Idea Factory Intensive Program #2

딥러닝 홀로서기

#1

이론강의/PyTorch실습/코드리뷰

딥러닝(Deep Learning)에 관심이 있는 학생 발굴을 통한
딥러닝의 이론적 배경 강의 및 오픈소스 딥러닝 라이브러리 PyTorch를 활용한 실습

Acknowledgement

Sung Kim's 모두를 위한 머신러닝/딥러닝 강의
- https://hunkim.github.io/ml/
- https://www.youtube.com/playlist?list=PLlMkM4tgfjnLSOjrEJN31gZATbcj_MpUm

Andrew Ng's and other ML tutorials
- https://class.coursera.org/ml-003/lecture
- http://www.holehouse.org/mlclass/ (note)
- Deep Learning Tutorial
- Andrej Karpathy's Youtube channel

WooYeon Kim & SeongOk Ryu's KAIST CH485 Artificial Intelligence and Chemistry
- https://github.com/SeongokRyu/CH485---Artificial-Intelligence-and-Chemistry

SungJu Hwang's KAIST CS492 Deep Learning Course Material

Many insightful articles, blog posts and Youtube channels

Facebook community
- Tensorflow KR (https://www.facebook.com/groups/TensorFlowKR/)
- Pytorch KR (https://www.facebook.com/groups/PyTorchKR/)

Medium Channel and Writers
- Toward Data Science (https://towardsdatascience.com/)

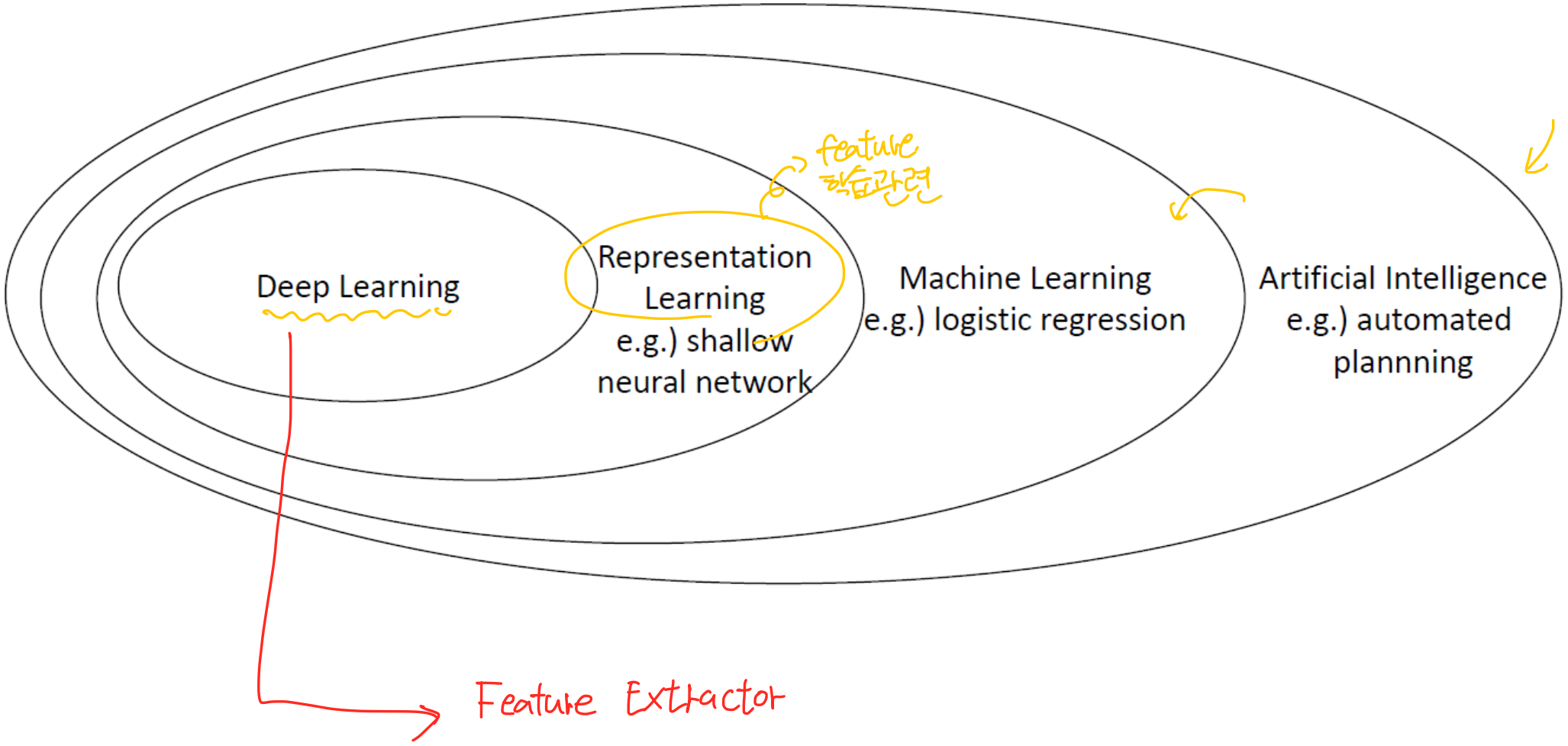
What is Machine Learning?

"A Field of study that gives computer the ability to learn without being explicitly programmed"

=> 일일이 rule-based로 짤 필요없이 자기가 알아서 학습을 한다.

– Arthur Samuel, 1959

Deep Learning, Machine Learning, Artificial Intelligence



Categories of ML Problems

학습법에 따른 분류

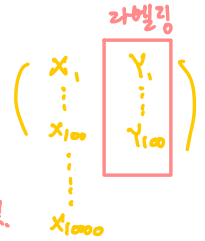
output space	Supervised Learning (암기)	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering (묵기)	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Supervised Learning
=> "x"라는 입력값을 주며 "y"라는 답을내야해!
같이 샘플을주며 학습시키는것

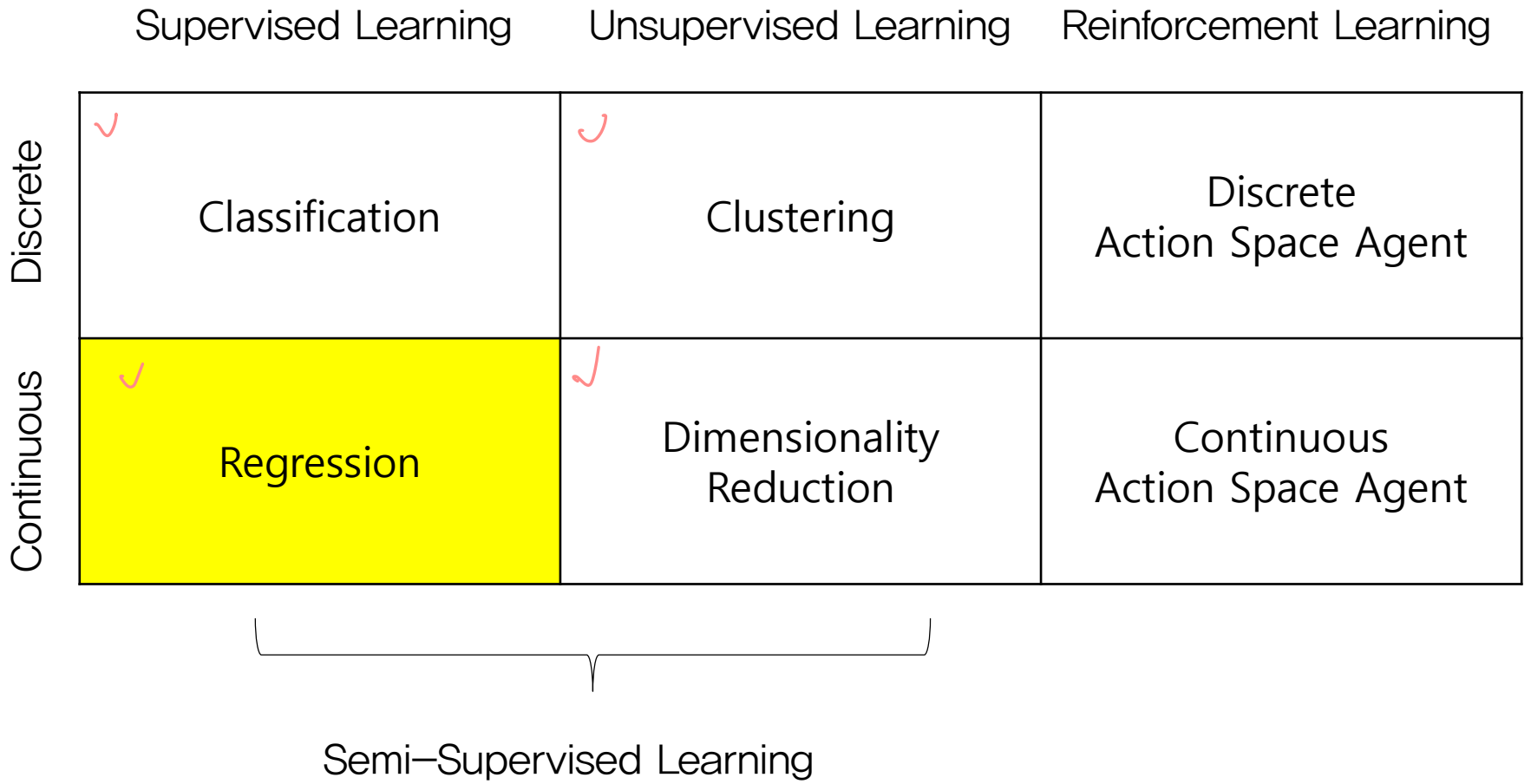
Unsupervised Learning
x만 있는 경우 활용할수 있는 방식
=> 비슷한것끼리 군집화를 자동으로 하고
이는 사람이 라벨링을 해주는것

