

SUB_PJT_01 (AI)

서울 1반 6팀 이민아

- 사전학습
 - [사전학습1 인공지능](#)
 - [사전학습2 회귀 및 경사하강법](#)
 - [사전학습3 신경망](#)
 - [사전학습4 파이썬 라이브러리](#)

사전학습1 인공지능

1. 인공지능

(1) AI

가장 광의의 개념으로 인텔리전트한 기계를 만드는 과학과 공학

(2) 머신러닝

머신(컴퓨터, 기계, 로봇)을 학습시키는 체계 및 플랫폼을 의미

- 지도학습 (Supervised Learning)
- 비지도학습 (Un-supervised Learning)
- 강화학습 (Reinforcement Learning)

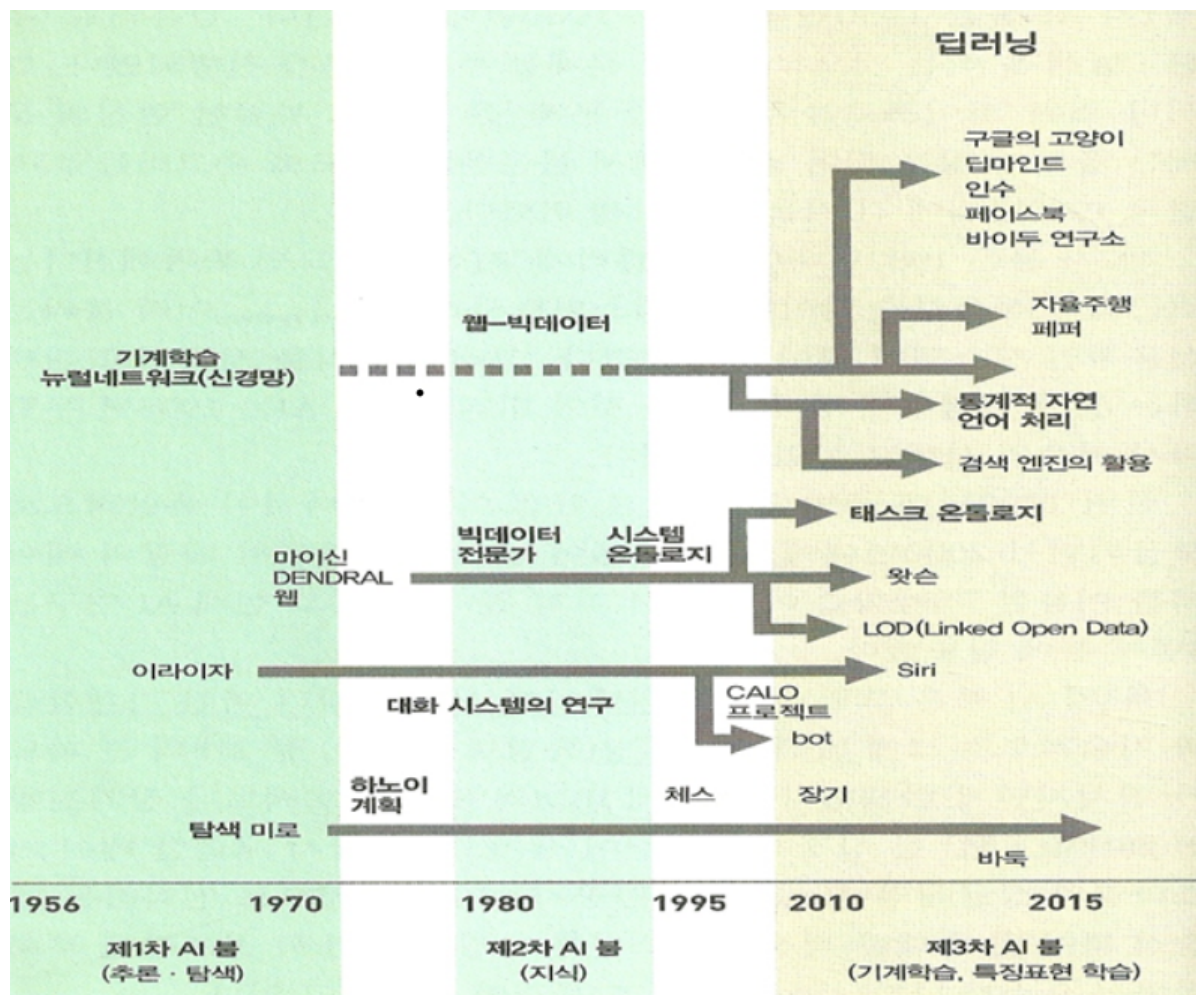
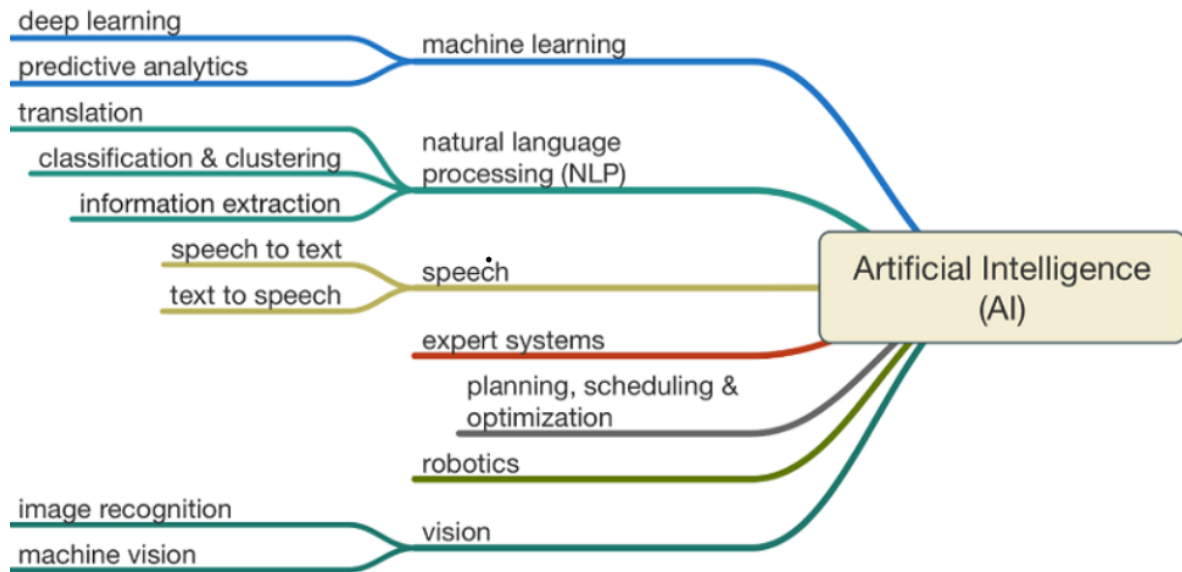
(3) 딥러닝

