

개발자를 위한 머신러닝

임백준



강사소개



- 현재 – Director of Engineering at adMarketplace
- 병렬처리 서버구축, 빅데이터 분석, 머신러닝
- 다수의 월스트리트 회사에서 실시간 트레이딩 시스템 구축
- 루슨트 테크놀로지스에서 네트워크 관련 소프트웨어 개발
- 『폴리글랏 프로그래밍』(2014), 『누워서 읽는 퍼즐북』(2010),
- 『프로그래밍은 상상이다』(2008), 『뉴욕의 프로그래머』(2007),
- 『소프트웨어산책』(2005), 『나는 프로그래머다』(2004),
- 『누워서 읽는 알고리즘』(2003), 『행복한 프로그래밍』(2003),
- 『프로그래머 그 다음 이야기』(2011) 등 출간

목차

왜 개발자를 위한 머신러닝인가



머신러닝을 익히는 사람의 20%는 수학자나 통계학자이고 80%는 개발자다.

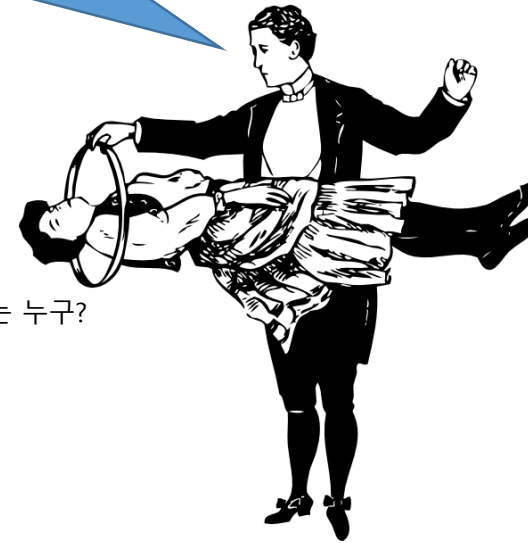
- 손 오웬, 클라우데라

머신러닝이란 무엇인가



나는 프로그래머

내가 바로 머신러닝 전문가일세

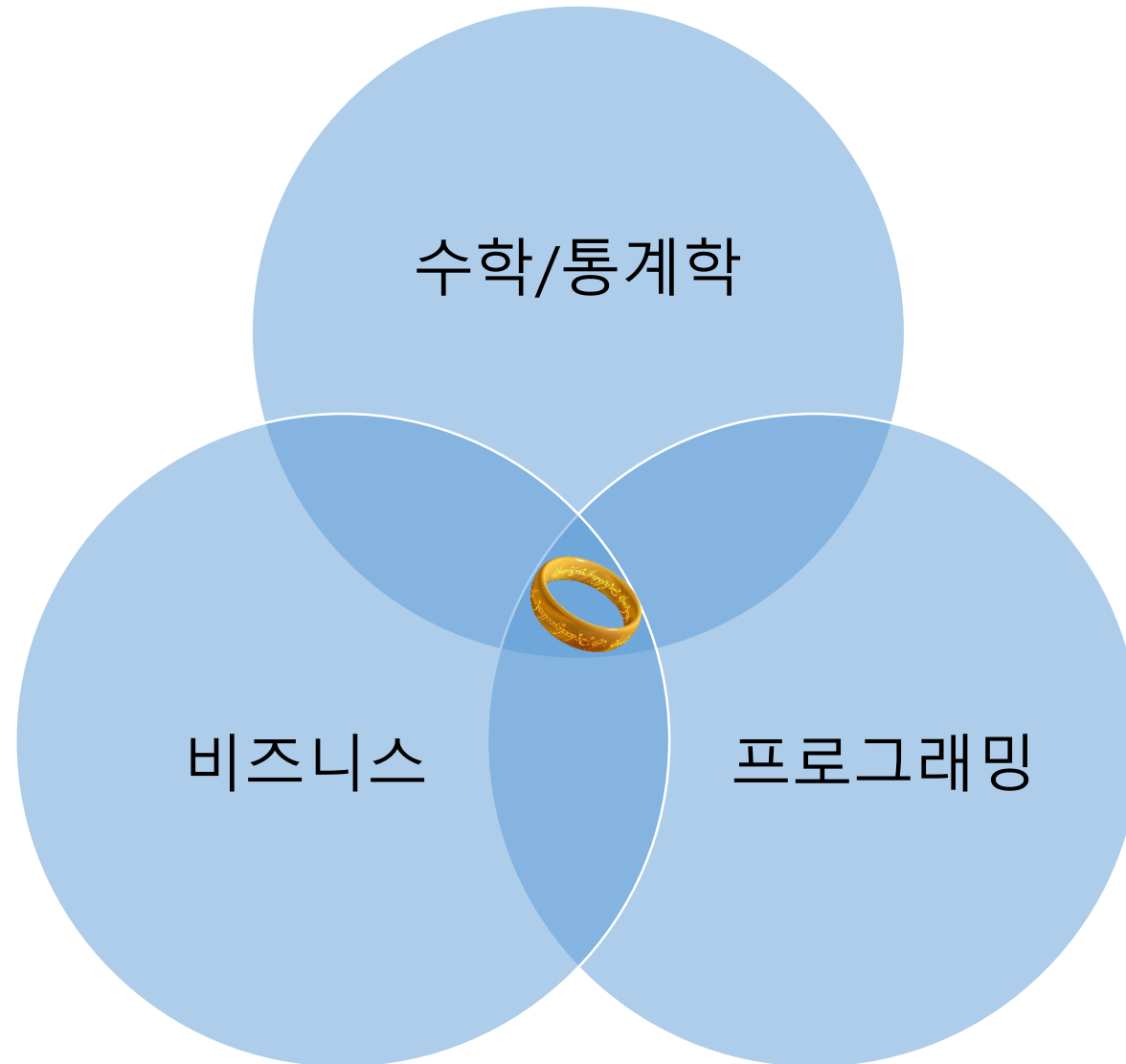


나는 누구?

