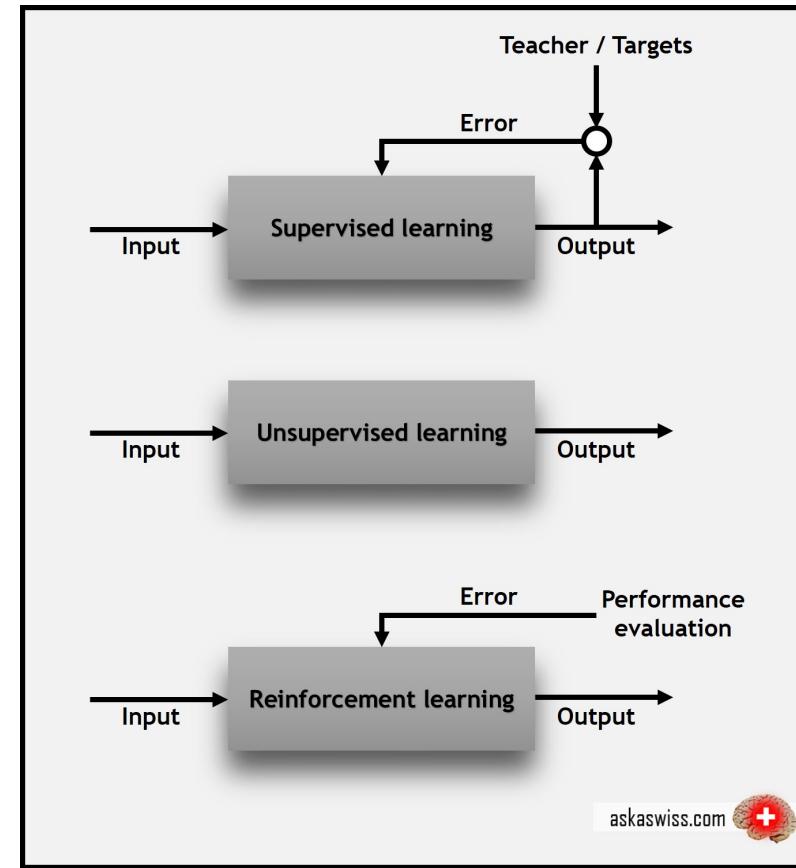
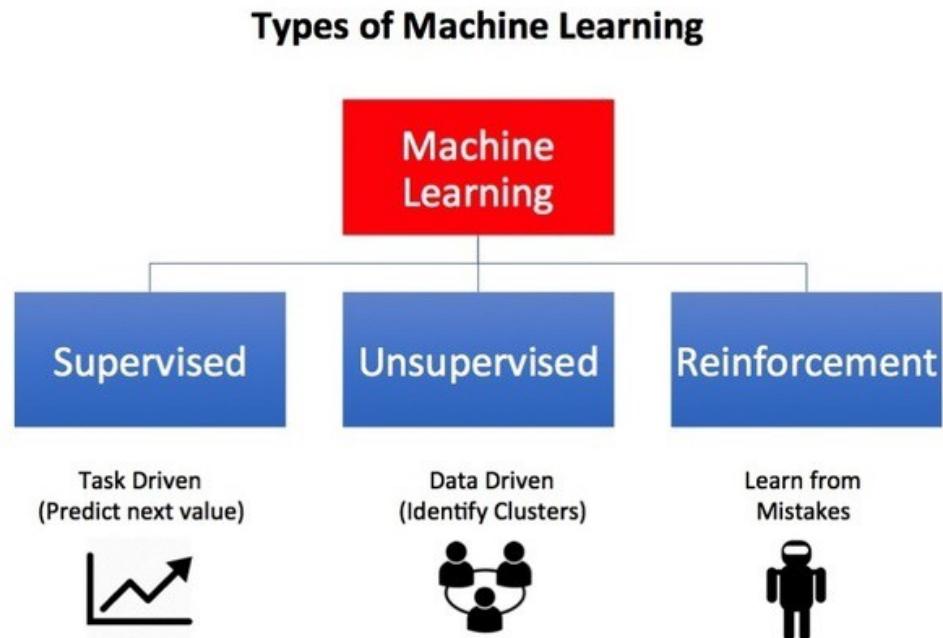