인간의 신경망을 차용하여 기계를 학습시킨다는 의미에서 협의의 AI로 인간의 신경망을 차용하여 학습시키는 머신러닝의 한 종류

- **입력층, 은닉층, 출력층, 활성화함수** 등
- **RBM(Restricted Boltzmann Machine)**
 - Hidden node와 Visual node로 구성된 무방향 그래프
 - DBN의 기본단위
- **DBN(Deep Belief Network)**
 - RBM이 적층되어있는 구조
 - MNIST(손글씨) 분류
- **DNN(Deep Neural Network)**
 - 입력층과 출력층 사이에 다중의 은닉층으로 구성된 신경망
 - 범용, 이미지 인식
- **CNN(Convolution Neural Network)**
 - 인간의 시신경구조 모방한 **vision처리** 수행모델
 - 이미지 인식, 컴퓨터비전
- **RNN(Recurrent Neural Network)**
 - 은닉층에서 출력층간 데이터의 저장 및 흐름 가능한 신경망
 - 언어모델링, 기계번역, 이미지캡션생성

(4) 모델

머신러닝을 통해 얻을 수 있는 최종 결과물이며 가설(Hypothesis)이라고도 불림



2. 머신 러닝

(1) 지도학습

- **정답을 주고** 학습시키는 머신러닝의 방법론
- 입력데이터에 **라벨링이 되어** 있고 출력값으로 사상되는 사상되는 함수(활성화함수)를 학습하여 성능을 향상시키는 방향
- **분류 (Classification)**
 - 값의 선택 및 분류 (**입력데이터**로 {입력, 정답}의 형태로 **정답은 '범주'의 형태**)
 - 주어진 데이터를 정해진 카테고리에 따라 분류하는 문제
 - 이메일이 스팸메일인지 아닌지를 예측한다고 하면 이메일은 스팸메일 / 정상적인 메일로 라벨링
 - 종양이 악성종양인지 / 아닌지로 구분할 수 있습니다. 이처럼 맞다 / 아니다로 구분되는 문제를 **Binary Classification**
 - 수능 공부시간에 따른 전공 학점을 A / B / C / D / F 으로 예측하는 경우도 있습니다. 이러한 분류를 **Multi-label Classification**
- **회귀 (Regression)**
 - 연관된 측정값을 통한 예측 (**입력데이터**로 {입력, 정답}의 형태 정답은 **'값'**의 형태)
 - 어떤 패턴이나 트렌드, 경향을 예측
 - 공부시간에 따른 전공 시험 점수를 예측하는 문제

