어떤 특징을 사용할 것인지 찾고 벡터로 변환해야함. 글자의 출현 빈도, 단어의 출현 빈도등을 구해 입력해야함

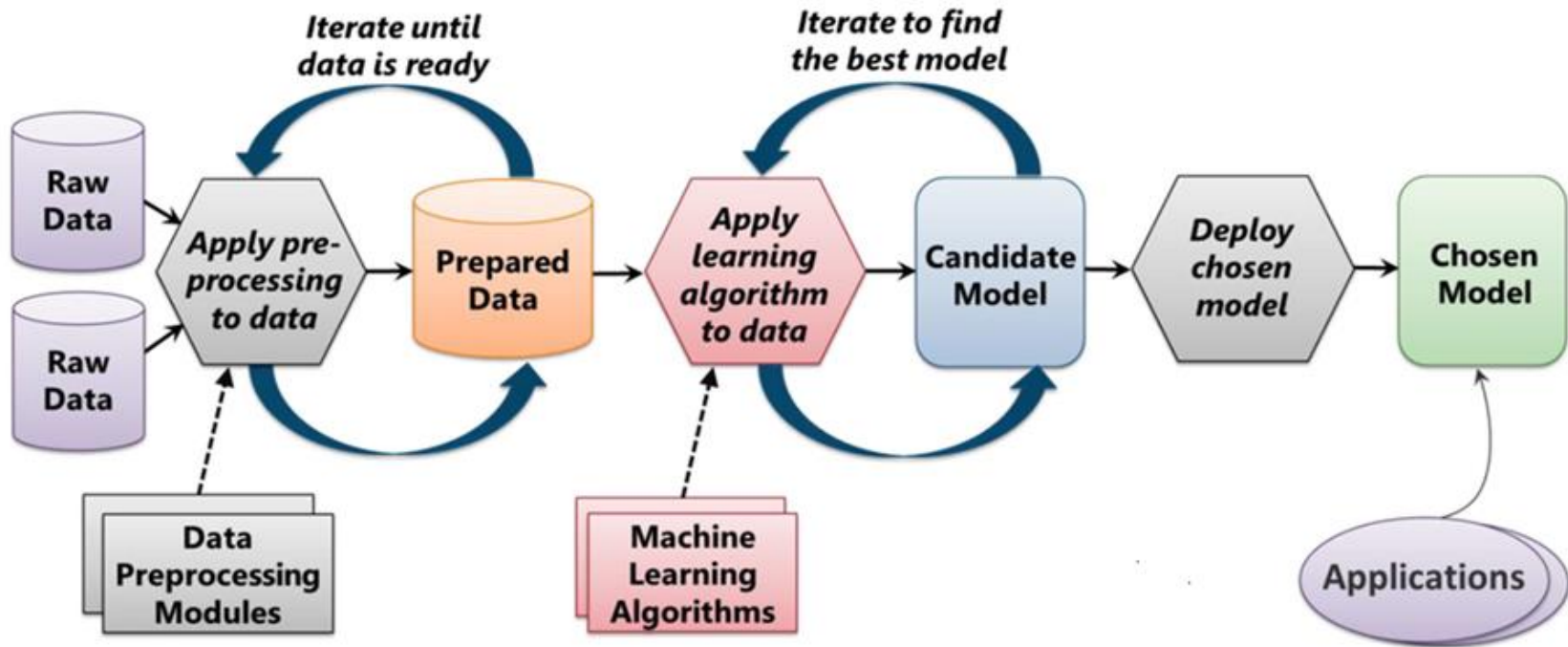
# 회귀분석이란?

- 회귀분석( regression analysis)이란 통계 용어임.
  - Y가 연동된 값일때  $Y=f(x)$ 와 같은 모델로 나타내는 것임
  - 회귀의 가장 기본적인 모델은  $Y=aX+b$ 임(선형회귀라고 함)
  - Y를 연속 측정의 종속변수(목표변수), X를 독립변수(설명변수)라고 부름.
  - X가 1차원이면 단순회귀, 2차원이상이면 다중회귀라고 함.

