

19주차 인공지능

인공지능 학습 팁



목차

01. 복습

02. 학습 모니터링

03. 오버피팅(Overfitting)

04. 실무 Q&A

05. 마무리

01

복습



✓ 데이터 수집

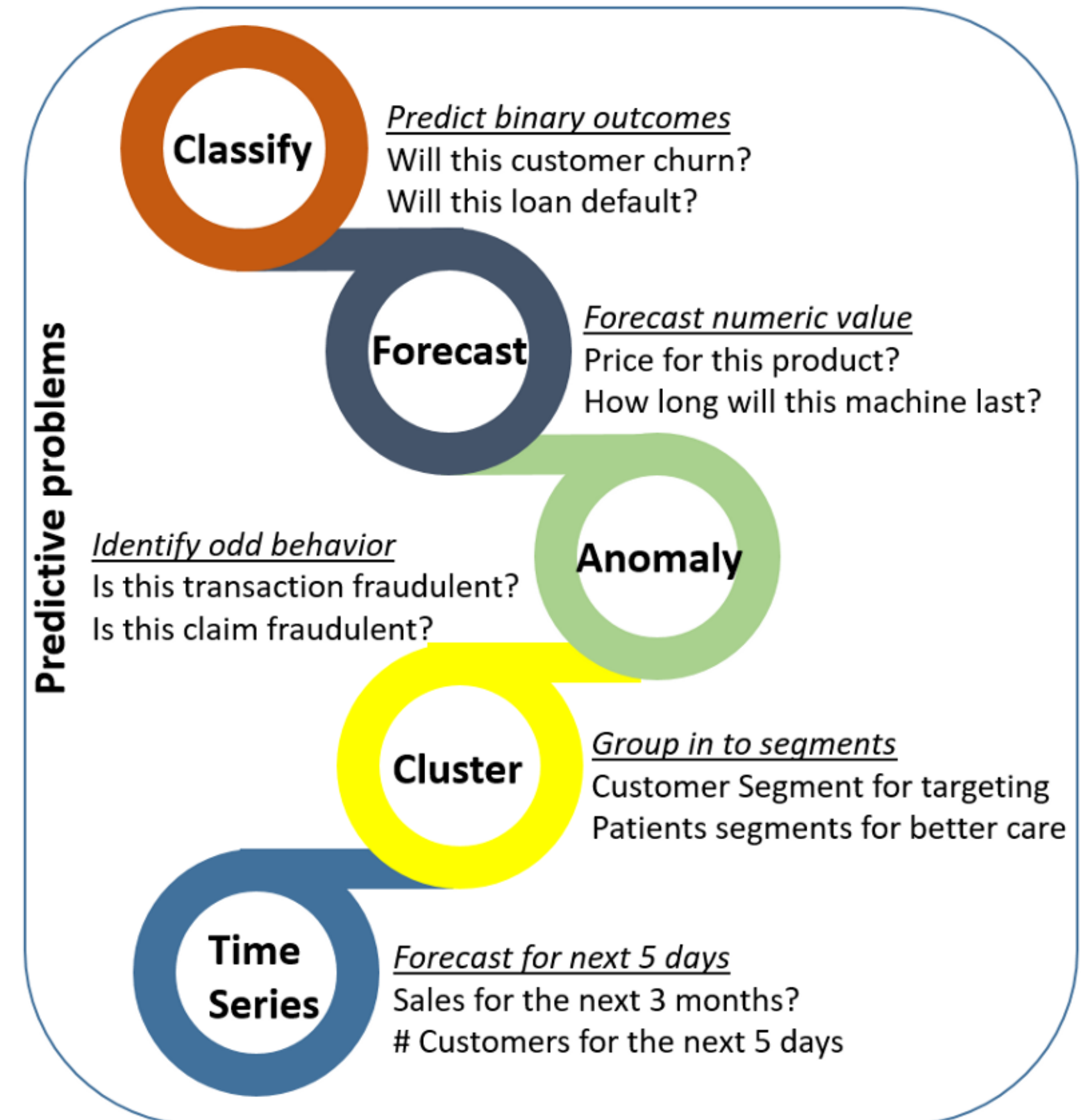
- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer: ??????

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 모델 입출력 정의:
 - 입력: 현재 및 이전 두 시즌에서 핫 한 패션 데이터
 - 출력: 다음 시즌 핫 한 패션 데이터

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 모델 입출력 정의:
 - 입력: 현재 및 이전 두 시즌에서 핫 한 패션 데이터
 - 출력: 다음 시즌 핫 한 패션 데이터
 - 모델 정의:
 - 시간에 따른 입력 데이터의 경향성을 파악하여 미래를 예측



✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 모델 입출력 정의:
 - 입력: 현재 및 이전 두 시즌에서 핫 한 패션 데이터
 - 출력: 다음 시즌 핫 한 패션 데이터
 - 모델 정의:
 - 시간에 따른 입력 데이터의 경향성을 파악하여 미래를 예측
 - 데이터 구축:
 - OO홈쇼핑 자체 데이터 구성 확인
 - 불충분할 시, 추가 데이터 확보 필요
 - 데이터 구성은 어떻게 할지?

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 데이터 구축:
 - OO홈쇼핑 자체 데이터 구성 확인 → 아래 화면처럼 데이터 존재
 - 모델에 맞는 데이터 구성 필요
 - 사용하는 모델에서 쓴 데이터의 구성에 맞게 가공



이미지 확대

586개 구매중

[무료배송] S/S 신상 원피스/블라우스/티셔츠/팬츠 外

★★★★☆ 4.2점 (4,898) >

루시앙 브랜드샵 >

7,900원 ~ ?

GS PAY (현대) 7% 즉시할인 (5만원 이상) ? GS PAY 등록 >

현대카드 5% 즉시할인 (5만원 이상) ?

무이자 무이자할부 특별행사 (최대 8개월) ?

배송비 무료배송

상품을 선택하세요.

옵션을 선택하세요.

찜 1

장바구니

바로구매

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 데이터 구축:
 - OO홈쇼핑 자체 데이터 구성 확인 → 아래 화면처럼 데이터 존재
 - 모델에 맞는 데이터 구성 필요
 - 사용하는 모델에서 쓴 데이터의 구성에 맞게 가공

시즌	이미지	가격	판매수량
2022 S/S	1.jpg	7900	50
2021 F/W	2.jpg	11200	132



이미지 확대

586개 구매중

[무료배송] S/S 신상 원피스/블라우스/티셔츠/팬츠 外

★★★★☆ 4.2점 (4,898) >

루시앙 브랜드샵 >

7,900원 ~ ?

GS PAY (현대) 7% 즉시할인 (5만원 이상) ? GS PAY 등록 >

현대카드 5% 즉시할인 (5만원 이상) ?

무이자 무이자할부 특별행사 (최대 8개월) ?

배송비 무료배송

상품을 선택하세요.

옵션을 선택하세요.

찜 1 장바구니 바로구매

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 데이터 구축:
 - OO홈쇼핑 자체 데이터 수량이 부족할 경우?

✓ 데이터 수집

- OO홈쇼핑(고객사): 요즘 가장 핫 한 패션을 기반으로 다음 시즌에 잘 팔릴 패션을 예측하고 싶어요.
- AI Engineer:
 - 데이터 구축:
 - OO홈쇼핑 자체 데이터 수량이 부족할 경우 → 데이터 크롤링 기반 문제 해결
 - 타 업체 명예훼손 하지 않는 선에서 데이터 크롤링으로 해결 가능
 - ##홈쇼핑, @@홈쇼핑 등 다양한 곳에서 같은 데이터 포맷으로 수집

시즌	이미지	가격	시리얼넘버	판매수량	데이터출처
2022 S/S	0001.jpg	7900	124HNK873	50	OO홈쇼핑
2021 F/W	0002.jpg	11200	125NKS123	132	OO홈쇼핑
...					
2021 F/W	0232.jpg	21900	AN32S123	11	##홈쇼핑
2020 F/W	0233.jpg	9900	AN32S125	1003	##홈쇼핑

✓ 데이터 구축 팁

- Feature를 잘 학습할 수 있도록 합리적인 학습 데이터 수량 필요
- Outlier가 존재하지 않게 수집
- 데이터의 출처가 다양할 경우, format 통일 필요
- Image or text 데이터의 경우, 꼼꼼한 preprocessing을 통해 model이 각 데이터의 핵심 데이터를 학습할 수 있도록 준비

✓ Data augmentation

- Data augmentation이란?
- Data augmentation을 하는 이유?

✓ Transfer learning

- Transfer learning이란?
- Transfer learning을 하는 이유?

