



인공지능 TensorFlow (Ch 1)

데이터사이언스 & A.I 정화민 교수



TensorFlow 소개

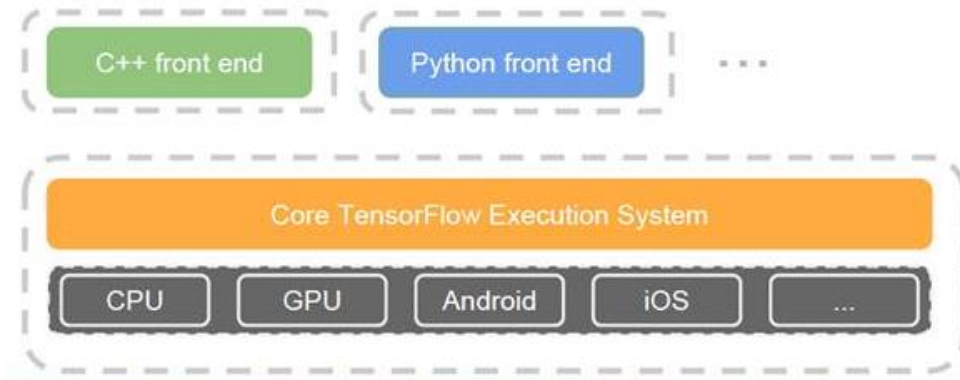


TensorFlow

- 개요 : 2015년 11월 2세대 분산식 머신러닝 프레임워크인 텐서플로우 소스 GitHub에 오픈소스 공개
(1 세대 분산식 머신러닝 프레임 워크 디스트빌리프(DistBelief))
2017년 1월 API 인터페이스가 안정된 1.0 알파 버전 발표, 안정화 버전 2019년 6월
- 개발목적 : 구글 브레인팀에서 머신러닝 연구속도 향상, 연구된 모델을 기반으로 상용화 목적
- 구글 오픈소스 공개 목적 : 커뮤니티 역량을 통해 텐서플로우 완벽하게 개선 목적
학계와 산업계 전반에서 딥러닝 효율성 제고, 효과적 활용 하는데 목적.
- <https://www.tensorflow.org> 텐서플로우 공식 홈페이지
- <https://github.com/tensorflow/tensorflow> 텐서플로우 GitHub 홈페이지
- <https://github.com/tensorflow/models> 텐서플로우 모델 웨어하우스 홈페이지

TensorFlow 소개

- 텐서플로우 기본구조

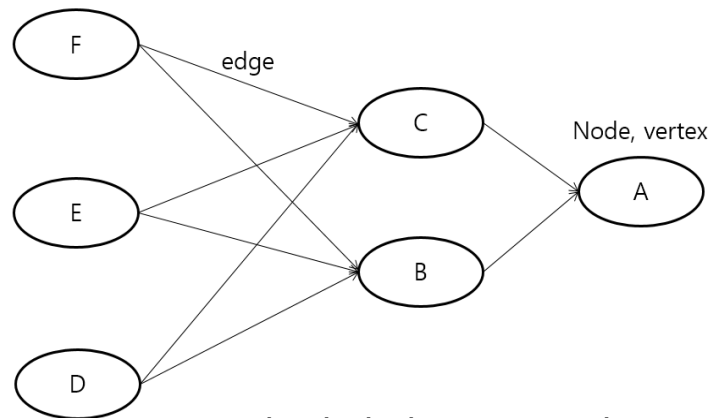


[source : https://github.com/zer0n/deepframeworks](https://github.com/zer0n/deepframeworks)

- 텐서플로우는 머신러닝 알고리즘 인터페이스 구현, 머신러닝 알고리즘 프레임워크 실행
- 텐서플로우 구조에서 프론트엔드는 파이썬 C++, Go, Java 등 다양한 언어를 사용한 개발지원
- 백엔드에서는 C++, CUDA 등을 사용한 개발지원, iOS, 안드로이드, CPU , GPU 클러스트 구조 시스템에도 쉽게 이식가능

