

TEXT MINING for PRACTICE

Python을 활용한 비정형 데이터 분석 - WEEK 12

전병진 FINGEREDMAN (fingeredman@gmail.com)

Part 11.

비정형데이터와 머신러닝

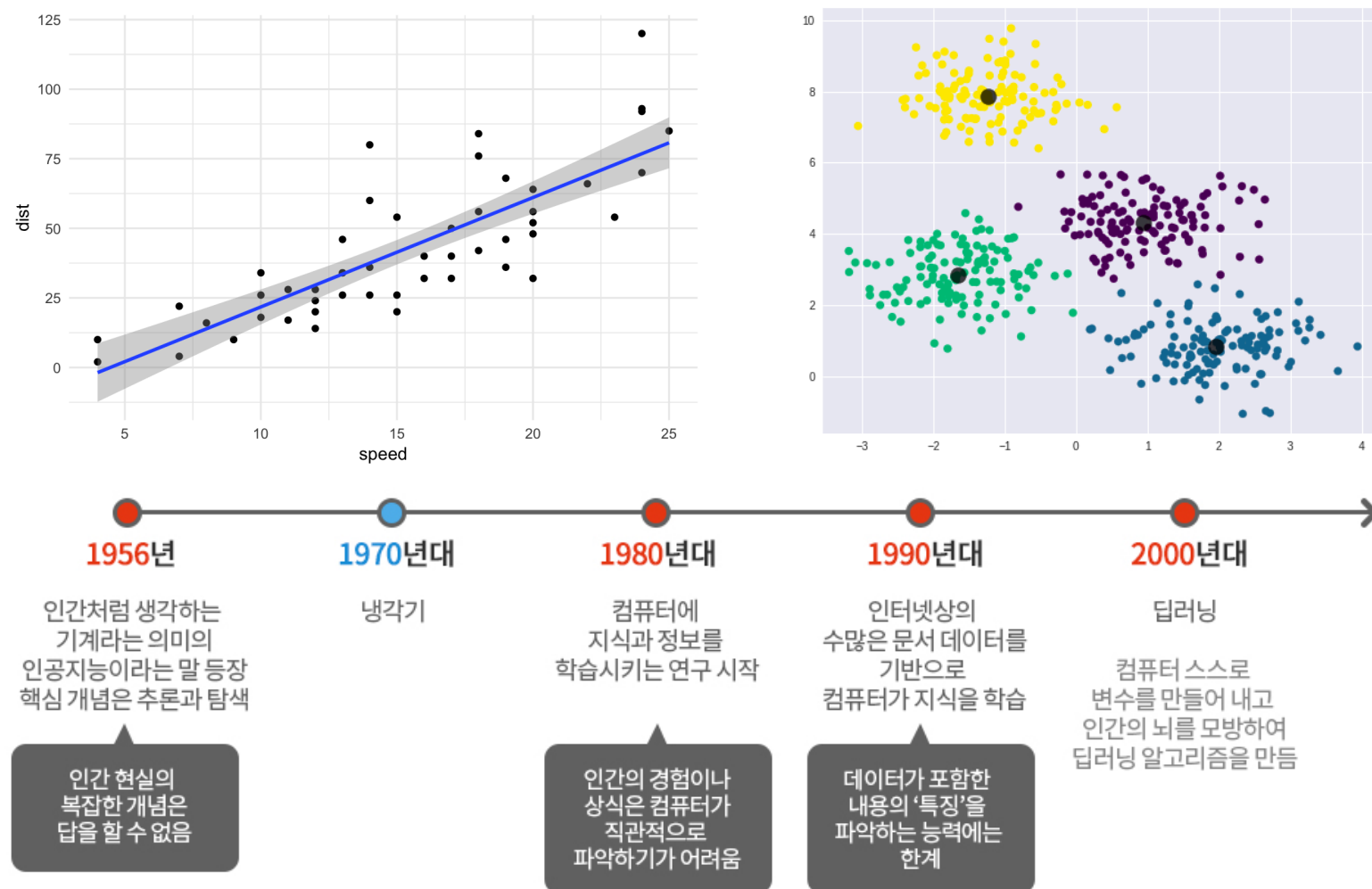
데이터 분석

Machine Learning & Deep Learning

Machine Learning

머신러닝 (Machine Learning) 이란?

