

WELCOME WELCOME WELCOME WELCOME WELCOME WELCOME WELCOME

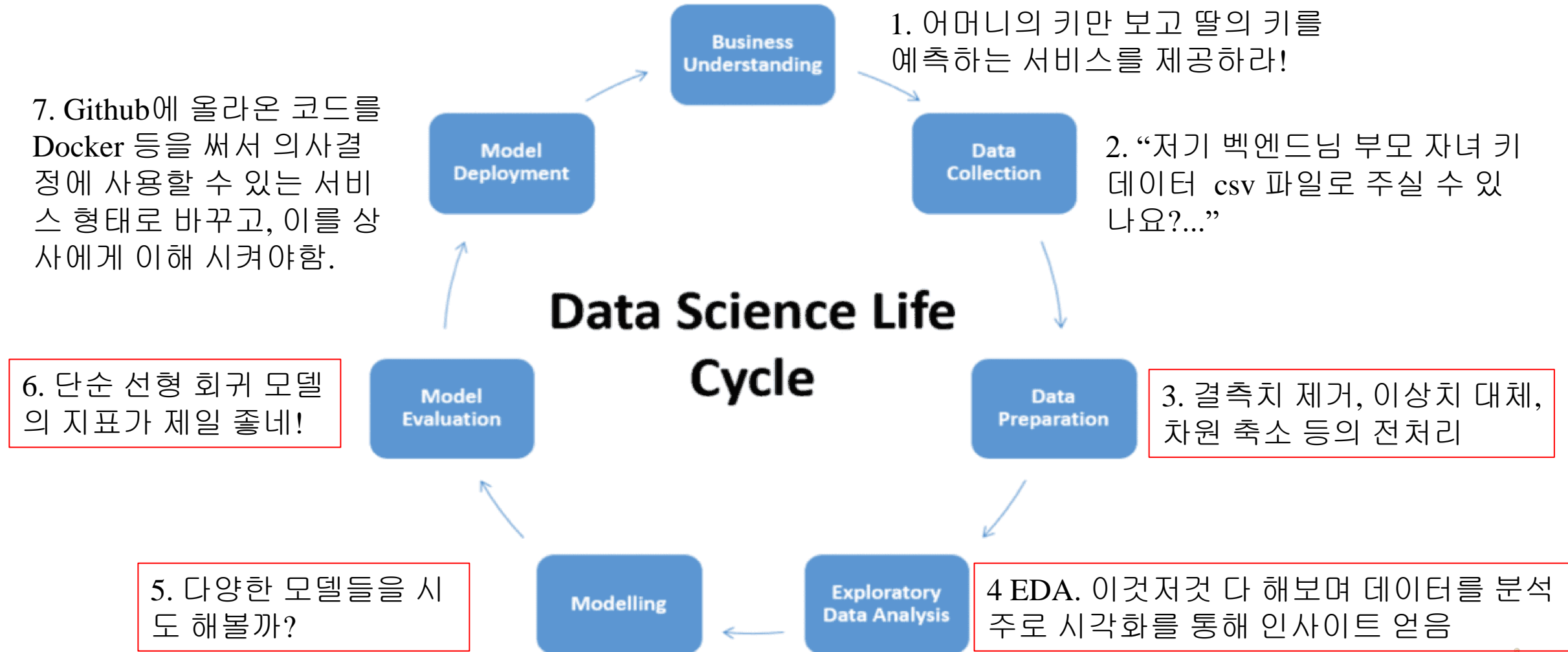
DAICON 스타디 1주차 이제윤



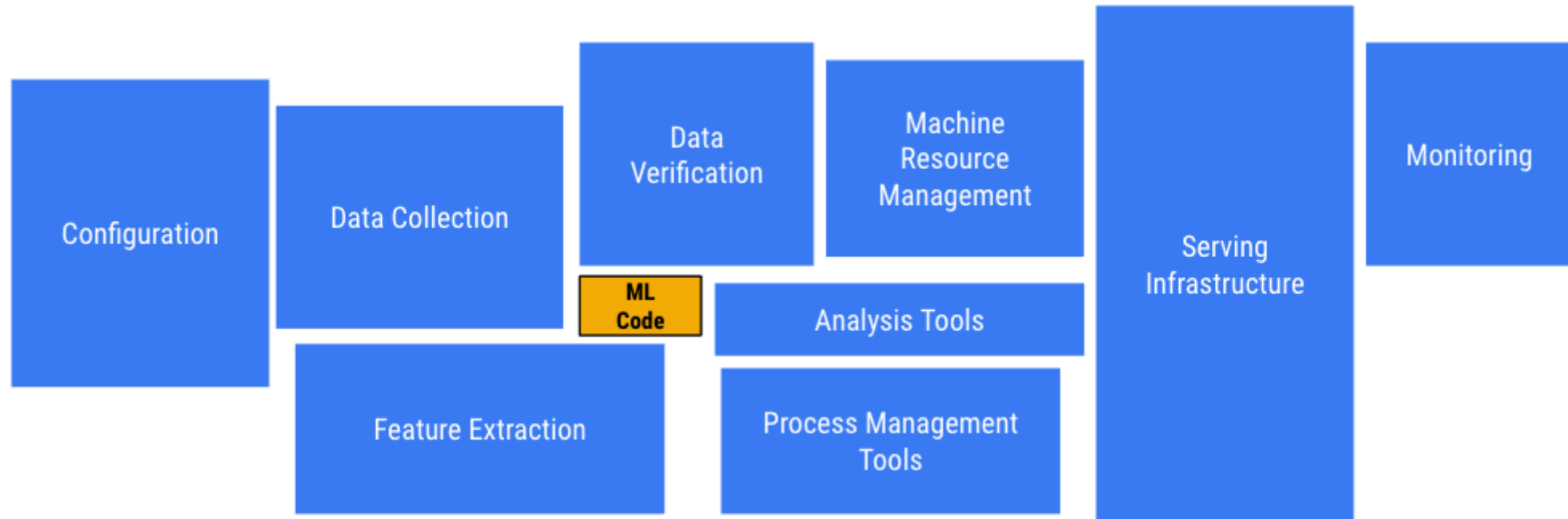
일정 및 계획

- 매주 화요일 3시30분부터 1시간 반 내외의 강의 진행
- 나머지 30 분 정도는 과제에 대한 공지와 설명
- 과제는 주로 Python으로 수행하고 결과는 Google Drive에 올리기.