I. 금융 빅데이터와 데이터 분석 (Big data in finance & financial time series data analysis)

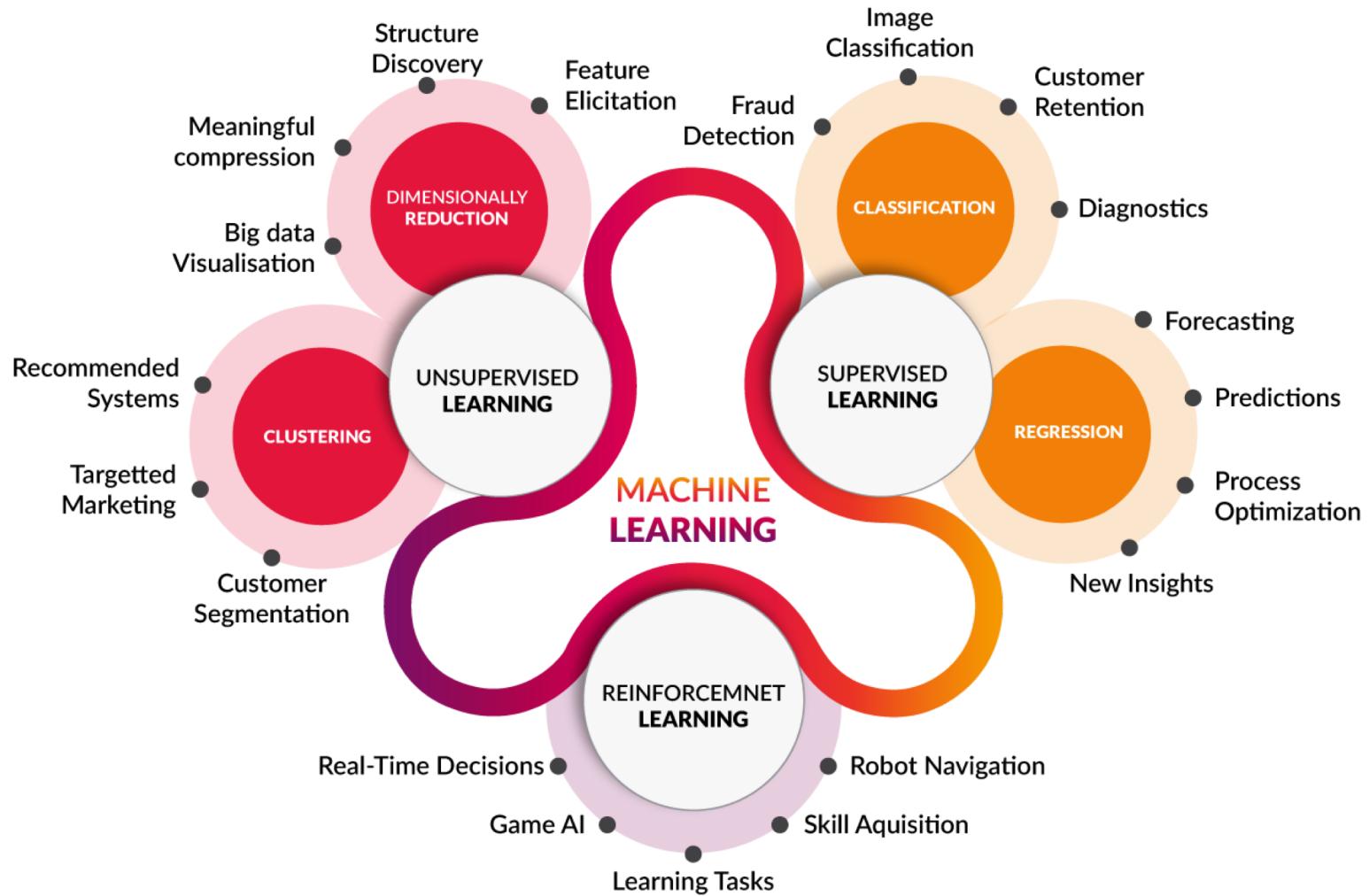
⑤ 머신 러닝의 유형

머신 러닝의 유형(데이터 학습)

- 지도 학습 (Supervised Learning)
- 비지도학습 (Unsupervised Learning)
- 강화학습 (Reinforcement Learning)



