

Azure Machine Learning

머신러닝 시작부터 예측모델 배포까지

<https://github.com/CloudBreadPaPa/azure-ml-busan>

<http://onoffmix.com/event/103412/>

Azure Machine Learning

개발자와 머신러닝(?)	01
어디에 어떻게 사용해야 하나	02
머신러닝 데모	03
학습모델 / 예측모델	04
지도 학습 / 비지도 학습 / 분석 알고리즘	05
예측모델 생성 데모	06
예측 모델을 API로 노출 및 Python 등에서 사용	07

개발자 & 머신러닝

“

뭐... 원데 그게?

”

deep dive: handling of time

extend our example to an RNN

$$h_1(t) = \sigma(W_1 x(t) + H_1 h_1(t-1) + b_1)$$

$$h1 = \text{Sigmoid}(w1 * x + H1 * \text{PastValue}(h1) + b1)$$

$$h_2(t) = \sigma(W_2 h_1(t) + H_2 h_2(t-1) + b_2)$$

$$h2 = \text{Sigmoid}(w2 * h1 + H2 * \text{PastValue}(h2) + b2)$$

$$P(t) = \text{softmax}(W_{\text{out}} h_2(t) + b_{\text{out}})$$

$$P = \text{Softmax}(w_{\text{out}} * h2 + b_{\text{out}})$$

$$ce(t) = L^T(t) \log P(t)$$

$$ce = \text{CrossEntropy}(L, P)$$

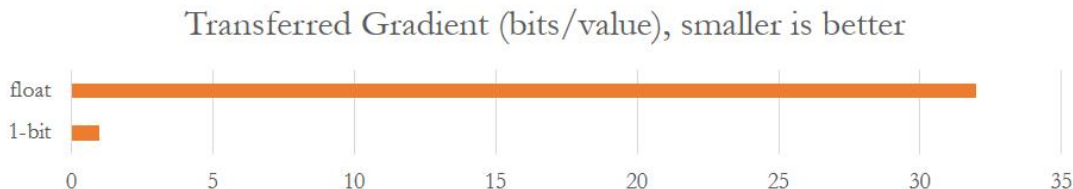
$$\sum_{\text{corpus}} ce(t) = \max$$

→ no explicit notion of time

deep dive: 1-bit SGD

- quantize **gradients** to but **1 bit per value** with **error feedback**
 - carries over quantization error to next minibatch

$$\begin{aligned}G_{ij\ell}^{\text{quant}}(t) &= \mathcal{Q}(G_{ij\ell}(t) + \Delta_{ij\ell}(t - N)) \\ \Delta_{ij\ell}(t) &= G_{ij\ell}(t) - \mathcal{Q}^{-1}(G_{ij\ell}^{\text{quant}}(t))\end{aligned}$$



“

여긴 어디?
난 누구?

”

“

OK.
잠시 방황한거 뿐이야.

”

“

어느 교수님 말씀 :

