머신 러닝의 이해와 지도학습을 이용한 분류

<머신 러닝>

인공지능이 스스로 똑똑해질 수 있게 만드는 기술

컴퓨터에게 사람이 직접 로직을 설명하지 않아도 데이터를 통해서 ‘학습’하고 일정한 규칙을 찾아내고, 이를 바탕으로 다른 데이터를 분류하거나 미래를 예측하는 것

<지도 학습>

훈련 데이터에 정답(Lable)이 있어 주어진 정답에 맞게 특징이 학습되기를 기대하는 경우에 사용

분류(Classification) -> 분류 개수에 따라 이중 분류(Bi-Classification), 다중분류(Multi-Classification)

이중 분류는 두 개의 군집으로 나누는 것을 의미, 텍스트 분석에서는 긍정과 부정을 구분하는 감정 분석(Sentiment Analysis)에서 활용

다중 분류는 이미지에서 사물이나 동물을 분류하는 이미지 태깅(Image Tagging)이나 텍스트 분석에서 토픽 모델링(Topic Modeling)에서 활용

회귀(Regression) -> 훈련 데이터의 값을 분석해 주어진 값에 대한 결과 수치를 예측

주어진 값에 따르는 최적의 예측 함수를 찾는 것

<비지도 학습>

훈련 데이터에 정답이 없고 주어진 특징 내에서 분류간 서로 구분이 잘 되도록 나누기 원하는 경우에 사용

군집화 -> 특징이 비슷한 데이터끼리 묶는 것

차원 축소 -> 특성이 많은 데이터를 특징의 수를 줄이면서 꼭 필요한 특징을 포함한 데이터로 표현하는 방법

<강화 학습>

학습(혹은 수집)되는 데이터에 정답은 없으나 동작하거나 반응하는 결과에 상과 벌을 주어서 스스로 진화할 수 있는 경우에 사용