- 팀과제는 4, 5, 6주차동안 진행할 생각 3명 * 2팀 임의로 배정

데이터분석 대회 나가서 1등 하고 싶다 [Data Science Life Cycle](#)



CDS(citizen data scientists) & EDS(expert data scientists)



CDS(citizen data scientists) & EDS(expert data scientists)

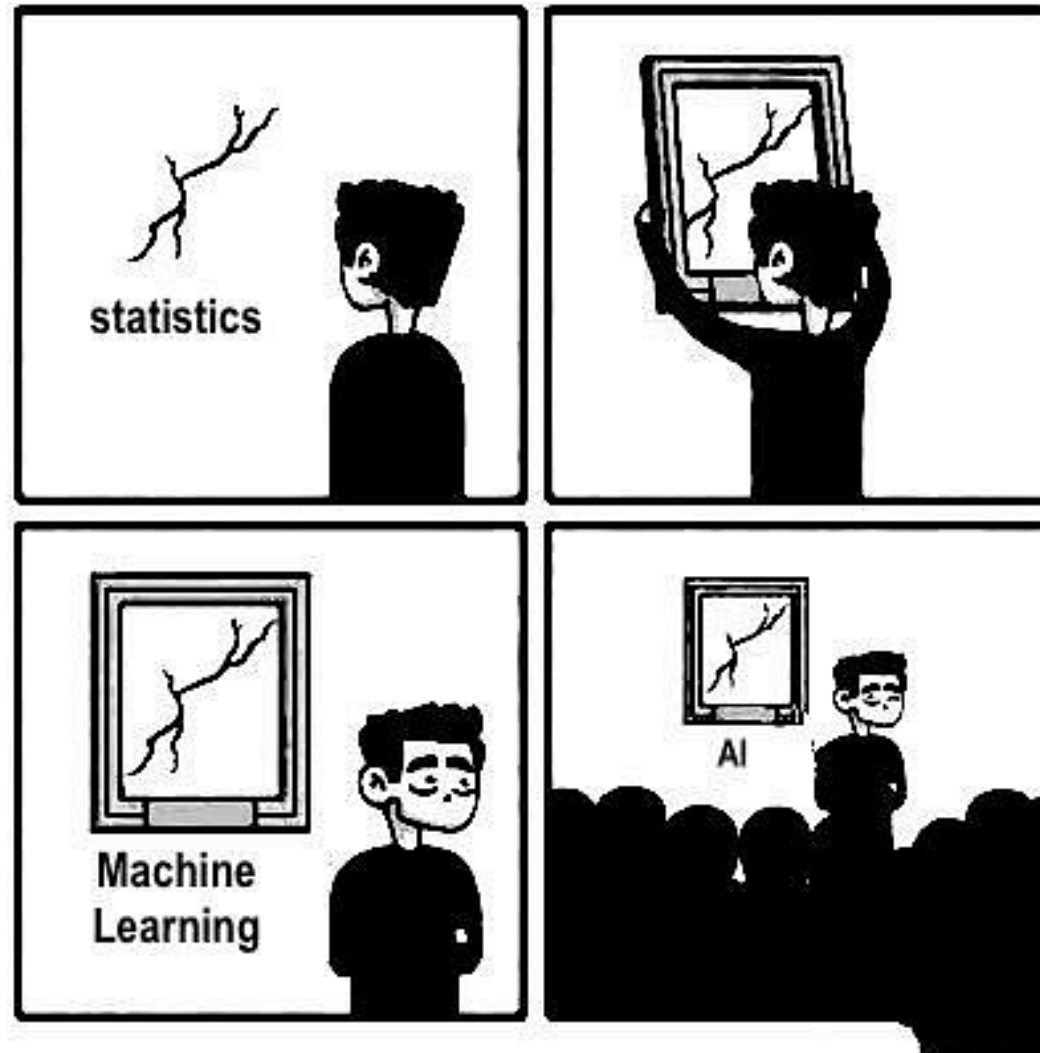
CDS(citizen data scientist)가 될 것인가 EDS(expert DS)가 될 것인가