단지, 우리와 단어가 다를 뿐이야

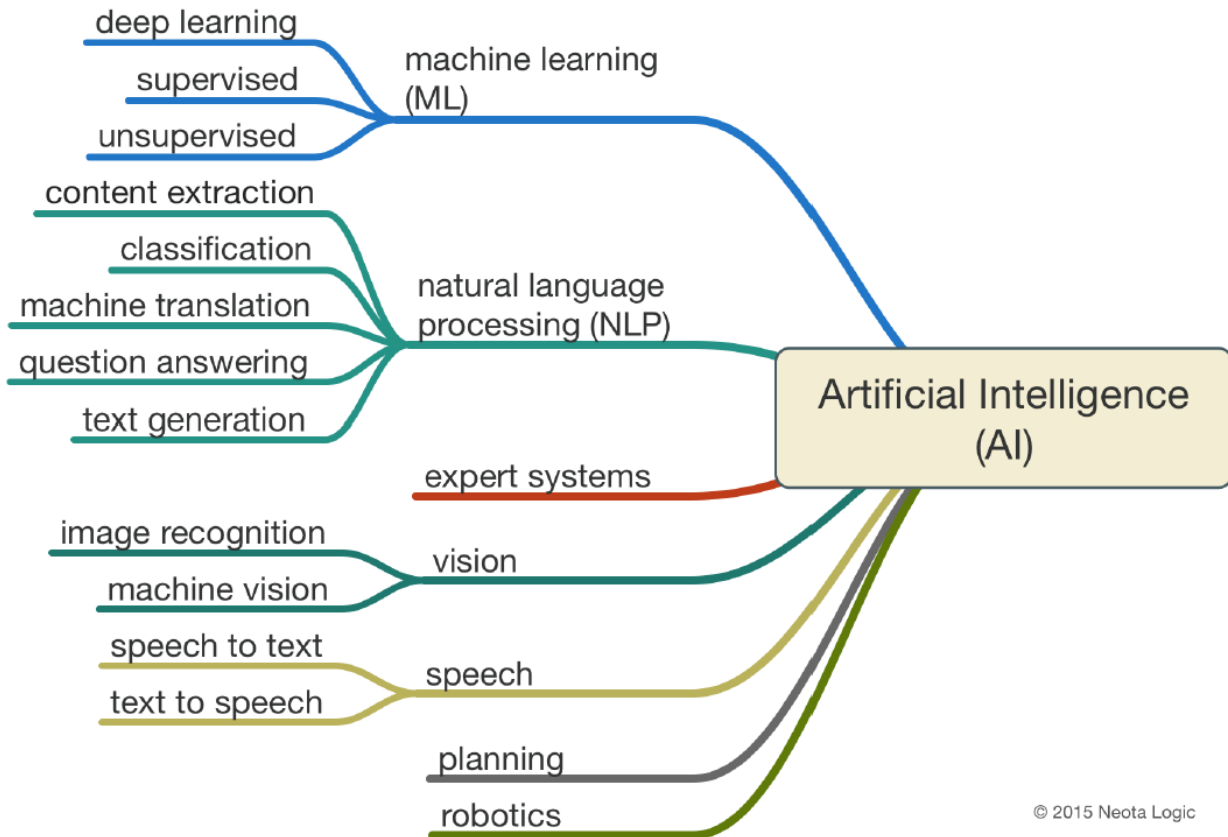
”

“

R, SAS, Python?
Tensorflow, CNTK, H2O, keras?
Scikit, Caret, fastcluster, party?
Azure ML, Google ML, Amazon ML?