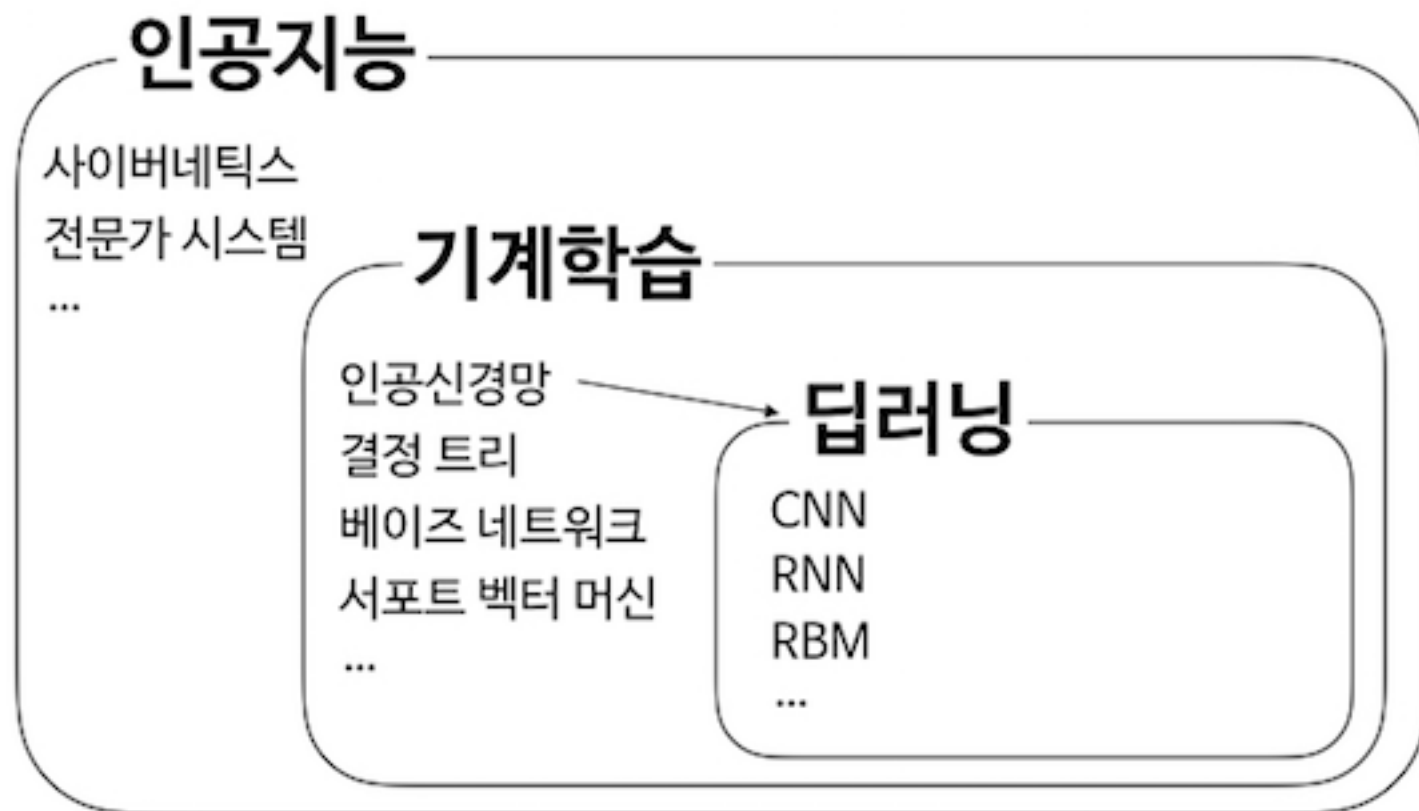
# 머신 러닝의 종류

- 교사학습(Supervised learning)
  - 데이터와 함께 답을 입력함.
  - 다른 데이터의 답을 예측함
- 비교사학습(Unsupervised learning)
  - 데이터는 입력하지만, 답을 입력하지 않음
  - 다른 데이터의 규칙성을 찾음
  - 클러스터 분석(Cluster analysis), 주성분 분석(Principal component analysis), 벡터 양자화(Vector quantization), 자기 조직화(Self organization)
- 강화학습(Reinforcement learning)
  - 부분적으로 답을 입력함
  - 데이터를 기반으로 최적의 답을 찾아냄. 현재 상태를 관찰하여 어떻게 대응해야할지와 관련된 문제를 다룸.
  - 행동의 주체와 환경(상황 또는 상태)가 등장함. 행동의 주체는 환경을 관찰하고 이를 기반으로 의사결정을 내려 행동함. 이때 어떤 환경이 변화하면서 행동의 주체가 어떤 보상을 받음.

# 머신러닝의 흐름



# 인공지능, 기계학습, 딥러닝



# 머신러닝 응용분야

- 클래스 분류(Classification)
  - 스팸 메일 분류, 필기 인식, 증권사기등등
- 클러스터링-그룹나누기(Clustering)
  - 값의 유사성을 기반으로 데이터를 여러 그룹으로 나눔. 사용자의 취향을 그룹으로 묶어서 취향에 맞는 광고 제공등등
- 추천(Recommendation)
  - 특정 데이터를 기반으로 다른 데이터를 추천하는 것임. 사용자가 인터넷 서점에서 구매한 책들을 기반으로 다른 책을 추천하는 경우
- 회귀(Regression)
  - 과거의 데이터를 기반으로 미래의 데이터를 예측하는 것임. 판매 예측, 주가 변동등을 예측함
- 차원축소(Dimensionality Reduction)
  - 데이터의 특성을 유지하면서 데이터의 양을 줄이는 것임. 특성을 유지한 상태로 고차원의 데이터를 저차원의 데이터로 변환함