<https://www.dataversity.net/citizen-data-scientists-where-do-they-belong/>

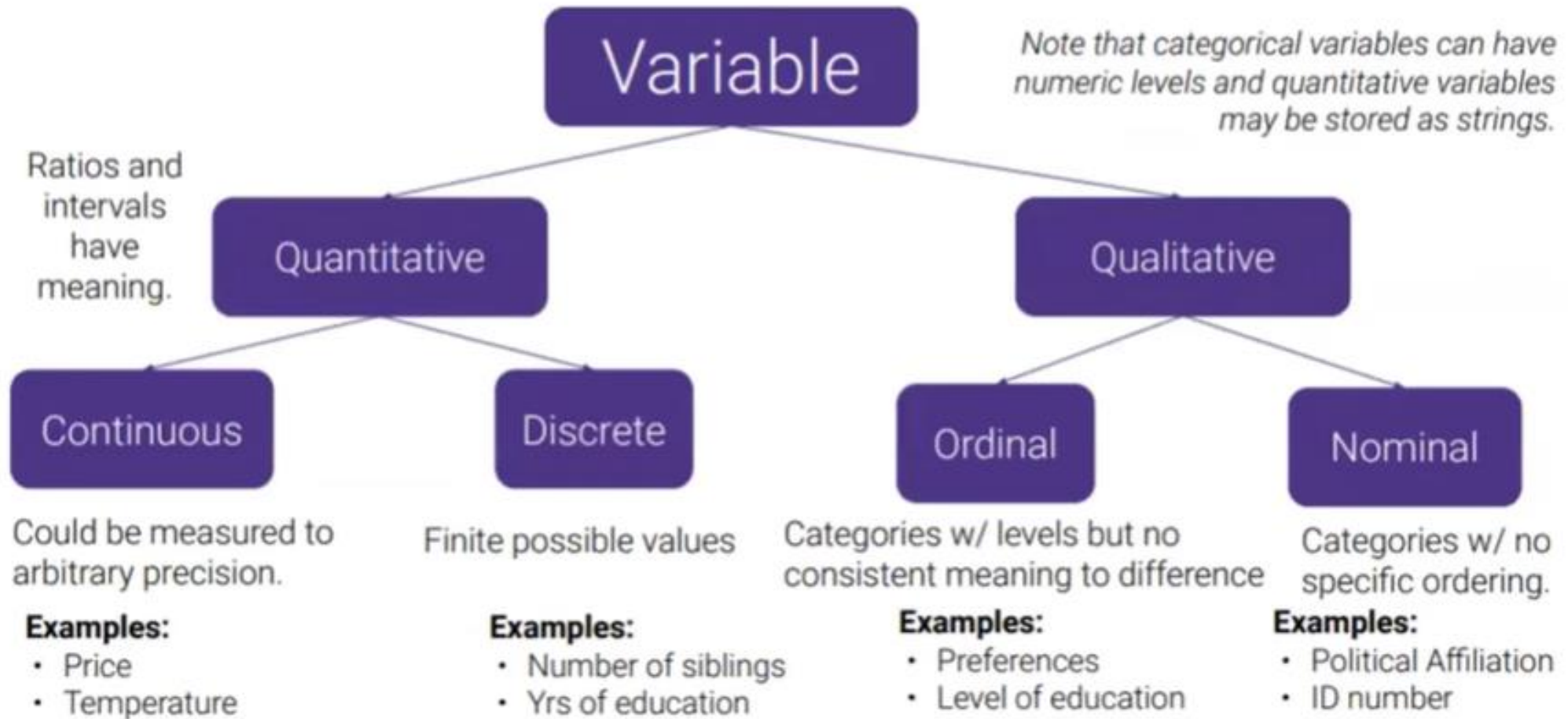
The Data Scientist vs. the Citizen Data Scientist

With **CDSs** growing “five times faster in numbers” than data scientists, there is a clear indication that automated ML packages and augmented analytics platforms are working. Author Kartik Patel, in *Analytics Translator vs. Citizen Data Scientist: What is the Difference?*, reiterates that Gartner has pointed out that the CDS creates models for predictive or prescriptive analytics, but this person’s main job function is outside of the realm of analytics or statistics. The CDS is not trained to be a computer scientist or data analyst, **but with the help of advanced data technologies and tools**, this role becomes empowered to analyze data and make data-driven, business decisions. The increasingly growing number of self-serve analytics and BI platforms with automated data preparation, built-in ML models, and augmented analytics capabilities empowers the CDS to conduct routine analytics and BI tasks **without the presence of a data scientist**.