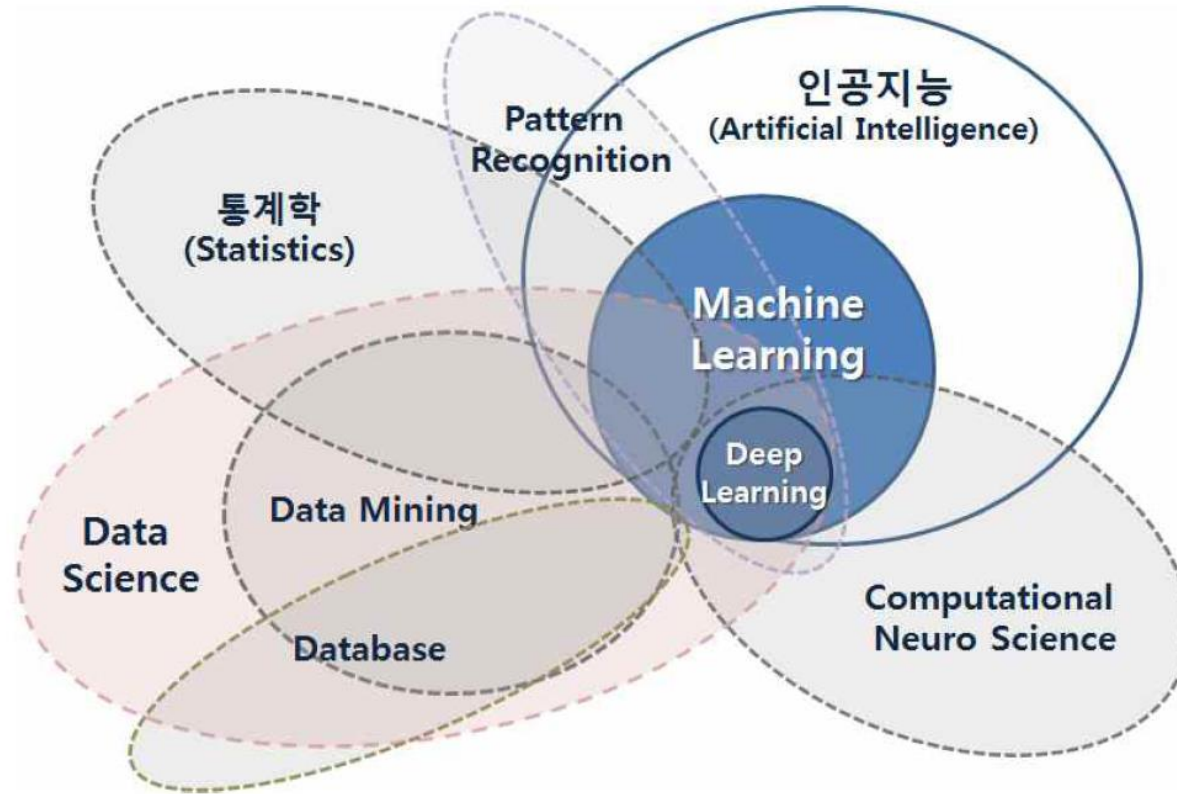
TensorFlow 소개

- 텐서플로우 프로그램 모델
 - 텐서 (Tensor) : 데이터를 의미함 (다차원 배열로 나타내는 데이터)
 - 플로우(Flow) : 데이터의 흐름
- 텐서플로우 연산 그래프 : 데이터 흐름 그래프에서 데이터플로우 그래프 (Dataflow graph)의 변(edge)은 어떤 노드에서 다른 노드로 흘러가는 (flow)데이터의 방향을 정하고, 그래프의 각 노드(node)는 하나의 연산을 나타내며, 입력값을 받아 다른 노드(node)로 전달할 결과값을 출력한다.
- 텐서플로우는 연결상태를 기반으로 연산을 최적화 하고 노드간의 의존관계(직접의존 direct dependency, 간접의존 indirect dependency)를 이용해 연산량을 최소화하고 노드간 의존관계를 파악할 수 있다.
- 노드 A는 C, B에 의해서 직접의존하고 있고, 노드 D, E, F에 간접의존 하고 있다.