✓ 질문(판서)

02

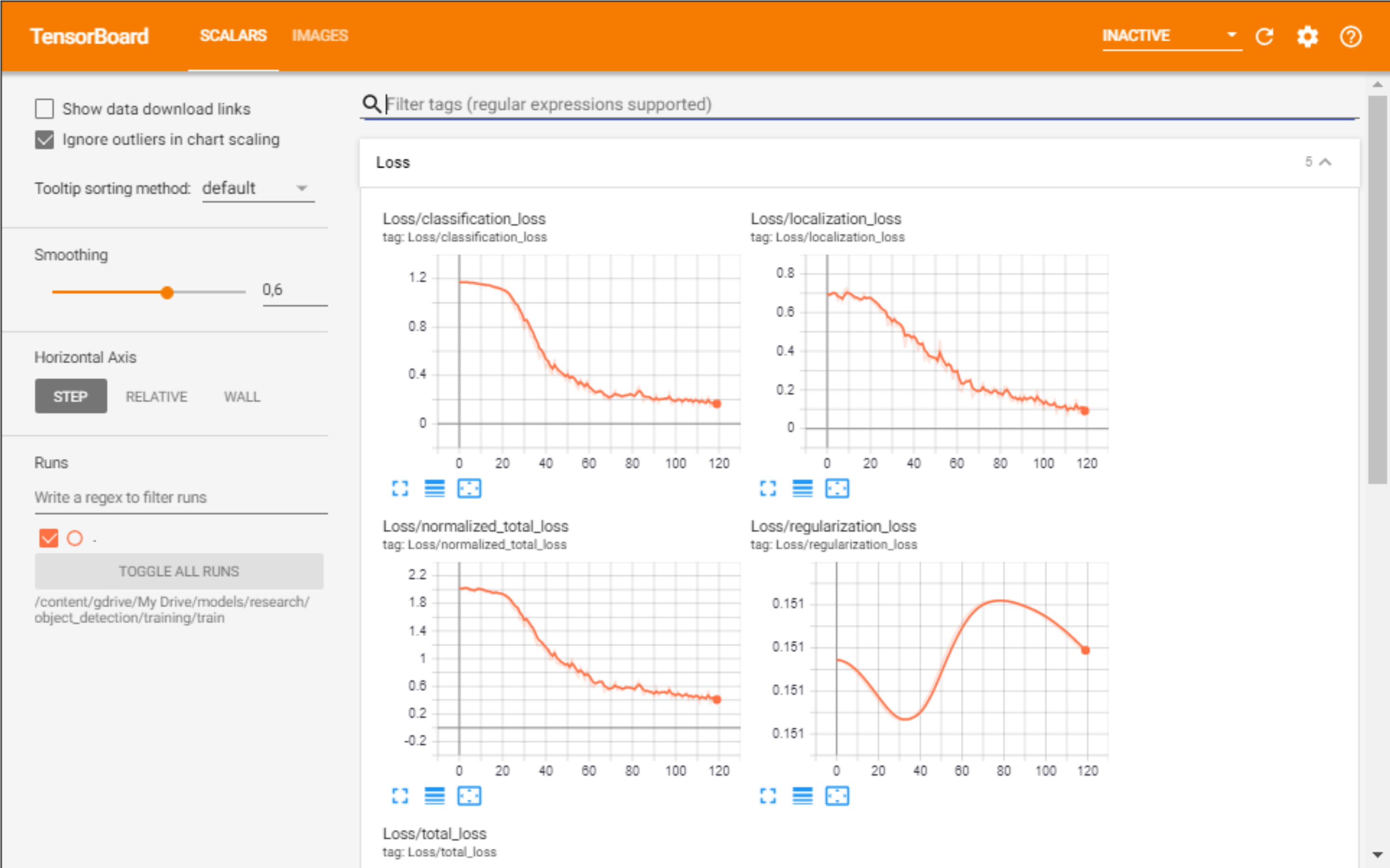
학습모니터링



✓ Tensorboard

- 텐서플로우의 시각화 도구
- 데이터 플로우 그래프를 분석하고 머신 러닝 모델을 시각화해 이해도를 높일 수 있도록 해주는 유용한 도구
 1. 하나, 변수에 대한 통계정보를 보여줌
 2. 둘, 그래프의 상세한 구조를 살펴볼 수 있도록 시각화해 나타냄 → 그래프는 매우 복잡해지는 경우가 많음(왜냐하면, 딥 뉴럴 네트워크는 종종 36000개 이상의 노드를 갖는다고 함)
 3. 셋, 인터랙티브한 시각화 기능을 통해 사용자가 그래프의 일부분을 확대 및 축소해 상세한 정보를 확인할 수 도 있음.
 4. 넷, 텐서 보드는 노드를 계층별로 접거나 펼칠 수 있고, 같은 형식의 데이터는 구별하기 쉽게 표시해 둠.

✓ Tensorboard



✓ Tensorboard

- 텐서플로우 튜토리얼: <https://neptune.ai/blog/tensorboard-tutorial>

✓ Tensorboard(실습)

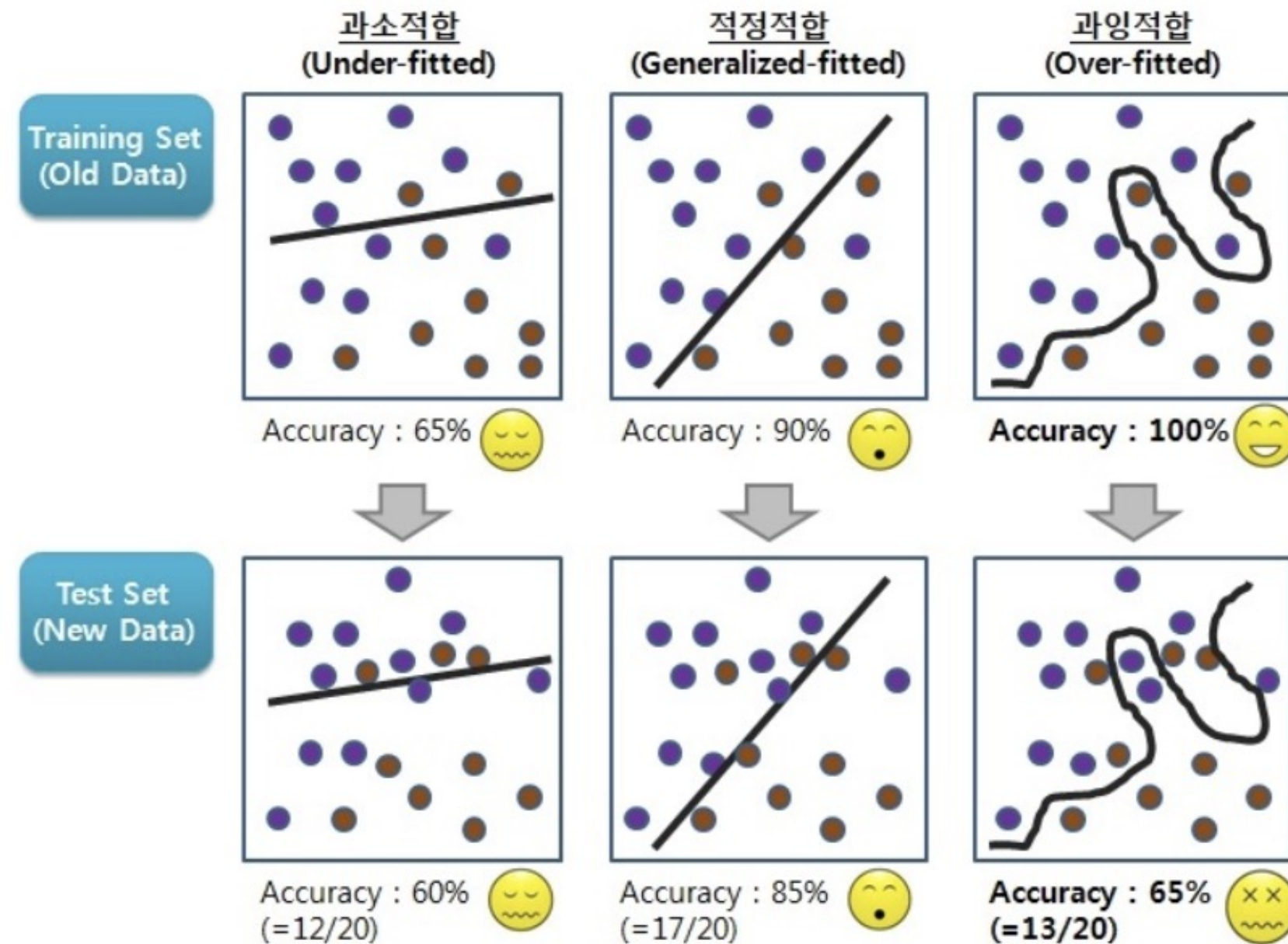
03

오버피팅(Overfitting)