머신러닝이란 무엇인가



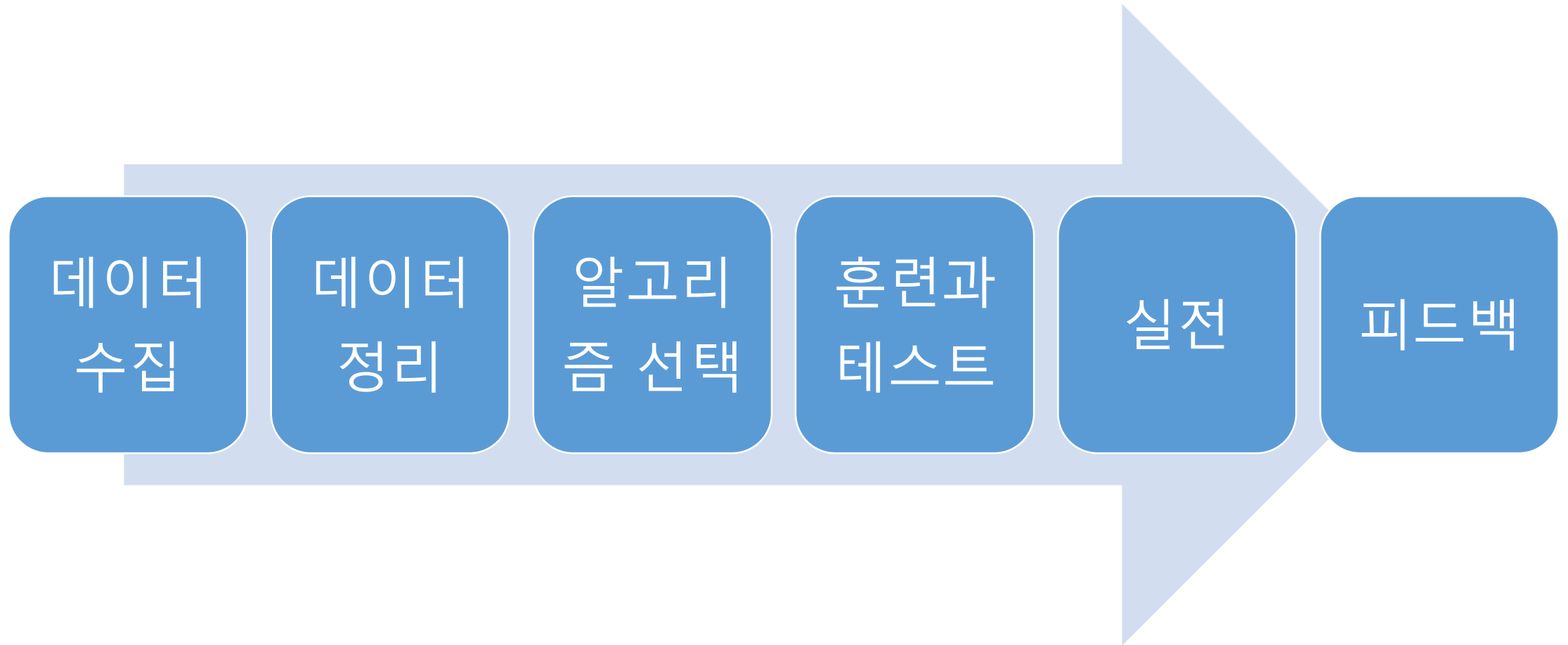
머신러닝이란 무엇인가



머신러닝의 사례

- 아마존 추천 알고리즘
- 넷플릭스 웹 페이지 개인화
- 구글 뉴스
- 우체국 우편번호 인식
- 구글 무인자동차, 로봇릭스
- thegrid.io 자동 웹 페이지 디자인
- 스팸, 악성코드, 신용카드 범죄 자동인식
- 주식거래. 금융시장 예측.
- 온라인 광고. 검색.
- IBM 왓슨 자연언어인식
- 얼굴인식
- 코넬 대학 음성인식기술로 고래보호
- IBM 의사 손 글씨를 해석해서 환자의 심장이상 징후를 예측하는 시스템 개발
- 영화 시나리오, 소설, 작곡, 미술
- 히트곡 예측 시스템
- 날씨 예측
- 고객 행동 예측
- 교통 통제
- 게임 소프트웨어
- 학습/교육 소프트웨어
- 선거 결과, 스포츠 시합 결과 예측

머신러닝의 단계



머신러닝과 개발자

개발자가 저지르는 5가지 잘못 – 제이슨 브라운리

1. 머신러닝은 학계의 전문가만 다룰 수 있다.
2. 머신러닝 코드를 스스로 작성해야 한다.
3. 데이터 처리에서 시각화에 이르는 과정을 직접 해결해야 한다.
4. 필요한 머신러닝 알고리즘과 해결책을 스스로 만들어야 한다.
5. 수학은 무시해도 좋다.

머신러닝의 종류



지도학습 (supervised learning)

- 회귀 (regression)
- 분류 (classification)



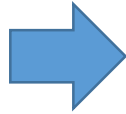
자율학습 (unsupervised learning)

- 군집화 (clustering)

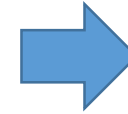
지도 학습 (supervised learning)

문제 + 정답
문제 + 정답
문제 + 정답

데이터 + 레이블



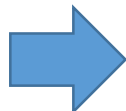
머신러닝 알고리즘



```
y''[t] == -9.81 +  $\frac{yDrag[x'[t], y'[t]]}{mass}$   
y'[0] == v Sin[θ], (*Initial vertical  
y[0] == 0, (*Initial altitude*)  
x''[t] ==  $\frac{xDrag[x'[t], y'[t]]}{mass}$ ,  
x'[0] == v Cos[θ], (*Initial horizont
```