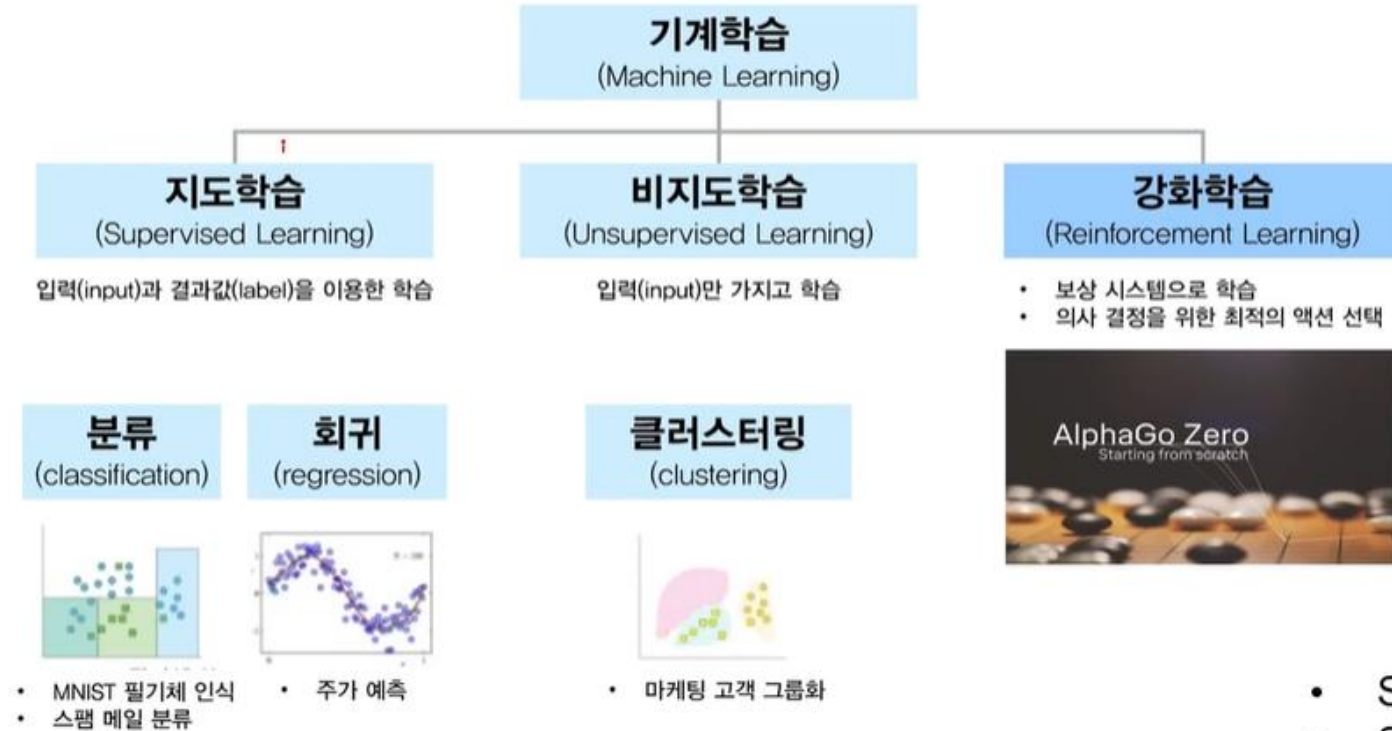
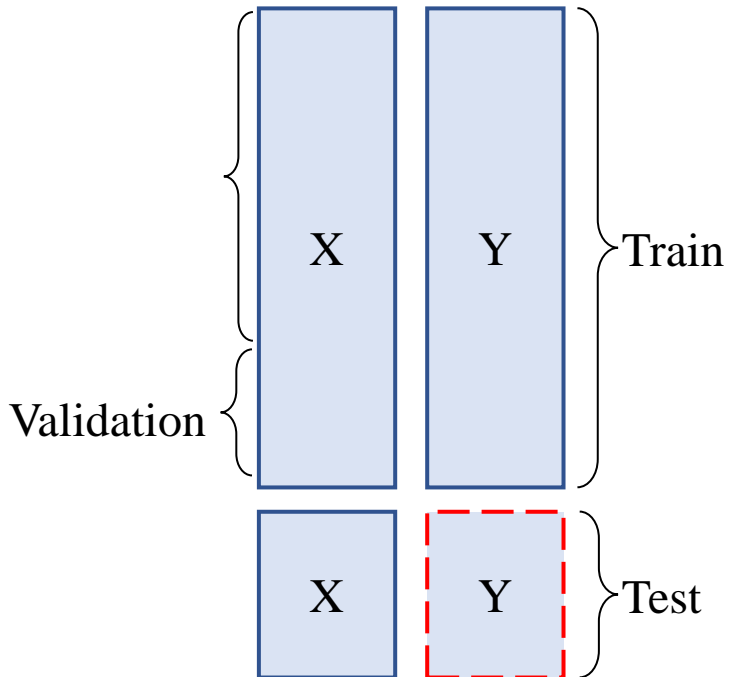
Machine Learning ???