Reinforcement Learning
=> 현재 상태 -> action -> 상태
이러한 과정에서
reward function을 통해
상태의 좋고 나쁨을 평가함으로써
자동으로 agent가 reward를 최대화하는
방식을 찾는것

(Semi-Supervised Learning)
<> Supervised와 Unsupervised를 섞어놓은것. => ex. 분류작업이 10000개 있으면 100개정도 라벨링을 해서 학습.



Categories of ML Problems



Regression Problem

Price Prediction Based on Gi-Young Style Chart Analysis

https://coinpan.com/free/22082900

Regression Problem

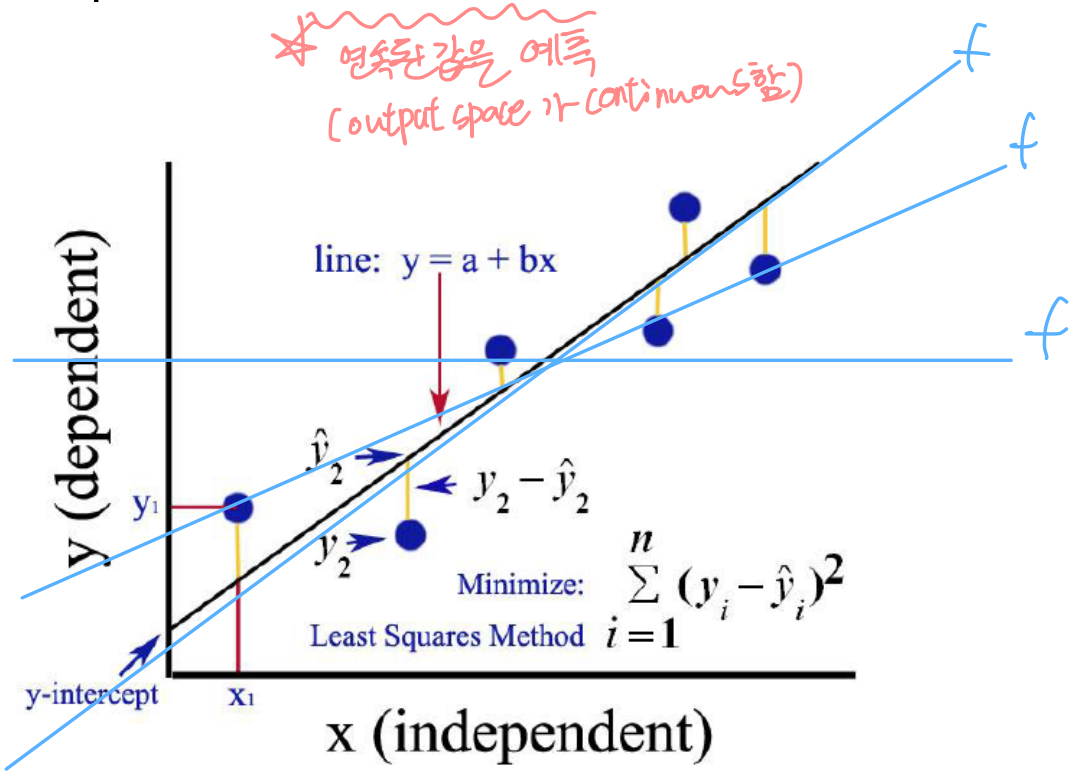
회귀문석

f(x)가 우리가 알는 x를 y에 매핑할 식을 알는경우, f(x)를 찾는 경우가 속함

X -> Y
f(x) = ax+b
↑ 사람이 form을 정해줌

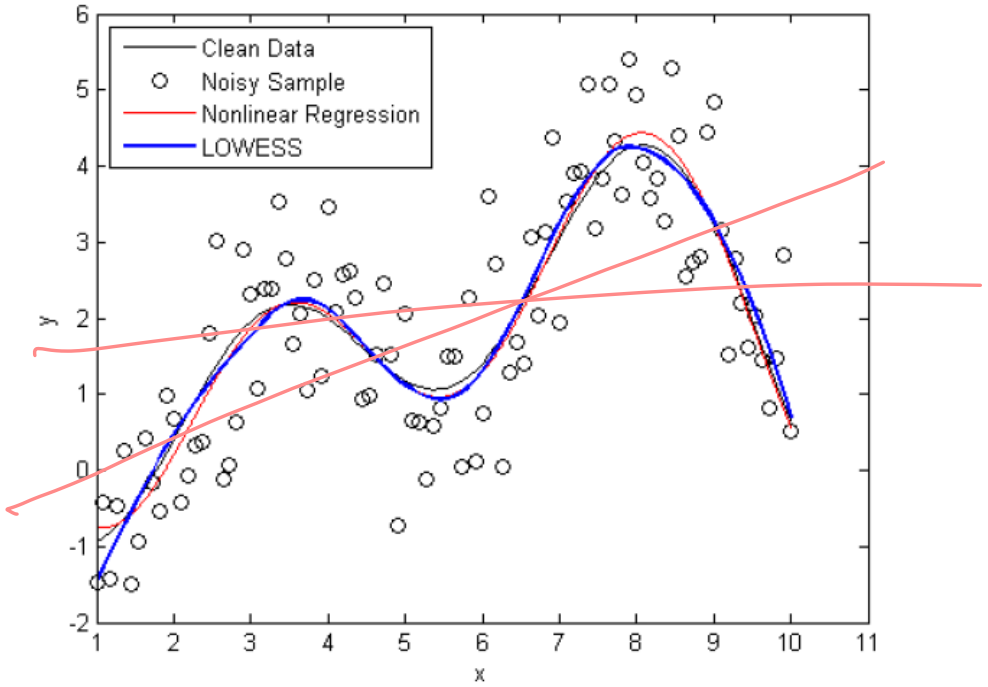
Fit the prediction function f(x) to the training data,
to predict continuous real value

★ 연속된 값을 예측 (output space가 continuous함)