모델

새로운 데이터



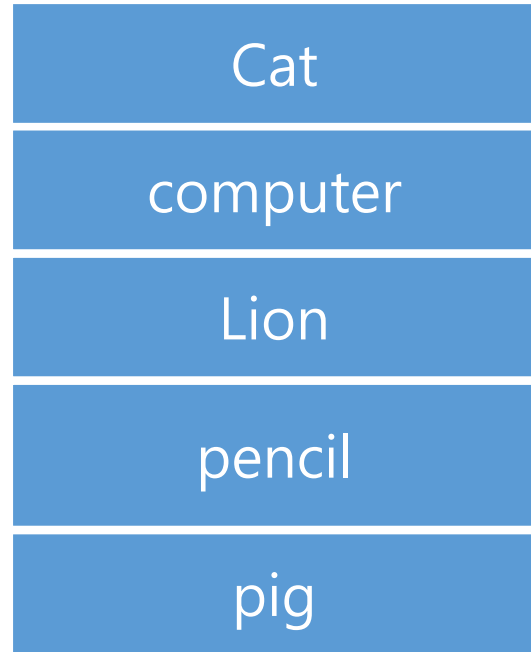
```
y''[t] == -9.81 +  $\frac{yDrag[x'[t], y'[t]]}{mass}$   
y'[0] == v Sin[θ], (*Initial vertical  
y[0] == 0, (*Initial altitude*)  
x''[t] ==  $\frac{xDrag[x'[t], y'[t]]}{mass}$ ,  
x'[0] == v Cos[θ], (*Initial horizont
```

모델

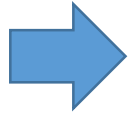


예측

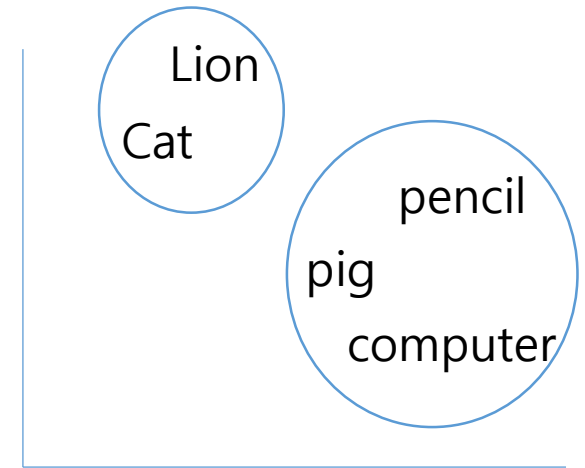
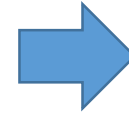
자율 학습 (unsupervised learning)



