

머신러닝 교과서 마수걸이

박종민

머신러닝 교과서 with 파이썬, 사이킷런, 텐서플로



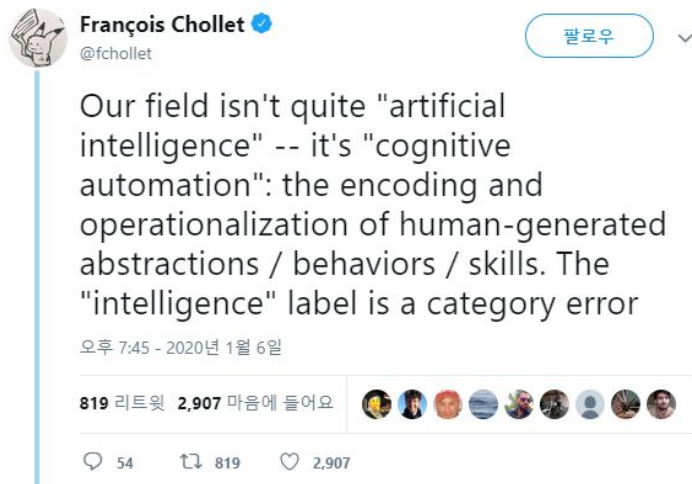
출처: [텐서플로우 블로그](#)

이 책에서 알아두면 좋은 자료

- [책 코드 저장소](#)
- [저자의 블로그겸 오타 수정 기록](#)
- [오타 신고](#)
- [핸즈온 머신러닝 1장 스터디](#)

머신러닝

- 데이터를 이해하는 알고리즘의 과학이자 애플리케이션
- 수동 분류 규칙 대신 효율적으로 지식을 추출하여 예측 모델을 만듦



출처: [Twitter, François Chollet](#)

머신러닝 종류

- 머신러닝 : 학습종류 = 수학의 정석 : 집합과 명제
- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
 - Clustering
 - Dimensionality Reduction
- Reinforcement Learning

기본 용어와 표기법

- **Row:** 데이터셋의 샘플
- **Column:** 데이터셋의 특성(feature)
- **Target:** 클래스 레이블, 정답

- 행렬로 표시
$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & x_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & x_4^{(150)} \end{bmatrix}$$

머신러닝 시스템 구축

- 전처리
 - 최적의 성능을 내기 위해 정규화, 차원 축소 등을 적용
- 예측 모델 훈련과 선택
 - 모델을 비교하기 위해 성능 측정 지표를 결정 (예: 정확도)
- 모델을 평가
 - 모델 훈련에서 사용되지 않았던 새로운 데이터를 사용하여 일반적으로 우수한지 검증
- 모델 서비스

Python 환경

- Anaconda?
 - 프로젝트에 따라 요구되는 Python 버전이 필요
 - 관리 및 설치가 편리
 - 불필요한 라이브러리 설치로 인한 혼동을 줄임
 - 독립 환경에서 필요한 라이브러리만 설치 가능
 - 참고: [Why Anaconda? How to control Anaconda?](#)

Colab 활용

- <https://colab.research.google.com/>
- Jupyter notebook 환경을 별도의 자원 필요없이 실행 가능
- 머신러닝에 필요한 대부분의 라이브러리 설치
- GPU, TPU 활용 가능

머신러닝의 역사

- 맥컬록-피츠(MCP) 뉴런
 - 간소화된 뇌의 뉴런 개념
- Perceptron
 - 뉴런 출력 값을 계산하여 신경망의 가중치를 업데이트
- 적응형 선형 뉴런(Adaline)
 - 가중치를 업데이트할 때 단위 계단 함수 대신 선형 활성화 함수를 사용 (p65)

머신러닝의 주요 방법

- Gradient Descent

- 연속적인 선형 활성화 함수를 사용하면 비용 함수가 미분 가능하고 접선의 기울기를 0으로 가깝게 가도록 가중치를 업데이트

- 표준화

- 특성별로 값의 범위가 다를 수 있고, 정규 분포 성질을 부여하면 수렴이 빨라짐

- SGD(Stochastic Gradient Descent)

- 전체 중 하나의 훈련 샘플을 기반

- Mini-batch Learning

- 절충 방법으로 훈련 데이터의 작은 일부분을 사용

실습

- Scikit-learn을 사용하는 방법과 동일하게 구성
- [Jupyter notebook viewer](#)
- [Colab](#)