Linear regression

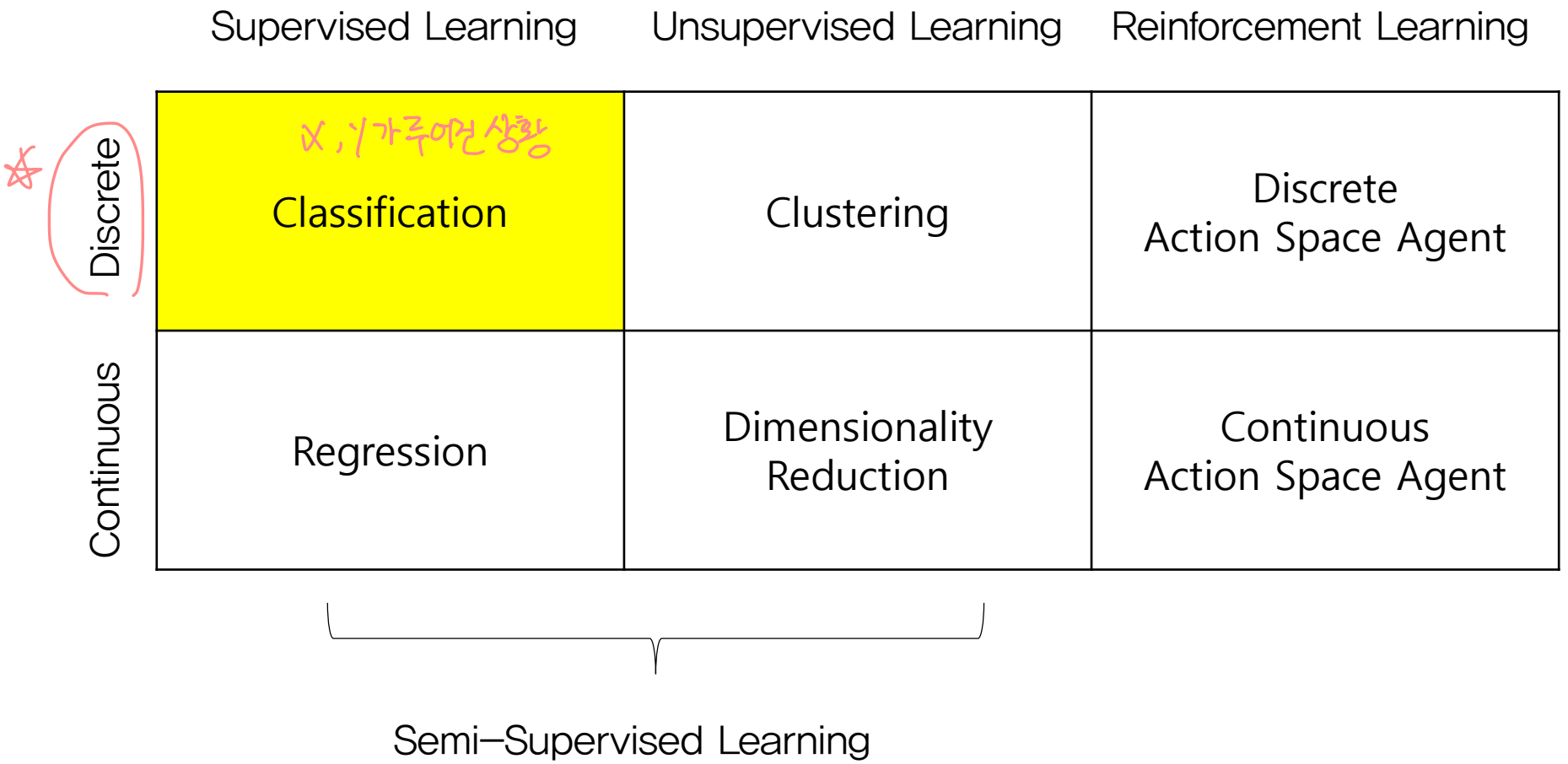
↑ 데이터들을 가장 잘 나타내는 함수를 찾는것



Nonlinear regression

↑ 직선으로 표현 안되는 nonlinear한 경우

Categories of ML Problems



Classification Problem

x : image

y : class [Ch, M]

Chihuahua or Muffin?

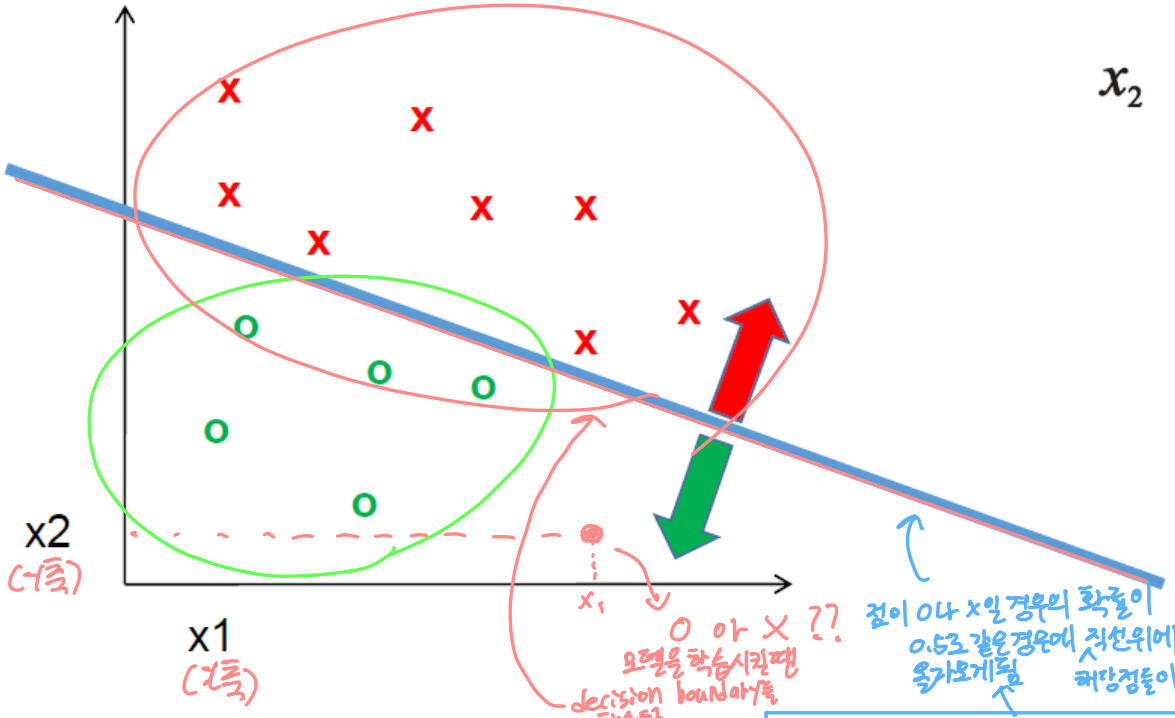
https://medium.freecodecamp.org/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d

