지도학습/비지도학습/강화학습

지도학습 : 정답을 알고 학습하는 유형의 데이터 분석

입력 값(x)과 정답(y, label)을 포함하는 훈련용 데이터를 이용하여 학습하고 그 학습된 결과를 바탕으로 미지의 데이터(test data)에 대해 미래 값을 예측하는 방법

정답에 해당하는 값을 목적변수(target variable) 또는 레이블(label)이라고 부른다.

회귀 : 수치 값

분류 : 카테고리 변수

예) 스팸메일 분류기의 학습 : 정답 샘플도 같이 주어져야 한다.

회귀 : 회귀 분석은 관찰된 연속형 변수들에 대해 두 변수 사이의 모형을 구한 뒤 적합도를 측정해 내는 분석 방법이다. 즉, 수치를 예측하는 것

회귀 분석의 응용 : 경제지표 예측, 사회학 연구, 마케팅 등

회귀 분석 알고리즘 : 선형 회귀, KNN, SVM, 로지스틱 회구, 랜덤 포레스트, 신경망

분류 : 어떤 항목이 어느 그룹에 속하는지를 판별

이진 분류 : 두 가지 카테고리를 나누는 작업

다중 분류 : 세 개 이상의 클래스를 나누는 작업

분류의 응용 : 스펨메일 분류, 우수 고객 분류 등

비지도 학습 : 정답(label)은 없고 입력 데이터만 있는 훈련용 데이터(training data)를 이용한 학습을 통해 정답을 찾는 것이 아닌 입력 데이터의 패턴, 특성 등을 발견하는 방법

기법 : 군집화(유사한 항목들을 같은 그룹으로 묶는다) / 시각화(데이터의 속성을 명확하게 시각화하기 위해서 고차원의 특성 값들을 2차원이나 3차원으로 차원을 축소하는 작업) / 데이터 변환(데이터를 분석하기 좋게 다른 형태로 변환) / 주성분 분석(PCA)(머신 러닝에 사용할 특성의 수를 줄인다)

연관 분석 : 어떤 사건이 다른 사건과 얼마나 자주 동시에 발생하는지 파악, 자주 발생하는 패턴 찾기(상품의 연관성, 취향의 연관성 등 분석), 같이 구매한 상품 분석, 상품으 진ㄴ열 배치 및 상품 프로모션(쿠폰 발행 등)에 활용

강화 학습 : 일정 기간 동안의 행동에 대해 보상을 해줌으로써 어느 방향으로 학습해야 하는지 방향성만 알려주는 학습 방법 / 시간이 흐르면서 모델이 바람직한 방향으로 가고 있는지를 알려줄 수 있고 이를 통해 학습하는 방법

응용 예) 게임의 경우 입력시마다 답을 주지는 못하지만 게임을 이기고 있는지 지고 있는지를 알려줌, 알파고(강화학습 사용)