

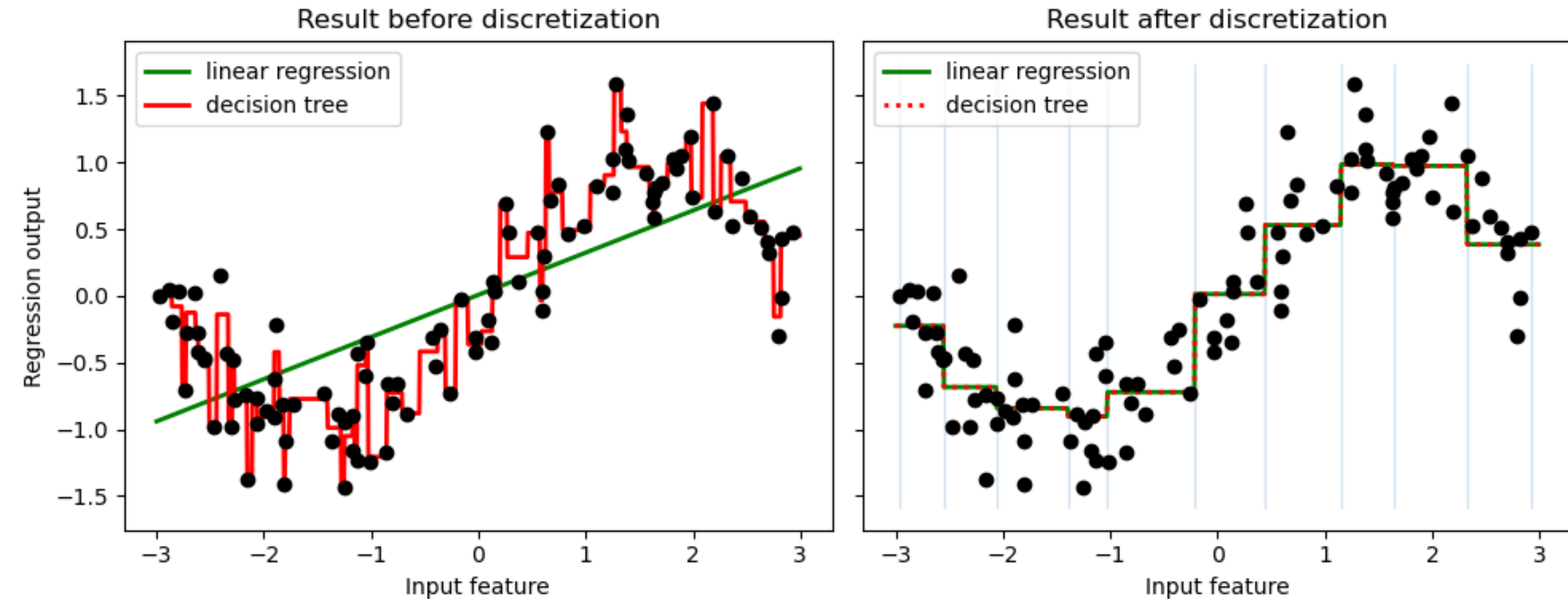
머신러닝 개념 기초

머신러닝의 정의와 학습의 원리

Contents

이번에 배울 내용은요

1. 머신러닝이란 무엇인가요
2. 컴퓨터는 어떻게 학습을 하나요
3. 컴퓨터가 학습을 하면 어떤 것들을 할 수 있나요
4. 머신러닝 필수 개념들에 대해 알아봅시다
5. 머신러닝 프로젝트 진행 방법



1

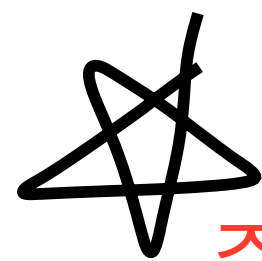
머신러닝이란 무엇인가요

머신러닝 (Machine Learning)

컴퓨터에게 학습이란?

- 머신러닝의 엄밀한 정의 (Formal Definition)

"A computer program is said to learn from *experience E* with respect to some *class of tasks T* and *performance measure P* **if its performance at tasks in T, as measured by P, improves with experience E.**"



주어진 task T에 대해 P를 높이는 일. (learning (E)을 통해)

머신러닝 (Machine Learning)

컴퓨터에게 학습이란?

- 머신러닝의 직관적인 의미
 - 사람이 학습하는 방식과 비슷하게 예시들(데이터)를 통해서 패턴을 찾아내는 방법
 - 컴퓨터에게 데이터를 바라보는 방법(모델)과 예시(데이터)를 주면 판단을 내리는 기준을 찾는 방법
 - 컴퓨터는 숫자(0, 1)밖에 이해하지 못하므로, 모든 것은 수학적 기준에 의해 판단

2

컴퓨터는 어떻게 학습을 하나요

‘prediction for unseen data’