Classification Problem

분류

Identifying which of a set of categories a new instance belongs

=> 새로운 x가 어느 클래스에 속할지 분류하는 문제

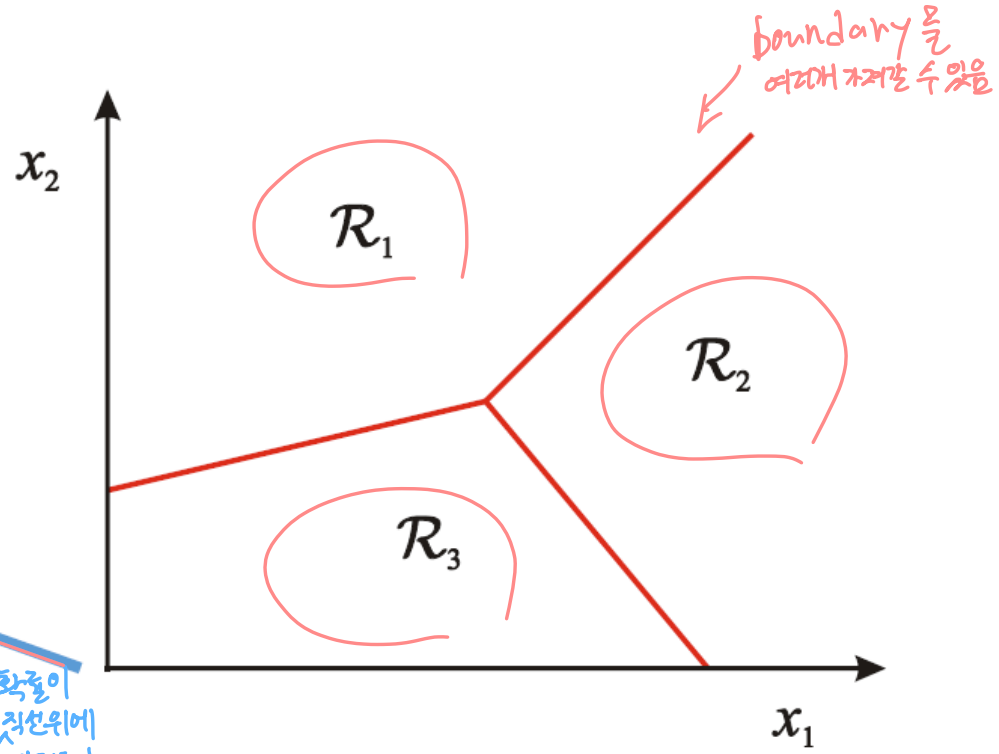


Binary Classification

input
x_1, x_2

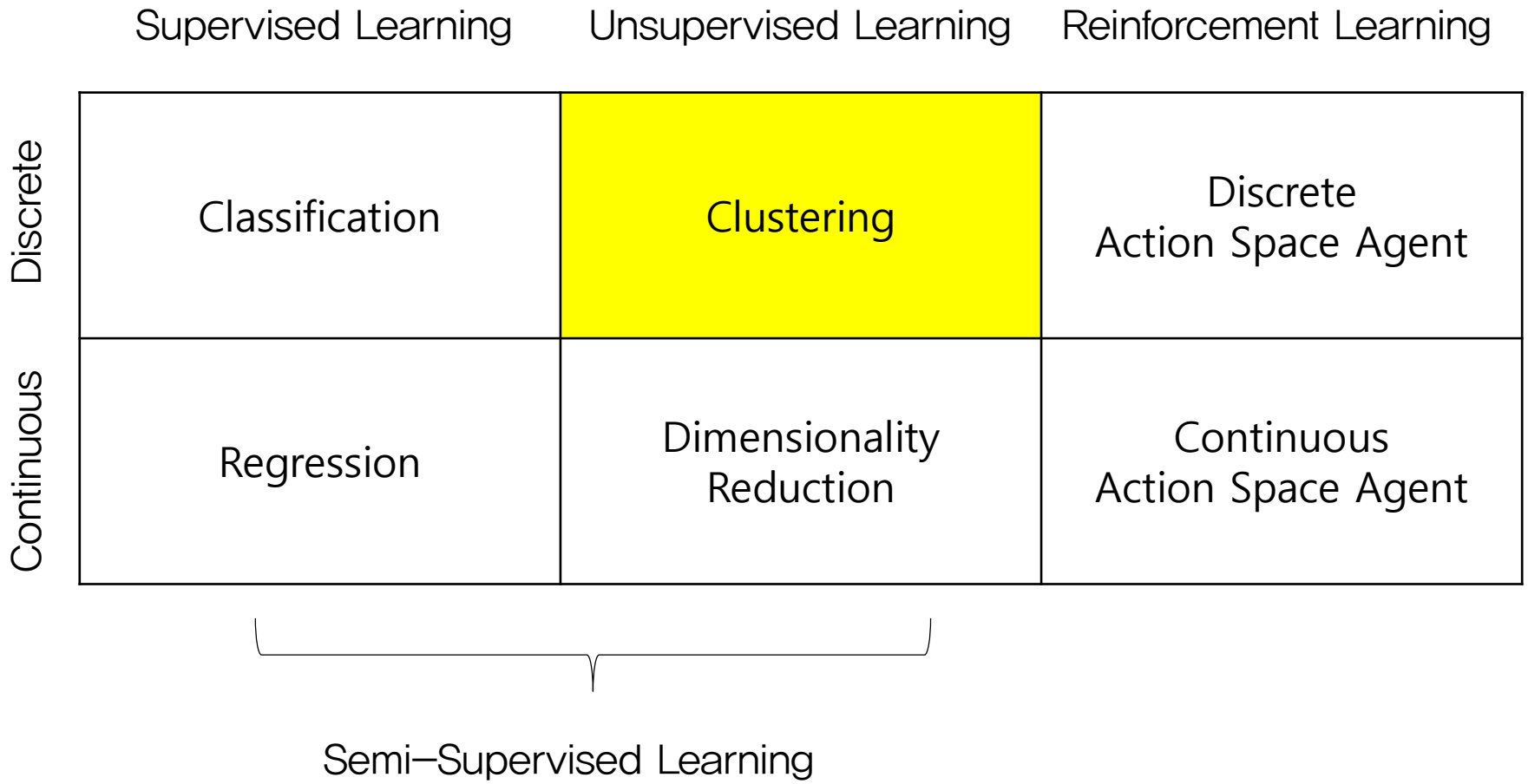
$P(O) = 0.5$

$P(X) = 0.5$

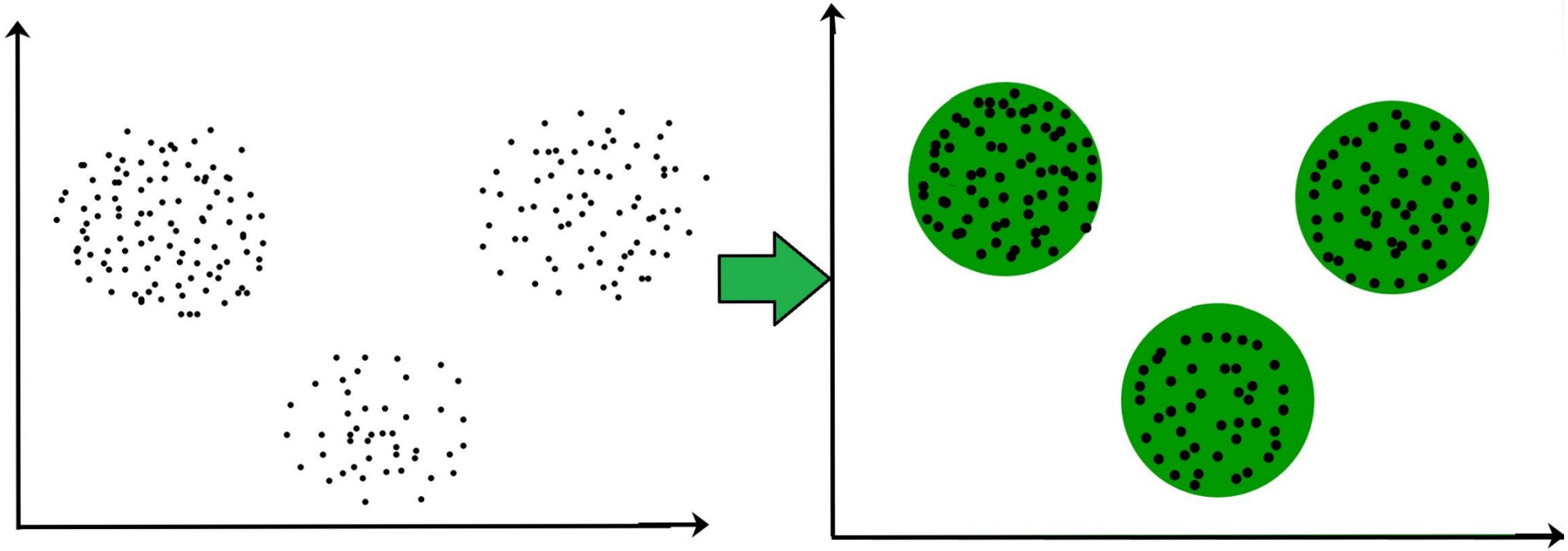


Multi-class Classification

Categories of ML Problems



Clustering Problem

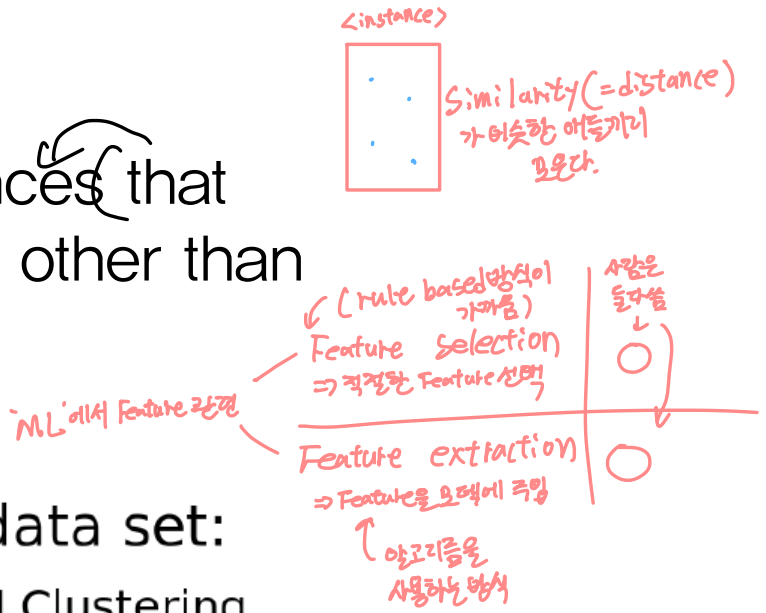


Grouping similar samples into K groups

