

41. 더 나아가기

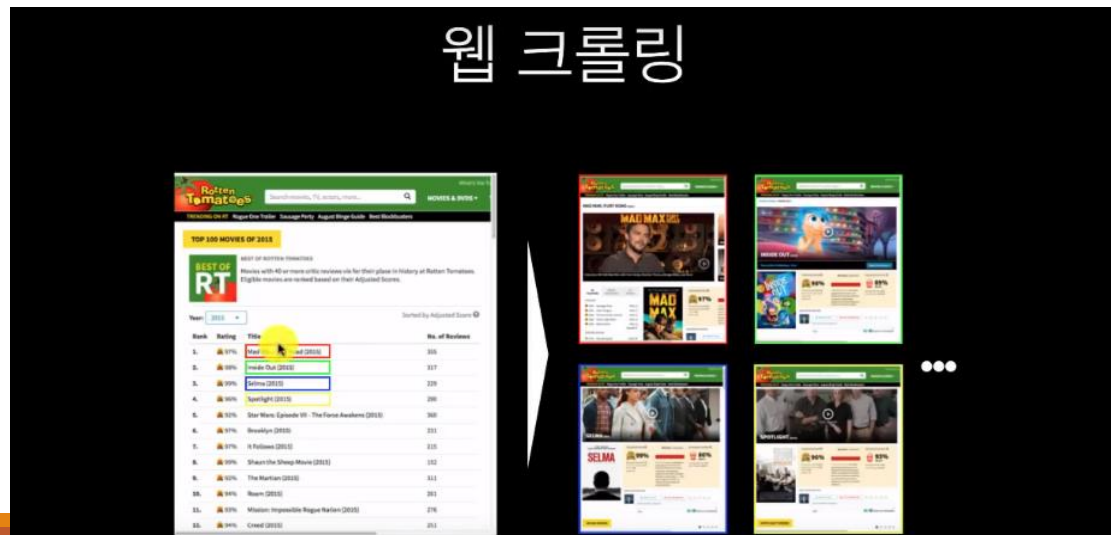
2018.12

일병 김재형

웹 크롤링과 스크래핑

정의

- 크롤링(crawling)
 - 웹 페이지의 하이퍼링크를 순회하면서 웹 페이지를 다운로드
 - 너무 많은 요청은 서버에 부하를 가하며,
DoS(Deny of Service, 서비스 거부 공격)이 될 수 있다.



웹 크롤링과 스크래핑

정의

- 스크래핑(scraping)
- 다운로드 한 웹 페이지에서 필요한 정보를 자동으로 추출

웹 스크래핑



항목	데이터
title	MAD MAX: FURY ROAD
rating	97
director	George Miller
Genre	Action & Adventure, Science Fiction & Fantasy
posterimg	https://resizing.flixster.com/Z-t-4ec5aR9VQpOT-BfHJER-0-720x305/v1bTwmTESMT1B4teqOZE3MTM5CzEYMDA7NTAGZc1Q



웹 크롤링과 스크래핑

사용 모듈

- lxml, BeautifulSoup, Scrapy, Selenium 등을 사용할 수 있다.



<https://www.digitalocean.com/community/tutorials/how-to-scrape-web-pages-with-beautiful-soup-and-python-3>