REVIEW about variables

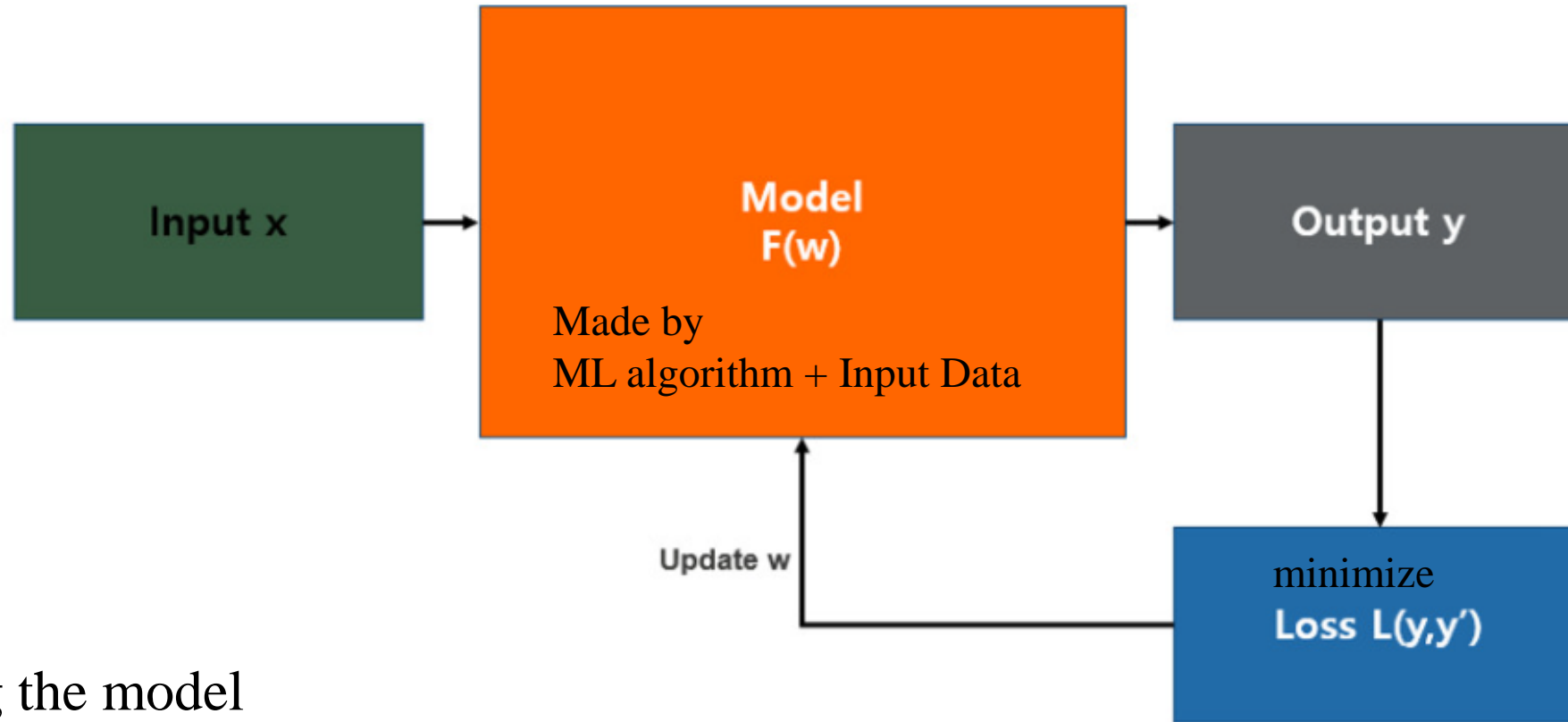


Types of Machine Learning



- Semi-Supervised
- Self-Supervised L

Machine “Learning”