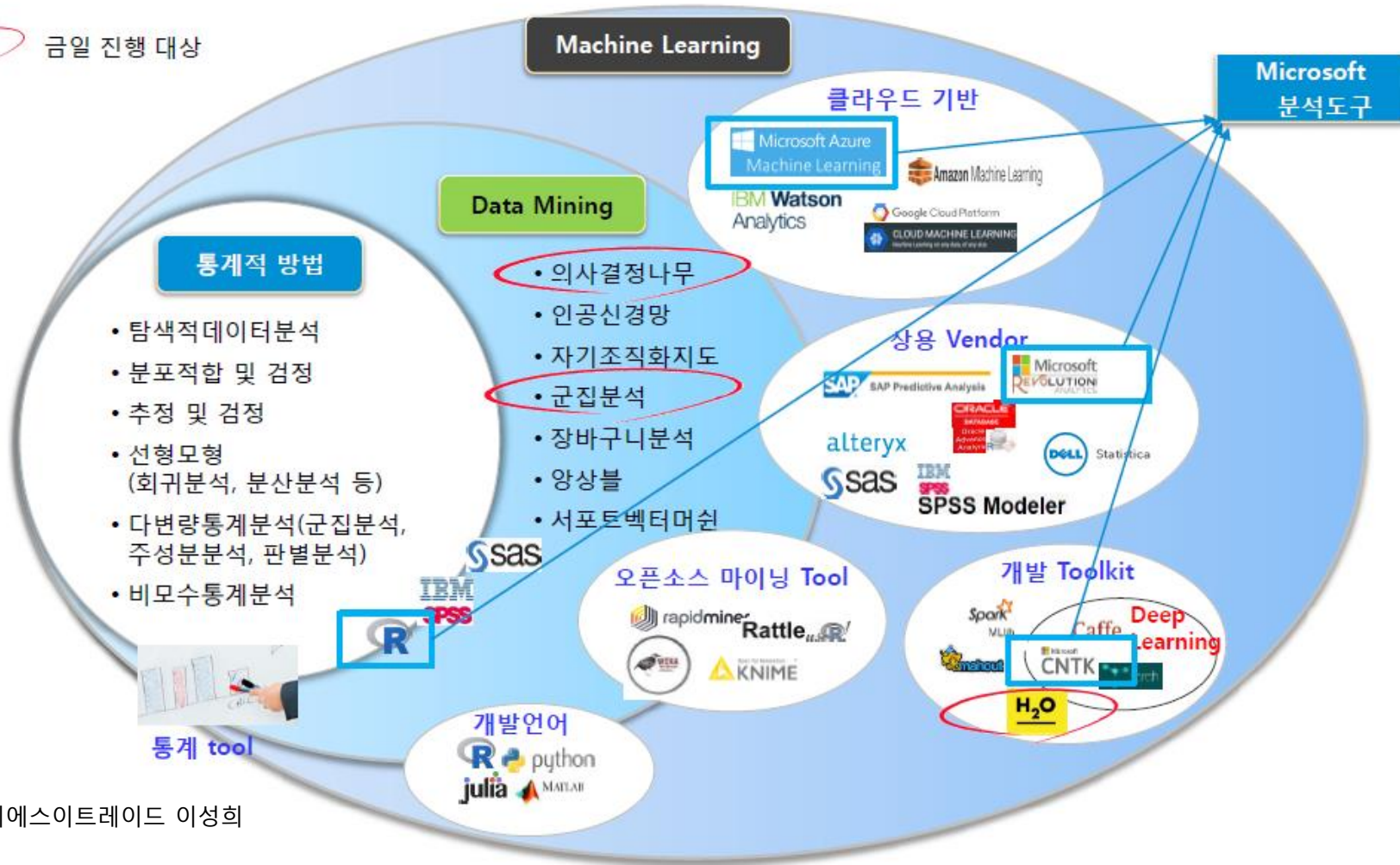
”

머신러닝과 인공지능, 인지서비스(Cognitive Service)



Machine Learning의 영역은 통계적 방법, 데이터마이닝 등 기존 분석기법들을 포괄하고 있음

금일 진행 대상



“

Machine Learning & Cloud(?)

”

“

ML로 태어나 Cloud에서 산다

”

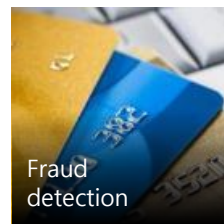
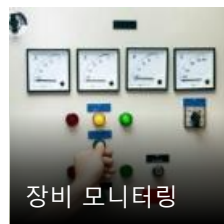
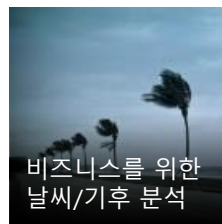
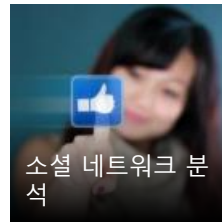
ML 알고리즘을 “개발”하는 개발자

ML 알고리즘을 “활용”하는 개발자

Machine Learning

어디에 어떻게 활용해야 하나?

예측 분석을 이용한
기술은 향후 모든
산업에 필요 충분 조건



산업별 Machine Learning 사례

금융

계정생성 시 검증

Fraud 방지

부정거래

예금관리

보험 설계

대출처리



유통

고객의 입장에서 분석

브랜드 분석

개인화 서비스 (가격 민감도 분석)