- ▶ 사람이 직접적으로 규칙을 지정하지 않아도, 컴퓨터가 직접 데이터를 통해 “학습 (learning)”하고 “경험”을 통해 자동으로 문제를 해결하는 방법
- ▶ 기계가 일일이 코드로 명시하지 않은 동작을 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘 등을 개발하는 분야



Machine Learning

머신러닝의 유형

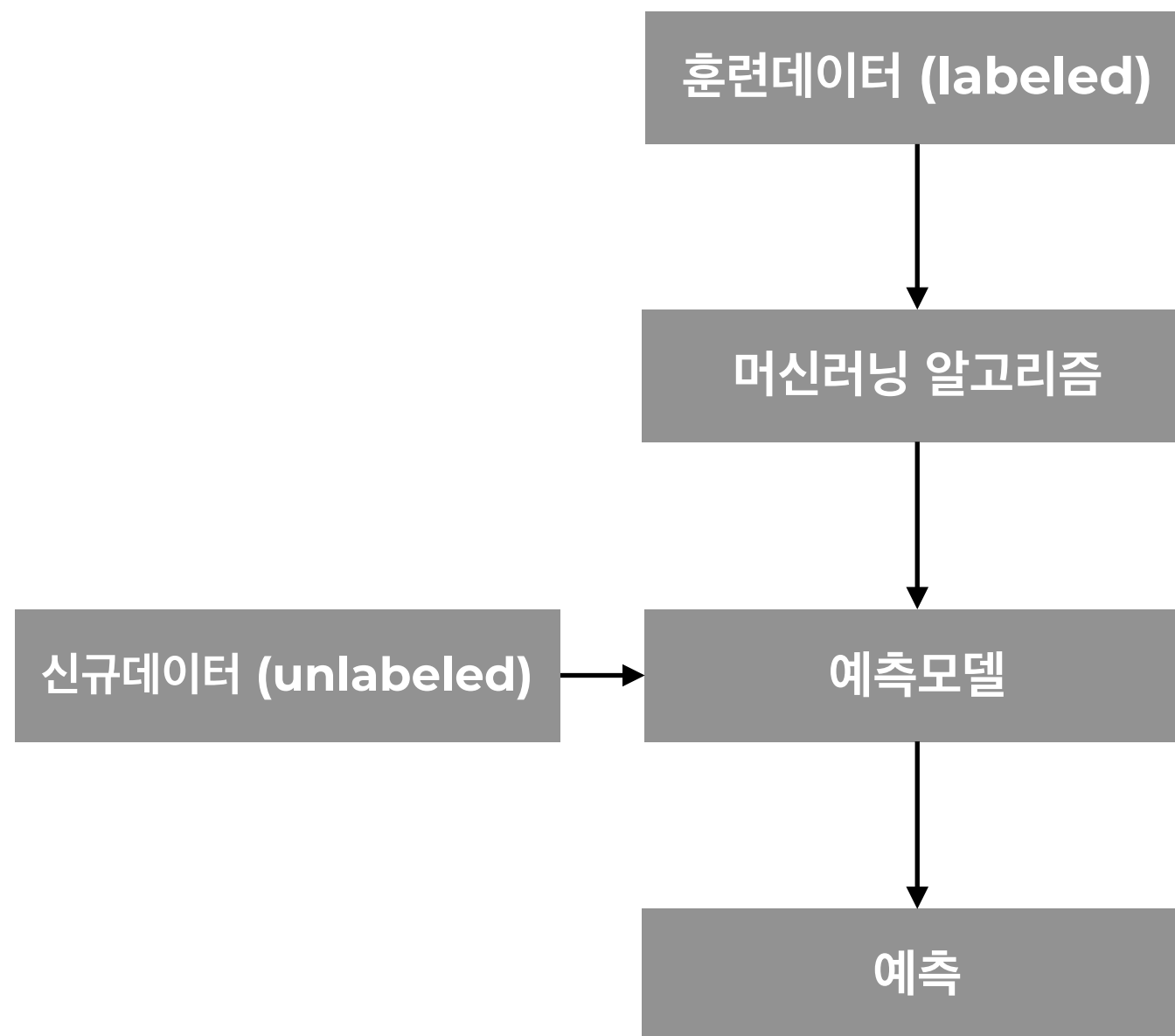
- ▶ 지도학습 (supervised learning) : 입력에 대한 정답을 알고있는 상태에서 대조하며 학습하는 방법
- ▶ 비지도학습 (unsupervised learning) : 정답이 없는 입력값들만 주어진 상태로 학습하는 방법