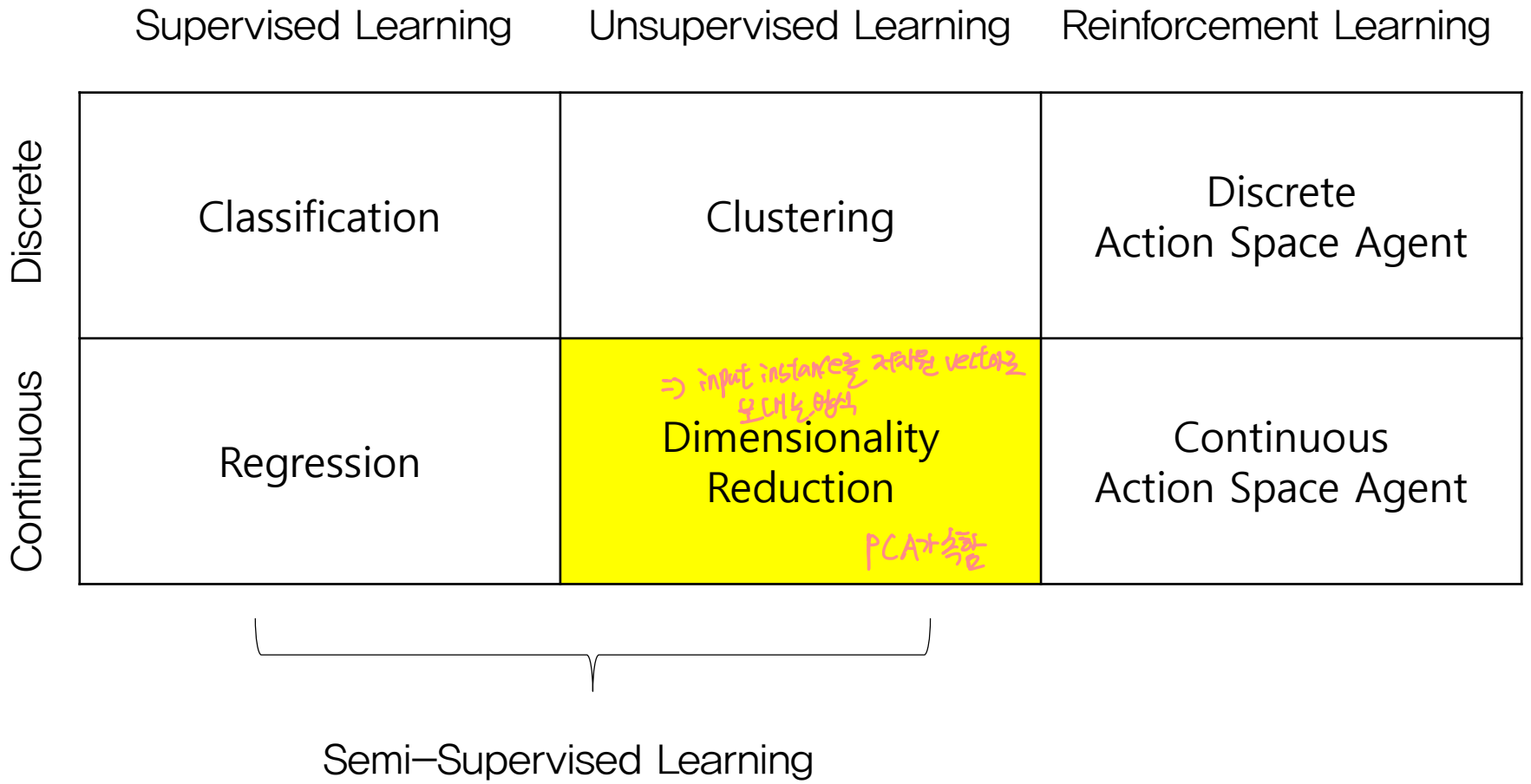
https://www.geeksforgeeks.org/clustering-in-machine-learning/

Clustering Problem

Automatic grouping of instances, such that the instances that belong to the same clusters are more similar to each other than to those in the other groups



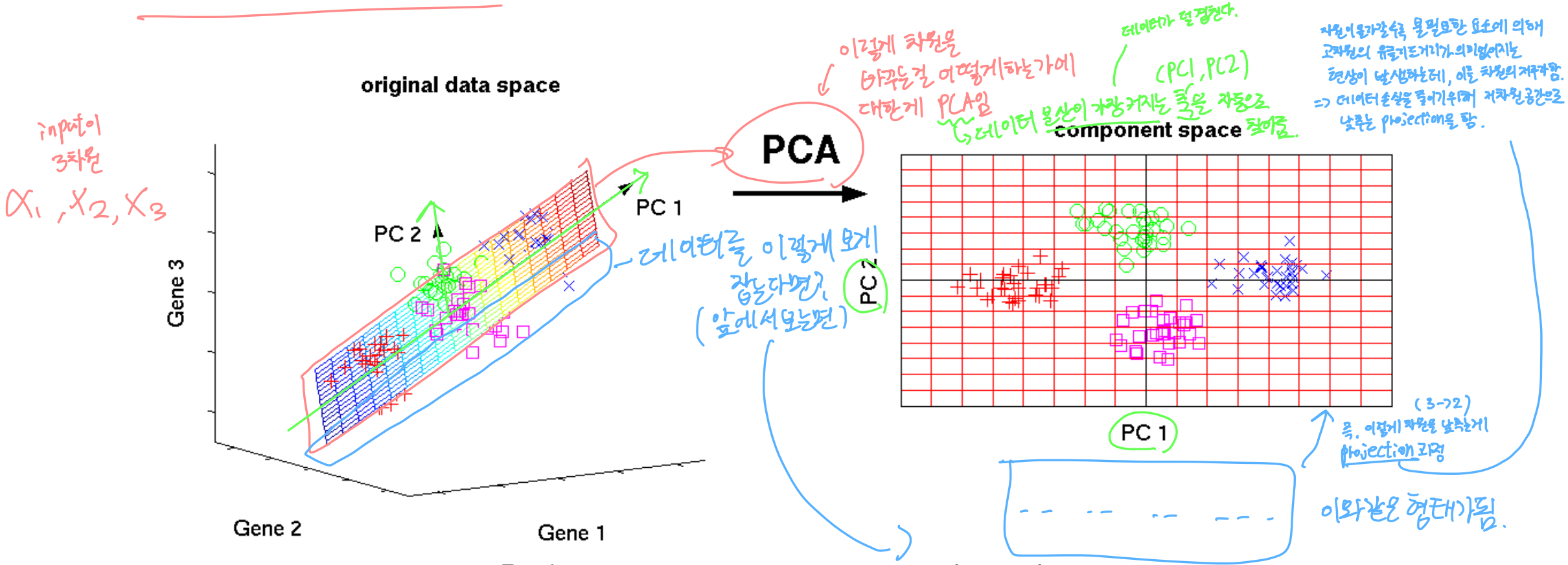
Categories of ML Problems



Dimensionality Reduction Problem

이 예제는 3차원 데이터를 2차원으로 낮추보는 시도를 하는중. => similarity나 distance 계산을 더 잘할수있음.