아직 보지않은 데이터에 대한 예측

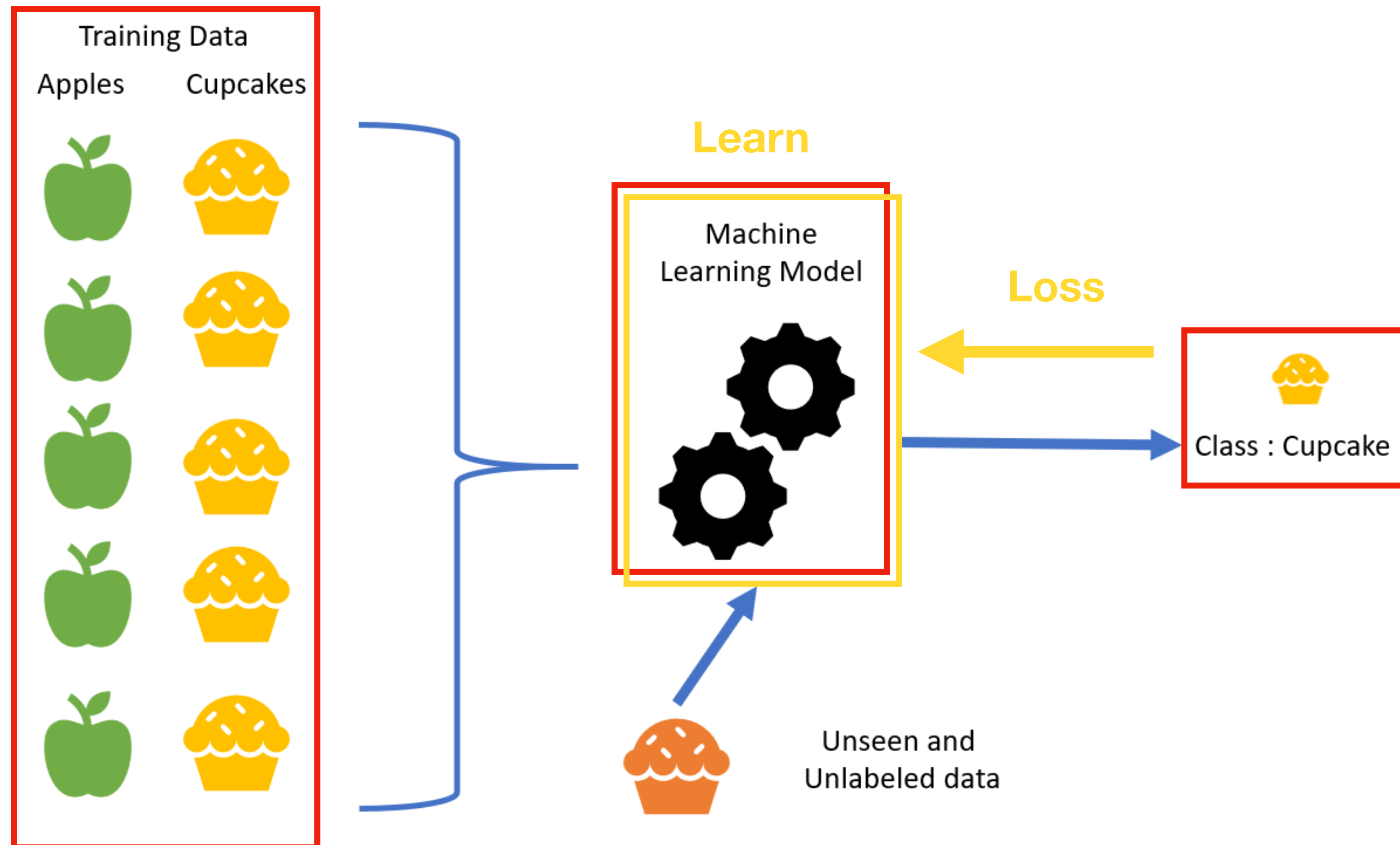
1. train/ test data

2. overfitting

3. generalization

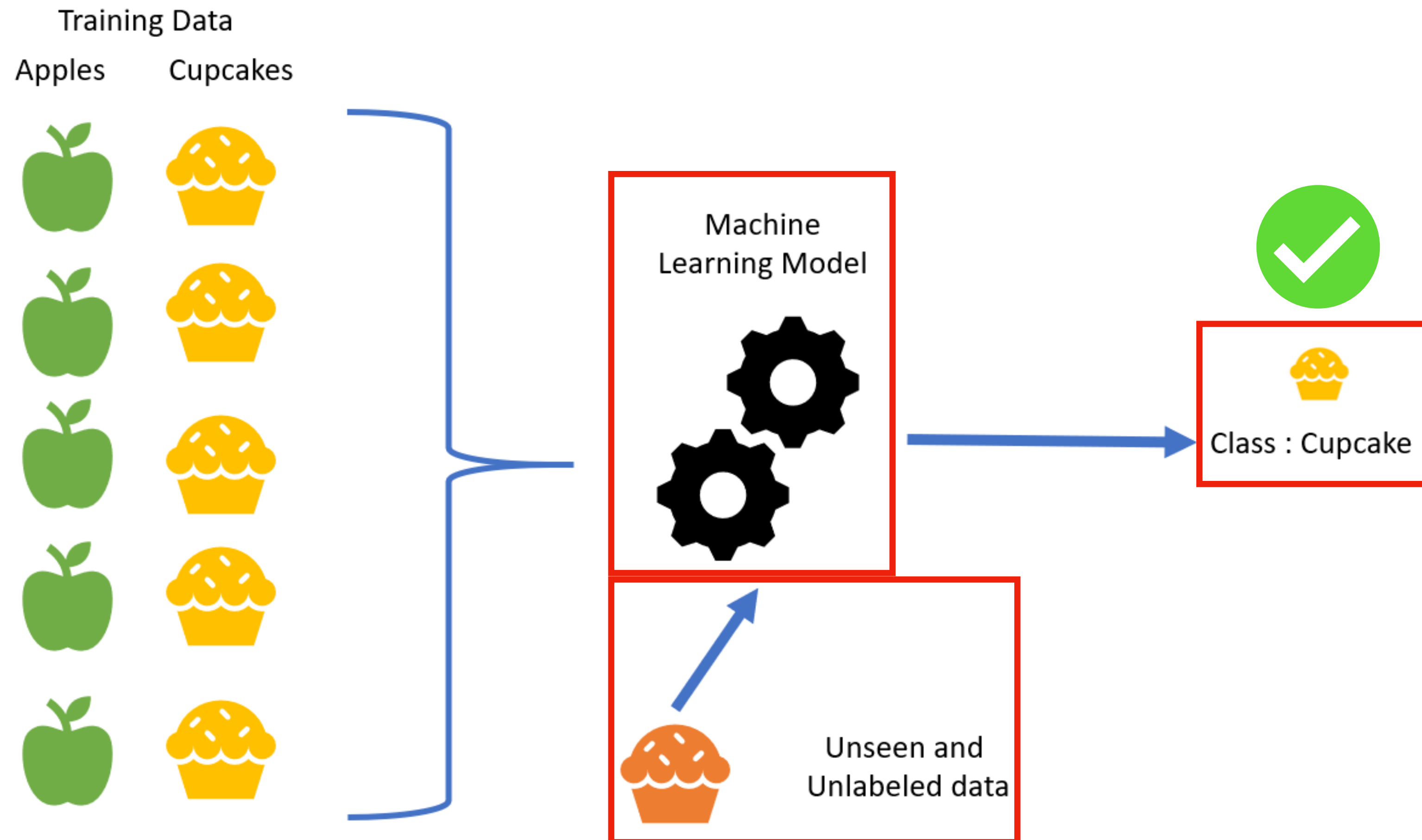
머신러닝 (Machine Learning)