Machine Learning Types	Tasks	Analysis methods/Algorithms
지도 학습 (Supervised Learning)	예측, 추정 (Prediction, Estimation)	<ul style="list-style-type: none"> Linear Regression Regression Tree, Model Tree SVM(Support Vector Machine) Neural Network, Deep Learning ARIMA, Exponential Smoothing
	분류 (Classification)	<ul style="list-style-type: none"> Decision Tree Logistic Regression, Discriminant Analysis k-NN(k-Nearest Neighbor), CBR(Case-Based Reasoning) Naïve Bayes Classification SVM, Neural Network Ensemble (Bagging, Boosting, Random Forest)
비지도 학습 (Unsupervised Learning)	패턴/구조 발견 (Pattern/Rule)	<ul style="list-style-type: none"> Association Rule Analysis, Sequence Analysis Network Analysis, Link Analysis, Graph theory Structural Equation Modeling, Path Analysis
	그룹화 (Grouping)	<ul style="list-style-type: none"> k-Means Clustering, Hierarchical Clustering, Density-based Clustering, Fuzzy Clustering SOM(Self-Organizing Map)
	차원 축소 (Dimension Reduction)	<ul style="list-style-type: none"> PCA(Principal Component Analysis), Factor Analysis, SVD(Singular Value Decomposition)
	영상, 이미지, 문자 (Video, Image, Text, Signal processing)	<ul style="list-style-type: none"> Wavelet/Fast Fourier Transformation, DTW(Dynamic Time Warping), SAX(Symbolic Aggregate Approximation), Line/Circular Hough Transformation Text mining, Sentiment Analysis

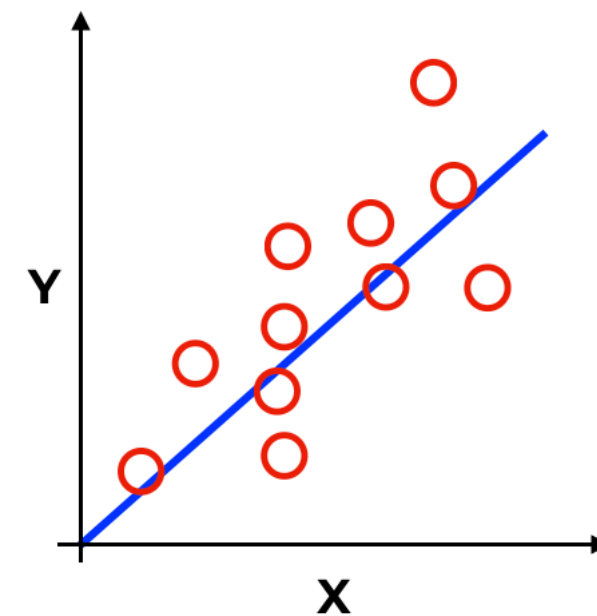


Machine Learning

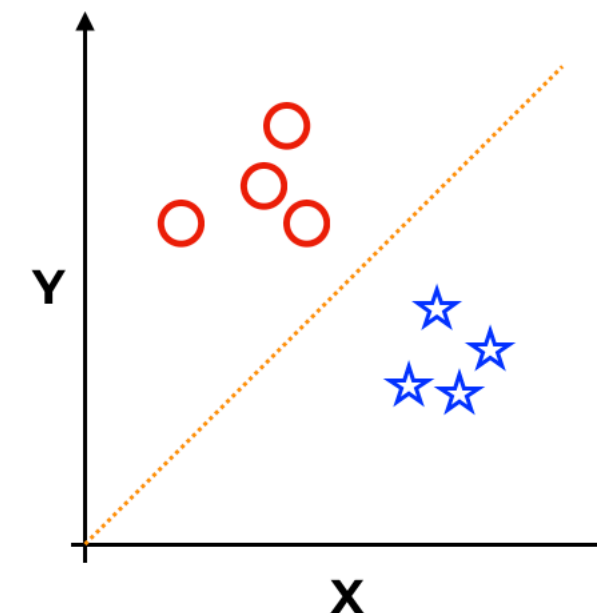
머신러닝의 유형: 지도학습 (Supervised Learning)