[그림]데이터 흐름 그래프

인공지능, 머신러닝, 딥러닝



[그림] 머신러닝과 여러 학문과의 관계

[Source: NCS 머신러닝기반 데이터 분석](#)

머신러닝

■ 머신러닝 개념

- 컴퓨터 과학의 영역에 속하는 인공지능의 한 분야
- 컴퓨터 프로그램이 어떤 것에 대한 학습을 통해 기존의 모델이나 결과물을 개선하거나 예측하게끔 구축하는 과정을 의미
- 훈련 데이터에서 여러 가지 통계적, 컴퓨터 알고리즘적, 수치적 방법 등을 통해서 규칙성이나 패턴을 찾은 뒤(이를 학습 과정이라고 함) 훈련에 사용하지 않은 새로운 데이터를 통해서 필요한 답을 얻어내는 것을 의미

머신러닝에 대한 다양한 정의	주요 제안자 및 출처
- 컴퓨터에게서 배울 수 있는 능력, 즉 코드로 정의하지 않은 동작을 실행하는 능력에 관한 연구 분야	Arthur Samuel (1959)
- 작업 T에 대한 성능을 P로 측정할 수 있고, 성능 P가 경험 E로 개선된다면 프로그램은 E,T,P로부터 배웠다고 할 수 있다.	Tom Mitchell (1996, 1997)
- 정확한 '작업(Task)' 을 성취할 수 있는 올바른 '모델(Model)' 을 구축하기 위해 올바른 '특성(Features)' 을 활용하는 것	Peter Flach (2012)
- 컴퓨터 프로그램이 어떤 것을 학습한 후에 최초 학습에 들인 시간과 노력보다 더 빠르고 수월하게 배운 것을 해낼 수 있게 하는 것	Jason Bell (2015)
- 인공지능의 한 분야로 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야	Wikipedia

[Source: NCS 머신러닝기반 데이터 분석](#)

머신러닝

- 지도 학습(supervised Learning)

- 지도 학습은 설명변수(혹은 독립변수, 특성(Feature) 등으로 표현)와 목적변수(혹은 반응변수, 종속변수, 목표변수, 출력값 등으로 표현) 간의 관계성을 표현
- 미래 관측을 예측해 내는 것에 초점이 있으며 주로 인식, 분류, 진단, 예측 등의 문제 해결에 적합하다.

분류	수치예측 (혹은 회귀)
- K- 최근접 이웃(K-nearest Neighbors)	선형 회귀 (Linear Regression)
- 로지스틱 회귀 (Logistic Regression)	확장된 회귀분석(ex: 다항회귀, 비선형, 회귀 등)
- 인공신경망 분석(Artificial Neural Network)	인공신경망 분석(Artificial Neural Network)
- 의사결정 트리(Decision Tree)	의사결정 트리(Decision Tree)
- 서포트 벡터 머신(Support Vector machine)	서포트 벡터 머신(회귀) (Support Vector machine)
- 나이브 베이즈 (Naïve Bayes)	PLS(Partial Least Squares)
-앙상블 기법(랜덤포레스트 등)	-앙상블 기법(랜덤포레스트 등)

[Source: NCS 머신러닝기반 데이터 분석](#)