머신러닝 예시 : 학습(training)



머신러닝 (Machine Learning)

머신러닝 예시 : 테스트(test, inference)

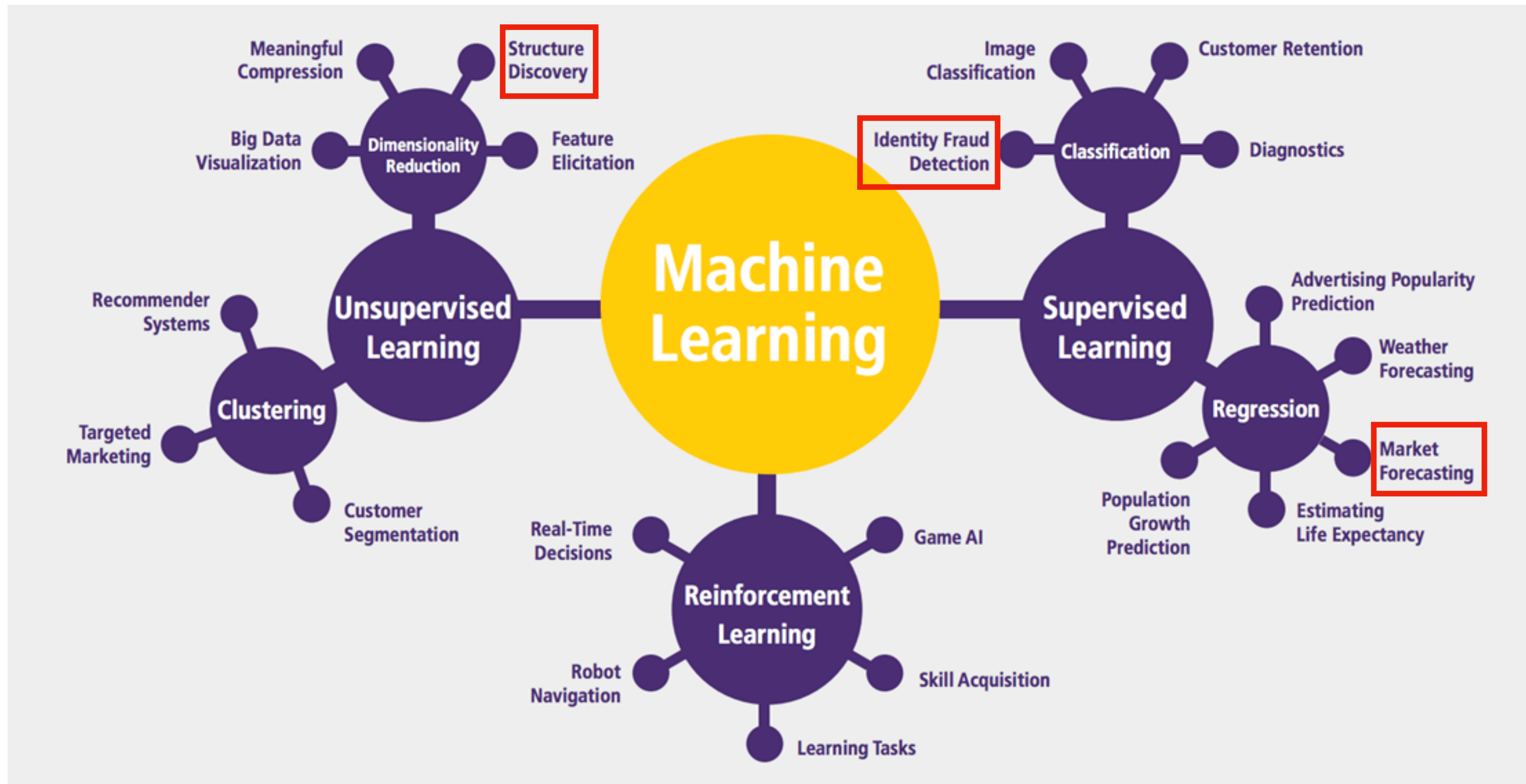


3

학습을 하면 어떤 것들을 할 수 있나요

머신러닝의 종류

Task of Machine Learning



머신러닝으로 할 수 있는 일들

Classification, Regression, Anomaly Detection