회귀분석 (Regression)

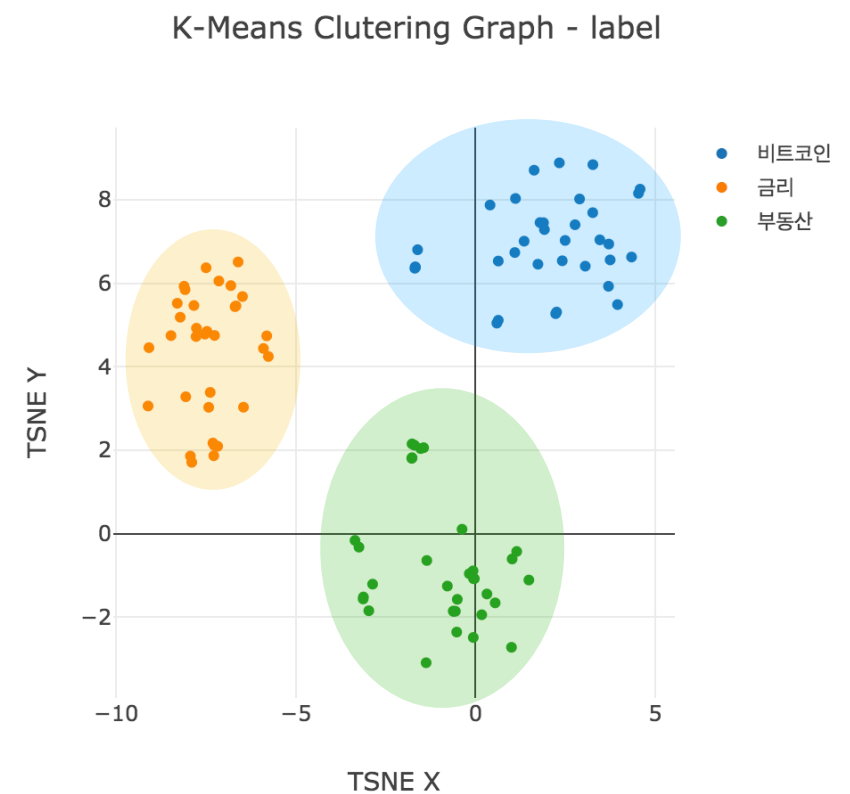
