

Big Data

데이터 분석 기획

류영표 강사

youngpyoryu@dongguk.edu

Copyright © “Youngpyo Ryu” All Rights Reserved.

This document was created for the exclusive use of “Youngpyo Ryu”.

It must not be passed on to third parties except with the explicit prior consent of “Youngpyo Ryu”.



류영표

Youngpyo Ryu

現 동국대학교 수학과/응용수학 석사수료

現 SD아카데미 국비과정 강사

現 Upstage AI X 네이버 부스트캠프 멘토

前 메가 IT아카데미(파이썬, 빅데이터) 강사

한국파스퇴르연구소 Image Mining 인턴(Deep learning)

前 (주)셈웨어(수학 콘텐츠, 데이터 분석 개발 및 연구인턴)

강의 경력

- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- 딥러닝 집중 교육과정 강사
- (재)윌튼블록체인 6일 과정 (파이썬기초, 크롤링, 머신러닝)
- 서울특별시 X AI 양재허브 X 모두의연구소 (중급 NLP과정) 보조강사
- SK아카데미_HLP(임원) 1차/2차 보조강사
- (주) 모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- LG전자 / LG 인화원 보조강사
- 인공지능 자연어처리(NLP) 기업데이터 분석 전문가 양성과정 멘토
- 고려대학교 선도대학 소속 30명 딥러닝 집중 강의

주요 프로젝트 및 기타사항

- 제1회 인공지능(AI)기반 데이터사이언티스트
전문가 양성과정 최우수상 수상(Q&A 챗봇)
- 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 1기 수료
- 제 1회 산업 수학 스터디 그룹 (질병에 영향을 미치는 유전자 정보 분석)
- 제 4,5회 산업 수학 스터디 그룹 (피부암, 유방암 분류)
- 빅데이터 여름학교 참석 (혼잡도를 최소화하는
새로운 노선 건설 위치의 최적화 문제)

데이터 분석 기법의 이해

- 데이터 분석은 통계에 기반을 두고 있지만, 통계지식과 복잡한 가정이 상대적으로 적은 실용적인 분야.
- 신규 데이터나 DW에 없는 데이터는 기존 운영시스템(legacy)에서 직접 가져오거나 운영데이터저장소(ODS)에서 정제된 데이터를 가져와서 DW의 데이터와 결합하여 활용
- 시각화 기법
 - 가장 낮은 수준의 분석이지만, 잘 사용하면 복잡한 분석보다 더 효율적, 대용량 데이터를 다룰 때와 탐색적 분석을 할 때 시각화는 필요
- 공간분석(GIS)
 - 공간적 차원과 관련된 속성들을 시각화하는 분석으로 지도 위에 관련된 속성들을 생성하고 크기, 모양, 선 굵기 등으로 구분하여 인사이트를 얻음.

데이터 분석 기법의 이해

➤ 탐색적 자료 분석(EDA, Exploratory Data Analysis)

- 다양한 차원과 값을 조합해 가며 특이점이나 의미있는 사실을 도출하고 분석의 최종 목적을 달성해가는 과정.

➤ 4가지 주제

- 저항성의 강조(Resistance) : 데이터 파손에 대한 저항성
- 잔차(Residual) 계산 : 개별 데이터가 주경향성에 얼마나 벗어났는지 확인
- 자료 변수의 재 표현(re-expression) : 분포의 선형성, 안정성, 대칭성
- 시각화(Graphical Representation) : 그래프

통계 분석

- 통계 : 어떤 현상을 종합적으로 한눈에 알아보기 쉽게 일정한 체계에 따라 숫자와 표, 그림의 형태로 나타나는 것.
- 기술통계(Descriptive statistics) : 모집단으로부터 표본을 추출하고, 표본이 가지고 있는 정보를 쉽게 파악할 수 있도록, 데이터를 정리하거나 요약하기 위해 하나의 숫자 또는 그래프의 형태로 표현하는 절차.
- 추측(토론)통계 (inferential statistics) : 모집단으로부터 추출된 표본의 표본 통계량으로 부터 모집단의 특성인 모수에 관해 통계적으로 추론하는 절차.

데이터 마이닝

대용량의 자료로부터 정보를 요약하고, 미래에 대한 예측을 목표로 자료에 존재하는 관계, 패턴, 규칙 등을 탐색하고, 이를 모형화 함.



인공지능

인간의 지적 능력을 컴퓨터를 통해 구현하는 기술



머신러닝

컴퓨터가 데이터를 통해 스스로 학습하여 예측이나 판단을 제공하는 기술

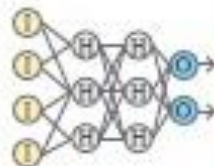
전문가
시스템

규칙기반
시스템

결정
트리

선형
회귀

퍼셉트론



딥러닝

깊은 인공신경망 알고리즘을 활용하는 머신러닝 기술

합성곱 신경망
(CNN)

심층
강화학습

순환 신경망
(RNN)



Thank you.

빅데이터 기초 / 류영표 강사
youngpyoryu@dongguk.edu

Copyright © “Youngpyo Ryu” All Rights Reserved.
This document was created for the exclusive use of “Youngpyo Ryu”.
It must not be passed on to third parties except with the explicit prior consent of “Youngpyo Ryu”.