

1. 지도 학습(Supervised learning)

지도 학습 알고리즘은 한 세트의 사례들을(examples) 기반으로 예측을 수행. 예를 들어, 과거 매출 이력(historical sales)을 이용해 미래 가격을 추산할 수 있다. 지도 학습에는 기존에 이미 분류된 학습용 데이터(labeled training data)로 구성된 입력 변수와 원하는 출력 변수가 수반되는데. 알고리즘을 이용해 학습용 데이터를 분석함으로써 입력 변수를 출력 변수와 매핑시키는 함수를 찾을 수 있다. 이렇게 추론된 함수는 학습용 데이터로부터 일반화(generalizing)를 통해 알려지지 않은 새로운 사례들을 매핑하고, 눈에 보이지 않는 상황(unseen situations) 속에서 결과를 예측한다.

- **분류(Classification):** 데이터가 범주형(categorical) 변수를 예측하기 위해 사용될 때 지도 학습을 ‘분류’라고 부르기도 한다. 이미지에 강아지나 고양이와 같은 레이블 또는 지표(indicator)를 할당하는 경우가 해당되는데. 레이블이 두 개인 경우를 ‘이진 분류(binary classification)’라고 부르며, 범주가 두 개 이상인 경우는 다중 클래스 분류(multi-class classification)라고 부른다.
- **회귀(Regression):** 연속 값을 예측할 때 문제는 회귀 문제가 된다. 시계열 데이터등.
- **예측(Forecasting):** 과거 및 현재 데이터를 기반으로 미래를 예측하는 과정이다. 예측은 동향(trends)을 분석하기 위해 가장 많이 사용된다. 예를 들어 올해와 전년도 매출을 기반으로 내년도 매출을 추산하는 과정.

2. 준지도 학습(Semi-supervised learning)

지도 학습은 데이터 분류(레이블링) 작업에 많은 비용과 시간이 소요될 수 있다는 단점을 지닌다. 따라서 분류된 자료가 한정적일 때에는 지도 학습을 개선하기 위해 미분류(unlabeled) 사례를 이용할 수 있는데. 이때 기계(machine)는 온전히 지도 받지 않기 때문에 “기계가 준지도(semi-supervised)를 받는다”라고 표현. 준지도 학습은 학습 정확성을 개선하기 위해 미분류 사례와 함께 소량의 분류(labeled) 데이터를 이용한다.

3. 비지도(자율) 학습(Unsupervised learning)

비지도 학습을 수행할 때 기계는 미분류 데이터만을 제공 받는다. 그리고 기계는 클러스터링 구조(clustering structure), 저차원 다양체(low-dimensional manifold), 희소 트리 및 그래프(a sparse tree and graph) 등과 같은 데이터의 기저를 이루는 고유 패턴을 발견하도록 설정된다.

- **클러스터링(Clustering):** 특정 기준에 따라 유사한 데이터 사례들을 하나의 세트로 그룹화 한다. 이 과정은 종종 전체 데이터 세트를 여러 그룹으로 분류하기 위해 사용되는데. 사용자는 고유한 패턴을 찾기 위해 개별 그룹 차원에서 분석을 수행할 수 있다.
- **차원 축소(Dimension Reduction):** 고려 중인 변수의 개수를 줄이는 작업. 많은 애플리케이션에서 원시 데이터(raw data)는 아주 높은 차원의 특징을 지니는데. 이때 일부 특징들은 중복되거나 작업과 아무 관련이 없다. 따라서 차원수(dimensionality)를 줄이면 잠재된 진정한 관계를 도출하기 용이해짐.

4. 강화 학습(Reinforcement learning)

강화 학습은 환경으로부터의 피드백을 기반으로 행위자(agent)의 행동을 분석하고 최적화한다. 기계는 어떤 액션을 취해야 할지 듣기 보다는 최고의 보상을 산출하는 액션을 발견하기 위해 서로 다른 시나리오를 시도다. 시행 착오(Trial-and-error)와 지연 보상(delayed reward)은 다른 기법과 구별되는 강화 학습만의 특징임.