

〈데이터분석과기계학습 14주차〉 앞으로 뭘 할까요?

인공지능융합공학부 데이터사이언스전공
곽찬희

오늘은

- 이번 학기 마지막 수업은 앞으로 공부/준비해야할 방향성에 대해 알아보시다.



공부의 방향성

- 다음에 답을 내봅시다
 - ✓ 내가 아는 것
 - ✓ 내가 할 수 있는 것
 - ✓ 내가 잘하는 것
 - ✓ 내가 (잘) 하고 싶은 것



노선을 설정합시다

- Data Engineer (Data 계의 Backend)

- ✓ 여러 tech stack 을 이용해 데이터가 잘 저장되고, 이를 분석 시스템에 연결하는 역할
- ✓ 주요 필요 역량: Hadoop, Spark, DB (SQL/NoSQL), AWS/GCP, Linux, ...
- ✓ 경력직만 주로 채용 (신입 채용을 본 기억이 없음)
- ✓ 수요 대비 공급이 매우! 적음 -> 수요공급의 mismatch -> 고연봉
- ✓ 다년 간의 수련/경력 필요
- ✓ DBA쪽에서 접근하기 쉬울 수도?



노선을 설정합시다

- Data Analyst / Scientist

- ✓ 데이터 분석계의 만능 역할
- ✓ 두루두루 골고루 알아야 함
- ✓ 주요 필요 역량: DB (SQL/NoSQL), 통계, ML/DL, 시각화
- ✓ 예전 대비 진입 장벽이 낮아지고 있음 (너도나도 데이터 공부)
- ✓ 수요 대비 공급 적정 혹은 과잉(좋은 자리인 경우) -> 경쟁 치열/ 저~중 연봉
- ✓ 경력과 신입의 차이가 큼



노선을 설정합시다

- Data Consultant

- ✓ COMMUNICATION!! 이 매우 중요
- ✓ 기술적인 스펙보다 외부 활동으로 증명 필요
- ✓ 핵심 역량: DB, 기본 분석, 발표자료 작성 및 발표 수행
- ✓ SI 혹은 컨설팅 펌
- ✓ 수요와 공급 비슷비슷



노선을 설정합시다

- ML/DL Researcher

- ✓ ML/DL을 단순히 적용하는 것이 아님. Domain 에 맞게 customize 해서 성능향상 (SOTA)를 노림
- ✓ 수요 >> 공급. 진입 장벽 있음
- ✓ 석박 우대 -> 작동 원리를 이해하고 design component 를 바꿔가며 실험 수행이 가능한 인재를 원함
- ✓ 핵심 역량: 논문 모델 구현 능력, 영어, 수학(미적분, 확률통계, 선형대수), 기타 전산 지식들



어떻게 골라요?

- **해보세요**

- ✓ 공부든, 프로젝트든, 공모전이든
- ✓ 먹어봐야 무슨 맛인지 알고, 무슨 맛인지 알아야 내가 좋아하는지 압니다
- ✓ 가만히 있으면 가마니가 됨

- **준비가 덜 되었다고요?**

- ✓ 절대! 완벽히 준비되는 사람은 없음

- **부딪쳐보세요**

- ✓ 공모전 처음부터 상 받는 사람 없습니다
- ✓ 예선 탈락도 해보고, 다른 사람 코드도 보고 따라하면서 성장하는겁니다.