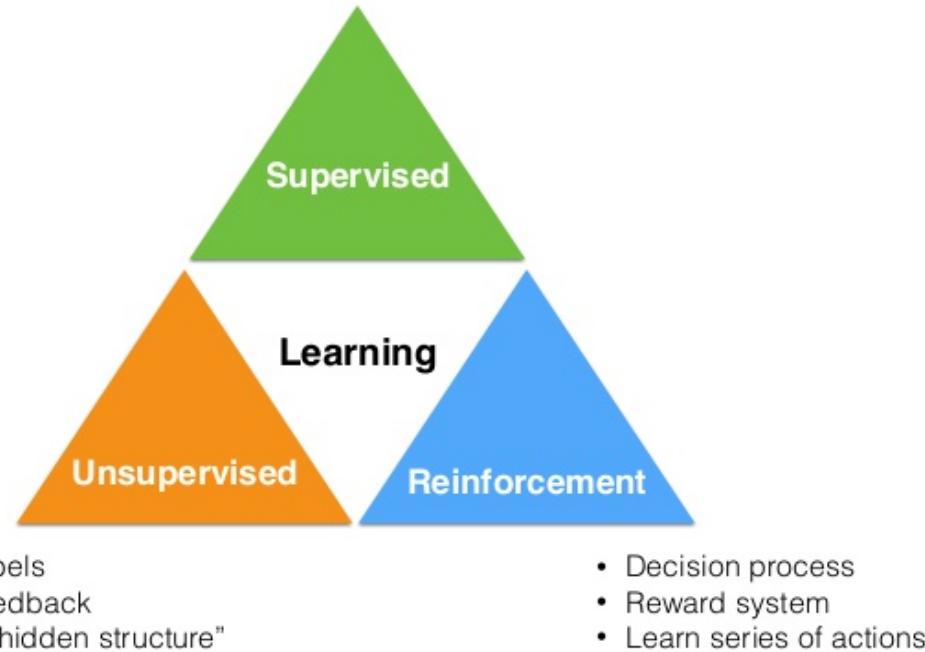
머신 러닝의 유형(데이터 학습)



머신 러닝의 유형(데이터 학습)

- **Supervised Learning**은 지도학습으로서 "정답"을 알 수 있어서 바로바로 피드백을 받으면서 학습하는 것
- **Unsupervised Learning**은 비지도 학습으로서 정답이 없는 "분류"와 같은 문제를 푸는 것
- **Reinforcement Learning**은 강화학습으로서 정답은 모르지만, 자신이 한 행동에 대한 "보상"을 알 수 있어서 그로부터 학습하는 것

- Labeled data
- Direct feedback
- Predict outcome/future



Machine Learning Algorithms *(sample)*

	<u>Unsupervised</u>	<u>Supervised</u>
Continuous	<ul style="list-style-type: none">• Clustering & Dimensionality Reduction<ul style="list-style-type: none">◦ SVD◦ PCA◦ K-means	<ul style="list-style-type: none">• Regression<ul style="list-style-type: none">◦ Linear◦ Polynomial• Decision Trees• Random Forests
Categorical	<ul style="list-style-type: none">• Association Analysis<ul style="list-style-type: none">◦ Apriori◦ FP-Growth• Hidden Markov Model	<ul style="list-style-type: none">• Classification<ul style="list-style-type: none">◦ KNN◦ Trees◦ Logistic Regression◦ Naive-Bayes◦ SVM

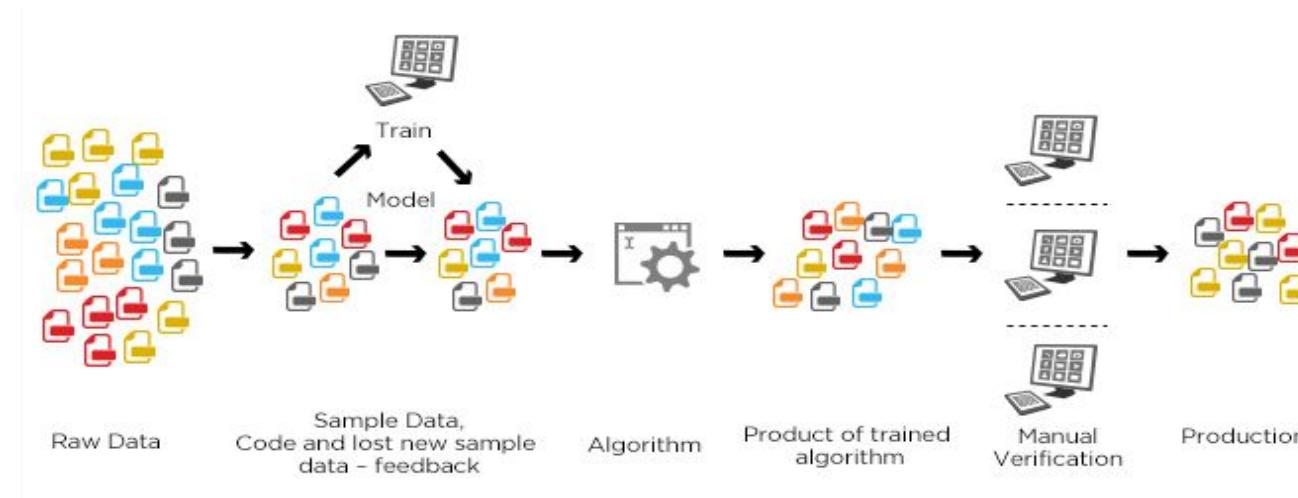
지도학습과 비지도학습

- 궁극적인 목표 : 모두 과거, 현재의 데이터로부터 미래를 예측하는 것
- 두 가지 방법의 차이점은 라벨링 된 데이터인지 아닌지임
- 만일 내가 가지고 있는 데이터가 라벨링 되어 있다면 지도학습, 그렇지 않다면 비지도학습