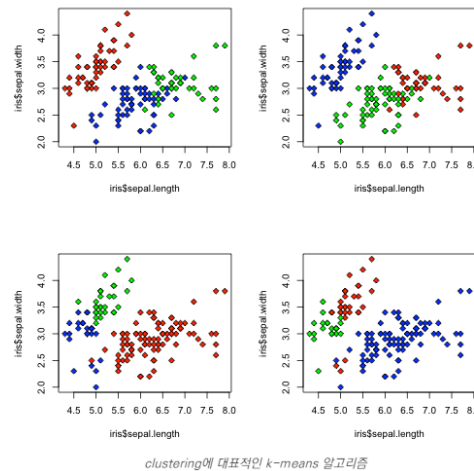
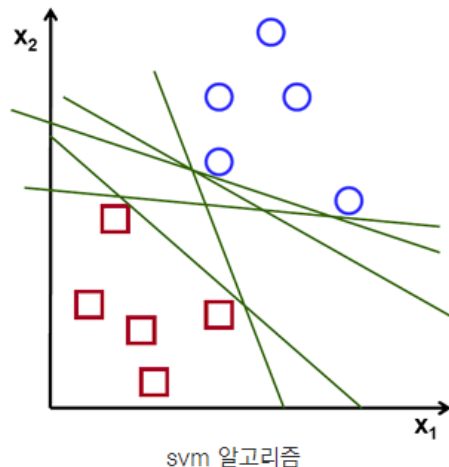
머신러닝

■ 비지도학습(Unsupervised Learning)

- 목적변수(혹은 반응변수, 종속변수, 목표변수, 출력값)에 대한 정보 없이 학습이 이루어지는 형태를 말하며, 예측의 문제보다는 주로 현상의 기술(Description)이나 특징 도출, 패턴 도출 등의 문제에 많이 활용
- 사전정보가 없는 상태에서 유용한 정보나 패턴을 탐색적으로 발견하고자 하는 데이터 마이닝의 성격이 더 강하다고 볼 수 있다.
- 중요알고리즘 : K- Means clustering, 계층적 clustering, 주성분분석(PCA)

■ 강화학습(Reinforcement Learning)

- 정답은 모르지만, 행동에 대한 보상을 알 수 있고 그 보상으로 최대의 보상을 받은 방법으로 학습



머신러닝 사례

- 마이너리티 리포트
 - Pre_Crime 기술로 세상에 존재하는 데이터를 분석하여 미래의 범죄를 예측하고 사전에 범죄 발생 차단
- 검색어 자동완성 / 연관검색어
 - 포털 사이트의 검색어 자동완성, 연관검색어 제공 서비스
- 아마존 상품추천 서비스
 - 사용자의 관심분야에 대한 체크리스트 작성하고 사용자의 상품을 구매한 누적데이터 바탕으로 정교한 상품추천으로 고객관계 마케팅
- 금융(주식 알고리즘 트레이딩, 고객 신용평가), 바이오(당뇨병 예측, 전자코 이용 폐암 발병 예측), 스팸 메일 분류 등 학습



amazon.com

Recommended for You

Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.



[Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop](#)