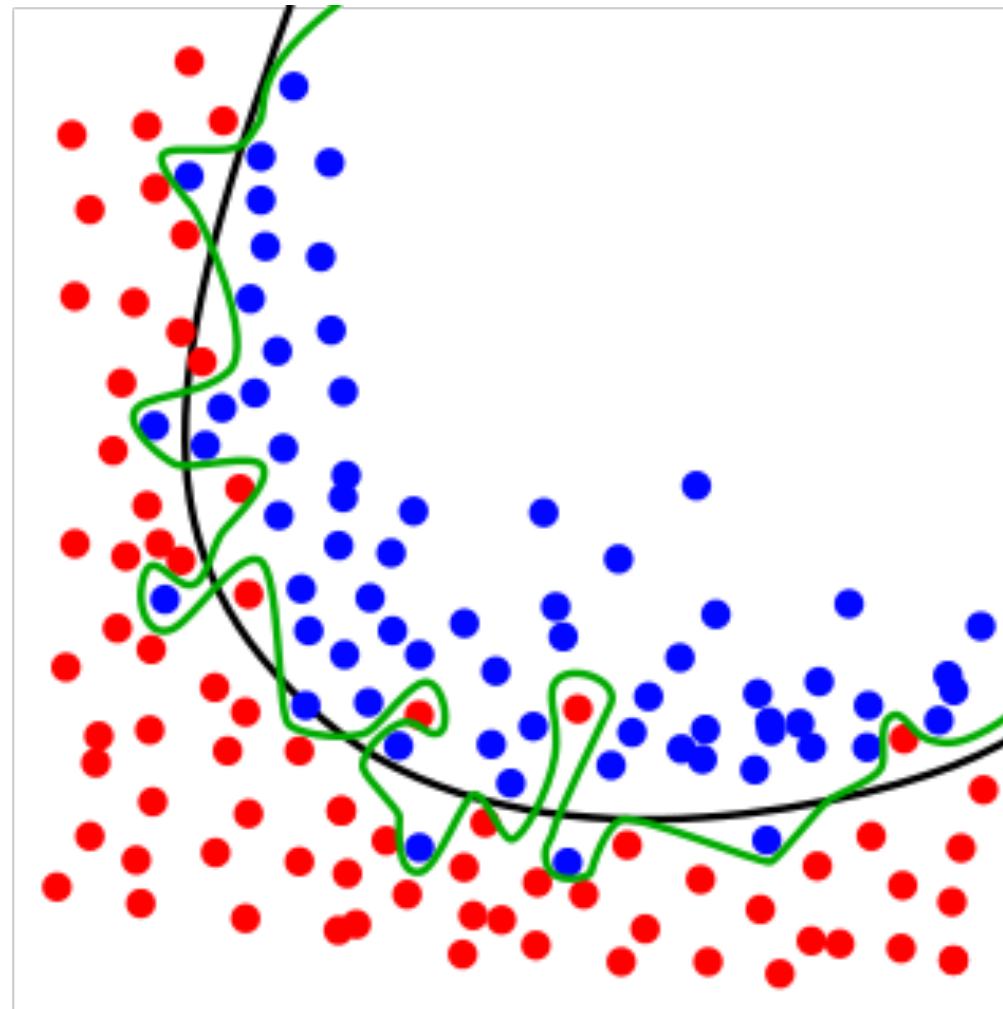
✓ 오버피팅

- 학습 데이터를 과하게 학습하는 것
- 일반적으로 학습 데이터는 실제 데이터의 부분 집합이므로 학습데이터에 대해서는 오차가 감소하지만 실제 데이터에 대해서는 오차가 증가하게 됨



✓ 오버피팅 발생 원인

- 오버피팅(과적합)은 데이터가 적을 수록 많이 발생
- 모델이 복잡한 경우도 발생한 경우도 발생한다고 하나, 이는 데이터가 많을 경우 발생하지 않음
 - ✓ 데이터 편향: 학습 데이터는 실제 데이터에서 편향된 부분만을 가질 수 있음
 - ✓ 데이터 오류: 학습 데이터에 오류가 포함된 값이 있을 수 있음
 - ✓ 차원의 저주: 고려하는 변수가 지나치게 많을 때 발생할 수 있음

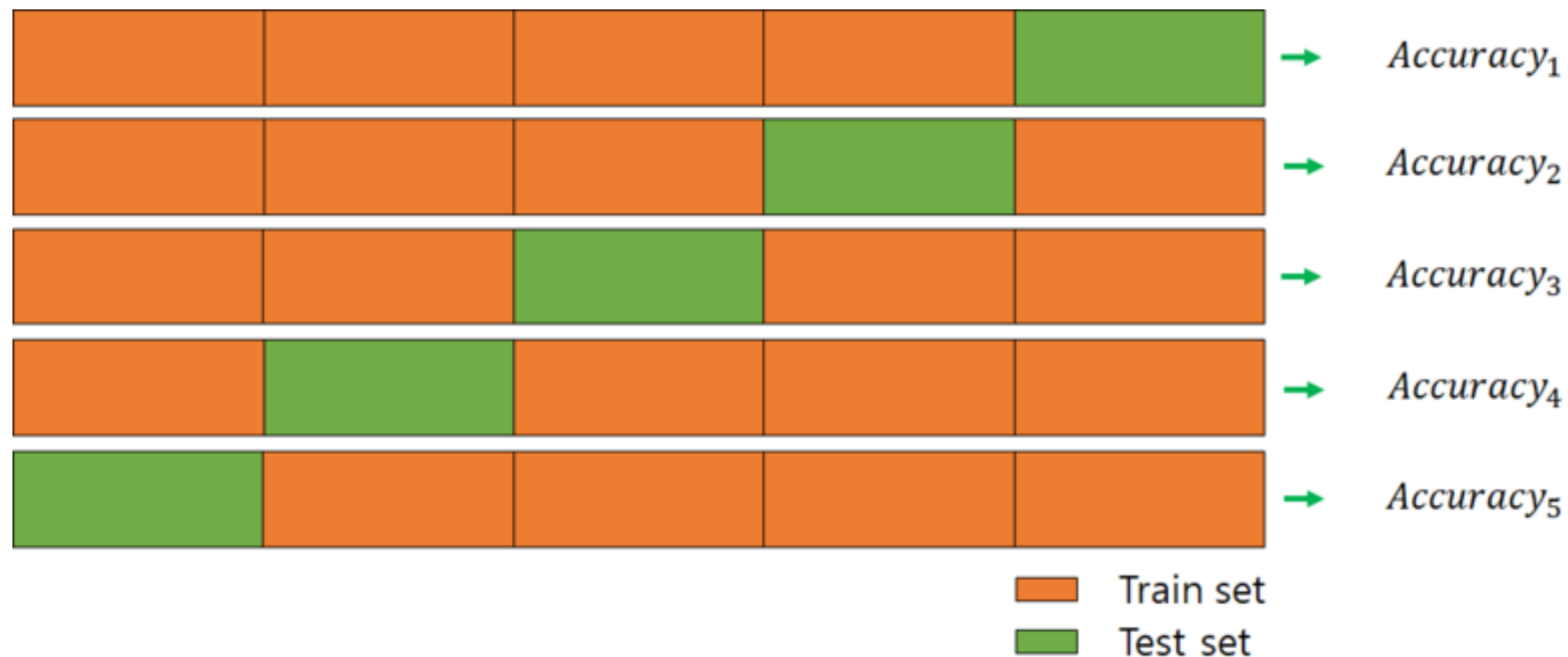


✓ 오버피팅 해결 방법

- Cross-validation
- Train more data
- Remove features
- Early Stopping
- Regularization
- Ensembling

✓ 오버피팅 해결 방법

- Cross-validation(교차검증): 기존 하나의 test set에 parameter를 최적화 시키는 방법이 아닌 여러 가지의 test set에 대해 일반적으로 parameter를 최적화 시킴