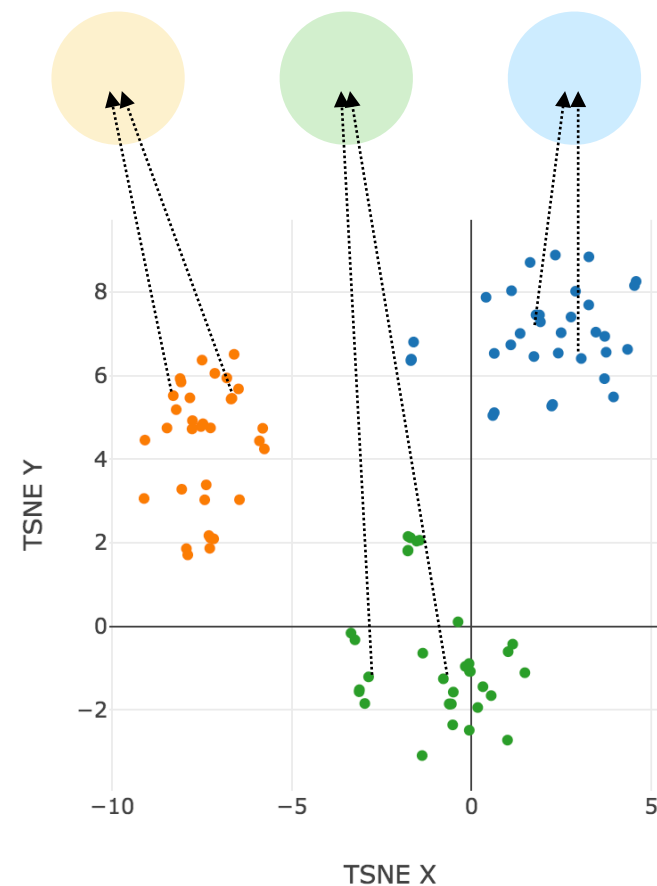
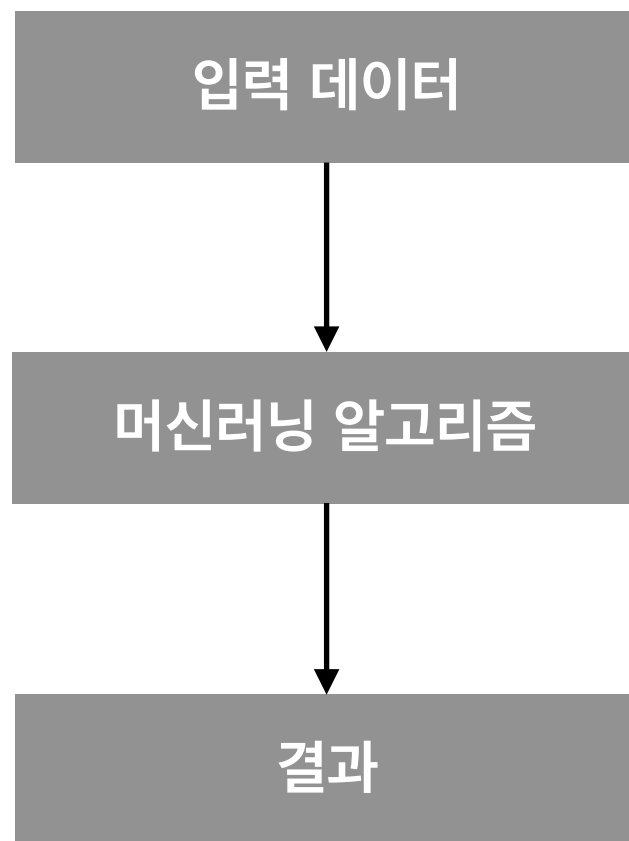