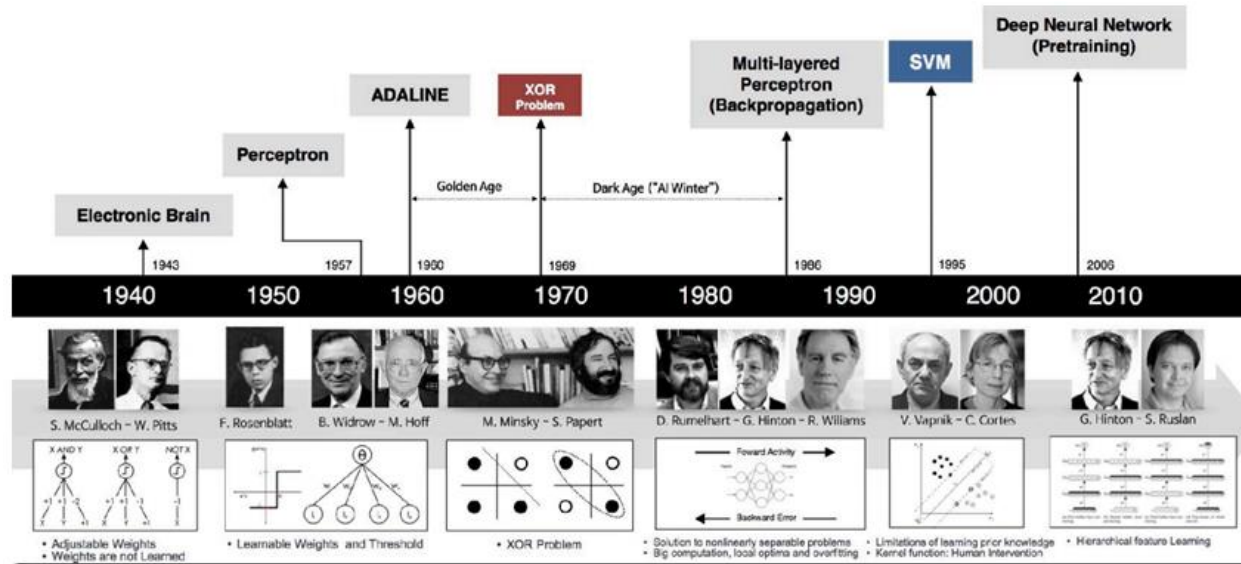


[Google Apps Administrator Guide: A Private-Label Web Workspace](#)



[Googlepedia: The Ultimate Google Resource \(3rd Edition\)](#)

딥러닝

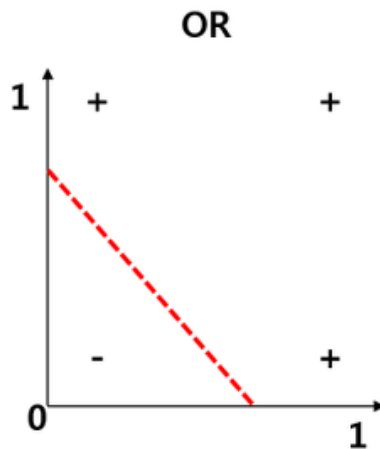


[그림] 딥러닝 역사

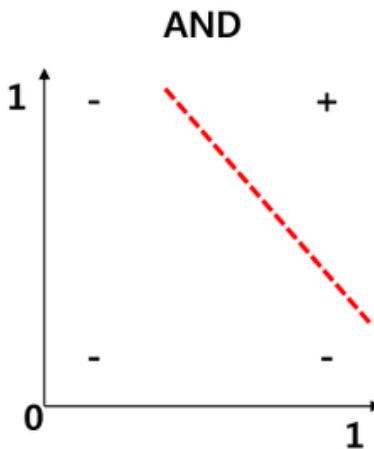
Source: <https://medium.com/ibm-data-science-experience/deep-learning-with-data-science-experience-8478cc0f81ac>

- 1960~1970년, 인공지능 황금기(딥러닝 초안은 Perceptron 알고리즘)
- 1970~1980년, 인공지능 암흑기 (XOR Problem 을 해결 못함)
- 1986년, 인공지능 암흑기의 끝 (Multi-Layer perceptron-G.Hinton 제프리 힌튼 교수)
- 1995년, SVM 개발
- 2006~2010년~, 딥러닝 부흥기 (음성인식, 이미지 인식 등)
- 2012년, AlexNet(음성인식, 이미지 인식의 에러율 16.4% 이하로 낮아짐)
- 2014년, 이미지넷 챌린지 GoogleNet 우승(22 layers) , 옥스포드대의 VGGnet 2등(19 layers)
- 2015년, 마이크로소프트 리서치(중국)의 RseNet (152 layers)