<https://blog.theodo.fr/2018/02/scrape-websites-5-minutes-scrapy/>

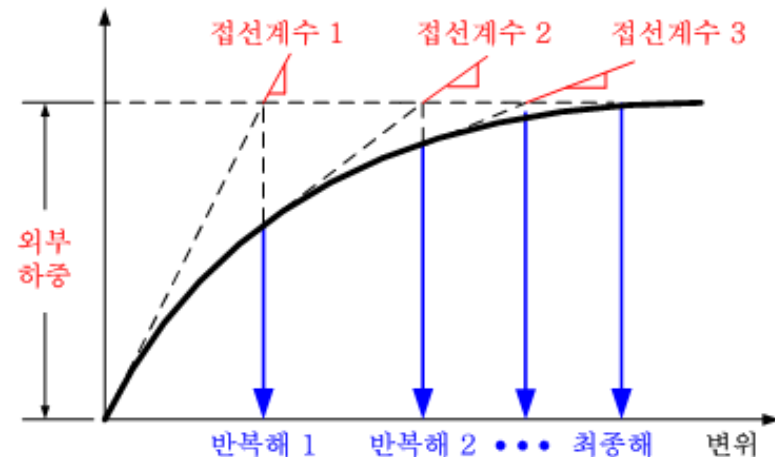
수치계산

정의

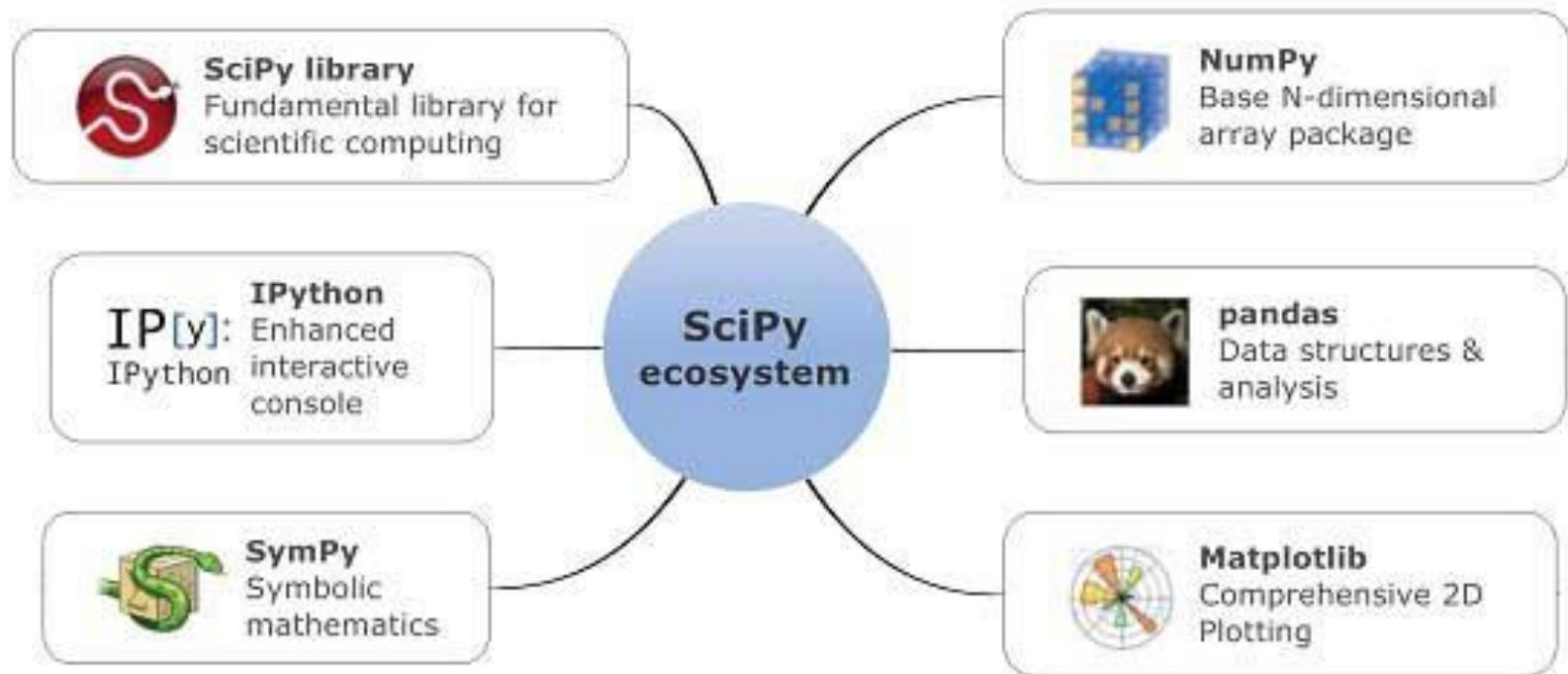
- 특정 미분 방정식과 같이 해가 존재하는 것이 알려져 있으나, 해를 구하는 방법이 나타나 있지 않는 경우가 있다.
- 수치를 대입하여 근사로 해를 구하는 방법
- Numpy: 행렬의 데이터 타입과 연산을 모아둔 패키지
- SciPy: 수치 계산 알고리즘과 툴 박스를 모아둔 패키지