분류 (Classification)



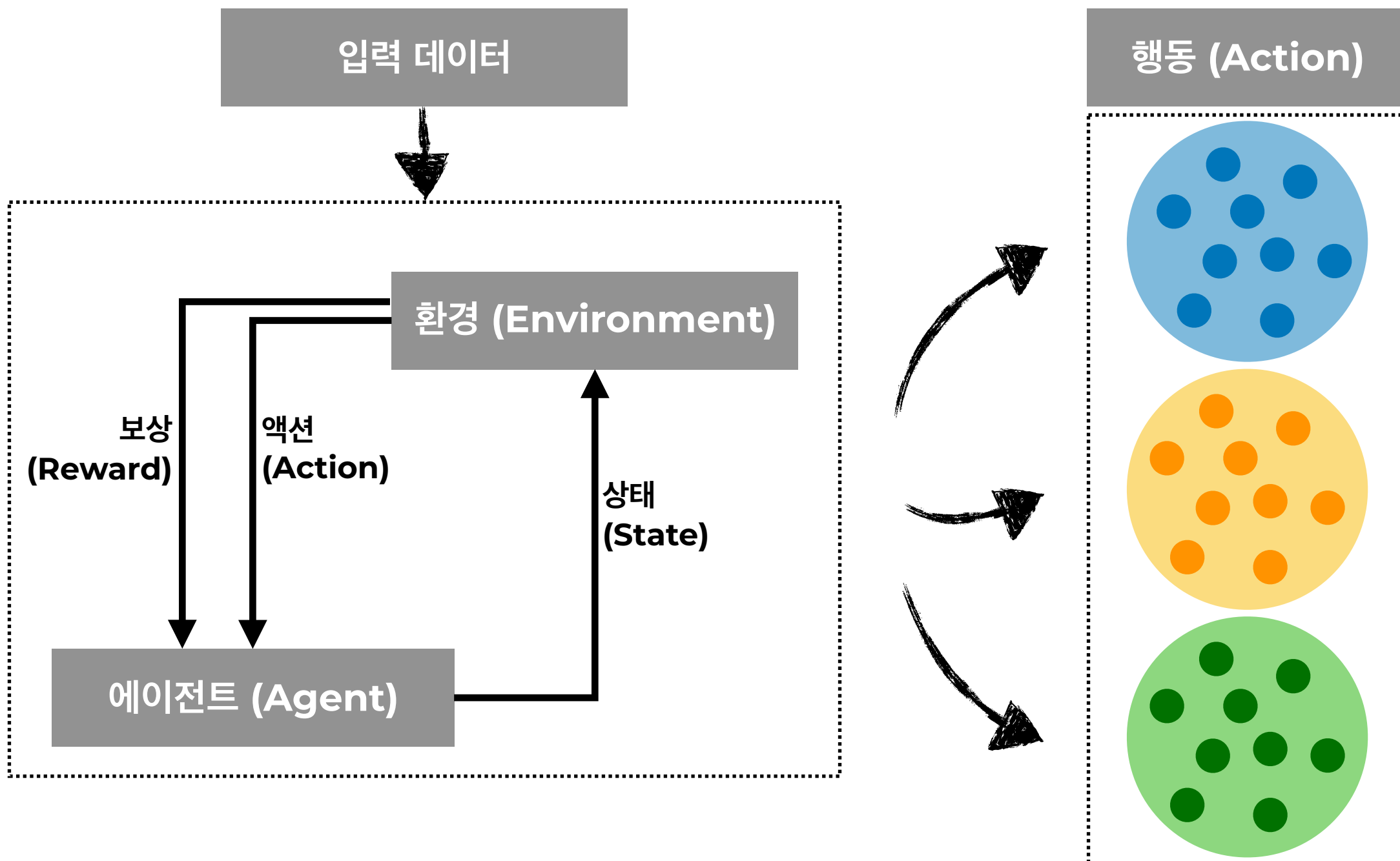
Machine Learning

머신러닝의 유형: 비지도학습 (Unsupervised Learning)



Machine Learning

머신러닝의 유형: 강화학습 (Reinforcement Learning)



Parametric Approach

Parametric Approach 란?

- ▶ 입력변수 (x)와 목표변수 (y) 사이의 복잡한 관계를 어떠한 파라미터 (w)와의 관계로 접근하는 방식
- ▶ 정답을 구하기위한 적절한 파라미터 (w)를 구하고 예측된 값 (pred.)과 정답 (label)과의 차이인 오류 (error, loss) 계산
- ▶ 머신러닝(인공신경망)의 궁극적인 목적은 오류의 평균을 최소화하는 적절한 파라미터 (w)를 도출

W_1	X_1	W_2	X_2	Y_n
?	0	?	1	2
	1		2	6
	1		1	4
	1.5		1	5



Parametric Approach

$$y = aX_1 + bX_2$$

a	x ₁	b	x ₂	y _n	Y	Y - y _n
?	0	?	1	y ₁	2	2 - y ₁
	1		2	y ₂	6	6 - y ₂
	1		1	y ₃	4	4 - y ₃
	1.5		1	y ₄	5	5 - y ₄
					AVG(Y - y _n)	

*Source : 진상형, Deep Learning: Introduction, 2017.9.30.

Parametric Approach

$$Y_n = W_1 \cdot X_1 + W_2 \cdot X_2$$

W_1	X_1	W_2	X_2	Y_n	Y	$Y - Y_n$
?	0	?	1	Y_1	2	$2 - Y_1$
	1		2	Y_2	6	$6 - Y_2$
	1		1	Y_3	4	$4 - Y_3$
	1.5		1	Y_4	5	$5 - Y_4$
	AVG($Y - Y_n$)					

*Source : 진상형, Deep Learning: Introduction, 2017.9.30.