인턴

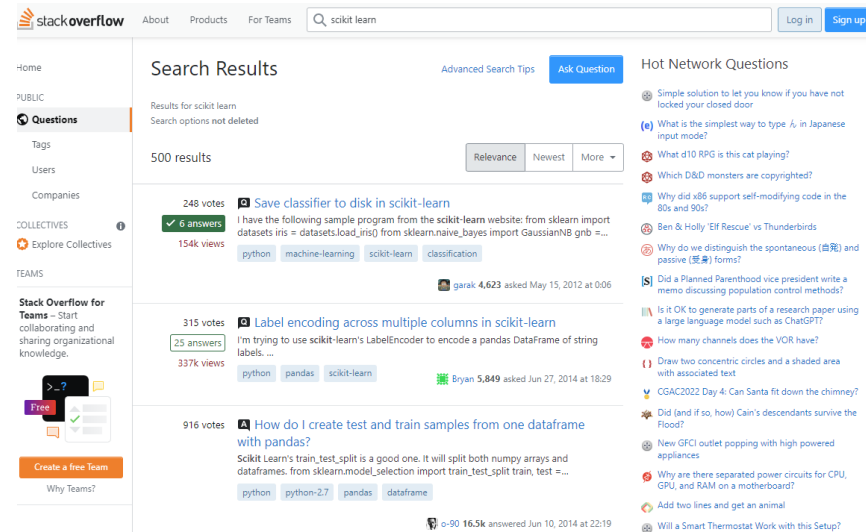
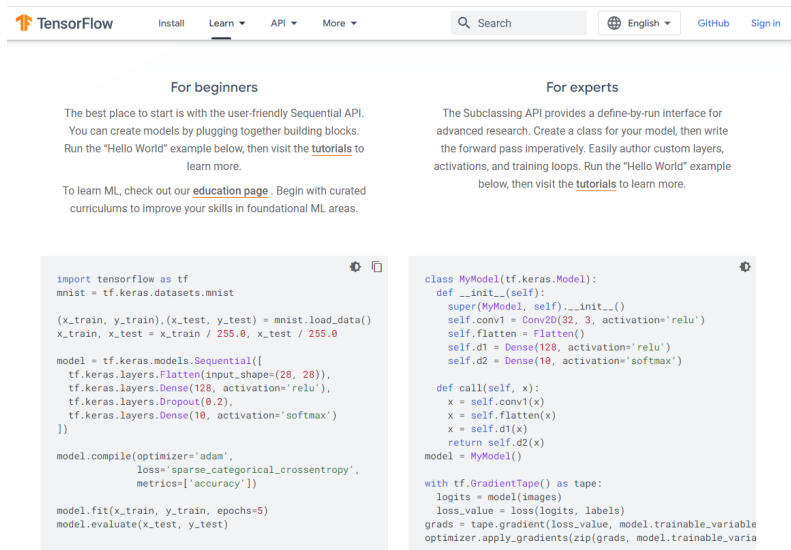
- 경험을 쌓을 수 있는 가장 좋은 기회
- 학교에서 혹은 혼자 공부할 때보다, 현업에서 구르면서 배우는 것이 학습의 속도와 양 측면에서 효율적임
- 3학년 2학기 이후에는 무조건 추천
- 학교 프로그램 활용 혹은 개별 회사 인턴십 지원
- 인턴 (채용) 연계형 부트캠프 참여도 좋은 방법



어느 곳으로 가든 일반적으로 필요한 것들

• 영어

- ✓ 나는 개발자가 될 것이기에 영어는 필요 없어? NONO
- ✓ 최신 개발 지식은 대부분 영어로 되어 있음 -> 영어에 익숙해지면 더 많은 지식 접근/검색 가능
- ✓ 점수를 위한 영어도 중요하지만, 개발문서를 읽을 수 있는 정도의 영어실력은 필수
- ✓ (대학원이라면) 고고익선 (토익 기준 850+)



어느 곳으로 가든 일반적으로 필요한 것들

- 알고리즘 / 자료구조 공부

- ✓ 대부분의 대기업에서 지원자 filtering 용으로 코딩 테스트 시행
- ✓ 코딩테스트: 사람의 언어로 표현된 문제를 프로그래밍 언어로 풀어낼 수 있는가?
- ✓ 문제 형식에 익숙해지는 것이 필요
- ✓ 백준 등 코딩테스트 사이트와 관련 서적, 기출문제 풀이 등

- ✓ 다음을 같이 고려해야 함

- 문제 풀이의 여러 가지 방법에 대해 설명하기
- 여러 방법 중 자신이 제안한 풀이가 왜 최선인지 설명하기
- 각 풀이법 장단점 분석하기



어느 곳으로 가든 일반적으로 필요한 것들

- 문서작업(Documentation)

- ✓ 많은 업무가 문서 기반해서 이뤄짐 (심지어 개발 또한!)
- ✓ 혼자 작업하는 일은 매우 드물고, 작업하더라도 문서로 남기고 보관해야 함
- ✓ 그래야 타인과 협업 가능(+나중에 code review 등 검수 가능)
- ✓ 깔끔한 글쓰기 필요 -> 좋은 글을 많이 읽읍시다

- 소통 (Communication)

- ✓ 타인과 말/글로 소통하는 역량 필요
- ✓ 공대생들의 아킬레스건
- ✓ Team Player 가 되는 훈련을 합시다



어느 곳으로 가든 일반적으로 필요한 것들

- 클라우드 (Cloud) + Linux

- ✓ 현업에서 윈도우 환경에서 개발을 하는 것은 매우 드문 일입니다.
- ✓ 자신의 컴퓨터에서 개발하는 것도 매우 드문 일입니다.
- ✓ 미리 AWS / GCP 를 익혀둡시다.
- ✓ 기초 Certificate 를 따두는 것도 좋음

- 자격증

- ✓ 생각보다 도움 안됨 (ADP / 빅데이터분석기사/ 정보처리기사 / SQLD정도만 의미 있음)
- ✓ 점수화되는 시험(ex. 공무원) 에서만 필수
- ✓ 나머지 분야는 선택적으로 취득