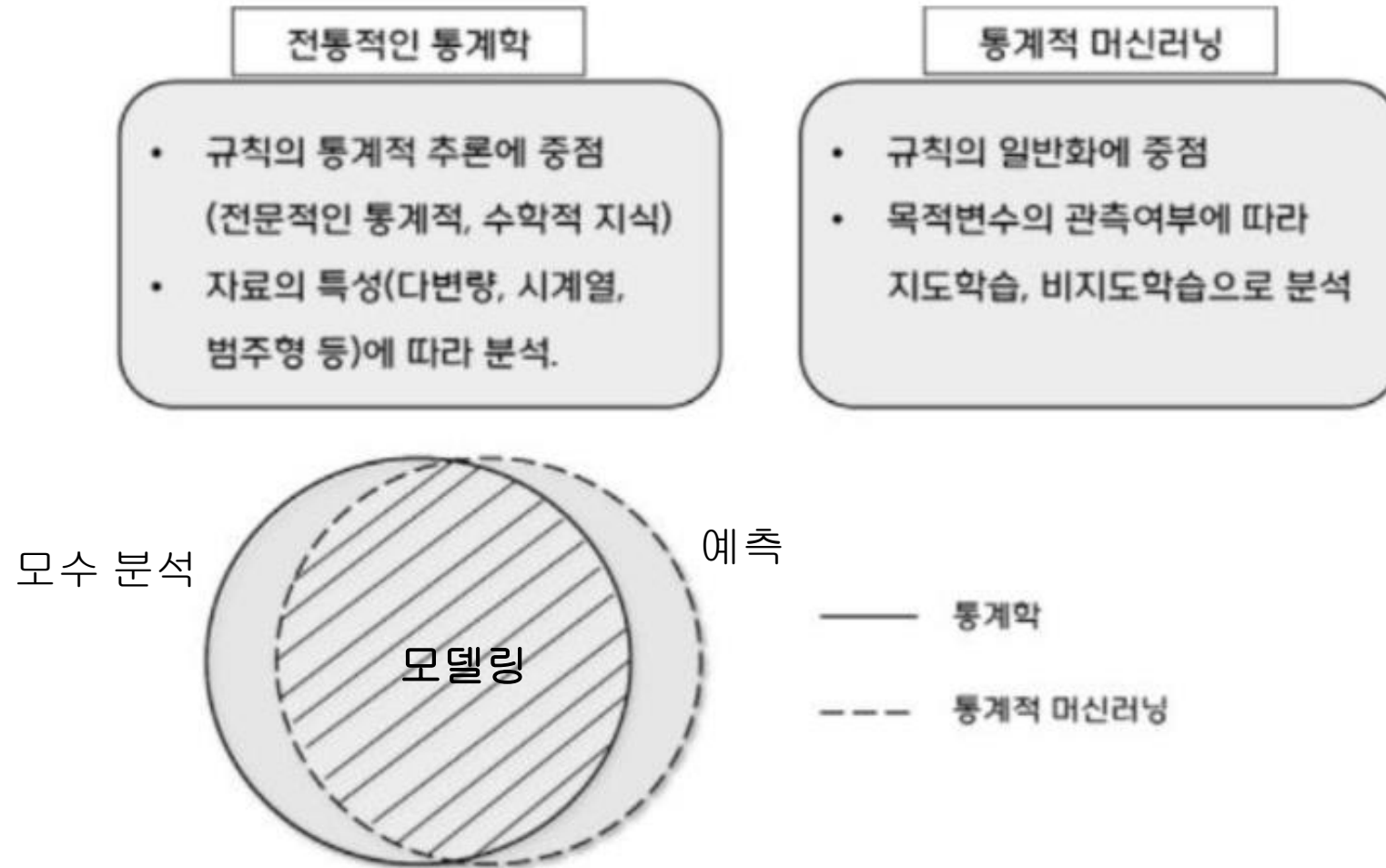


Optimizing the model

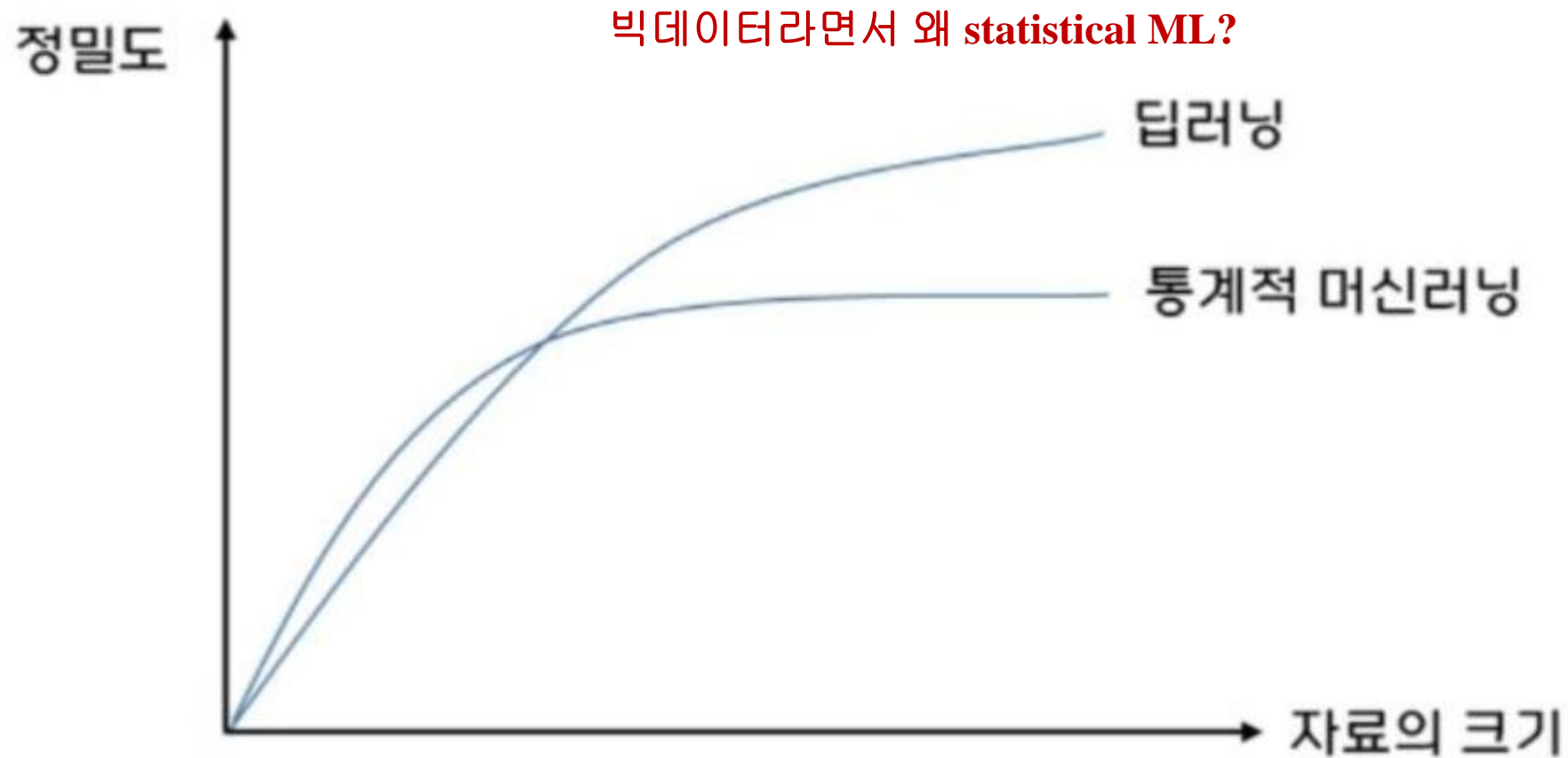
기준: Loss Function을 최소화하는 방향으로

어떻게: Newton method (R) / Gradient Descent (Python scikit-learn)

What is Statistical ML compared with Traditional Statistics



Machine Learning vs Deep Learning



Machine Learning vs Deep Learning

구분	통계적 머신러닝	딥러닝
데이터 크기	중/소 크기	빅데이터
분석자료 형태	2차원 텐서	2차원 텐서이상
강점을 갖는 자료	정형화된 자료	비정형자료
특성변수	특성변수를 만들어야 함	특성변수가 만들어짐
특성변수의 정규화 및 표준화	선택	필요
모형	매우 많음	기본적으로 3 개의 모형
최적화	일반적으로 전체 데이터 사용	배치데이터
해석여부	해석이 쉬움 (단, SVM과 boosting 제외)	어렵거나 불가능
하드웨어	중급	고성능(GPU 요구)
실행요구시간	최대 시간 단위	최대 주단위 시간

Let's go

기계가 학습한다는 건 무엇일까?

<https://brunch.co.kr/@gdhan/5>

<https://starcell.github.io/ai/ml-basic/>

기계학습에서 학습은 이 데이터셋을 가지고 다음과 같은 방식으로 진행이 됩니다.

- 계수 초기화
- 초기화된 계수와 입력데이터를 가지고 출력값을 계산
- 이렇게 계산해서 구한 출력값과 데이터셋에서 가지고 있던 출력값을 비교
- 이 두 값의 차이를 줄이는 쪽으로 계수들을 수정(개선)
- 새로 개선된 계수와 입력값을 가지고 출력값을 계산
- 데이터셋에서 가지고 있던 출력값과 비교
- 이러한 절차를 많은 데이터셋에 대하여 반복적으로 시행

Loss Function과 Likelihood Function 설명 예정

Let's minimize a Loss function (a view of machine learning)

손실함수(loss function)

- 손실 함수(loss function)란?