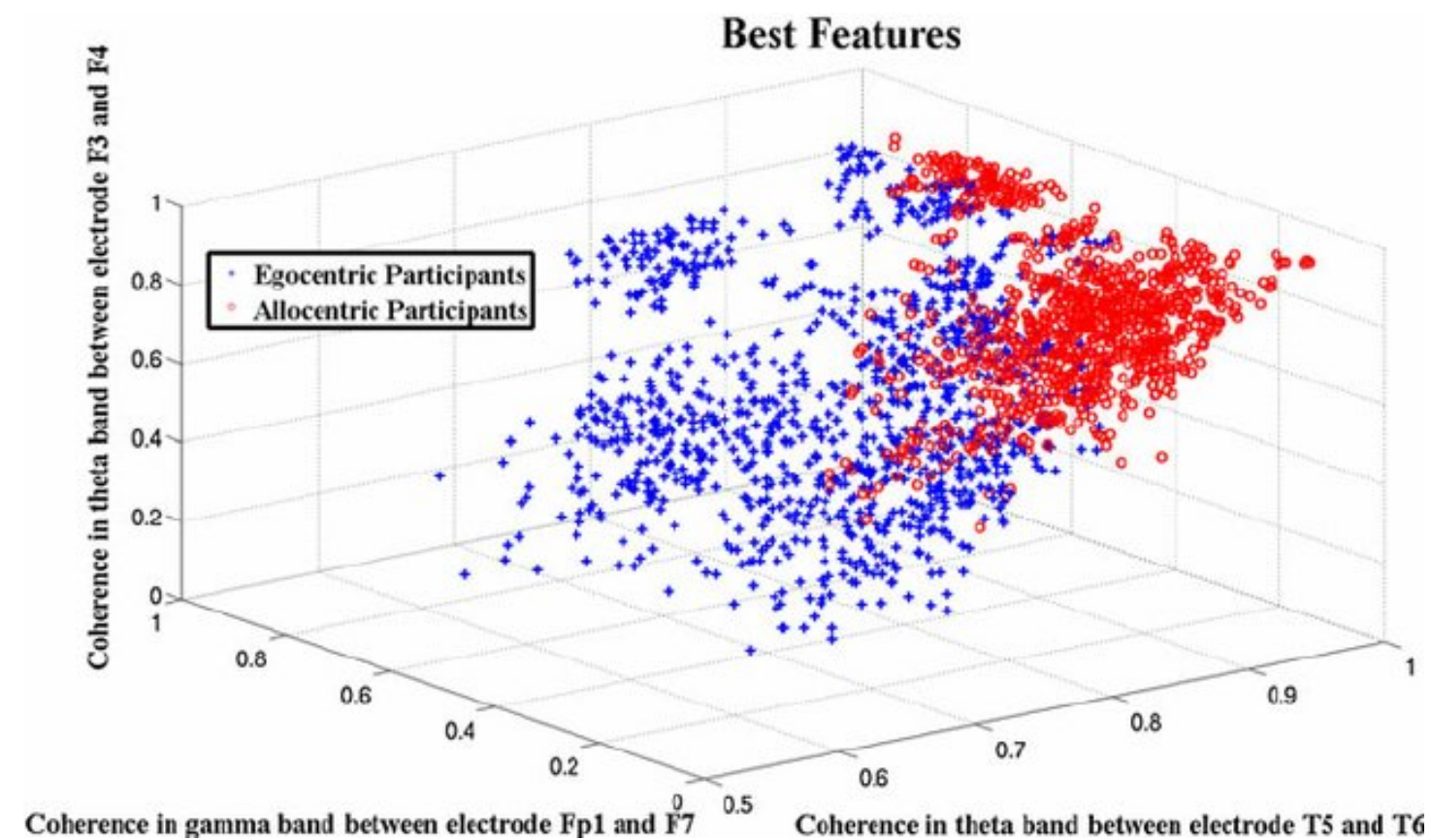
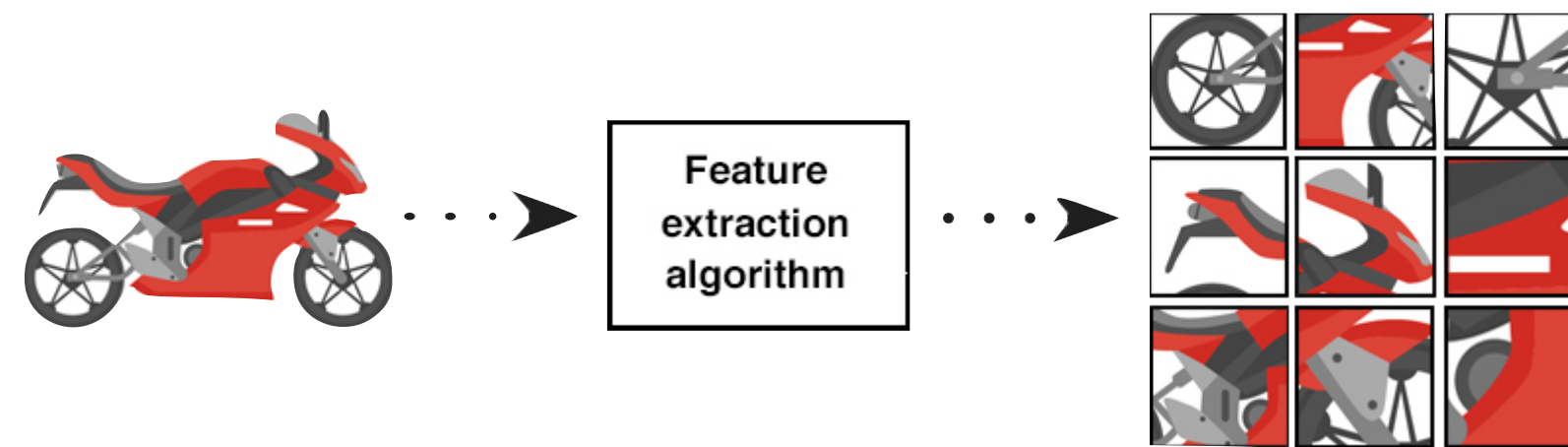
- 고객 파산 예측하기 (Binary Classification)
- 다이아몬드 가격 예측하기 (Regression)
- 단백질 구조 분류하기 (Anomaly Detection)

4

머신러닝 필수 개념에 대해 알아보시다.