$$Accuracy = Average(Accuracy_1, \dots, Accuracy_k)$$

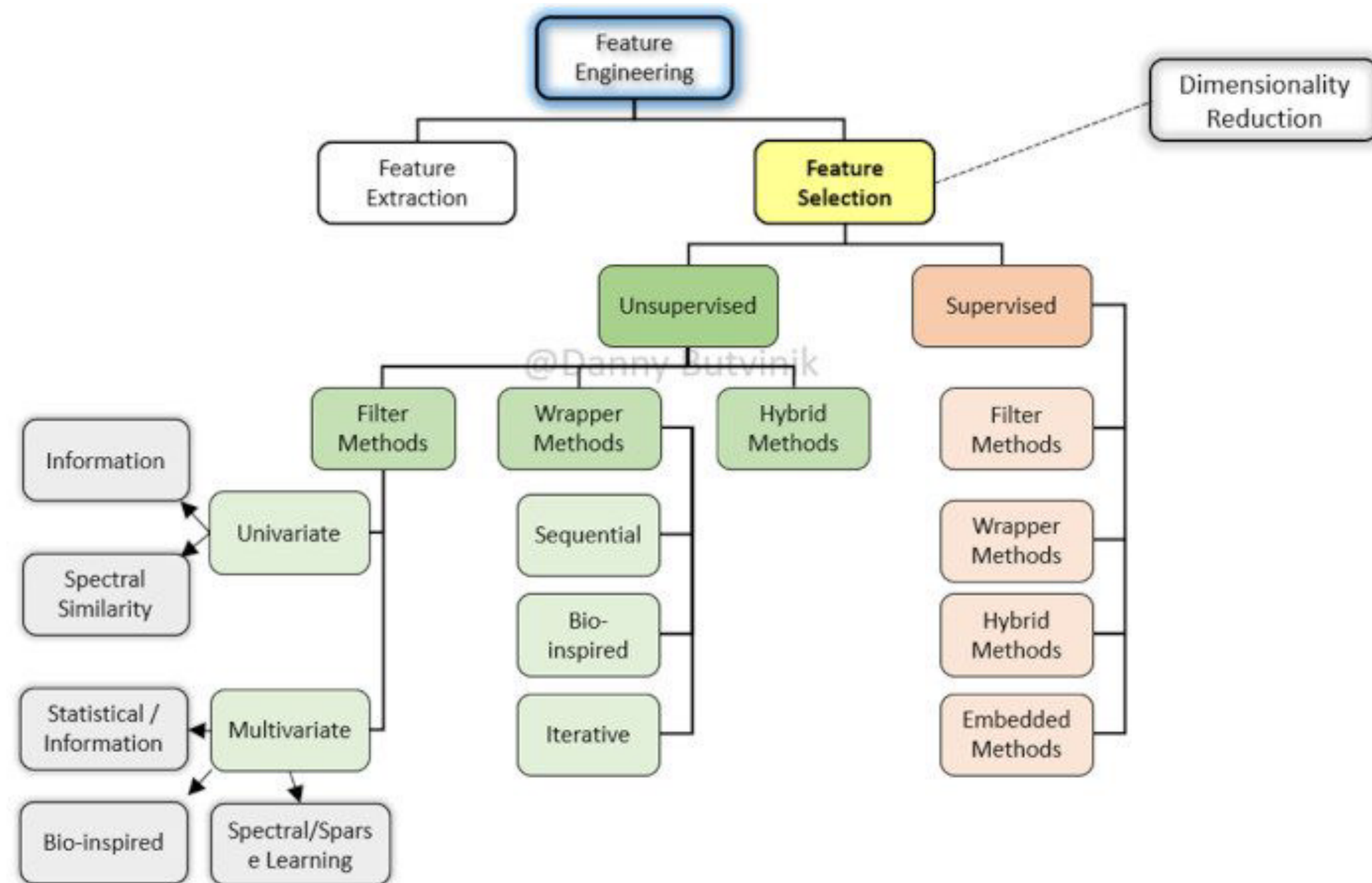
✓ 오버피팅 해결 방법

- Train more data: Data augmentation 방법 또는 실제 데이터를 더 수집하여, 많은 데이터로 모델을 학습 시킴



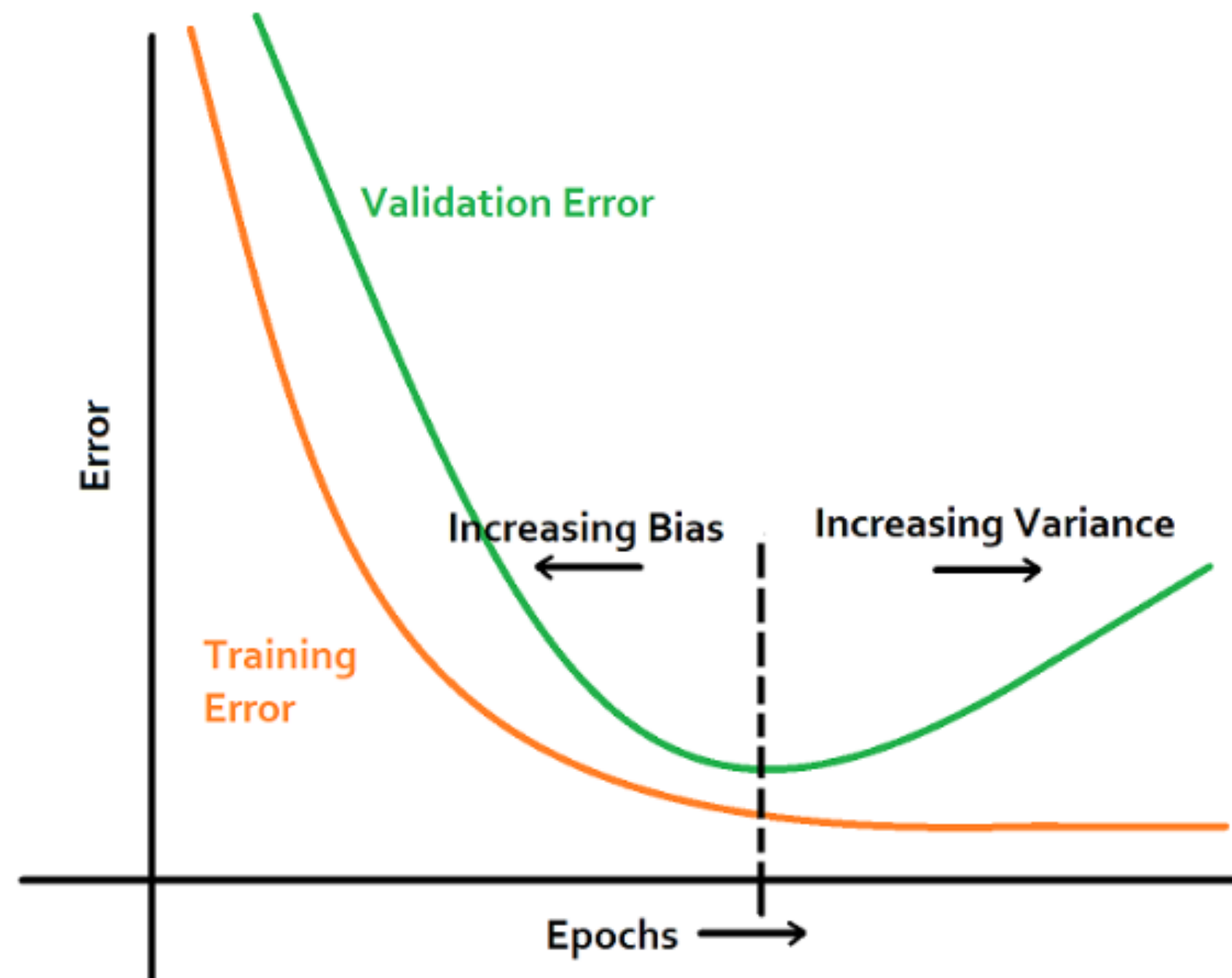
✓ 오버피팅 해결 방법

- Remove features(특징 제거): 가져다 쓰는 모델에서 feature selection 파트가 내장되어 있는데, 본인의 데이터에 맞는 feature를 적당히 선택할 수 있게 수정



✓ 오버피팅 해결 방법

- Early stopping(조기 종료): Neural Network가 과적합을 회피하도록 만드는 정칙화(regularization) 기법 중 하나. 훈련 데이터와는 별도로 검증 데이터(validation data)를 준비하고, 매 epoch마다 검증 데이터에 대한 오류(validation loss)를 측정하여 모델의 훈련 종료를 제어



✓ 오버피팅 해결 방법

- Regularization(일반화): regularization은 신경망이 범용성을 갖도록 함

Regularization 종류

- ① Weight decay(L2 regularization): overfitting은 일반적으로 가중치 값이 커서 발생하는 경우가 많으므로, 가중치 값이 클수록 큰 페널티를 부과해 과적합 억제함. 가중치 각각의 Loss function에 $1/2 * \lambda * W^2$ 을 더하는 방식으로 페널티 부과
- ② Dropout: 훈련 시에 각 계층마다 일정 비율의 노드를 임의로 선택해 weight parameter update시 누락시켜 진행하는 방법
- ③ Hyperparameter optimization: 대략적인 범위를 설정하고, 그 범위에서 무작위로 값을 선택해 정확도를 평가한 후 좋은 정확도를 내는 쪽으로 범위를 축소하는 방식을 반복

✓ 오버피팅 해결 방법

- **Ensemble(앙상블)**: 앙상블은 여러 개별 모델의 예측을 결합하기 위한 기계 학습 방법. 앙상블에는 몇 가지 다른 방법이 있지만 가장 일반적인 두 가지는 아래와 같음.
- **Bagging**은 복잡한 모델을 과적합할 가능성을 줄이려는 시도
- **Boosting**은 단순 모델의 예측 유연성을 향상시키려는 시도
- 더 알고 싶은 분: <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f>

✓ 오버피팅 해결 방법

- Cross-validation
- Train more data
- Remove features
- Early Stopping
- Regularization
- Ensembling

✓ Overfitting(실습)

04

실무 Q&A



✓ 질문

- AI 공부 방법
- AI 취업 방법
- AI 서비스 제공 방법
- 기타

✓ AI 공부 방법

- 온라인 강의
 - ✓ [Naver 부스트코스](#)
 - ✓ [K-MOOC](#)
 - ✓ [모두를 위한 머신러닝/딥러닝 강의](#)
 - ✓ [DeepLearningAI Andrew Ng 교수님 강의](#)
 - ✓ Youtube
 - ✓ Udemy
 - ✓ Coursera
- 논문
 - ✓ 주제 + review paper
 - ✓ 논문 제목 + 리뷰(review)

✓ AI 공부 방법

- 책
 - ✓ [책 추천 링크](#)
 - ✓ [국내 도서 추천](#)
- 스터디/온라인
 - ✓ [가짜연구소](#)
 - ✓ [모두의연구소](#)
- 부트캠프
 - ✓ Google Machine Learning Bootcamp 2022 3기
 - ✓ 지원 마감일: 5월 29일 오전 9시
 - ✓ 프로그램 참여 기간: 2022년 6월 22일부터 9월 19일까지
 - ✓ 공고: https://goo.gle/mlbc22_blog
- 정부 지원
 - ✓ 미국 CMU AI 집중 교육프로그램 파견 교육생 모집
 - ✓ 공고:
<https://scc.sogang.ac.kr/front/cmsboardview.do?currentPage=1&searchField=ALL&searchValue=&searchLowItem=ALL&bbsConfigF>
[K=5451&siteId=globalaiedu&pkid=879425](https://scc.sogang.ac.kr/front/cmsboardview.do?currentPage=1&searchField=ALL&searchValue=&searchLowItem=ALL&bbsConfigF)