<https://terms.naver.com/entry.nhn?docId=1115909&cid=40942&categoryId=32204>

http://nfx.co.kr/techpaper/keyword_view.asp?idx=139



과학 기술 컴퓨팅



<https://medium.com/esri-insider/strengthening-the-link-between-gis-and-science-4d6e31307de0>

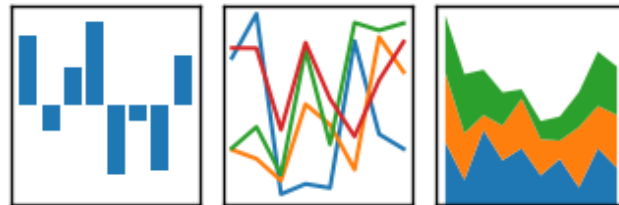
과학 기술 컴퓨팅

pandas

- 데이터 분석에서 사용되는 데이터 구조 및 데이터 조작 함수를 제공하는 패키지

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



자연어 처리

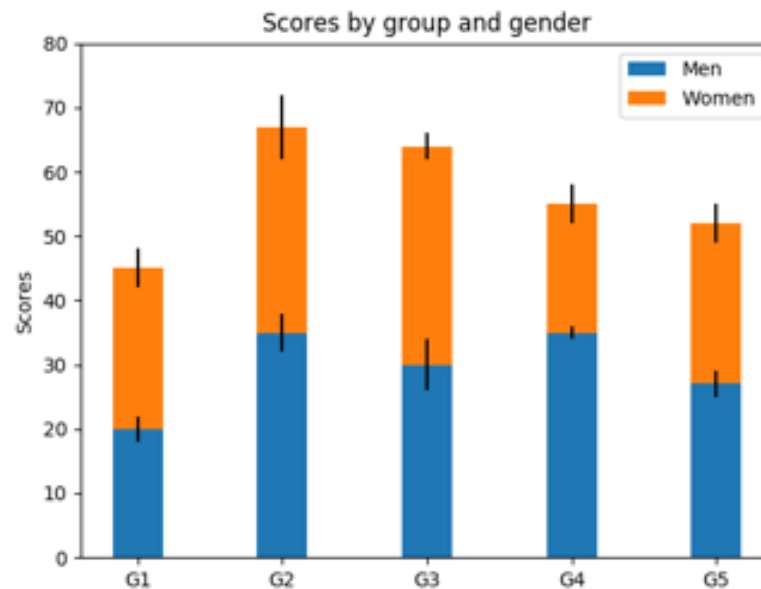
정의

- 인간의 언어를 컴퓨터와 같은 기계가 이해할 수 있게 만드는 분야

데이터 시각화

matplotlib

— 2차원 그래프 및 3차원 그래프를 그리는 패키지



wordcloud