Feature Vector

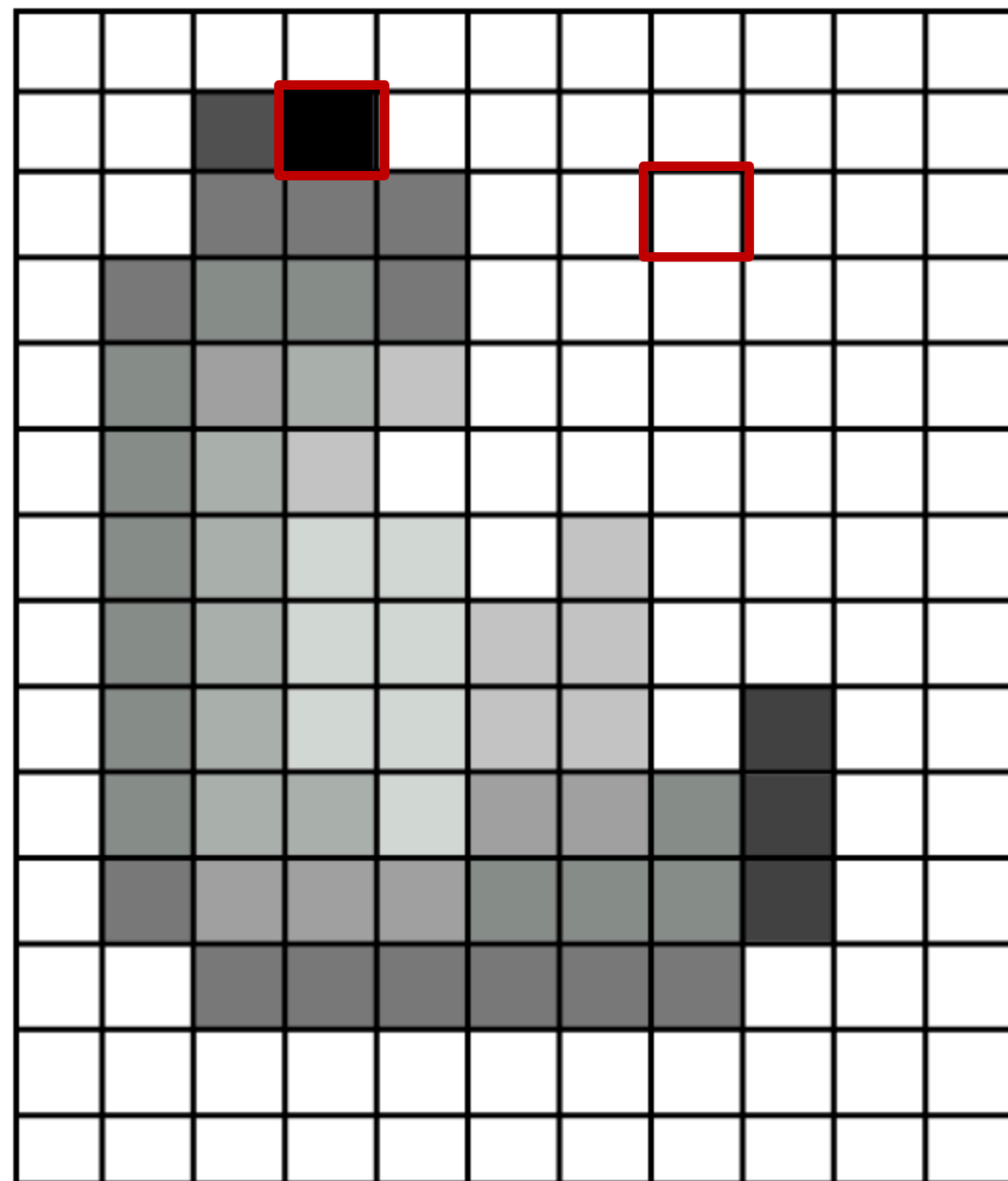
데이터를 어떻게 표현할 것인가?



- 데이터의 특징을 벡터로 나타낸 것 (정의)
- 어떤 feature를 사용하는지가 굉장히 중요합니다.
 - feature engineering
- feature vector가 존재하는 공간을 feature space라고 합니다.

Image Feature Vector

이미지를 벡터로 나타낸다면?