Parametric Approach

$$Y_n = 1 \cdot X_1 + 2 \cdot X_2$$

W_1	X_1	W_2	X_2	Y_n	Y	$Y - Y_n$
1	0	2	1	2	2	0
	1		2	5	6	-1
	1		1	3	4	-1
	1.5		1	3.5	5	-1.5
					AVG($Y - Y_n$)	-0.875

Parametric Approach

$$Y_n = 2 \cdot X_1 + 2 \cdot X_2$$

W_1	X_1	W_2	X_2	Y_n	Y	$Y - Y_n$
2	0	2	1	2	2	0
	1		2	6	6	0
	1		1	4	4	0
	1.5		1	5	5	0
					AVG($Y - Y_n$)	0

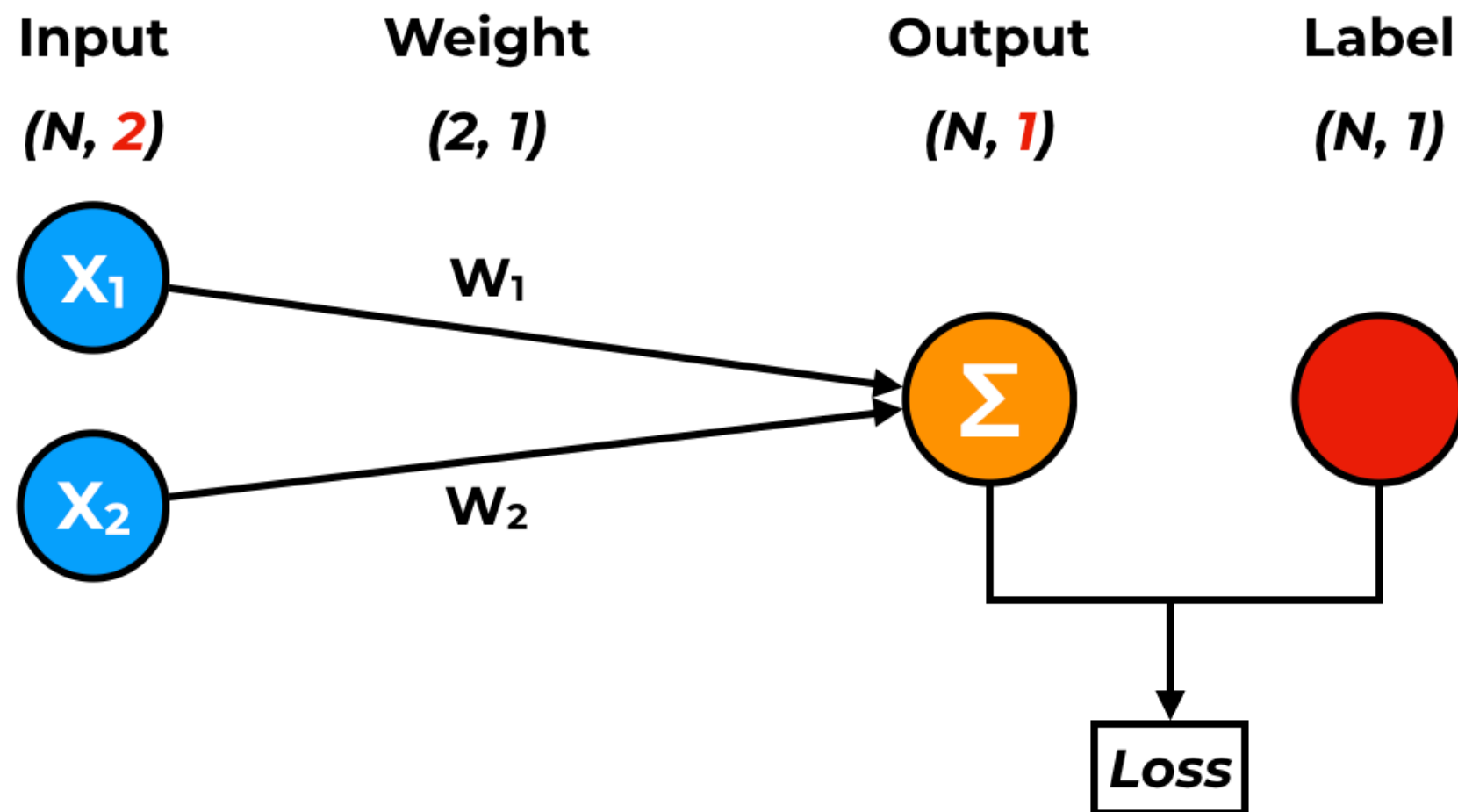
Parametric Approach

$$\text{Pred.} = 2 \cdot X_1 + 2 \cdot X_2$$

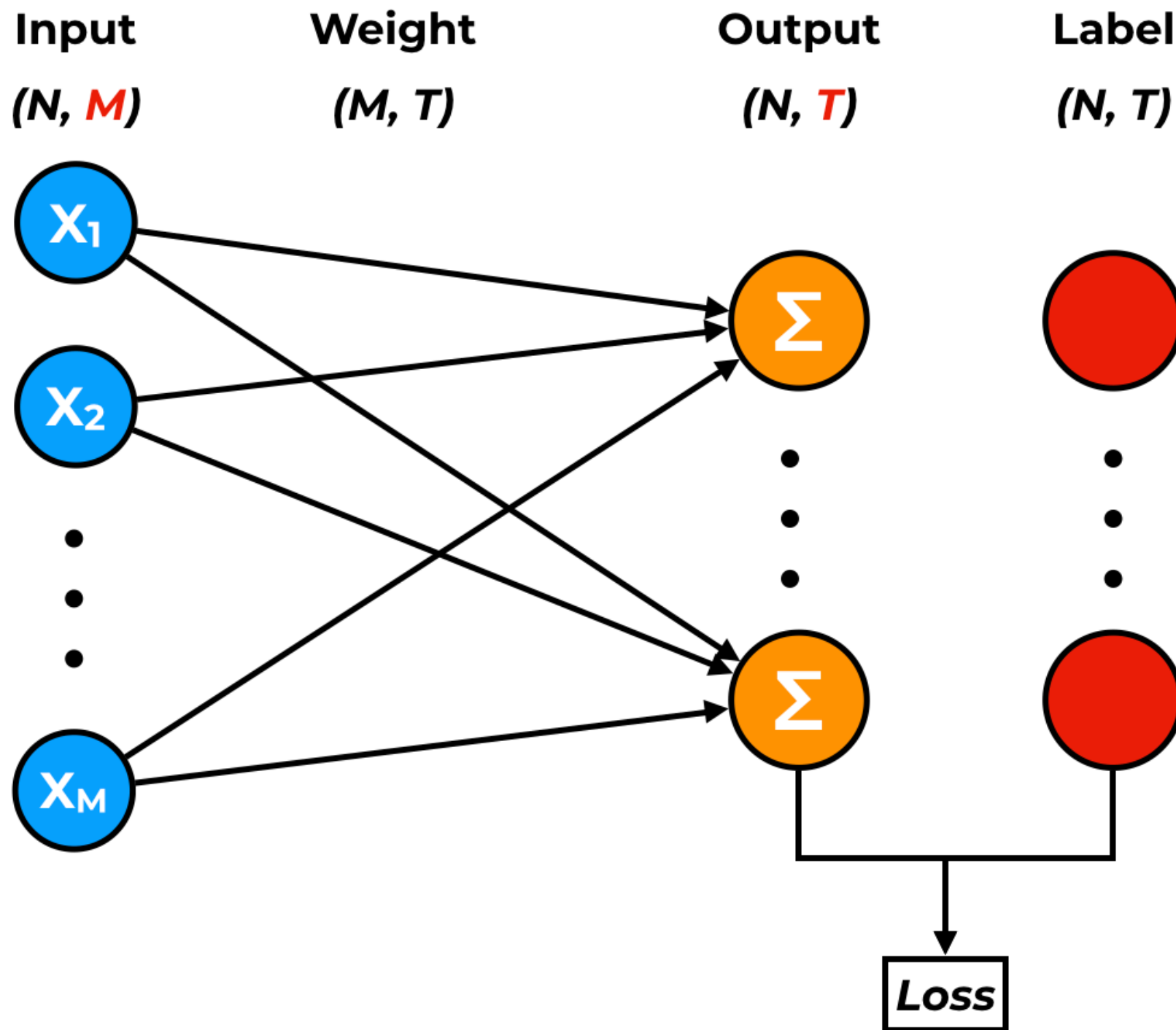
W_1	X_1	W_2	X_2	Pred.	Label	Error
2	0	2	1	2	2	0
	1		2	6	6	0
	1		1	4	4	0
	1.5		1	5	5	0
					Loss	0

Artificial Neural Network

$$Y_n = W_1 \cdot X_1 + W_2 \cdot X_2$$



Artificial Neural Network



E.O.D