

ML

박정웅

이 문서는 나눔글꼴로 작성되었습니다. [설치하기](#)

머신러닝

- 데이터를 이용해서 명시적으로 정의되지 않은 패턴을 컴퓨터로 학습하여 결과를 만들어내는 학문 분야
- 중요 요소
 1. 데이터 : 머신러닝은 알고리즘이 아니라
데이터 학습을 통해 실행 동작이 바뀜
 2. 패턴인식 : 데이터를 보고 패턴을 찾기가 핵심
 3. 컴퓨터를 통한 계산

머신러닝 분류

- 풀고자 하는 목표에 따른 분류

- 지도학습 : 주어진 데이터와 레이블(정답)을 이용해서
미지의 상태나 값을 예측하는 학습 방법
: 대부분의 머신러닝이 여기에 해당

- 회귀 : 숫자값을 예측. 보통 연속된 숫자(실수)를 예측

- 분류 : 입력 데이터들을 주어진 항목들로 나누기

- 랭킹 / 추천

- : 데이터들의 순위를 예측

머신러닝 분류

- 풀고자 하는 목표에 따른 분류
 - 비지도 학습 : 데이터 그 자체에서 유용한 패턴을 찾는 학습방법
 - 군집화 / 토픽 모델링
 - 밀도 추정 : 데이터 분포 예측 기법
 - 차원 축소 : 데이터의 차원을 낮추는 기법
 - 강화학습 : 기계(에이전트)가 환경과의 상호 작용 (선택과 피드백[보상] 반복)을 통해 장기적으로 얻는 이득을 최대화하도록 하는 학습방법

머신러닝 분류

- 기법에 따른 분류
 - 통계
 - 딥러닝

머신러닝 주요 개념

- 모델 : 데이터를 바라보는 시점 혹은 가정
 - 간단한 모델 : 선형 모델(회귀)
 - 복잡한 모델 : 결정 트리 모델
 - 구조가 있는 모델
 - 1) 순차 모델 : 연속된 관측값이 서로 연관성이 있을 때 주로 사용
: 문서의 텍스트, 시간과 관계된 데이터 분석에 주로 사용
 - 2) 그래프 모델 : 그래프를 이용해서 순차 모델보다 더 복잡한 구조를 모델링
: 문서 텍스트의 문법 구조(보통 트리 형태)를 직접 모델링
하거나 이미지의 픽셀 사이 관계를 그래프로 표현하여 모델링

머신러닝 주요 개념

- 손실함수(Loss Function)

- : 모델의 수식화된 학습 목표
- : 모델이 실제로 데이터를 바르게 표현했는지 혹은 얼마나 예측이 정확한지 수학적으로 표현하는 방식

- 1) 산술 손실 함수

- : 모델로 산술값을 예측하고 데이터에 대한 예측값과 실제 관측값을 비교하는 함수

- 2) 확률 손실 함수

- : 모델로 항목이나 값에 대한 확률을 예측하는 경우에 사용

- 3) 랭킹 손실 함수

- : 모델로 순서를 결정할 때 사용

- 4) 모델 복잡도와 관련된 손실함수

- : 위 3가지 손실함수들과 결합하여 모델이 필요 이상으로 복잡해지지 않도록 방지하는 손실함수

머신러닝 주요 개념

- 최적화

- : 실제로 학습을 하는 방법

- : 손실함수의 결과값을 최소화 하는 모델의 인자를 찾는 과정

- 1) 경사하강법

- : 임의의 지점에서 시작해서 경사를 따라 내려갈 수 없을 때 까지 반복적으로 내려가며 최적화를 수행

- 2) 뉴턴/준 뉴턴 방법

- 3) 확률적 경사하강법 - 뉴턴/준 뉴턴 방법(계산시 자원을 많이 사용)을 보완

- : 일부 데이터만 이용해서 손실함수와 1차 미분값을 근사적으로 계산하는 방법

- 4) 역전파 - 딥러닝에서 많이 사용하는 방법

머신러닝 주요 개념

- 모델 평가

- : 실제 활용에서 성능을 평가

- : 학습 데이터 뿐만 아니라 새로운 데이터가 들어왔을 때
도 잘 동작하는지 평가

모델의 주요 개념

- 데이터의 유형

- 1) 텍스트 데이터

- 2) 수치 데이터

- 3) 이미지 데이터 / 비디오 데이터

- 4) 음성 데이터

- 5) 복합 데이터

모델 주요 개념

- 문제 유형과 해결 방법

- 1) 회귀 문제 (가장 기본)

- 입력을 받아 가장 적합한 숫자값을 예측하는 문제
 - 기법 : 선형 회귀, 가우시안 프로세스 회귀, 칼만 필터

- 2) 분류 문제 (기본)

- 보통 손실함수를 직접 최적화해서 푸는 방법을 많이 사용
 - 기법 : 로지스틱 회귀, 서포트 벡터 머신, 신경망

- 3) 군집화 문제

- 기법 : k-평균 군집화, 토픽 모델링, 평균 이동 군집화

- 4) 표현형 학습(임베딩 학습)

- : 풀고자 하는 문제에 적합한 표현형(representation)을 데이터로부터 추출하는 방법

- 기법 : word2vec 모델과 그 파생모델, 행렬 분해



이 문서는 나눔글꼴로 작성되었습니다. [설치하기](#)