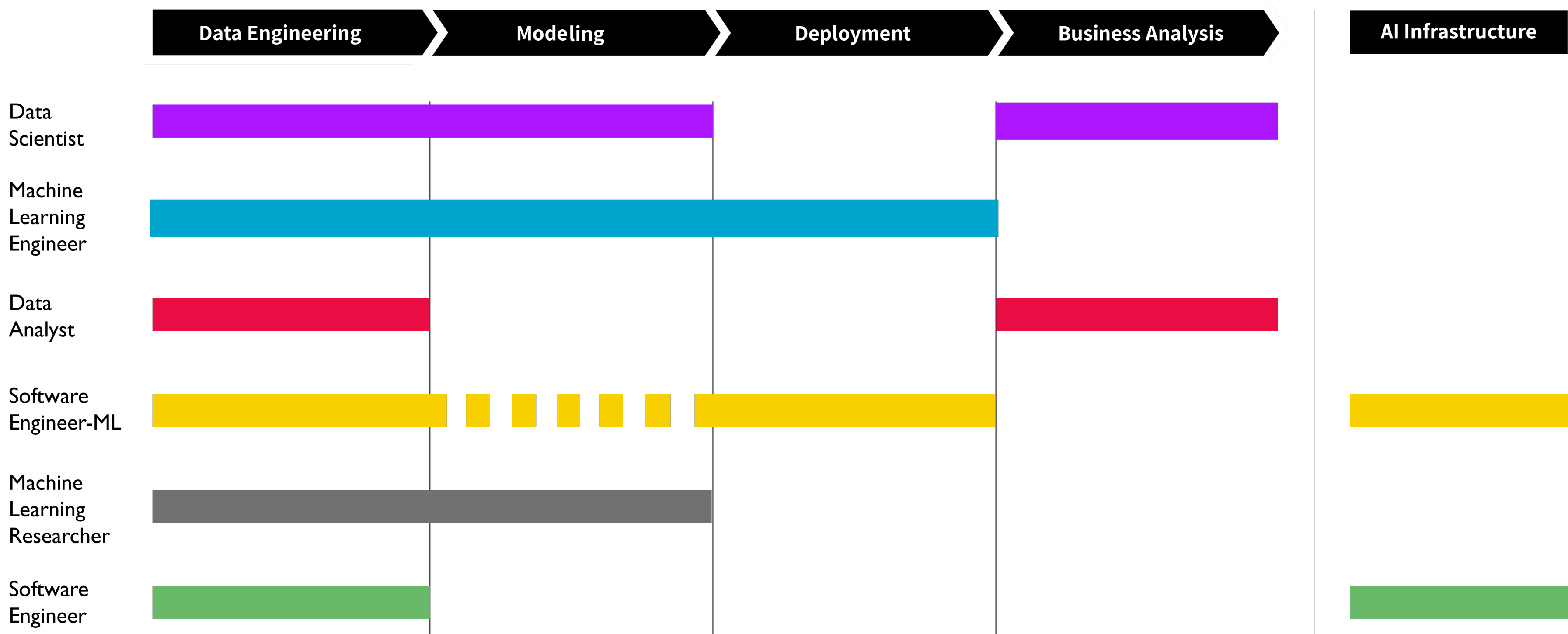
✓ AI 취업 방법

- 학부: 전자공학과 (2018년 졸업)
- 석사: 전자공학과 (2020년 졸업)
 - ✓ 인공지능 관련 SCI 논문 1편
- 박사: 컴퓨터공학과 (2021 입학, 휴학 예정)

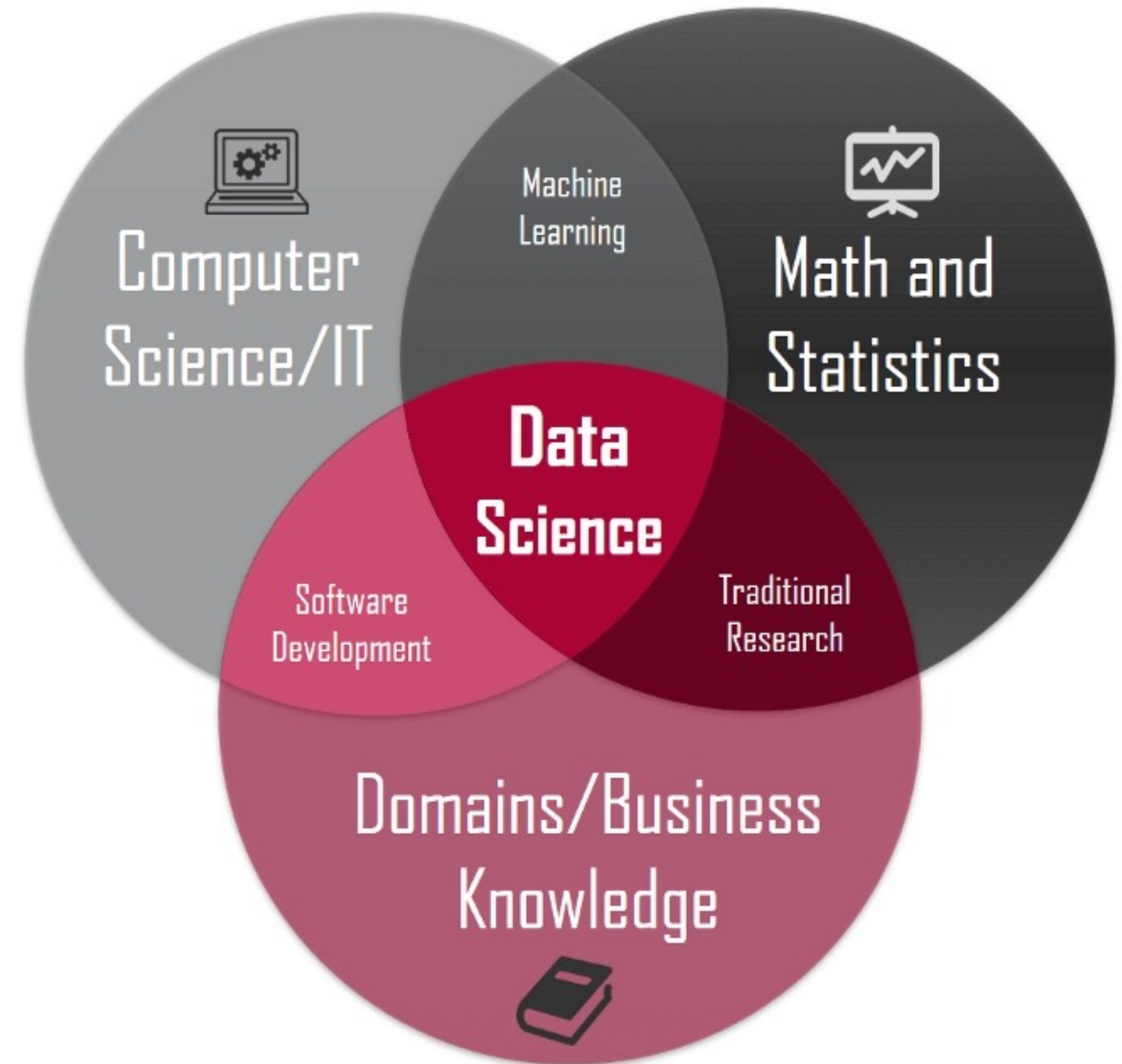
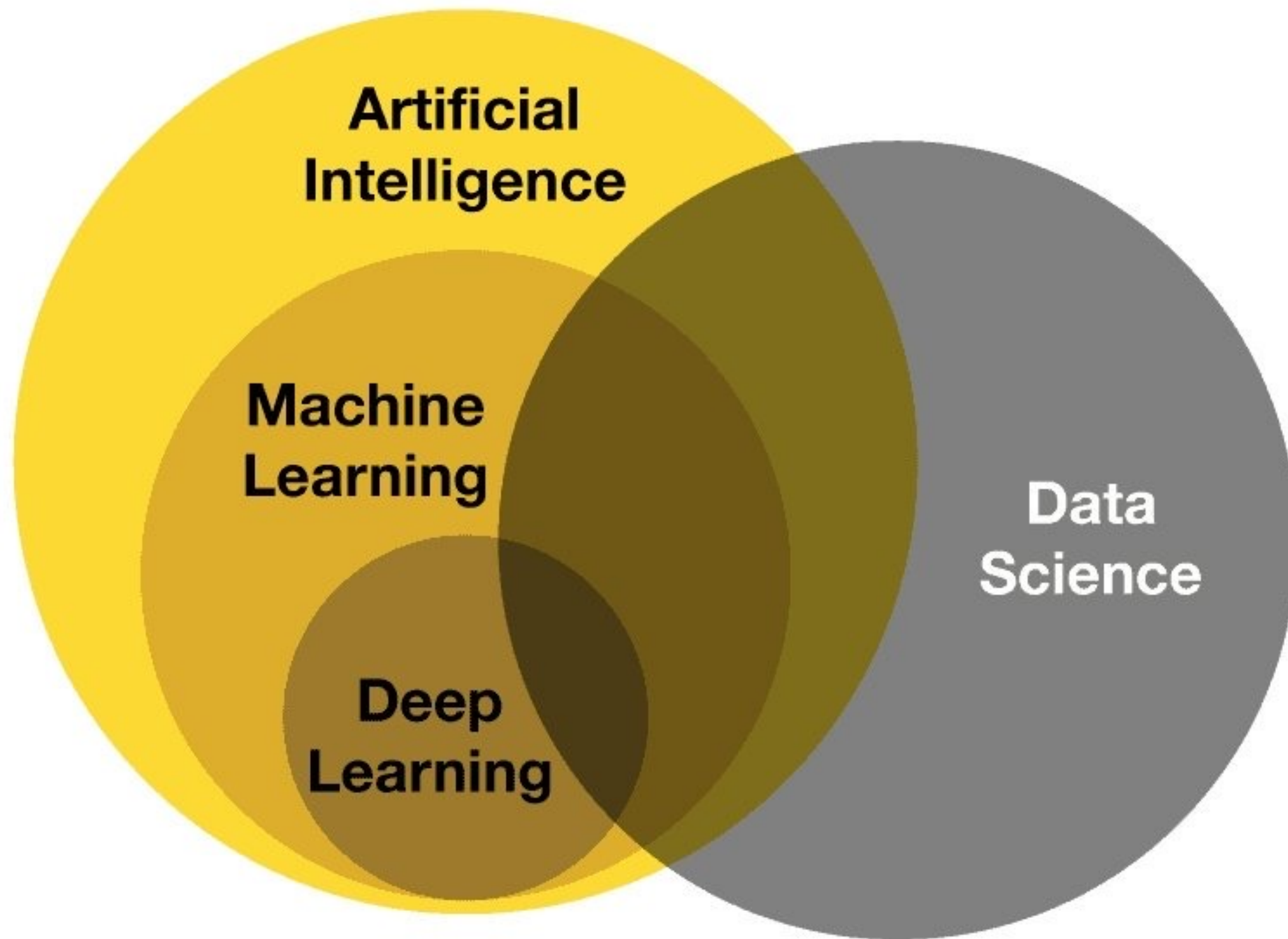
- CMU AI 집중 교육프로그램 파견 2019년도 1기 참여
- Google Machine Learning Bootcamp 2019 1기 참여