개인 프로모션 및 지역화

웹 사이트 최적화

매장의 전시계획



통신

CDR(Call detail record)분석

인프라 투자 계획예측

차기 제품 구매

실시간 대역폭 할당

신제품 개발

Azure 투자 계획 및 실행



제조

재료공급 시기

SCM 및 물류

조립라인의 품질검증

사전 품질 관리

장치의 오류 분석 및 AS시기예측



의료

유전자 데이터 분석 실험

환자 상태 모니터링

재발율 감소

의료 데이터의 저장

약품의 개발

질병 패턴 분석



석유화학

원전 분석

에너지 개발 및 수요량 분석/예측

컴플라이언스 보고서

능동적인 장치수리

이미지 프로세싱



공공 서비스

공공 자료 분석

재난재해 위험상황의 예측

자원이 낭비와 소모감시

사회시설에 대한 모니터링과 예산 편성

각종 통계성 업무 활용



학습모델 / 예측모델

“

바보(머신)에게 공부할 기회를
= 학습모델

”

“

바보(머신)에게 배운거 물어볼까
= 예측모델

”

모델 구축 데모 + 코드

Data Munging(Wrangling)

ETL

Set by task & Record oriented task

“

우리 개발자들은!?

”

Python - NumPy, Pandas

R - array, Datafarme

...

또는, 완소 SQL + ETL 도구들!

지도 학습

(Supervised Learning)

비지도 학습

(Unsupervised Learning)

구분	Reinforcement Learning	Machine Learning (Supervised Learning)	비고
목적함수	보상을 최대화 (또는 손실을 최소화)	오차를 최소화 (오차 = 추정 - 실제)	
산출방식	순차적으로 현재 스테이지의 보상과 총 보상을 산출하여 "총 보상"이 최대화 되도록 함	실제 사례를 기반으로 사례와 가장 유사하게 모사하도록 함수를 구성하도록 함	
산출 방법론	Optimization	분류문제와 예측문제로 구분되며 다양한 알고리즘 존재	
데이터 구성	State 별 Action Matrix (모든 가능한 State 각각에 대한 모든 실행 가능한 Action과 확률)	State vs. Action에 대한 성공과 실패 사례	
특징	규칙기반으로의 설계가 용이함 (Heuristic 설계 용이)	- 데이터마이닝: 규칙(if/else) 기반 설계 용이 - 기계학습: 규칙 파악이 어려움	기계학습은 정확도 향상이 주 목표임
구현 방법	최적화 엔진 (Dynamic LP 등)	통계 소프트웨어 또는 기계학습 엔진	

Unsupervised Learning

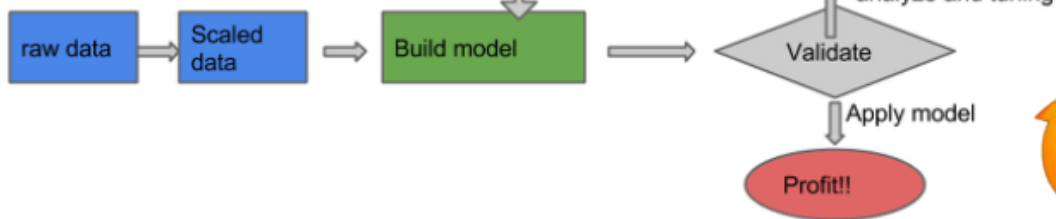
Supervised Learning

학습대상이 있는지(Target의 여부)에 따라

→ Unsupervised Learning(自律학습)과 Supervised Learning(指導학습)으로 구분

Unsupervised Learning

Unsupervised learning(자율학습)



Clustering

- Hierarchical Clustering
- K-Means
- Mixture Modeling

Supervised Learning

Supervised learning(지도학습)



Target

Continuous

- Decision Tree
- Boosting Trees
- Random Forest
- SVM(Support Vector Machine)
- Neural Networks

Discrete

- Naïve Bayes
- K-Nearest Neighbors
- Logistic Regression
- SVM

“

산업 시나리오

”

“

게임 User가 게임을 이탈
{할지 | 안할지} : 예측

게임 User를 위한 아이템 추천
{ i1 | i2 | i3 ... } : 예측

”

Machine Learning

Game user churn prediction
In-Game item suggestion

예측모델 생성 데모

API로 노출

Python 등에서 API 사용 DEMO

Restful Front-End를 이용한
Machine Learning API 호출
+ 대량 Batch 분석
DEMO

Q & A