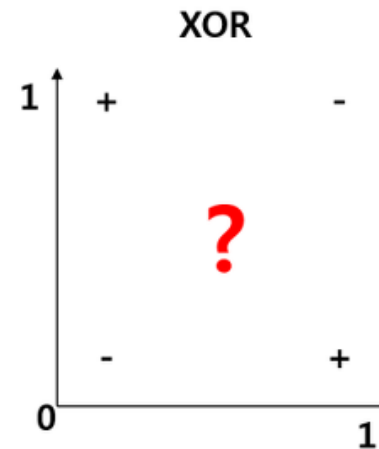
딤러닝



x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1



x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

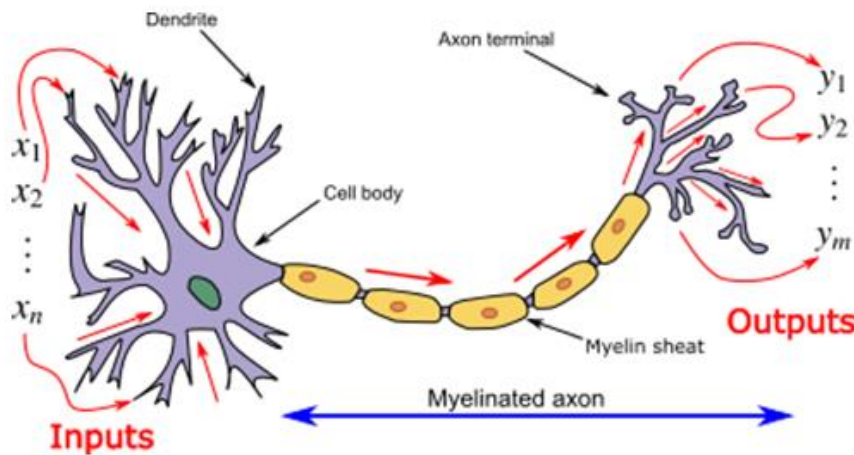
- OR (논리합) : 하나 이상의 입력값이 1이면 1출력
- AND(논리곱): 곱하기 처럼 모든 입력값이 1일 때만 1출력
- XOR 연산: 입력값이 같지 않으면 1을 출력, 같으면 0 출력

-> XOR Problem : OR, AND 그래프에서는 데이터를 Linear로 분류할 수 있으나, XOR에서는 불가능

딥러닝

■ 딥러닝(Deep Learning)

- 기본적으로 인공신경망(Artificial neural network, ANN)을 대상으로 입력계층, 히든계층, 출력 계층으로 구분하여 분류하고 시냅스에 해당하는 각 노드(node)간의 연결부분의 가중치 값을 변경하여 가장 근사치에 맞는 결과를 찾아내는 것으로 시작.
- 중간계층의 히든 계층의 복잡도에 따라서 알고리즘 수행결과를 연산하는데 시간이 달라짐.
- 딥러닝은 사람의 뇌와 유사한 인공신경망을 구성하여 처리
- 딥러닝 계산은 복잡하지 않지만, 많은 양의 데이터를 대상으로 분류하고, 병렬연산을 가능하게 해주는 GPU(Graphic Processing Unit)를 활용하면 대량의 데이터 연산에도 연산 시간을 단축



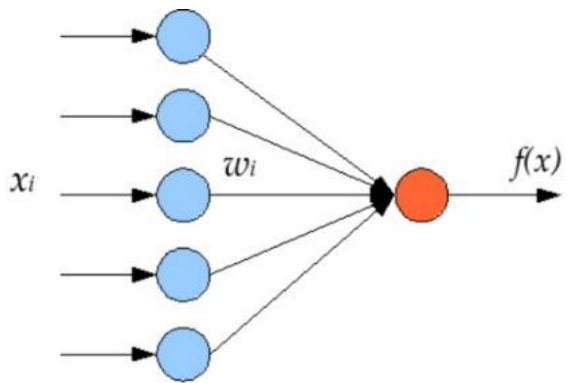
[그림] 생물학적 신경망

생물학적 신경망	인공신경망
Soma	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight

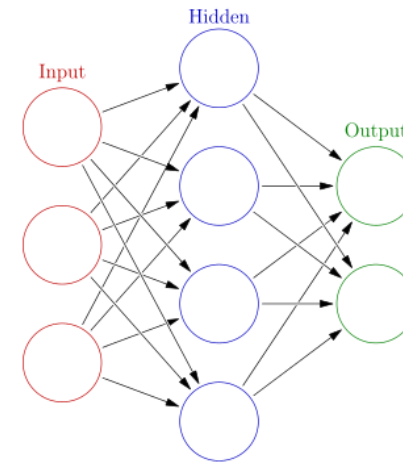
딥러닝

- 딥러닝(Deep Learning)