기법	설명
신경망(Neural Network)	인간의 뉴런구조 차용한 모델. 입력층, 은닉층, 출력층, 활성화함수
은닉마르코프모델(HMM)	전이확률(노드간 상태전이 확률), 발생확률(이벤트발생확률) 곱. 확률적 모델
의사결정트리	트리형성, 정지규칙, 데이터의 분할로 이뤄짐. 예측목적
다층신경망(MLP)	다층 구조. 뉴런, 활성화함수, 잡음에 견고한 구조
지지벡터머신(SVM)	마진을 최대화하는 분류알고리즘. 빠른 학습
베이지언망(Bayesian Network)	확률적 모델. 변수간의 의존관계 학습

(2) 비지도학습

- **정답없는 데이터를 어떻게** 구성되었는지를 알아내는 머신러닝의 학습 방법론
- 입력데이터는 **라벨링이 되어있지 않고** 입력된 패턴의 공통적인 특성을 파악하는 것이 목적
- **군집** : 그룹핑 (비슷한 데이터들끼리 묶어주는 기능)
- **특징** : 특징 도출

기법	설명
K-means 알고리즘	군집별 중심에서 거리를 기반으로 그룹분류. K개의 군집.
군집화(Clustering)	데이터 그룹핑. 데이터집합을 군집화
계층적 군집화	전체 데이터 계층 분할
자기조직지도(SOM)	층구조 신경망. 데이터의 가시화.
주성분분석(PCA)	분산큰 데이터의 차원 축소. 특징 도출.
독립성분분석(ICA)	독립성분 분석. 요인 분석
EM(Expectation Maximisation)	통계모델. 최대가능도의 획득

(3) 강화학습

- 입력된 데이터(행동)의 선택에 따른 보상에 따라 보상치를 극대화

기법	설명
몬테카를로 시뮬레이션	난수를 사용하여 함수의 값을 확률적으로 계산
Q-Learning	유한 마르코프 과정의 최적정책 결정

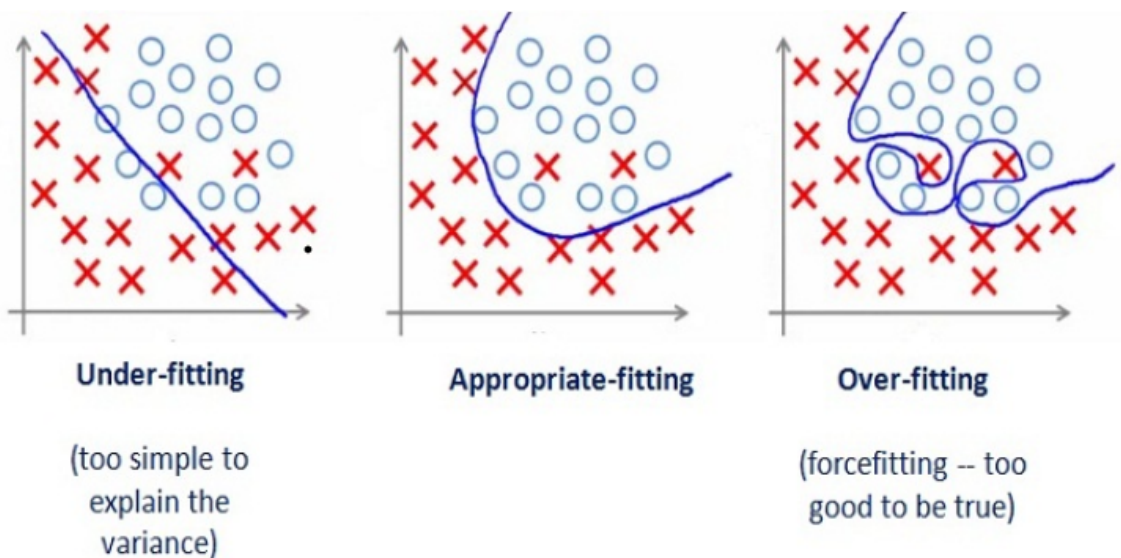
(4) 프로세스

- 데이터 수집
- 입력 데이터 정제
- 데이터 분석(규칙, 가설)
- 모델(알고리즘) 선택, 훈련
- 검증, 테스트 진행
- 사용 및 배포

(5) 과적합 (Overfitting)