머신러닝 혹은 딥러닝 모델의 출력값과 사용자가 원하는 출력값의 오차를 의미

손실함수는 정답(y)와 예측(\hat{y})를 입력으로 받아 실숫값 점수를 만드는데, 이 점수가 높을수록 모델이 안좋은 것

손실함수의 함수값이 최소화 되도록 하는 가중치(weight)와 편향(bias)를 찾는 것이 딥러닝 학습의 목표

손실함수의 종류

예시

$RSS, MSE, MAE, MAPE, R^2$

Let's minimize a Loss function (a view of machine learning)

손실 함수는 데이터를 토대로 산출한 모델의 예측 값과 실제 값과의 차이를 표현하는 지표입니다. 조금 더 쉽게 설명하면, 손실 함수는 모델 성능의 '나쁨'을 나타내는 지표로, “현재의 모델이 데이터를 얼마나 잘 처리하지 못하느냐”를 나타내는 지표라고 할 수 있습니다. “얼마나 나쁘냐”를 어떤 방식으로 표현하느냐에 따라 여러 가지 손실 함수가 존재합니다.

<https://brunch.co.kr/@mnc/9>

통계학적 모델은 일반적으로 회귀(regression)와 분류(classification) 두 가지 종류로 나뉘지는데, 손실 함수도 그에 따라 두 가지 종류로 나뉘집니다. 회귀 타입에 쓰이는 대표적 손실 함수는 MAE, MSE, RMSE 가 있으며, 분류에 쓰이는 손실 함수는 Binary cross-entropy, Categorical cross-entropy 등이 있습니다. 다음은 언급된 손실 함수 종류와 각각의 특징에 대해서 설명하도록 하겠습니다.

MAXIMIZE a Likelihood function (a probabilistic view)

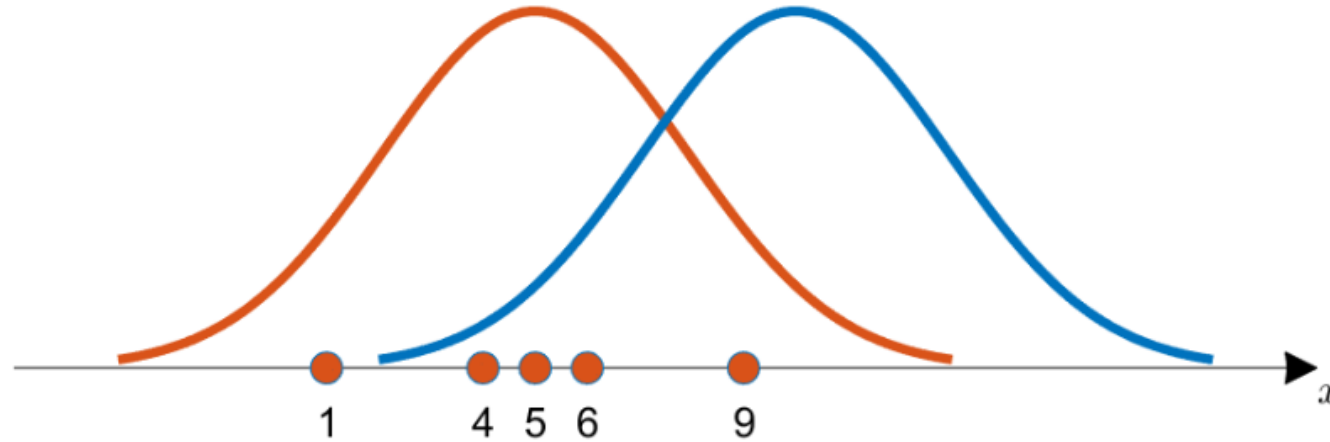


그림 1. 획득한 데이터와 추정되는 후보 분포 2개(각각 주황색, 파란색 곡선으로 표시)

눈으로 보기에다 파란색 곡선 보다는 주황색 곡선에서 이 데이터들을 얻었을 가능성이 더 커보인다.

왜냐면 획득한 데이터들의 분포가 주황색 곡선의 중심에 더 일치하는 것 처럼 보이기 때문이다.

이 예시를 보면, 우리가 데이터를 관찰함으로써 이 데이터가 추출되었을 것으로 생각되는 분포의 특성을 추정할 수 있음을 알 수 있다. 여기서는 추출된 분포가 정규분포라고 가정했고, 우리는 분포의 특성 중 평균을 추정하려고 했다.

A KEY Table of the class !! (a Loss & Likelihood relationship)

	Model	Loss Function	Dist. of Likelihood Function	활성화 함수
분류 Classification	이진 분류 Binary	Binary Cross-Entropy Loss	Bernoulli Dist. Logistic Regression	Sigmoid function
	다중 분류 Multiclass	Categorical Cross-Entropy Loss	Multinomial Dist. Softmax Regression	Softmax function
회귀 Regression	선형 회귀 Linear	주로 MSE, RSS, MAE 등	Gaussian dist. (Normal dist.) Linear Regression	Identity function

Cross Entropy

~2:24

<https://www.youtube.com/watch?v=r3iRRQ2ViQM> about Entropy

Watch all

<https://www.youtube.com/watch?v=Jt5BS71uVfI> about Cross Entropy

Cross Entropy

Cross Entropy는 두 개의 확률분포 p 와 q 에 대해 하나의 사건 X 가 갖는 정보량으로 정의된다. 즉, 서로 다른 두 확률분포에 대해 같은 사건이 가지는 정보량을 계산한 것이다. 이는 q 에 대한 정보량을 p 에 대해서 평균낸 것으로 볼 수 있다 [4].

Cross Entropy:

$$H_{p,q}(X) = - \sum_{i=1}^N p(x_i) \log q(x_i)$$

Cross entropy는 기계학습에서 손실함수(loss function)을 정의하는데 사용되곤 한다. 이때, p 는 true probability로써 true label에 대한 분포를, q 는 현재 예측모델의 추정값에 대한 분포를 나타낸다 [13].