알려지지 않은 데이터



머신러닝 알고리즘



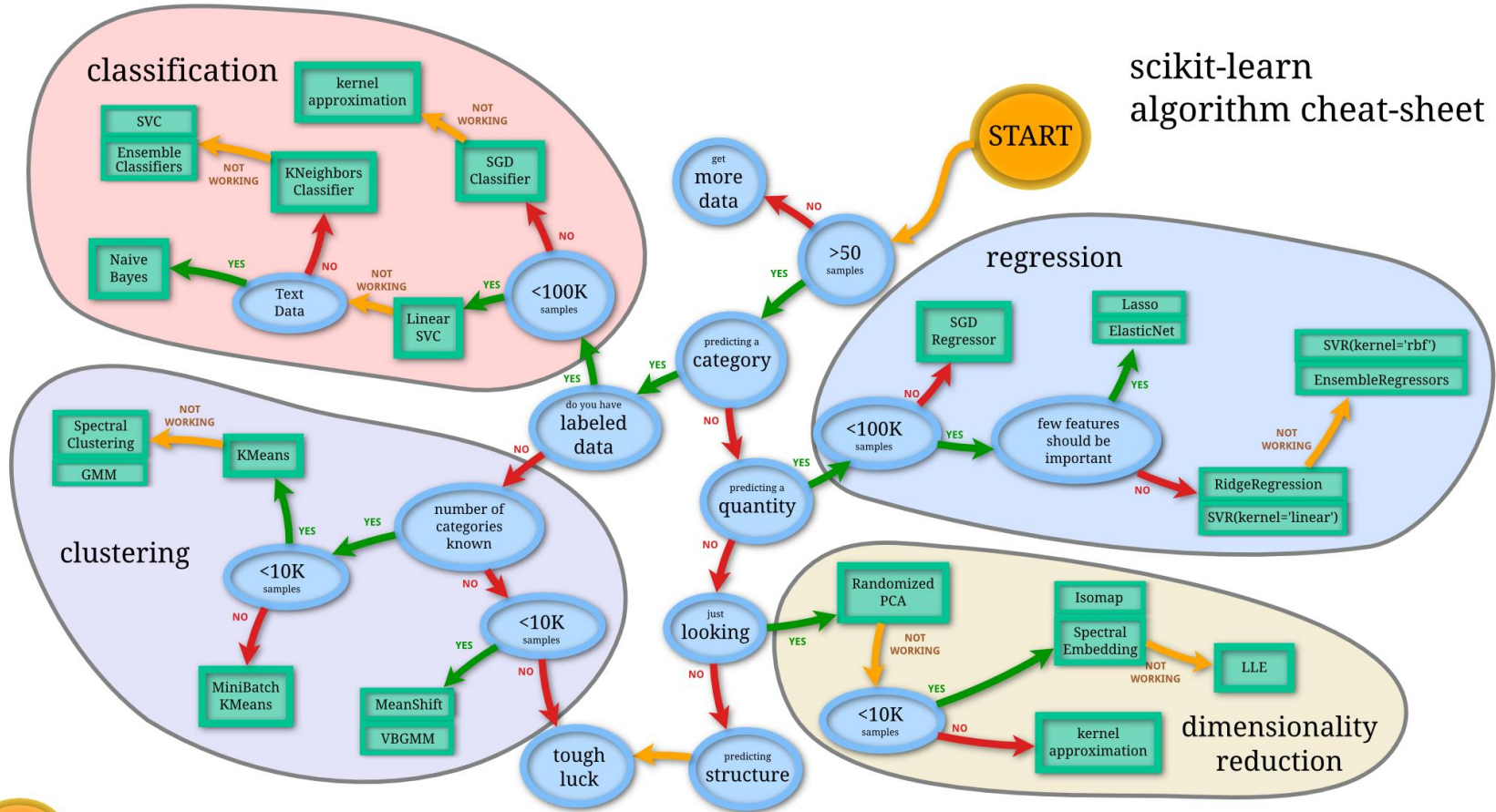
그룹

머신러닝 알고리즘 종류



	자율 학습 (unsupervised)	지도 학습 (supervised)
연속 변수 (continuous)	분류화 (clustering)	회귀 (regression) <ul style="list-style-type: none">• 결정 트리• Random Forests• Boosting Trees• Neural Networks• Support Vector Regression
범주 변수 (categorical)	분류화 (clustering)	분류 (classification) <ul style="list-style-type: none">• Naïve-Bayes• K-Nearest Neighbors• Logistic Regression• Support Vector Machine• Trees

알고리즘 선택 커닝 페이퍼



머신러닝 서비스와 라이브러리

머신러닝 서비스 = 클라우드 + 빅데이터 + 라이브러리

- Azure ML
- Google Prediction API
- IBM Watson Analytics API
- PredictionIO on Amazon AWS

머신러닝 라이브러리

- Apache Spark Mllib
- Apache Mahout
- Scikit-learn

학습자료

- 머신러닝 인 액션, 피터 해링턴, 제이펍
- Machine Learning, Andrew Ng, coursera.org
(<http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=MachineLearning>)
- Machine Learning: The Basics, youtube.com
- How to Get Started with Machine Learning, youtube.com
- 머신러닝 기초, <http://www.whydsp.org/237>
- 서비스, 라이브러리 홈페이지