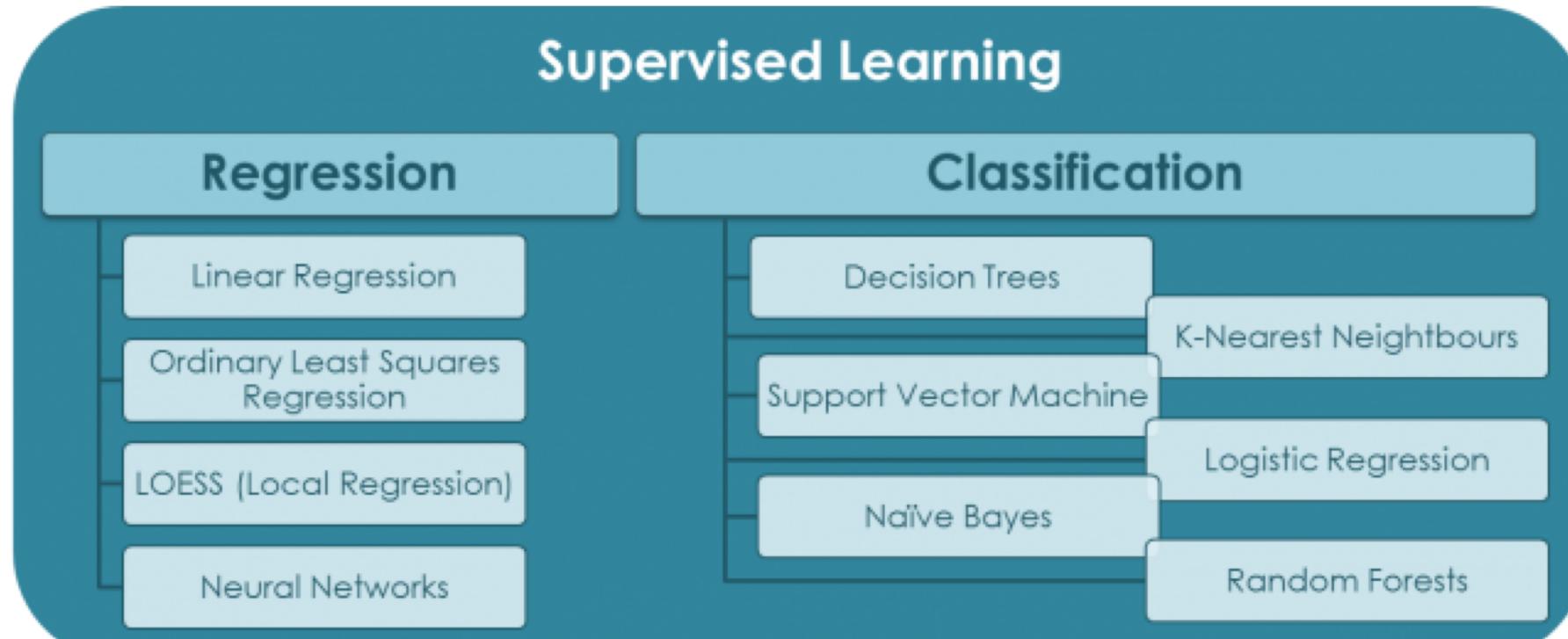
* 라벨링 된 데이터(Labeled data)란 데이터에 대한 답이 주어져 있는 것 (평가가 되어 있는 것) 을 말함

지도학습 (Supervised Learning)

- 훈련 데이터에 결과가 미리 주어진 상태에서 학습의 결과가 주어진 값 (목표 값)이 나오도록 학습 시키는 방법. 학습이 완료되면 결과를 알려 주지 않아도 스스로 추론한 결과를 제시함. 주어진 입력 값이 어느 부류에 속하는지 (**Classification**), 입력 값으로 예측이나 추정 (**Regression**)하는데 주로 사용됨.
- 훈련 데이터에는 결과 값이 같이 주어짐. 학습 과정에서 학습 결과와 주어진 결과와의 오차가 최소가 되도록 반복 학습함. 학습 결과를 이용하여 결과가 주어지지 않은 시험 데이터의 결과를 추정함