Portfolio

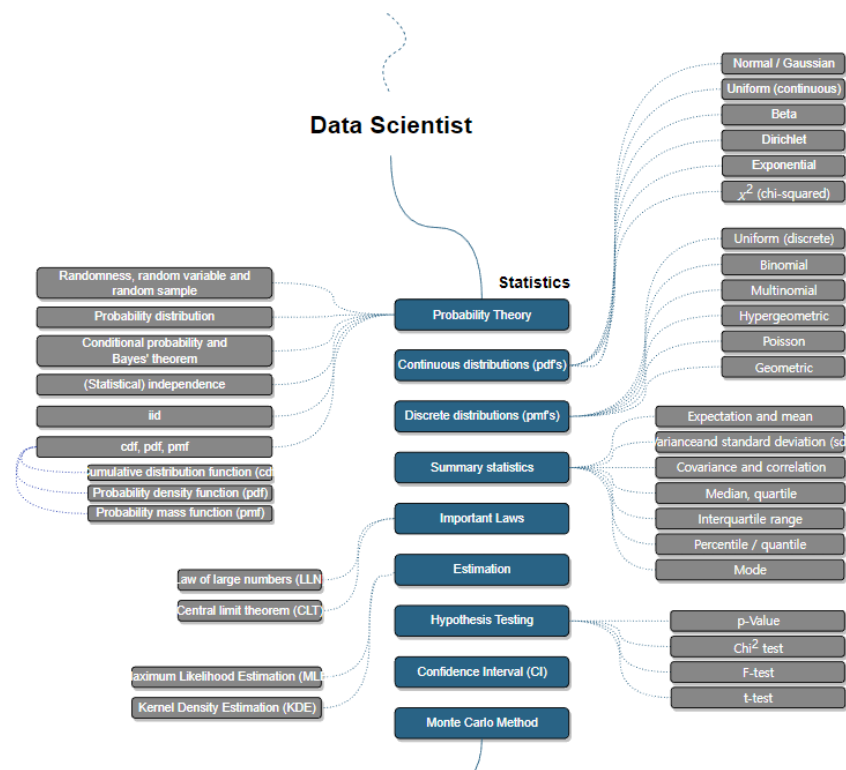
- 포트폴리오 준비의 시작은 코드 정리
- 처음에는 수업/프로젝트 등 수행한 모든 것을 일기장처럼 정리하는 형식으로 접근
- 후로 갈수록 간략한 버전 준비
- 하나의 프로젝트는 다음 문장처럼 스토리로 정리가 되어야 함
 - ✓ A라는 데이터로 B라는 과제를 수행했는데, 이 때 이러이러한 이유로 C라는 기법을 선택했다. 그 결과 D를 얻어서 E라는 시사점을 도출했다.
- 프로젝트가 외부에서 인정받는 결과로 마무리되는 것이 BEST
 - ✓ 공모전 수상 (가급적 교외, 큰 대회)
 - ✓ 논문 / 학술대회 발표



이제 기술 얘기 해봅시다.

- Google 에 Machine Learning / AI roadmap 이라고 치면 여러 자료가 나옴

✓ <https://i.am.ai/roadmap/>



ML/DL 로 가고 싶다!

- 대학원 진학 여부를 먼저 판단
 - ✓ 학점 + 영어 + 포트폴리오
- 차근차근 논문 구현해보고, 활용해보기
 - ✓ <https://paperswithcode.com>
- 무턱대고 최신 모델 사용에 집착하지 말고, 기초부터 차근차근



추천 자료 (이게 다는 아니지만)

- 딥러닝 기초

- ✓ 혼자 공부하는 머신러닝+딥러닝 (한빛미디어)
- ✓ 밑바닥부터 시작하는 딥러닝 1~3 (한빛미디어)
- ✓ 머신러닝 교과서(길벗)

- 수학

- ✓ 딥러닝을 위한 선형대수학 딥러닝을 위한 선형대수학 (한빛아카데미)
- ✓ 모두의 인공지능 기초 수학 누구나 쉽게 시작하는 인공지능 수학 (길벗)

- 유튜브 강의

- ✓ <https://www.youtube.com/@statquest>
- ✓ <https://www.youtube.com/@3blue1brown>
- ✓ <https://www.youtube.com/@dsba2979>
- ✓ <https://www.youtube.com/@user-yu5qs4ct2b>

전문 분야 정하기

- 모든 것을 잘할 수는 없습니다
- 무엇을 깊게 공부할 것인가?
 - ✓ Image
 - ✓ Vision
 - ✓ Audio / Video
 - ✓ NLP
 - ✓ Generative Model
 - ✓ XAI
 - ✓ ...
- 전문 지식과 포트폴리오도 이에 맞춰야 합니다.

DS / 분석가로 가고 싶다

- 많이 경험해보는 것이 중요합니다.
- 다양한 데이터를 다루면서 새로운 기법 익히기
 - ✓ 예) 시계열 데이터를 다루며 관련 전처리/기법을 익힘
- 캐글/데이콘/각종 경진대회를 부지런히 나가세요
 - ✓ “준비가 안된 것 같아요” 같은 변명은 거절합니다.



결국

- 공학자에게 실력 == 연봉 == 나은 워라밸 == 나은 커리어
- 기왕 공학의 길을 걸어볼 작정이라면, 한번쯤은 자신의 모든 걸 걸어보세요.
- 꾸준한 노력과 도전만이 살길입니다.