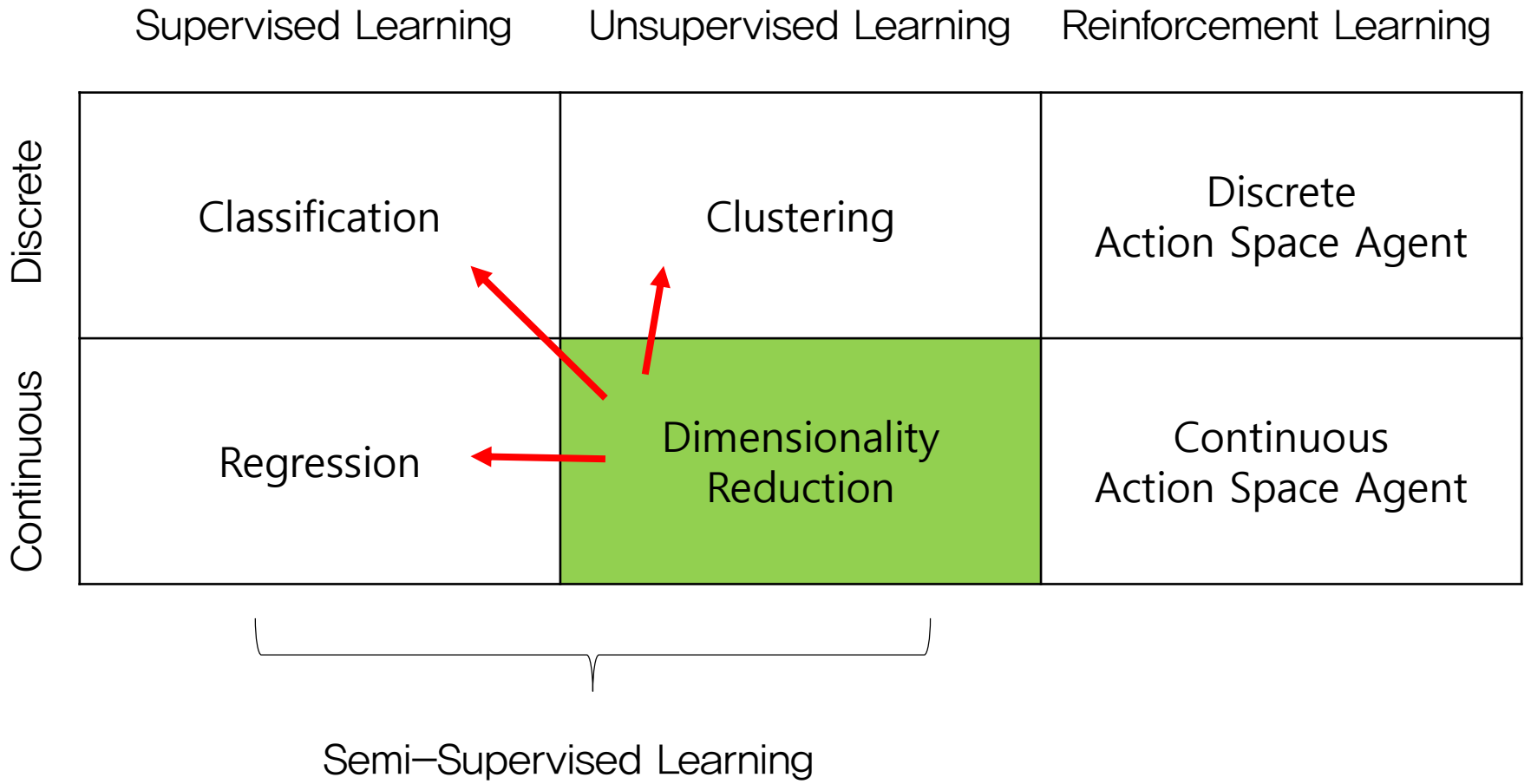
Reduce the dimension of input data, to avoid the effect of the curse of dimensionality



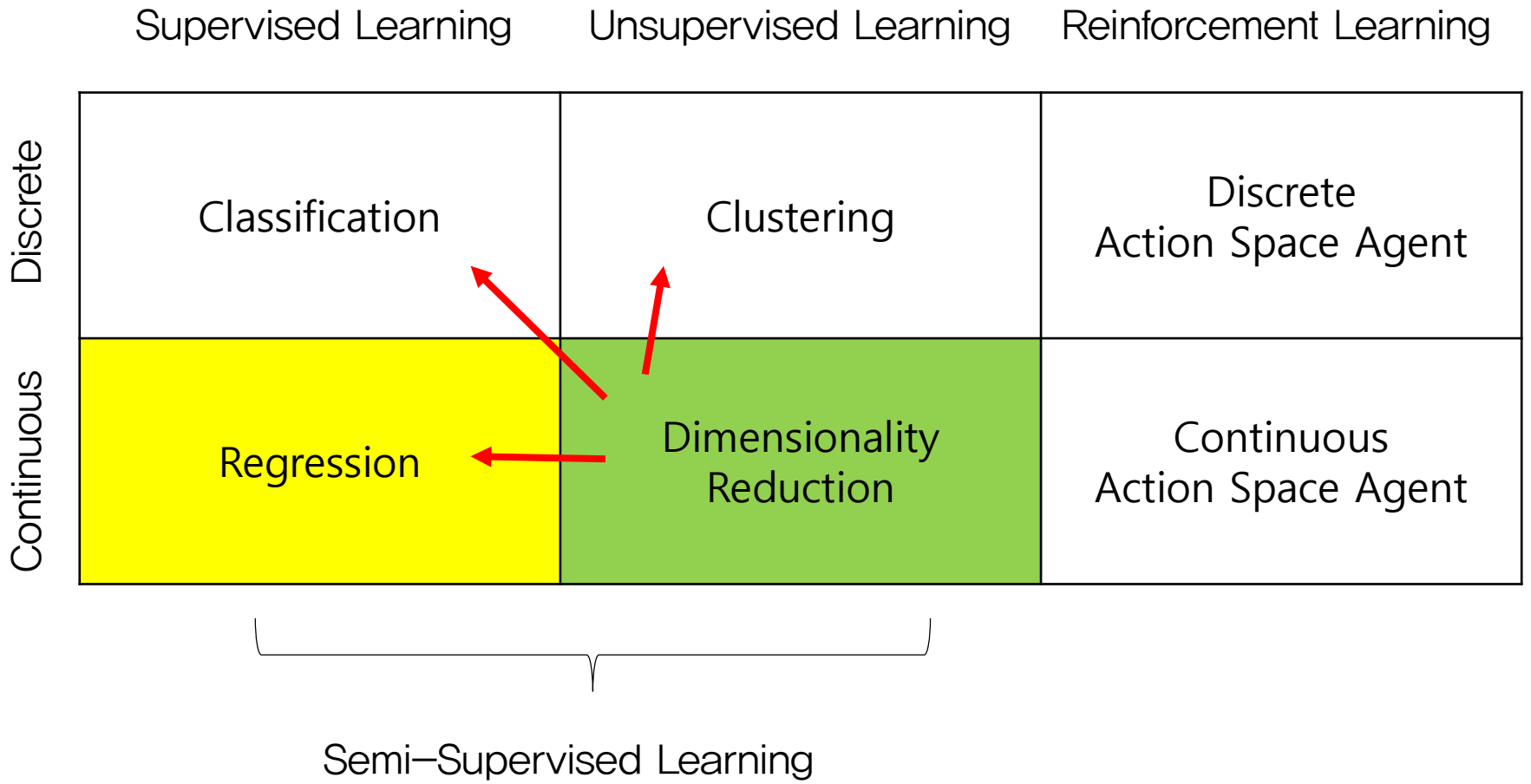
Reduce unnecessary representation axis

https://hackernoon.com/a-laymans-introduction-to-principal-components-2fca55c19fa0

