

머신러닝 1강

특징벡터 : 분석 대상의 성질을 나타내는 데이터

로지스틱 회귀는 '분류할 규칙'을 만들어 내는 목적으로 이용됨

1.2 머신러닝 알고리즘 분류

1.2.1 분류 : 클래스 판정을 산출하는 알고리즘

퍼셉트론, 로지스틱 회귀 : 분류를 위한 알고리즘.

여러 클래스로 분류된 기존의 데이터를 토대로 새로 수집된 데이터가 어느 클래스에 속하는가를 예측하는 규칙을 만들어 냄.

ROC(Receiver Operating Characteristic) curve : 알고리즘 정확도 분포곡선

회귀분석 : 수치를 예측하기 위함에 그 목적이 있음.

기존에 수집된 데이터의 배경에 어떠한 함수가 숨어 있을 것이라고 예상하고 그 함수가 무엇인지 추측한다.

1.2.3 클러스터링 : 지도자 없이 그룹화하는 알고리즘

지도학습(Supervised Learning) : 문제의 답이 데이터에 주어져 있는 상태에서 그 데이터를 가지고 일반적인 판정 규칙을 만들어 내는 수법

ex) 휴대폰 갈아타기, 손글씨 자동인식

비지도학습(Unsupervised Learning) : 기계학습의 일종으로, 데이터가 어떻게 구성되었는지를 알아내는 문제의 범주에 속한다. 이 방법은 지도학습 혹은 강화학습

과는 달리 입력값에 대한 목표치가 주어지지 않는다.

cf) 휴대폰 갈아타기 : 특별한 목적을 미리 설정해서 분류한 것이 아닌 프로필 정보나 이용 상황 등을

기준으로 하여 자연스럽게 몇 가지 그룹(클러스터)이 형성됐는지 발견하는 것이 비지도학습의 예시..

cf) 손글씨 자동인식 : 수많은 손글씨 데이터(이미지 데이터)를 수집해서 각 글씨 이미지가 나타내는 문자를 인쇄체로 태그에 써서 붙인 후에 컴퓨터가 글씨 이미지와 태그를 함께 읽으며 자동 인식 규칙을 생성해 가는 방식 //지도학습!!

그러나 수많은 손글씨 데이터를 사람이 일일이 판단해서 태그를 붙인다는 것은 매우 힘든일이기에 컴퓨터가 손글씨 이미지 데이터만을 읽어 분류하는 것 //비지도학습!!

1.2.3 정리 : 이처럼 각 그룹의 특징을 확인한다면 새로운 서비스나 약정을 만들어 낼 때에 참고가 될 것임!!!

또한 각각의 그룹에서 신규 서비스를 이용할 확률을 구하는 등 새로운 머신러닝의 소재로 이용할 수도 있음!!!

1.2.4 그 밖의 알고리즘

1) 유사성 매칭

새로 수집한 데이터가 기존의 어떤 데이터와 유사한지 판정하는 것.

ex) 기업에서 새로운 고객을 찾아 영업을 시작할 때, 이 새 고객이 기존의 어떤 고객과 유사한지를 조사하여 그 조사 결과에 맞춰 영업 방침을 정할 수 있을 것이다. 기존의 최우수 고객과 유사한 고객을 찾을 수 있다면 영업 실적을 더욱 높일 수 있을 것

2) 동시발생 분석

연관성 분석이라고도 불림. 동시에 발생하는 사건을 기존 데이터에서 발견하는 것으로 '비지도 학습'의 한 종류이다.

3) 링크 예측

데이터 사이에 잠재적으로 존재하는 '연결고리'를 예측하는 기법임.

ex) SNS에서 인간관계를 예측하기 위해 이용됨. 친구로 등록되어 있지 않은 A와도 친구, B와도 친구인 사람들이 많다면 'A와 B도 실생활에서도 친구이지 않을까?' 하고 예측하는 것이다.