

< LEARNING TYPE >

1. Supervised Learning (지도학습)

입력, 정답(label) 이 있는 data 학습

Linear Regression 집 크기/방 개수 보고 가격 예측
 Logistic Regression Email Spam 유무 분류
 Decision Tree/ Random Forest 환산 나이/혈압/혈액 수치 보고 심장병 여부 예측
 Support Vector Machine 문서의 단어 빈도수로 뉴스기사 Category 분류
 Neural Network < 손글씨 숫자 (Mnist) 분류
 기온/습도/바람 보고 다음날 기온 예측

2. Unsupervised Learning (비지도학습)

정답(label) 없이 data 구조, 패턴 학습

K-Mean Clustering 신상품 소비패턴 분석 → 3가지 (절약형/일반/소비) 자동 군집화
 Hierarchical Clustering 유전자 발현 데이터 분석 → 유사 유전자 군집 구성
 Principal Component Analysis 얼굴분석 (300차원) → 2차원 축소해 비슷한 얼굴들이 비슷한 위치에 분포되게 ...
 Autoencoders 흐릿한 MRI (노이즈 낀 이미지)를 깨끗하게 복원

3. Semi-Supervised Learning (준지도학습)

일부만 label 있고, 나머진 없음

Label Propagation label이 일부만 있는 소셜Network에서 사람들의 정치성향 추정
 Semi-Supervised SVM 기사 일부만 주제 명시 → 나머지 자동 주제 분류
 Mean Teacher, Fix Match labeled 500 장 + unlabeled 5000 장 이미지로 폐경 기간 모델 훈련.

4. Self-Supervised Learning (자기지도학습)

data에서 정답 생성 (deep learning ...)

Masked Language Modeling (BERT) - 문장에서 단어 일부 가려 예측
 Contrastive Learning (SimCLR, MoCo) - 두 장의 augment 된 이미지를 같은 것으로 인식
 Next Token Prediction (GPT) - 수많은 문장 기반으로 다음 단어 예측

5. Reinforcement Learning (강화학습)

Agent가 환경과 상호작용하며 보상 최대화

Q-Learning ex) 게임AI가 플레이어와 스코어 전략 학습 (AlphaGo)
 Deep Q-Network
 Proximal Policy Optimization ex) 자율주행 자동차가 도로 주행행위 학습
 : ex) 로봇이 물건 잡는 행동 학습

< TASK TYPE >

1. Classification : data를 여러 class 중 하나로 분류

Logistic Regression, SVM, Decision Tree, Neural Network

ex) 개/고양이 이미지 분류, Email Spam, 질병 유무 예측

2. Regression : 연속적인 수치값 예측

Linear Regression, Ridge, SVR, Neural Network

ex) 집값 예측, 주식 가격 예측, 온도 예측

3. Clustering : label 없는 data를 유사성 따라 그룹화

K-Mean, DBSCAN, Hierarchical Learning

ex) 뉴스기사 1000개를 토픽(unlabeled) 기반으로 자동 군집화

4. Dimensionality Reduction : 고차원 data를 저차원으로 축소

PCA, t-SNE, Autoencoder

ex) 시각화 - 300d 임베딩된 벡터를 2D로 줄여서 유사한 단어가끼리 뭉치도록 시각화

전처리 - text에서 불필요한 단어 제거한 후 주요단어들로 구성된 핵심 특징만 남겨 학습시간 ↓

노이즈제거 - 오래된 흑백사진에서 배경의 noise 제거하고 인물 윤곽만 남기는 ... etc

5. Anomaly Detection : 정상 pattern 벗어난 data 감지

One-Class SVM, Isolation Forest.

ex) 금융사기 탐지.