- [한국전자기술연구원](#) (2020.10 ~ 2022.05) CV Engineer, 계약직
 - ✓ 인공지능 관련 국내 학회 논문 3편
 - ✓ 특허 2건 작성
 - ✓ 실서비스 3개 개발
- Webarter Inc. (2021.04 ~ 현재) CTO, Co-founder
- 뷰메진(2022.06 ~) CV Engineer, 팀장

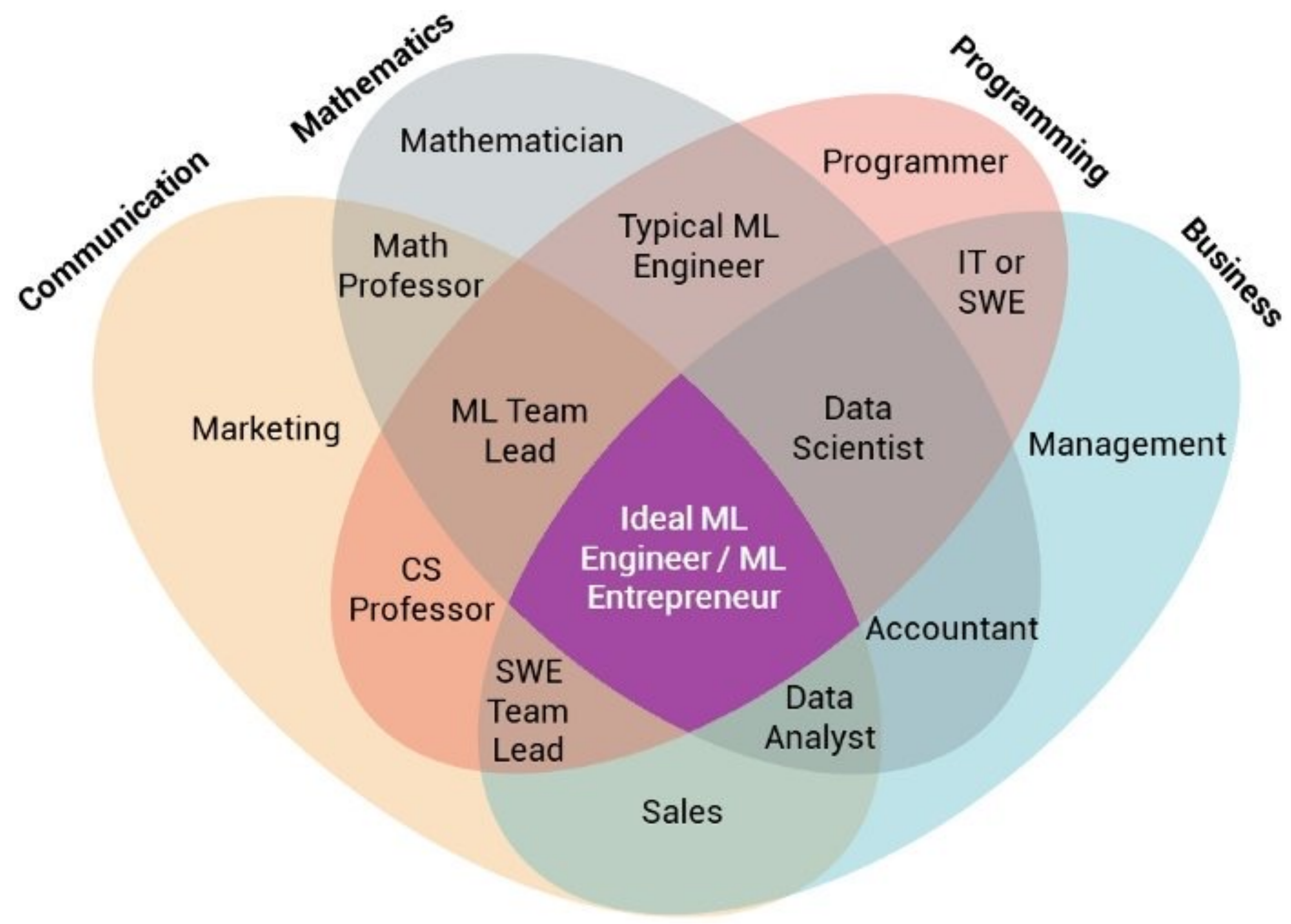
✓ AI 취업 방법



✓ AI 취업 방법



✓ AI 취업 방법



✓ AI 취업 방법

- AI Engineer

✓ AI 취업 방법

- AI Researcher

✓ AI 취업 방법

- 기타

① 도메인 지식

② 공부의 정도

✓ AI 서비스 제공 방법

- AI 공부 방법
- AI 취업 방법
- AI 서비스 제공 방법

✓ AI Challenge

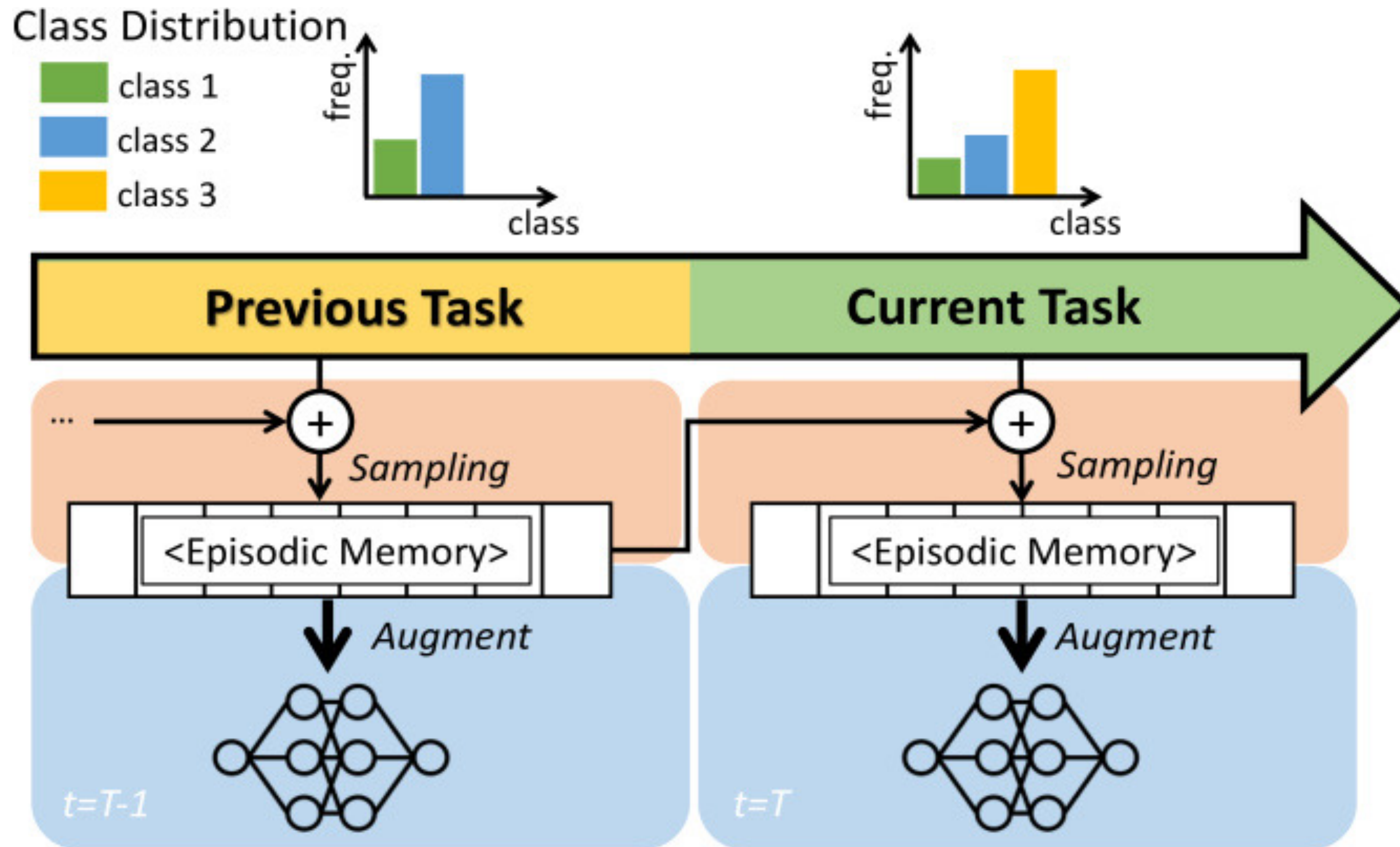
- Kaggle(pdf 같이 보기)
- CLOVA AI RUSH
- AI Challenge
 - ✓ AI Hub
 - ✓ CVPR Workshop
 - ✓ 등..

✓ 하드웨어

- Colab
- [AWS 무료 크레딧](#)
- [GCP 무료 크레딧](#)
 - ✓ [Google cloud study jam](#)
- Desktop + GPU
- Laptop + GPU
- GPU 빵빵한 회사 자원..