- **Overfitting**

학습 데이터에 너무 최적화를 하다보니, 실제 데이터와 차이가 많이 발생하는 모델 생성



- **Regularization**

- 학습 데이터를 조금 희생하더라도 모델을 최대한 간단하게 만들어서 **Overfitting 방지 기법**
- 눈으로 보고 쉽게 확인할 수 있지 않기 때문에 학습 과정에서의 Overfitting 판단 어려움
- **Validation**
 - **Overfitting 여부를 손쉽게 판단하기 위해**

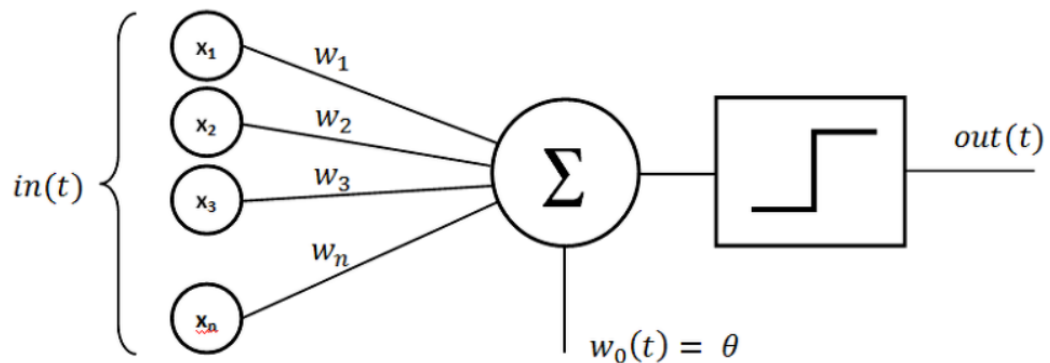
(6) 데이터

- 훈련데이터(Training Data)
- 시험데이터(Test Data)

3. 딥러닝

(1) 퍼셉트론

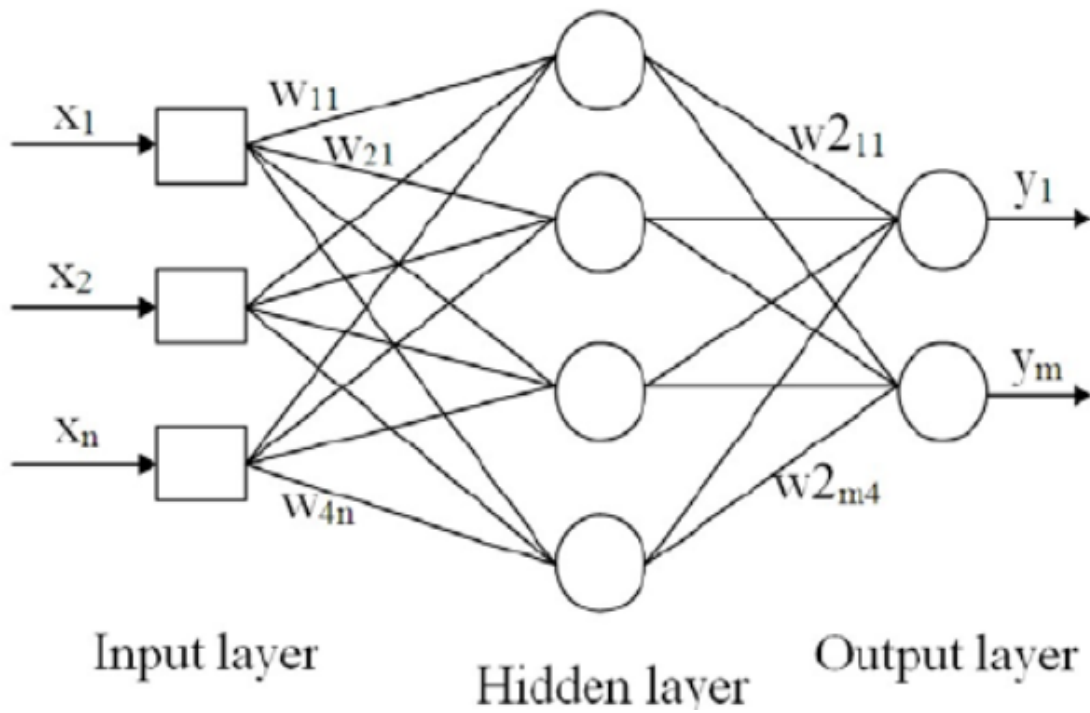
- 퍼셉트론
 - 1957년 미국의 심리학자 프랑크 로젠블라트(Frank Rosenblatt)에 의해 고안된 **인공신경망 최초의 알고리즘**
 - 다수의 입력과 하나의 출력
 - 뉴런이라고도 부르며 뉴런은 입력, 가중치, 활성화함수, 출력으로 구성