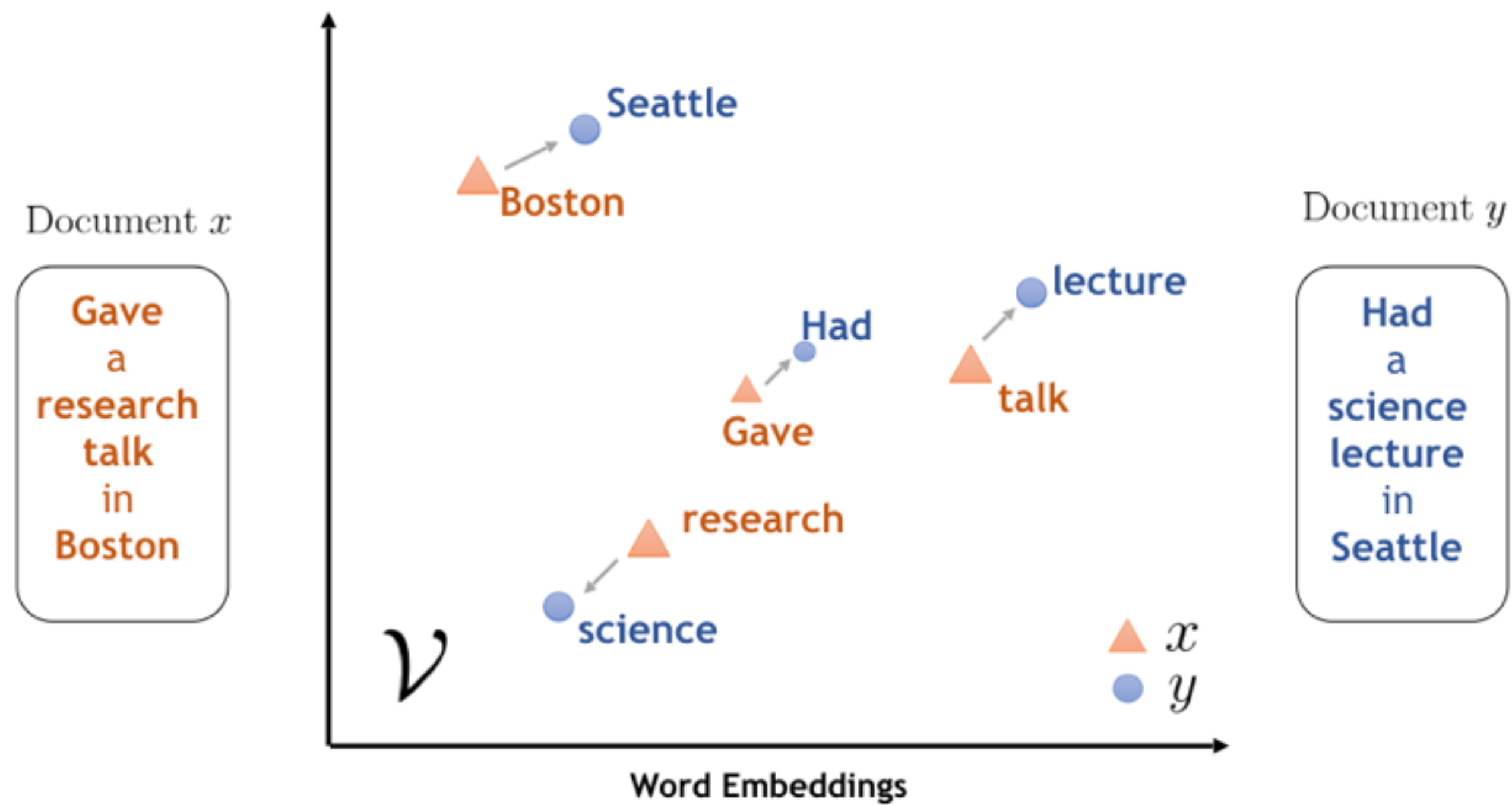
=

255	255	255	255	255	255	255	255	255	255	255
255	255	20	0	255	255	255	255	255	255	255
255	255	75	75	255	255	255	255	255	255	255
255	75	95	95	75	255	255	255	255	255	255
255	96	127	145	175	255	255	255	255	255	255
255	127	145	175	175	175	255	255	255	255	255
255	127	145	200	200	175	175	95	255	255	255
255	127	145	200	200	175	175	95	47	255	255
255	127	145	145	175	127	127	95	47	255	255
255	74	127	127	127	95	95	95	47	255	255
255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255

0 = black; 255 = white

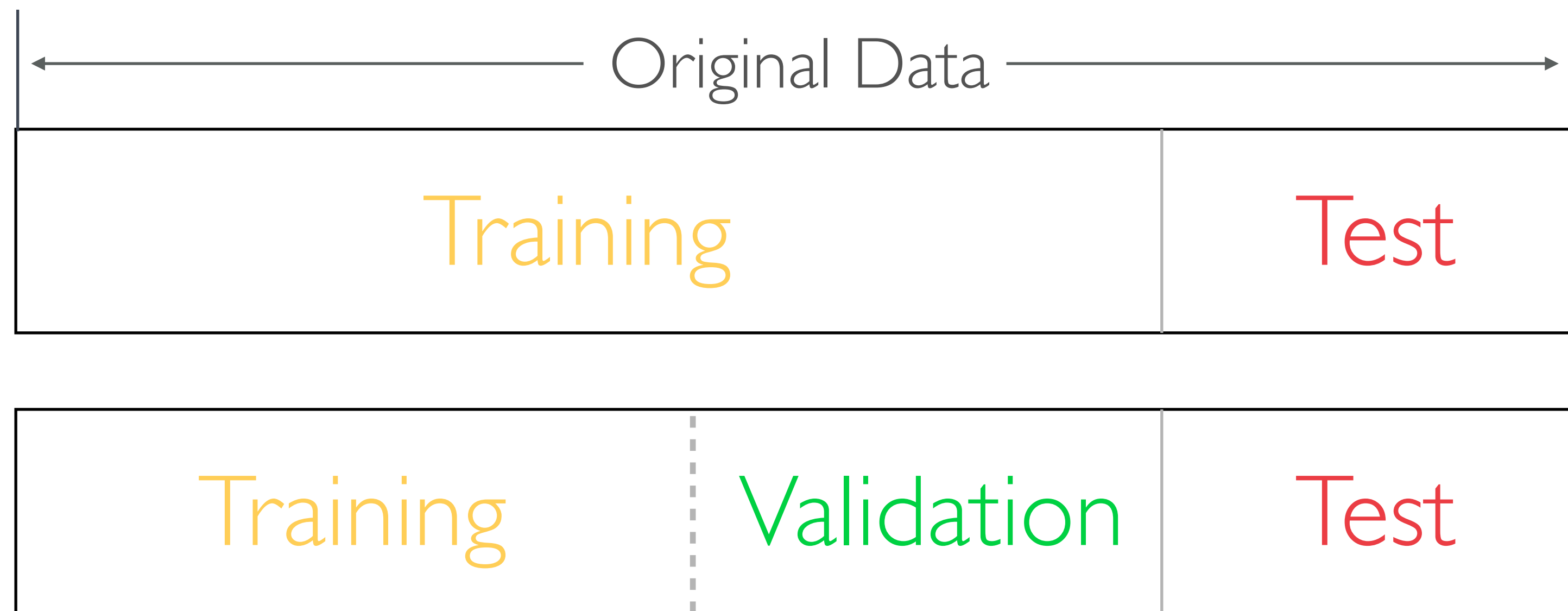
Text Feature Vector

텍스트를 벡터로 나타낸다면?