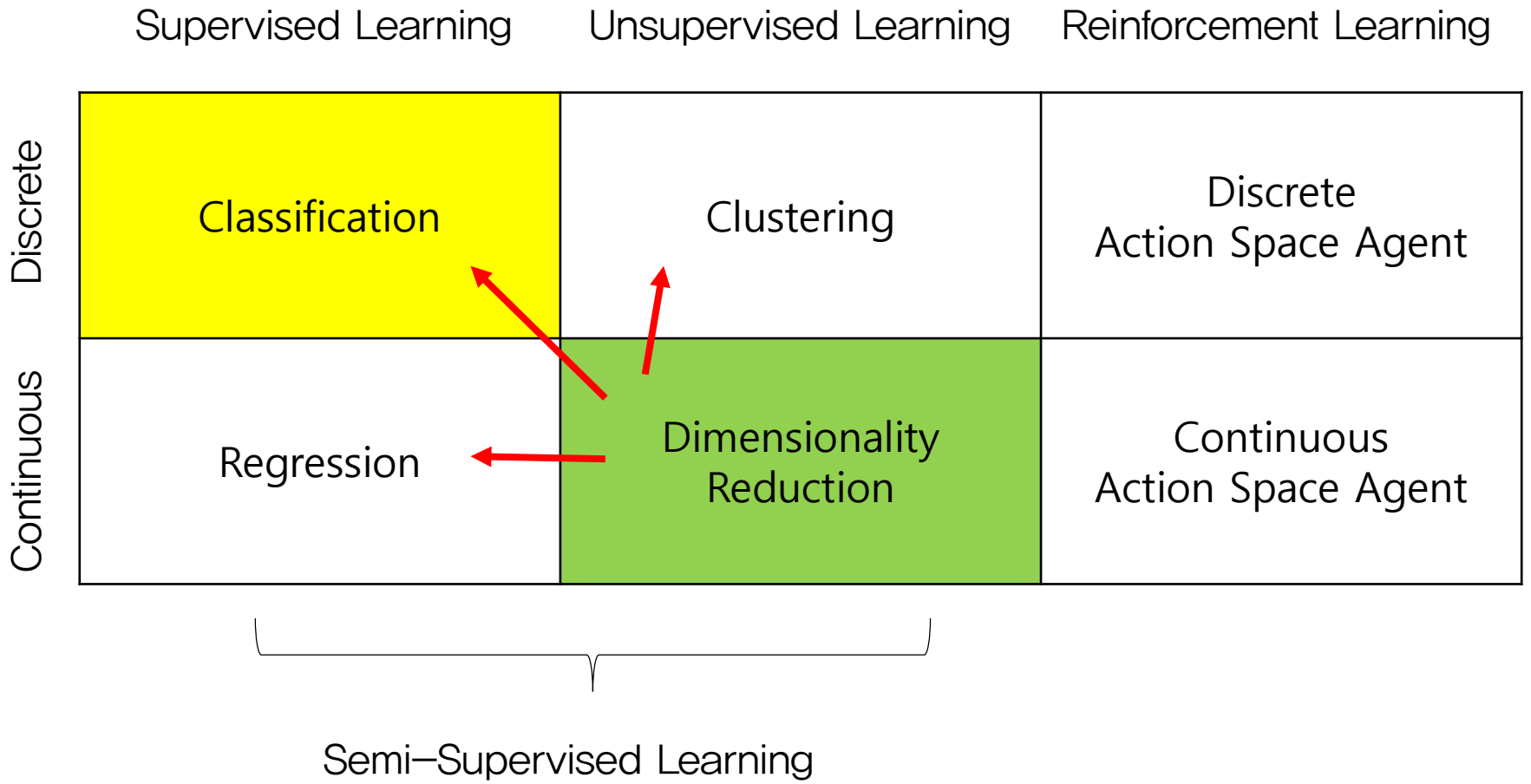
Categories of ML Problems



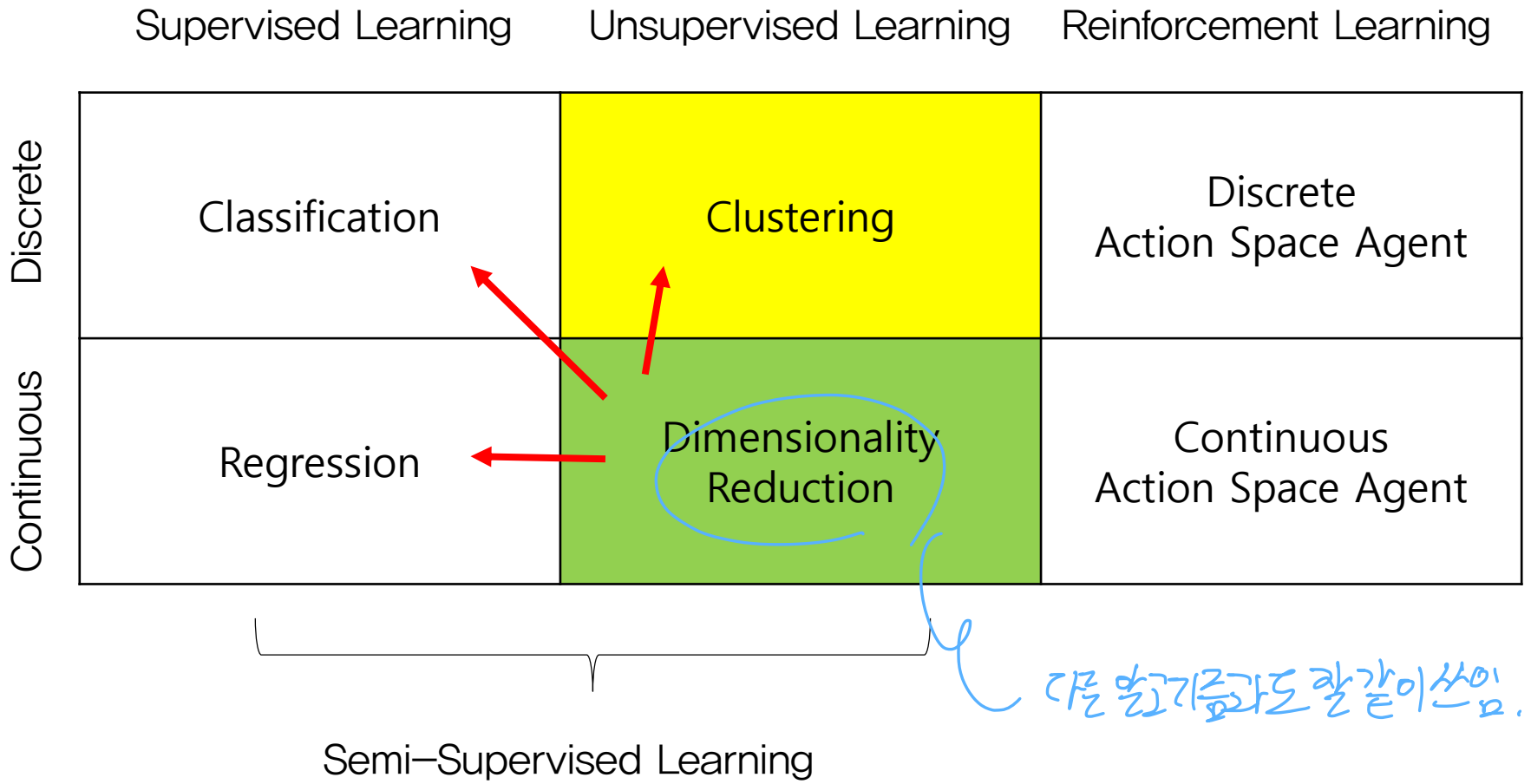
Categories of ML Problems



Categories of ML Problems

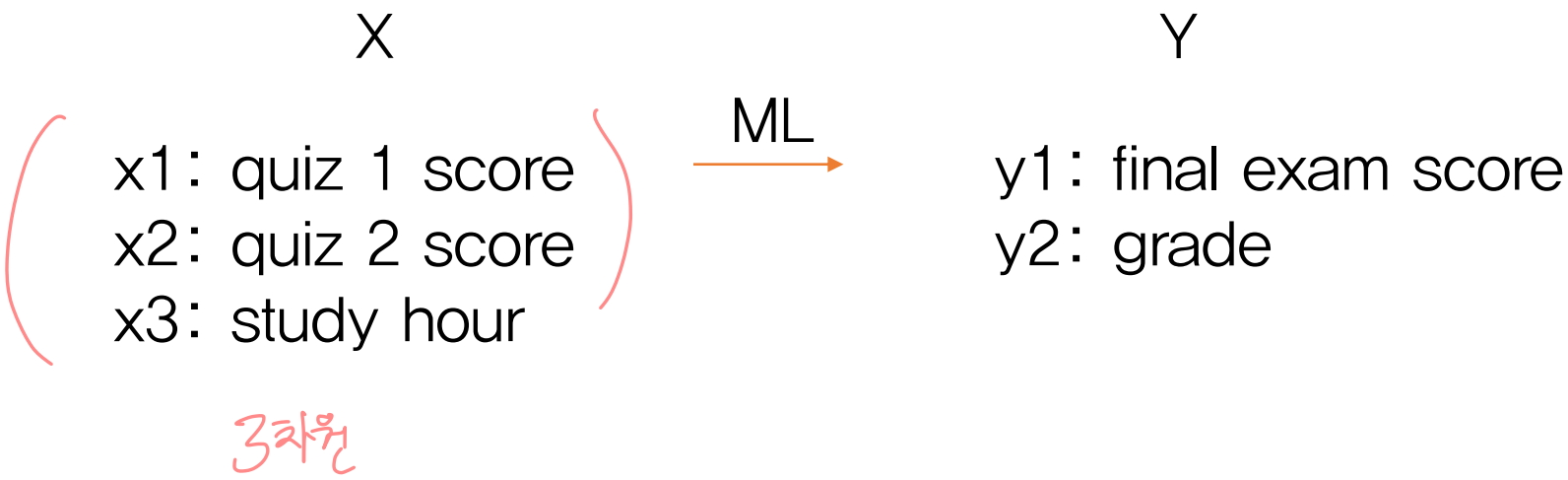


Categories of ML Problems

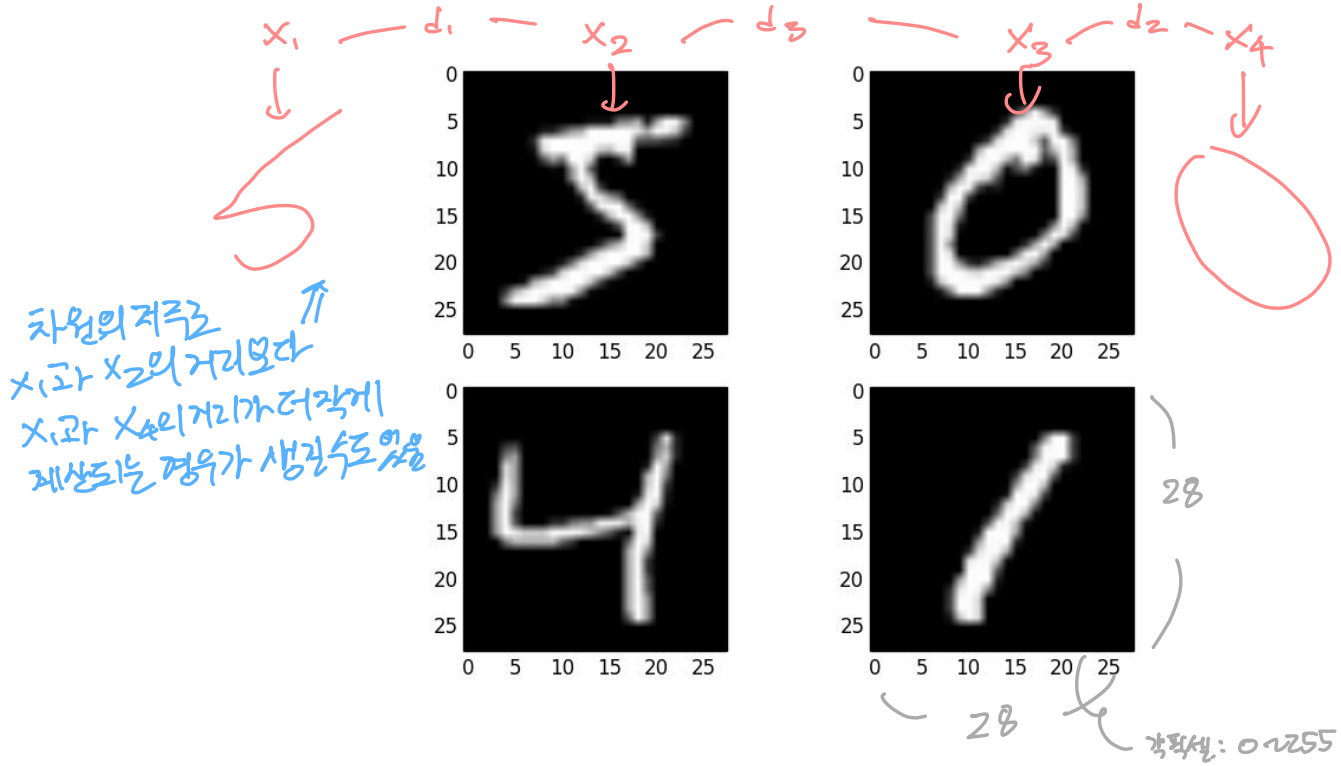


Feature & Data Representation

Case 1



Feature & Data Representation



차원의 거리로 x1과 x2의 거리보다 x1과 x4의 거리가 더 작게 계산되는 경우가 생길수도 있음

Case 2

X

x1: first pixel value
x2: second pixel value
x3: third pixel value
...
x784: 784th pixel value

Y

y1: digit

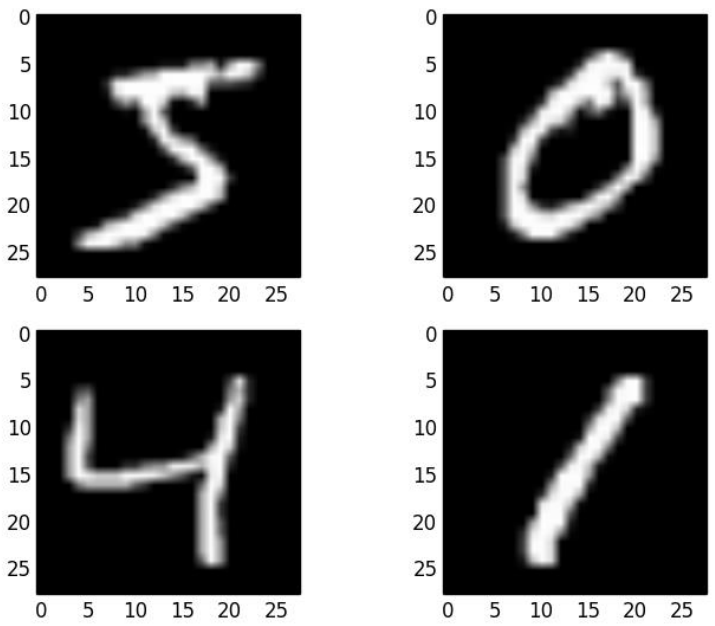
0~9

10개의 class가 아닌 descriptive space로 매핑하고싶은 경우를 생각해보자

각 사진 1개 1개가 784차원 vector의 vector로 표현할 수 있음

https://machinelearningmastery.com/handwritten-digit-recognition-using-convolutional-neural-networks-python-keras/

Feature & Data Representation



Feature Extraction selector 와 이름이 비슷.

그차원으로 변경

Case 2

X → h → Y

DL

x1: first pixel value
x2: second pixel value
x3: third pixel value
...
x784: 784th pixel value

h1: feature1
h2: feature2
h3: feature3

ML

y1: digit

https://machinelearningmastery.com/handwritten-digit-recognition-using-convolutional-neural-networks-python-keras/