

# Fast AI - 2주차 발표

▼ Class	
🕒 Created	@Sep 6, 2020 5:13 PM
🔗 Materials	
☑ Reviewed	<input type="checkbox"/>
▼ Type	

[The Practice of Deep Learning](#)

[Starting Your Project](#)

[The State of Deep Learning](#)

[Computer Vision](#)

[NLP](#)

[The Drivetrain Approach](#)

## The Practice of Deep Learning

딥러닝은 적은 코드로 많은 문제를 신속하게 해결할 수 있음

하지만 딥러닝은 마술이 아니다.

딥러닝에서 제약조건을 과소평가하고, 딥러닝 기능을 과대평가하면 실망스러운 결과를 얻을 수 있다.

## Starting Your Project

현재 수행할 프로젝트가 있는가?

자신의 분석 프로젝트에서 모델을 만들고, 적용되는 실제 경험을 얻을 수 있다.

프로젝트를 선택할때 가장 중요한 것은 "데이터 가용성" 이다.

데이터를 찾으려 수년을 낭비하기도 하는데 중요한 것은 우리의 목표는 완벽한 데이터셋을 찾는것이 아니라 분석을 시작하고 반복하는 것이 중요

이 모든 것을 합리적인 시간 안에 완료 해야한다.

조직에 속해있다면 동료들에게 실제 작동하는 프로토타입을 보여줌으로써 아이디어가 실제로 구현되고 작동된다는 것을 보여줘야한다.

실제 프로젝트를 할때도 데이터에 대한 의사결정이 가장 어렵고, 프로젝트와 결과의 반 이상을 차지하는것을 볼수 있다.

보통은 이러한 의사결정 중요하기 때문에 경험이 많은 회사 고위직 분들과 현업 / IT관련 분들의 협의로 인해 진행이 된다.

## The State of Deep Learning

### Computer Vision

컴퓨터 비전은 물체에 대한 인식과 같다.

컴퓨터 비전은 이미지 라벨링이 느리고 비용이 많이 들 수 있다.

이것을 돕는 방법은 밝기 및 대비를 변경하는 등 다양한 입력데이터를 생성하는 것인데 텍스트 및 기타 유형의 데이터셋에도 해당 방법이 적합하다. 음성인식도 가능

### NLP

컴퓨터는 스팸분류같은 작업 수행을 잘 한다.

짧은 문서든 긴 문서든 분류하는데 능숙함

댓글 분석도 되고, 특정 사용자의 스타일도 모방 가능

그러나 아직 NLP로 완전히 올바른 응답을 추천받지는 못한다. 예를 들어 의료정보의 지식 기반에 따른 딥러닝 모델을 만든다고 할때 신뢰성이 높다고 볼 순 없다. 완전히 틀린 콘텐츠를 만들 수 있기 때문에 주의 필요

그 외의 분야에서도 올바른 답변보다 허위정보나 불안을 유발하는 텍스트들을 추천하는 경우가 더 많을 수 있다.

그러나.. NLP는 많은 응용 분야를 가지고 있다.

번역이나 요약때 오역을 할수 있지만 그래도 성능이 괜찮은 편이기 때문에 Google 번역 같은 서비스들은 사용자들이 많이 신뢰하고 이용한다.

Google 번역도 딥러닝 기반..

텍스트와 이미지의 결합

위에서 말한 동일한 문제를 가지고 있고,

사용을 위해 모델과 사람이 상호 작용하여 프로세스의 일부로 사용하는 것이 좋다.

## The Drivetrain Approach

누구에게나 유용한 모델을 만들기 위한 지침?

예1) 자율자동차

자율자동차는 A에서 B지점으로 가고자 하는 것이 목적

물론 운전자는 많은 모델과 엄청난 데이터를 인지하면서 차를 타는 것이 아니지만  
정교한 예측분석이 잘 작동되도록 체계적인 설계접근방식이 필요하다.

우리는 데이터를 사용하여 더 많은 데이터를 생성할 수 있고, 실행가능한 결과를 만들어 낼 수 있기 때문..

이것이 Drivetrain 접근방식의 목표

예2) 구글의 검색엔진

구글은 검색엔진을 만들때

검색때 사용자의 주요목표가 무엇인지를 고려함

그래서 가장 연관성이 높은 검색 결과를 표시할 수 있게 만듦

거기서 그치지 않고, 검색이 더 잘 되도록 어떤 조치를 취할 수 있는지 고려한다.  
그때 검색 우선순위를 생성하는데 필요한 새로운 데이터를 고려

그래서 다른 페이지에 링크된 페이지에 대한 암시적 정보가 검색에 영향을 준다는 것을 깨닫고 이 데이터를 적용하여 검색 예측 모델을 구축

! 이렇게 추가데이터는 추가적으로 구축할 수 있는 모델을 결정한다.

### 예3) 추천시스템

추천시스템의 목적은 추천없이 구매하지 않았을 항목에 대한 추천으로 고객들에게 구매 유도.

여기서 래버는 추천 사항의 순위

새로운 판매를 유발할 추천 사항을 생성하려면 새로운 데이터 수집이 필요

이를 위해 다양한 고객의 광범위한 추천 사항에 대한 데이터를 수집해야하고, 무작위 실험이 수행되어야함

이것은 소수의 조직이 취하는 방식이지만, 꼭 수행해야하는 작업이다.

마지막으로 추천을 보거나 보지 않는 조건에 따라 구매확률에 대한 두가지 모델을 구축할 수 있다.

이 두가지 확률차이는 고객에게 주어진 추천에 대한 효용함수로 나타난다.

예를 들어 이미 고객이 거절된 친숙한 책을 추천하는 것과 추천 없이 구매할 책에 대해서 추천 알고리즘이 추천할 경우는 낮다.

모델의 실제 구현을 위해 단순한 모델 학습 이상의 것들이 필요함

더 많은 데이터를 수집하고, 실험하고, 모델을 통합해야함

다음번에서는 프로젝트때 사용할 데이터를 찾는 방법에 대해 배울 예정