대표적인 예로써 **BINARY Classification**에 대해 설명!

$$y_1 = (1,0), y_2 = (0,1)$$

승인/ 거부

남 / 여

유 / 무

참 / 거짓

cf) Categorical Classification은 One-hot encoding 필요

$$y_1 = (1,0,0), y_2 = (0,1,0), y_3 = (0,0,1)$$

빨강/ 초록 / 파랑

A / B / 그 외

고양이/강아지/그 외

아이폰/갤럭시/그 외

Binary Cross-Entropy Loss VS Bernoulli dist. log Likelihood

$$\begin{aligned} H_{p,q}(Y|X) &= - \sum_{i=1}^N \sum_{y \in \{0,1\}} p(y_i | x_i) \log q(y_i | x_i) \\ &= - \sum_{i=1}^N [p(y_i = 1 | x_i) \log q(y_i = 1 | x_i) + p(y_i = 0 | x_i) \log q(y_i = 0 | x_i)] \\ &= - \sum_{i=1}^N [p(y_i = 1 | x_i) \log q(y_i = 1 | x_i) + \{1 - p(y_i = 1 | x_i)\} \log \{1 - q(y_i = 1 | x_i)\}] \\ &= - \sum_{i=1}^N [p(y_i) \log q(y_i) + \{1 - p(y_i)\} \log \{1 - q(y_i)\}] \end{aligned}$$



Binary Cross-Entropy Loss VS Bernoulli dist. log Likelihood

Cross Entropy via KL Divergence

KL Divergence를 변형하면 cross entropy에 대한 식으로 정리된다 [4].

$$\begin{aligned} D_{KL}(p\|q) &= \sum_{i=1}^N p(x_i) \left(\log \frac{p(x_i)}{q(x_i)} \right) \\ &= \sum_{i=1}^N p(x_i) \log p(x_i) - \sum_{i=1}^N p(x_i) \log q(x_i) \\ &= -H_p(X) + H_{p,q}(X) \end{aligned}$$

이를 $H_{p,q}(X)$ 에 대해 정리하면 다음과 같다.

$$H_{p,q}(X) = D_{KL}(p\|q) + H_p(X)$$

즉, cross entropy를 최소화하는 것은 KL Divergence를 최소화하는 것과도 같다. 그럼으로써 p 를 근사하는 q 의 확률분포가 최대한 p 와 가까질 수 있도록 예측모델의 파라미터를 조정하게 된다.



Binary Cross-Entropy Loss VS Bernoulli dist. log Likelihood

파라미터 π 를 따르는 어떤 확률분포를 $f(Y; \pi)$ 라고 할때, 관측데이터 y 에 대한 베르누이분포는 다음과 같다.

$$f(Y = y; \pi) = \pi^y (1 - \pi)^{1-y}, y \in \{0, 1\}.$$

이 함수는 $Y = 1$ 의 입력에 π 를, $Y = 0$ 의 입력에는 $1 - \pi$ 를 반환한다.

$$\begin{aligned} l(\pi | y) &= \log (L(\pi | y)) \\ &= \log \left(\prod_{i=1}^n f(y_i; \pi) \right) \\ &= \sum_{i=1}^n \log (f(y_i; \pi)) \\ &= \sum_{i=1}^n \log (\pi^{y_i} (1 - \pi)^{1-y_i}) \end{aligned}$$

$$= \sum_{i=1}^n (y_i \log(\pi) + (1 - y_i) \log(1 - \pi))$$



주어진 관측데이터에 대해 likelihood function $L(\pi | x)$ 를 가장 크게 하는 파라미터 π 를 추정하는 것을 *maximum-likelihood estimation*이라고 한다 [2]. 즉, 관측된 데이터를 기반으로 분포를 추정하는 것이다.

$$\arg \max_{\pi} L(\pi | y)$$

Likelihood를 최대화하는 π 는 또한 loglikelihood를 최대화하므로, likelihood 대신 loglikelihood를 사용할 수 있다.

$$\arg \max_{\pi} l(\pi | y)$$

여기서 loglikelihood에 -1을 곱하고 동일한 솔루션을 얻기 위해 argmax문제를 argmin문제로 변환한다.

$$\arg \min_{\pi} -l(\pi | y)$$

위 식을 전개해보자.

$$\arg \min_{\pi} - \sum_{i=1}^n (y_i \log(\pi) + (1 - y_i) \log(1 - \pi))$$

이로써 binary cross entropy를 최소화하는 문제가 정의되었다. Binary cross entropy는 파라미터 π 를 따르는 베르누이분포와 관측데이터의 분포가 얼마나 다른지를 나타내며, 이를 최소화하는 문제는 관측데이터에 가장 적합한(fitting) 베르누이분포의 파라미터 π 를 추정하는 것으로 해석할 수 있다.

과제!

- Hands on Machine Learning 54 ~ 109p 연습
 - 2. End-to-End Machine Learning Project 챕터에 해당하는 내용 전반적인 프로젝트 과정에 익숙해져 오는 것이 목표~

Reference

- About Loss function and Likelihood

[손실 함수\(loss function\) 정의 - MAE, MSE, MLE](#)

[\[손실 함수\] Binary Cross Entropy](#)

[\[ML101\] #3. Loss Function](#)

[\[AI 응용 ML\] 이론 | 3.3 RSS, MSE, MAE, \$R^2\$ \(결정 계수\)](#)

[\[머신러닝\] 손실함수의 종류](#)

[최대우도법\(MLE\) - 공돌이의 수학정리노트](#)

- KUBIG 2022-SPRING Machine Learning Study

- Hands on Machine Learning

- [#오늘의 파이썬 #1일1오파 #파이썬 #python – DAICON](#)

방학에 할 거
없으면 이거 해