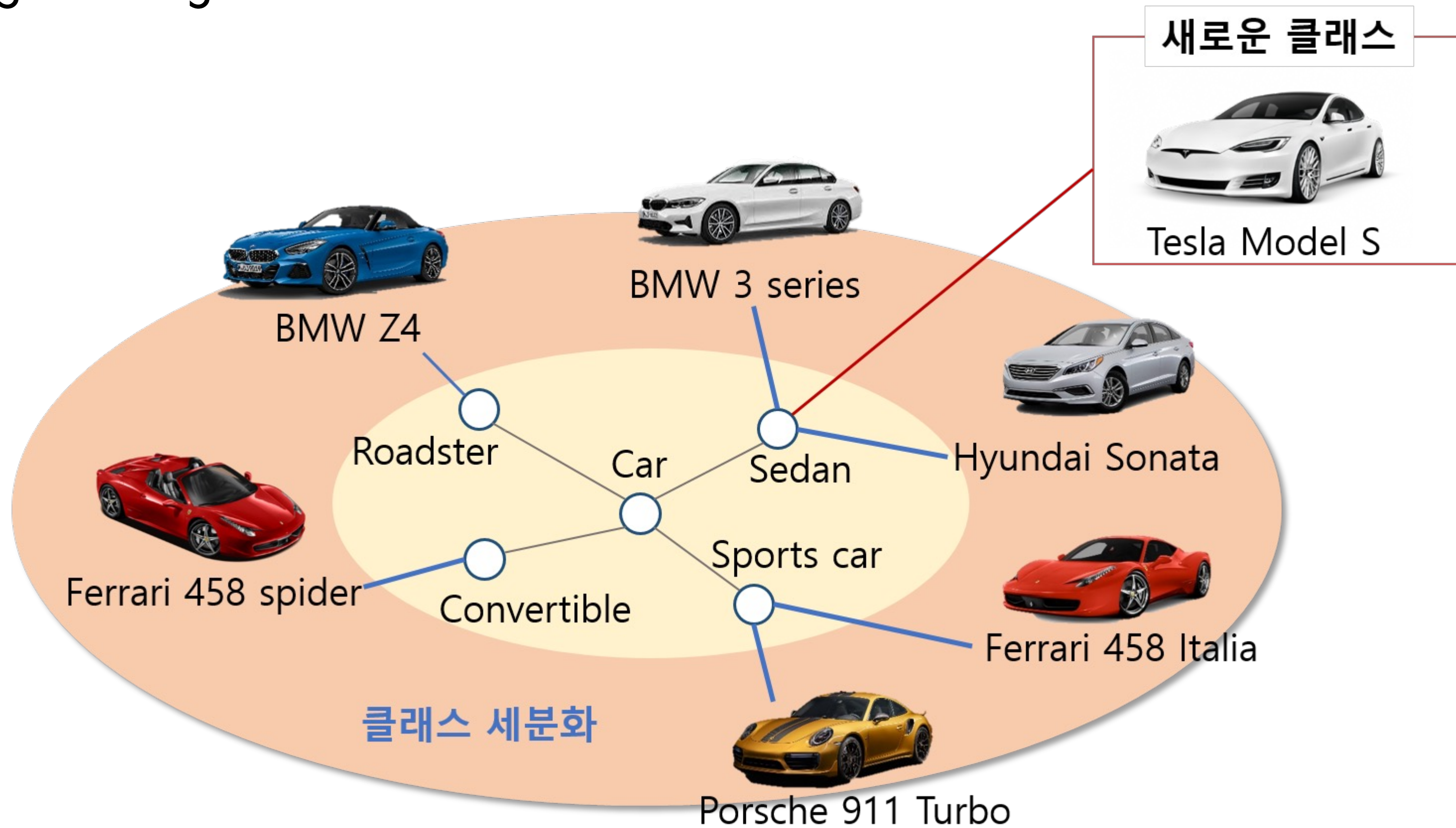
✓ Continual Learning, Lifelong Learning

• Continual Learning



✓ Continual Learning, Lifelong Learning

• Lifelong Learning



✓ Continual Learning, Lifelong Learning

- MLOps

5 Best Practices to optimize your MLOps lifecycle on Azure:

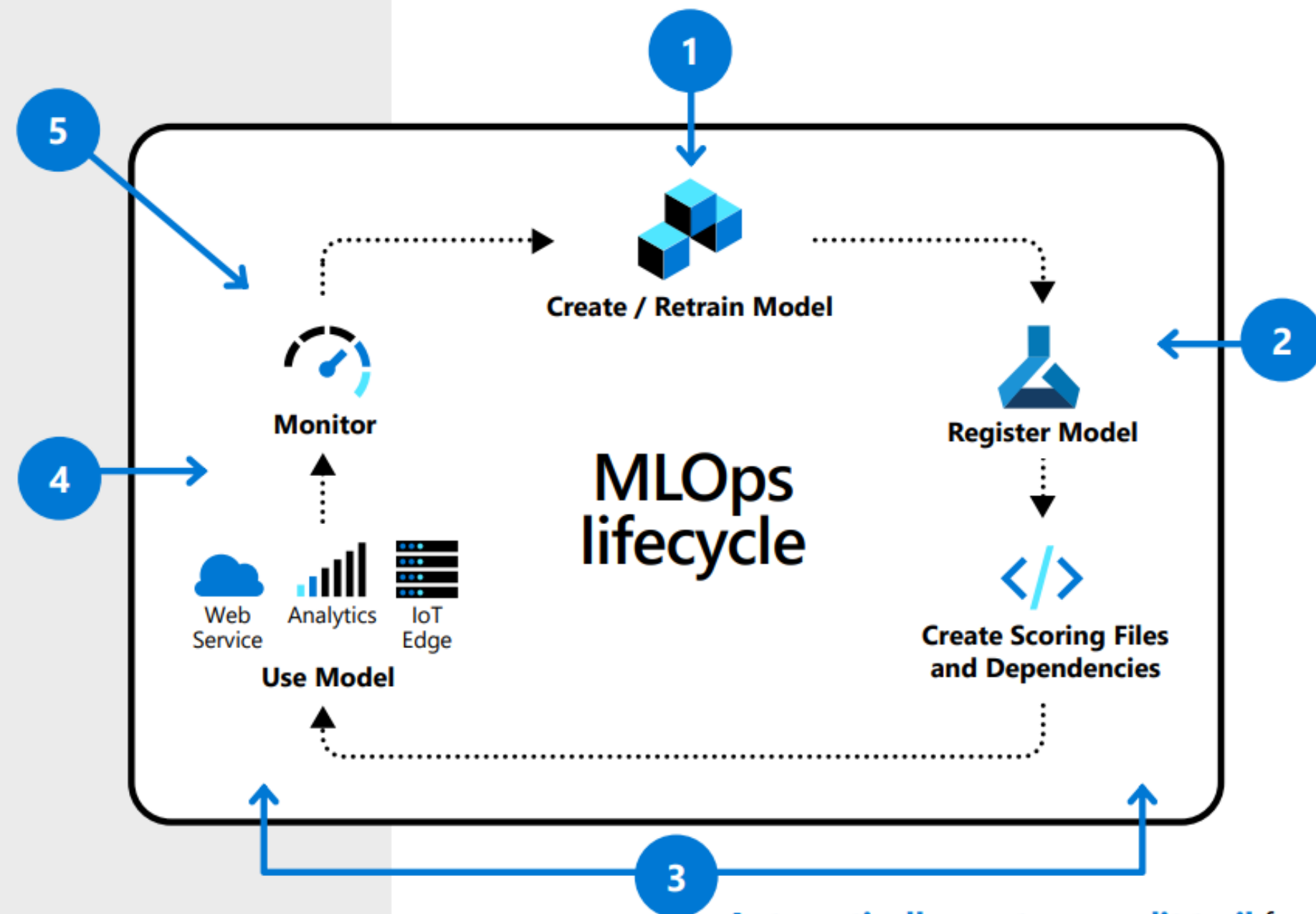
Observe data drift and feed back model information to improve future training.

Deploy and monitor performance so you can release models with confidence and know when to retrain.

Create models with reusable ML pipelines using the Azure Machine Learning extension for Azure DevOps. Store your code in GitHub so it automatically integrates into your MLOps pipeline.

Automate your MLOps rollout using Azure DevOps + Azure Machine Learning for version models with rich metadata and event management.

Automatically create an audit trail for all artifacts in your MLOps pipeline ensure asset integrity and meet regulatory requirements.

