

[ML,DL] 머신러닝에 대한 간단한 개념들과 사용 하는 주요 패키지 — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2020-11-15 오후 10:31

URL: <https://continuous-development.tistory.com/159?category=736685>

ML,DL

[ML,DL] 머신러닝에 대한 간단한 개념들과 사용 하는 주요 패키지

2020. 10. 27. 00:39 수정 삭제 공개

머신러닝이란?

머신러닝이란 애플리케이션을 수정하지 않고도 데이터를 기반으로 패턴을 학습하고 결과를 추론하는 알고리즘 기법

현실 세계의 매우 복잡한 조건으로 인해 기존의 소프트웨어 코드만으로 해결하기 어려운 문제들이 많이 생겼다. 이걸 머신러닝을 통해 해결해나가고 있다.

머신러닝이 필요한 이유

현재 시대에서 다양한 복잡도를 사람이 할 수 없는 정도의 방대한 코드가 되었다.

이것을 사람이 직접 알고리즘을 짜기에는 어렵다.

그래서 컴퓨터에 데이터를 넣고 숨겨진 패턴을 발견해 컴퓨터가 알고리즘을 직접만들어 예측을 하게 한다.

머신러닝의 종류

머신러닝은 지도학습 / 비지도 학습 / 강화 학습이 있다.

지도 학습

- 분류, 회귀, 시각/음성 감지/인지

지도 학습은 정답이 포함되어 있는 데이터를 특정한 알고리즘을 통해 학습시킨다. 이것을 통해 새로운 데이터가 들어왔을 때 회귀(예측)와 분류가 가능하다.

비지도 학습

- 군집화(클러스터링), 차원 축소

비지도 학습은 말 그대로 정답이 없다. 이런 상태에서 학습을 통해서 군집화를 하겠다는 뜻이다.

강화 학습

강화 학습은 행동에 대해 보상을 준다. 이렇게 보상을 최대화하는 방법을 찾는 게 강화 학습이다.

머신러닝의 단점

데이터에 의존적이다. 안 좋은데 데이터를 넣을 경우 좋지 않은 결과가 나온다.

과적합 되기 쉬움, 그래서 이걸 위해 다양한 기법들을 사용한다.

논리적인 이해가 어렵다(블랙박스) 결과만 나올 뿐 왜 이렇게 되었는지 알 수 없다.

R과 python 비교

R- 통계전용 프로그램 (분석 툴) / 다양한 시각화를 가지고 있다.

python - 확장성 / 유연성이 매우 뛰어나며 지금 머신러닝과 딥러닝 쪽에 상당한 지분을 가지고 있다.

대부분의 deep learning은 python으로 제공되고 있다.

텐서 플로우가 2.0으로 넘어가면서 케라스를 흡수했다.

파이썬 머신러닝 생태계를 구성하는 주요 패키지



scikit learn - 머신러닝 패키지로서 머신러닝을 하는 데 있어서 다양한 패키지를 제공해준다.

numpy, Scipy - numpy와 scipy를 통해 배열과 선형대수, 통계 패키지 등을 통해 데이터를 다룰 수 있다.

pandas - 데이터 내가 원하는 형태로 가공할 수 있다.

matplotlib, seaborn - pandas로 핸들링 한 데이터를 시각화해주는 툴이다.

'ML,DL' 카테고리의 다른 글

[ML/DL] 데이터 인코딩 - Label Encoding / One-hot Encoding/ dummies

[ML/DL] 파이썬(python)을 이용한 분류(Classification)하기

[ML/DL] 대체법의 종류와 다중 대체법 사용법

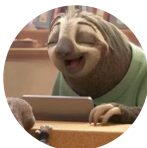
[ML,Dl] 머신러닝에 대한 간단한 개념들과 사용 하는 주요 패키지

[ML/DL]결측치의 종류와 결측치 처리 가이드라인

[ML,DL] 머신러닝(Machine larning)과 딥러닝(Deep larning)의 정의와 차이점

⌕

머신러닝 기본 개념



나아무늘보

혼자 끄적끄적하는 블로그 입니다.