- 인공신경망(Artificial neural network, ANN) 중 을Hidden-layer 3개 이상 들어간 Multi-layer로 Regression / Classification 문제를 푸는 것을 딥러닝이라고 함.
- y값이 (-) 양수 ~ (+) 무한대의 범위를 가지는 Regression 문제
y값이 0 또는 1이 나오는 Classification 문제



[그림] Single layer perceptron



[그림] Multi layer perceptron

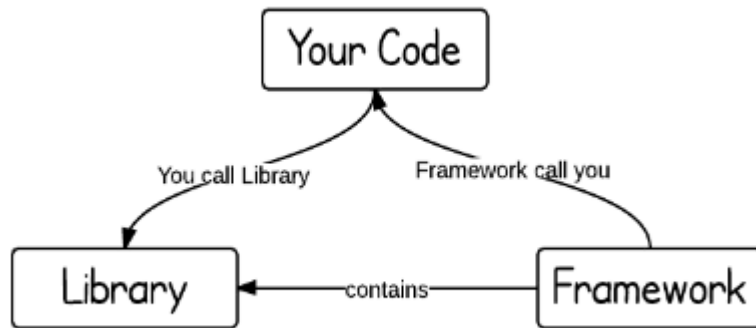
딥러닝

■ 딥러닝 프레임워크(Deep Learning Framework)

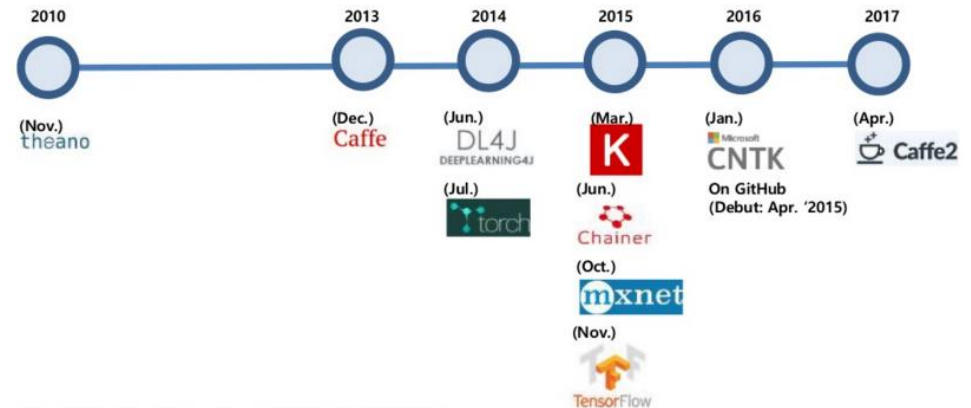
- 인공지능 라이브러리(library), 머신러닝 라이브러리 (library)라고 부르는 도구를 말함.
- 프레임워크(Framework)는 뼈대나 기반 구조를 의미하며 소프트웨어 프레임 워크는 복잡한 문제를 해결하거나 서술하는데 사용되는 기본 개념구조

■ 프레임워크와 라이브러리의 차이점

- 프레임워크는 전체적인 흐름을 직접 제어하고 사용자는 필요한 소스코드를 만들어서 추가, 라이브러리는 사용자가 전체적인 흐름을 만들어 나가면서 소스코드를 작성하여 필요한 부분에서 라이브러리를 가져다 씀.