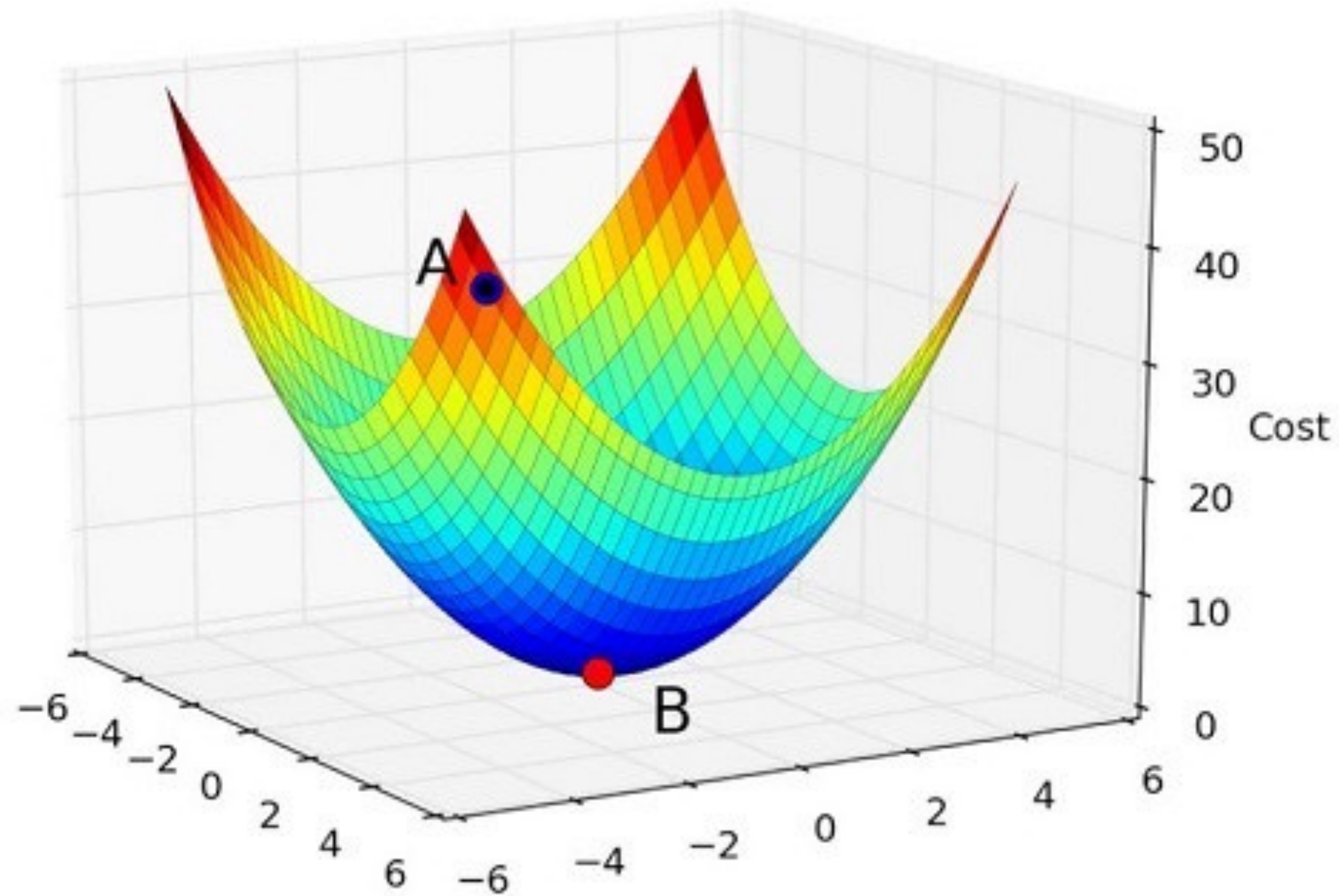


05

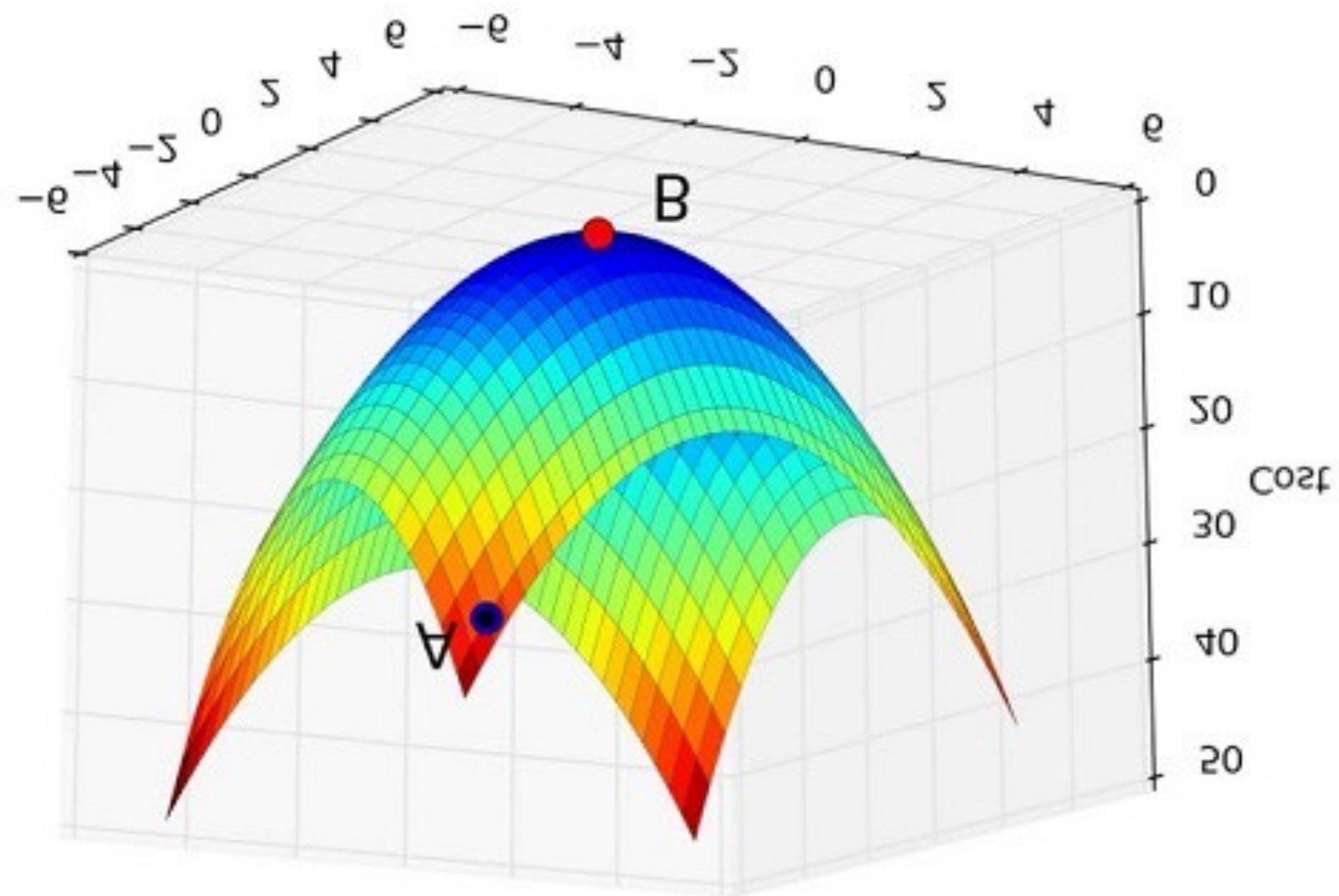
마무리



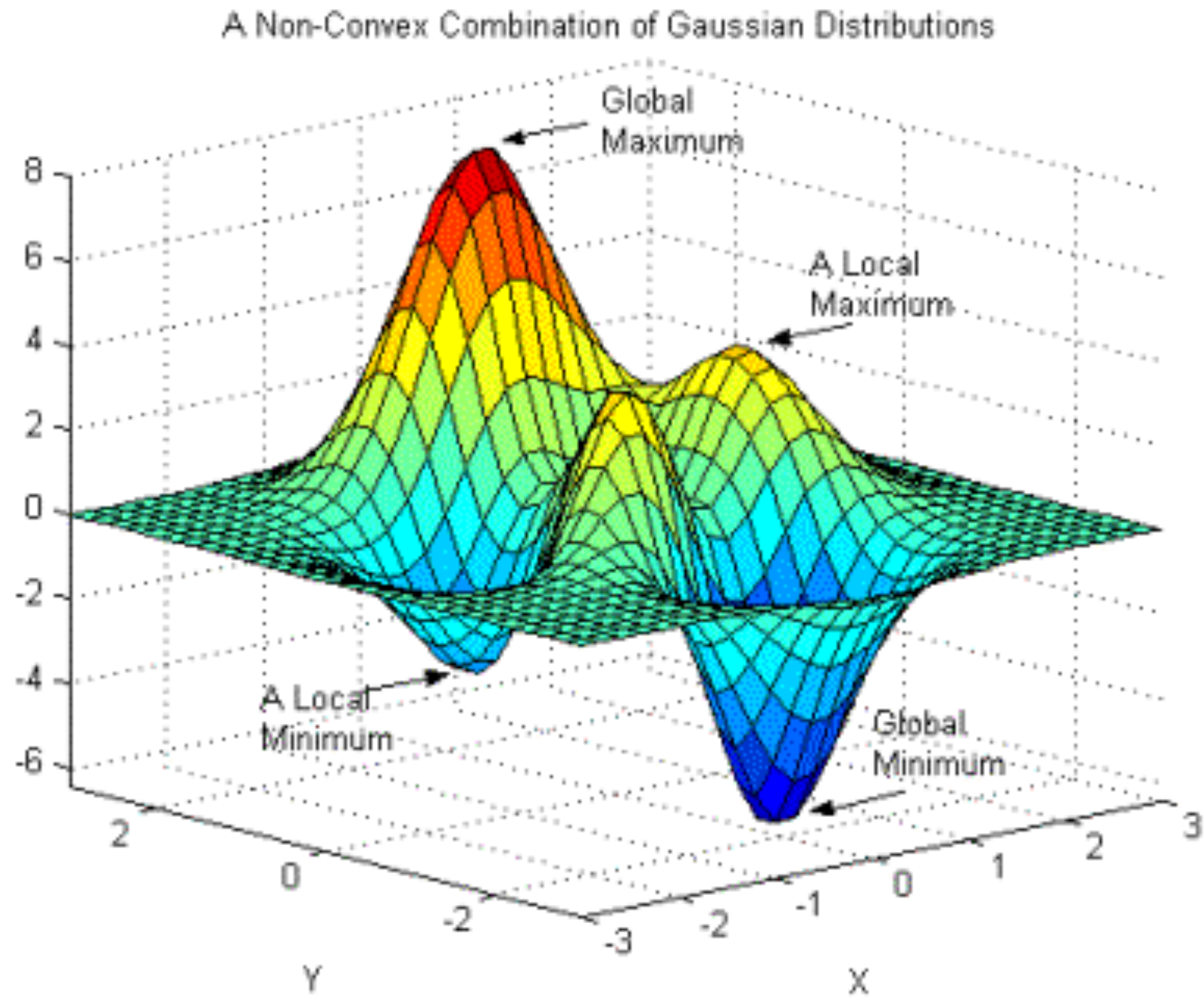
✓ 한마디..



✓ 한마디..



✔ 한마디..



✔ Contact



크레딧

/* elice */

코스 매니저

콘텐츠 제작자
한서우

강사
한서우

감수자

디자이너
한서우

연락처

TEL

070-4633-2015

WEB

<https://elice.io>

E-MAIL

contact@elice.io