- 다층 퍼셉트론
 - 퍼셉트론은 AND, OR, NANA 같은 선형문제는 해결(분류) 가능하지만 **XOR 같은 비선형 문제**는 해결 불가능
 - XOR 문제를 분류하기 위해서는 직선이 아닌 **곡선으로 분류** 가능함
 - 단층 퍼셉트론에서 층을 증가시킴으로써, XOR 문제 해결 가능

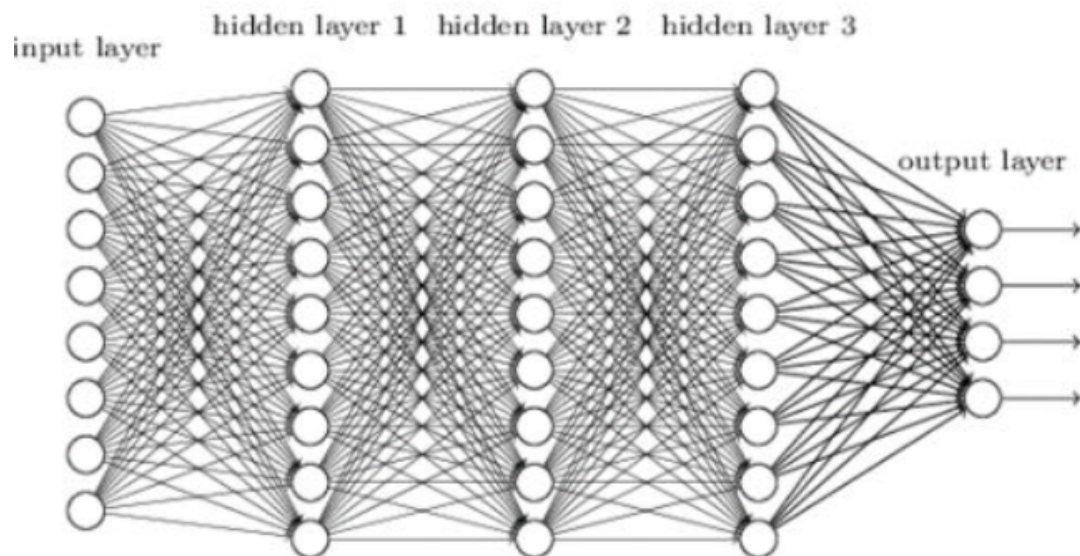
(2) 인공신경망 (ANN)

- 단층 퍼셉트론을 기반으로 생물학적 신경망 영감을 받은 **통계 학습 알고리즘**
- 심층신경망(DNN)의 기초



(3) 심층신경망 DNN(Deep Neural Network)

- 정의
 - 인공신경망에서 은닉 계층을 증가시킨 신경망
 - 일반적으로 은닉층이 3개 이상일 경우, 심층신경망이라고 표현
 - 딥러닝에 사용
- 프로세스
 - 학습에 사용할 데이터를 로드한다. (**Matrix** 형태로 변환)
 - 입력층에서 가중치와 편향 값을 **활성화 함수**와 **경사 하강법**을 통해 출력값을 은닉층으로 전달한다.
 - 출력층까지 연산을 반복한다.
 - 출력값을 바탕으로 **오차 역전파**를 수행한다.
 - **최적의 손실 함수 값**을 구할때까지 반복한다.



(4) 자연어처리 (NLP)

- 인간의 언어 형상을 컴퓨터와 같은 기계를 이용해서 모사할 수 있도록 연구하고 이를 구현하는 인공지능의 주요 분야 중 하나

(5) 딥러닝 적용 사례

<https://brunch.co.kr/@itschloe1/23>

4. One-Hot-Encoding (1-of-Encoding)

- 결과가 0과 1이 아닌, 3가지 이상의 범주를 가질 때
- 출력 노드를 범주 개수만큼, 그리고 각 자릿수마다 범주를 나타내도록 0과 1로 표현하는 방식

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

France = [0, 0, 0, 1, 0, 0, ..., 0]