Data Split

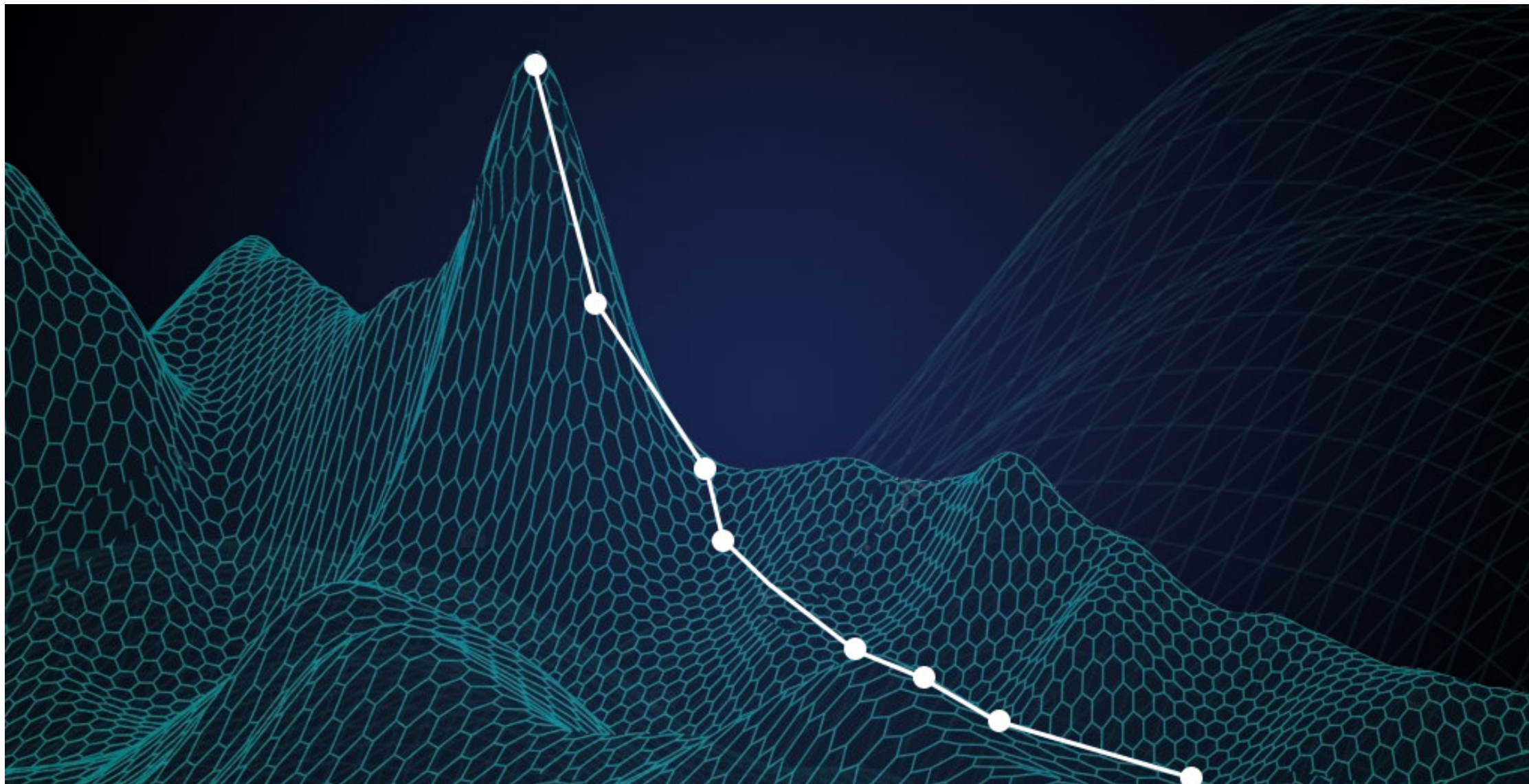
학습에 필요한 데이터



- 학습에 사용하는 데이터 : training data
- hyperparameter tuning에 사용하는 데이터 : validation data
- 최종 성능확인에 사용하는 데이터 : test data

Loss Function

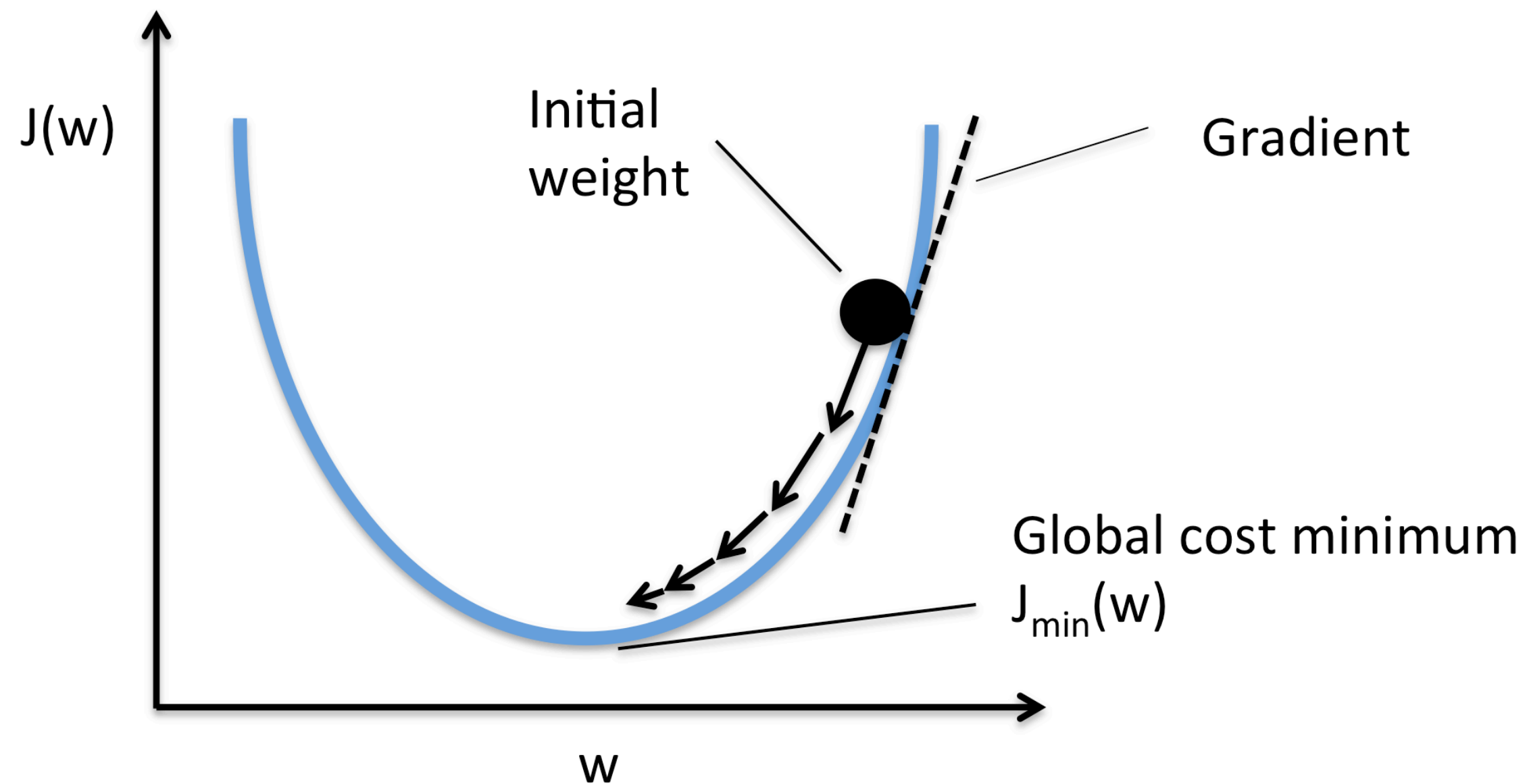
학습 목적은 무엇인가?