[그림] 인공지능 프레임 워크 vs 라이브러리



[그림] 딥러닝 프레임 워크

딥러닝

■ 딥러닝 프레임워크(Deep Learning Framework)비교

프레임워크	내용
TensorFlow	가장 인기있는 딥러닝 라이브러리 중 하나인 Tensorflow는 Google Brain 팀에서 개발했으며 2015년 오픈소스로 공개 '2세대 머신러닝 시스템'으로 불리는 Tensorflow는 Python 기반 라이브러리로, 여러 CPU 및 GPU와 모든 플랫폼, 데스크톱 및 모바일에서 사용. 또한 C++ 및 R과 같은 다른 언어도 지원하며 딥러닝 모델을 직접 작성하거나 Keras와 같은 래퍼 라이브러리를 사용하여 직접 작성
Theano	최초의 딥러닝 라이브러리 중 하나인 Theano는 Python 기반이며 CPU 및 GPU의 수치 계산에 매우 유용 Tensorflow와 마찬가지로 Theano는 저수준 라이브러리로, 딥러닝 모델을 직접 만들거나 그 위에 래퍼 라이브러리를 사용하여 프로세스를 단순화 할 수 있음. 그러나 다른 확장 학습 프레임워크와 달리 확장성이 뛰어나지 않으며 다중 GPU 지원이 부족함 범용적으로 딥러닝을 할 때 전 세계의 많은 개발자가 여전히 선택하고 있습니다
Keras	Keras는 효율적인 신경망 구축을 위한 단순화 된 인터페이스로 개발 Keras는 Theano 또는 Tensorflow에서 작동하도록 구성 할 수 있으며 Python으로 작성되었으며, 매우 가볍고 배우기 쉬움 비교적 새로운 라이브러리임에도 불구하고 아주 좋은 문서를 가지고 있으며 몇 줄의 코드에서 Keras를 사용하여 신경망을 만들 수 있습니다.
Caffe	표현, 속도 및 모듈성을 염두에 두고 개발된 Caffe는 Berkeley Vision and Learning Center (BVLC)에서 주로 개발한 최초의 딥러닝 라이브러리 중 하나. 이것은 또한 Python 인터페이스를 가지고 있는 C++ 라이브러리이며, CNN(Convolutional Neural Networks)을 모델링 할 때 기본 애플리케이션으로 찾음. 이 라이브러리를 사용하면 얻을 수 있는 주요 이점 중 하나는 바로 Caffe Model Zoo에서 미리 훈련된 여러 네트워크를 바로 사용할 수 있다는 점 Caffe를 페이스북은 최근 고성능 개방형 학습 모델을 구축 할 수 있는 유연성을 제공하는 새로운 가벼운 모듈 식 딥러닝 프레임 워크인 Caffe2를 공개했음.
Torch	DeepLearning4j (또는 DL4J)는 Java로 개발된 인기있는 딥러닝 프레임워크이며 다른 JVM 언어도 지원. 상업, 산업 중심의 분산 딥러닝 플랫폼으로 널리 사용됩니다. DL4j를 사용하면 Apache Hadoop 및 Apache Spark와 같은 널리 사용되는 Big Data 도구를 기반으로 구현할 수 있으므로 전체 Java 생태계의 힘을 결합하여 효율적인 딥러닝을 수행 할 수 있다는 장점
MxNet	MXNet은 R, Python, C++ 및 Julia와 같은 언어를 지원하는 딥러닝 프레임워크 중 하나입니다. 이것은 이 언어들 중 어느 것을 알고 있다면, 편안하게 학습 영역을 쉽게 접근할 수 있기 때문에 유용합니다. 백엔드는 C++과 cuda로 작성되었으며 Theano와 같이 자체 메모리를 관리 할 수 있습니다. MXNet은 확장성이 좋고 다중 GPU와 컴퓨터로 작업 할 수 있기 때문에 대중적입니다. 기업용으로 매우 유용합니다. 이것이 아마존이 MXNet을 딥러닝을 위한 참조 라이브러리로 사용한 이유 중 하나이기도 함.