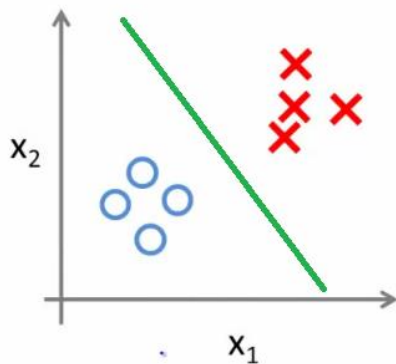
-

딥러닝

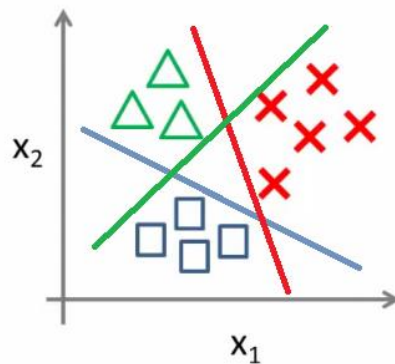
머신러닝과 딥러닝

- 여러 변환기법을 통해 핵심적인 내용 또는 기능을 요약하는 것(추상화)을 시도하는 기계학습 (인공지능)
- 사람의 사고방식을 컴퓨터에게 가르치는 한 분야

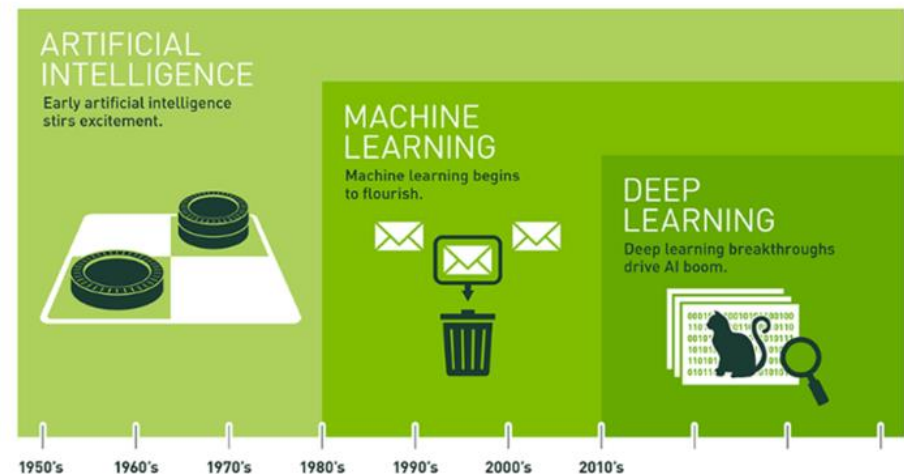
Binary classification:



Multi-class classification:



https://ko.wikipedia.org/wiki/%EB%94%A5_%EB%9F%AC%EB%8B%9D

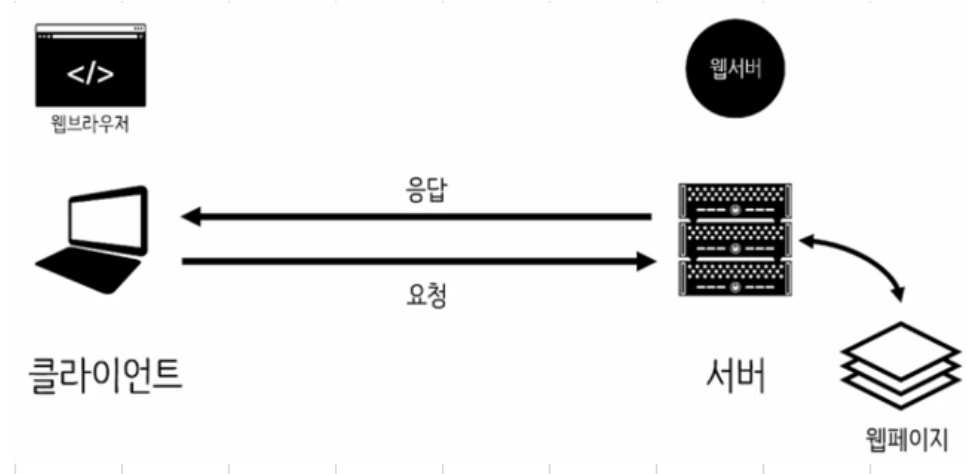


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

웹 프로그래밍

웹 프로그래밍

- 웹 서버를 구축하여
웹 서비스를 사용자에게 제공하는 것



이 외에도...

네트워크 프로그래밍

병렬 컴퓨팅

봇 생성하기

게임 만들기 등이 가능

수업을 마무리하며