- Loss Function = Target Function
- $\text{Loss} = y - \hat{y}$
- Q1. Loss가 0이 되면 어떤 의미일까요?
- A1. _____
- Q2. Loss가 0이 되는 지점을 어떻게 찾을까요?
- A2. _____

Loss Function Optimization

학습 목적의 최적 포인트를 찾자



- 골짜기 가장 밑바닥으로 내려갈 수 있는 현실적인 방법
- 각 지점에서 기울기(gradient)가 가장 큰 방향 (가파른) 방향으로 내려갑니다.
- 한 번에 얼마나 내려가느냐(learning rate)에 따라 내려가는 속도가 결정됩니다.

Q. 그냥 함수를 미분해서 0이되는 지점을 찾으면 안되나요?

5

머신러닝 프로젝트 진행 방법, ML Workflow

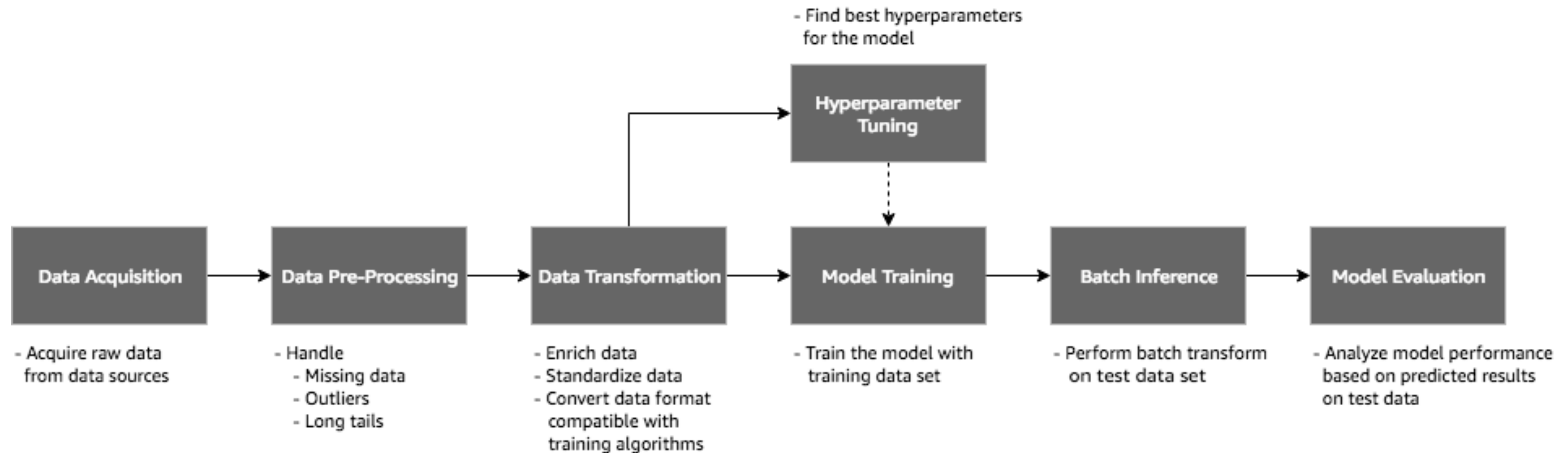
ML Workflow Basic Concept

7 steps of ML Workflow

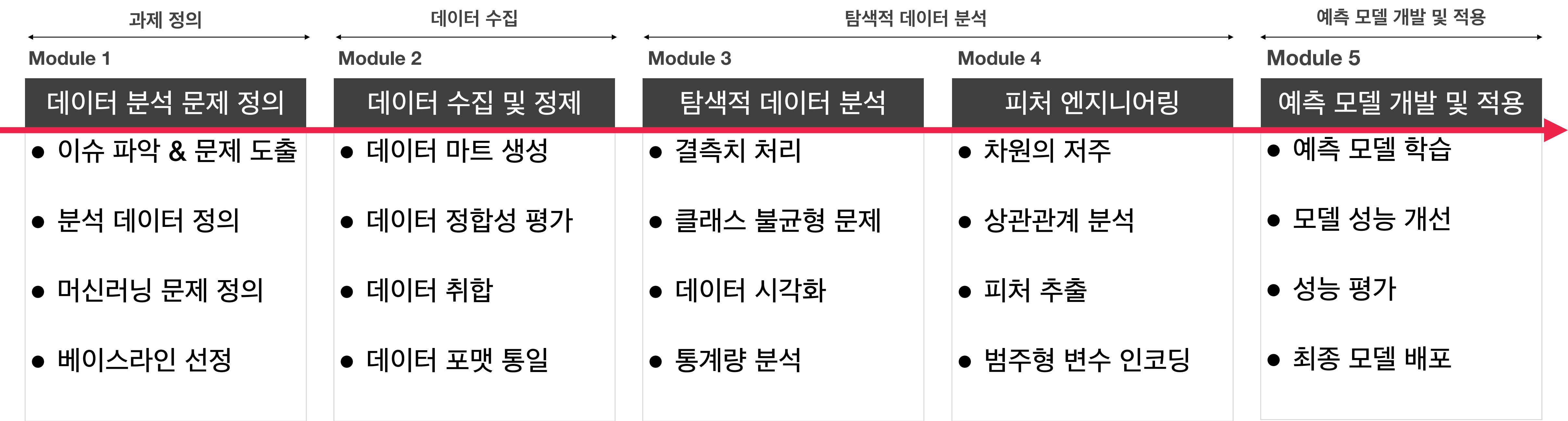


ML Workflow Diagram

7 steps of ML Workflow



ML Workflow



End of Slides