지도학습 (Supervised Learning)의 유형

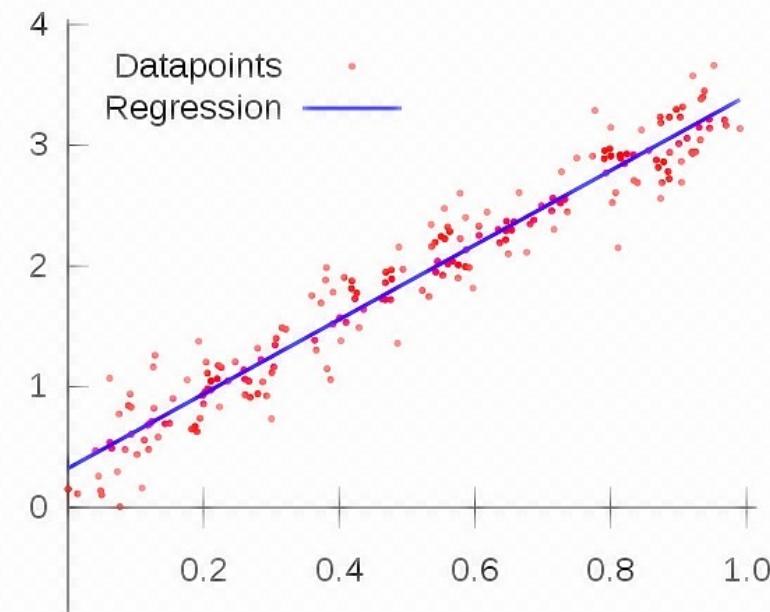


지도학습 (Supervised Learning)의 유형

1) Regression

- 연속된 값을 예측하는 문제
- 종속 변수 [*Output*]와 독립 변수 [*Input*]사이의 관계를 조사하는 예측 모델링 기술의 한 형태
- 선형 그래프
- 날씨, 시계열 모델링, 프로세스 최적화를 예측하는 데 사용됨
- 예 : 주택 가격 예측. 객실 수, 지역성, 교통 편리 성, 집의 나이, 집의 면적 예측, 공부시간에 따른 전공 시험 점수를 예측하는 문제

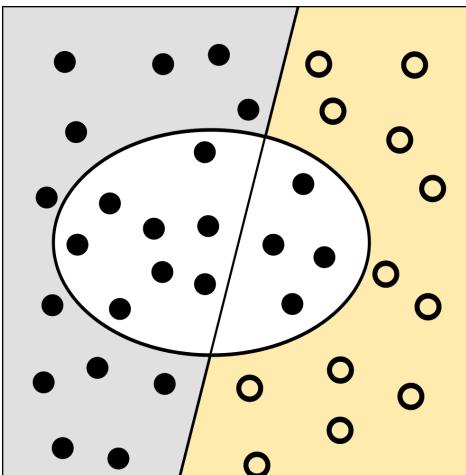
- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Ridge Regression
- Lasso Regression



지도학습 (Supervised Learning)의 유형

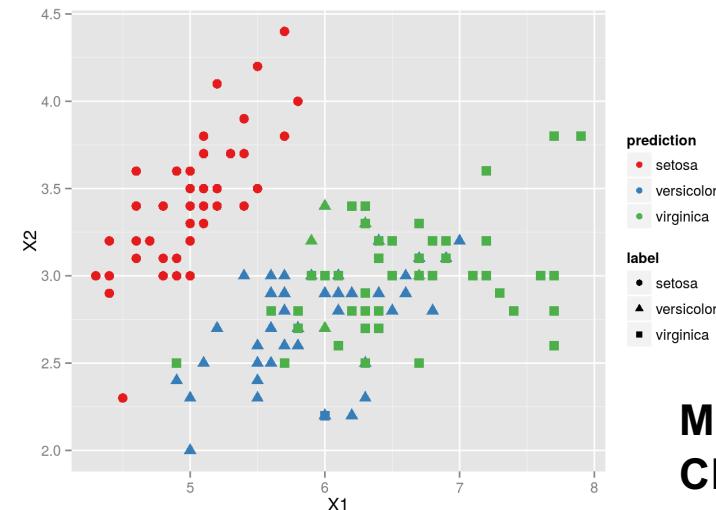
2) Classification

- 주어진 데이터를 정해진 카테고리에 따라 분류하는 문제
- 맞다/아니다로 구분되는 문제 : **Binary Classification**
- A / B / C / D / F 으로 예측하는 문제 : **Multi-label Classification**



Binary Classification

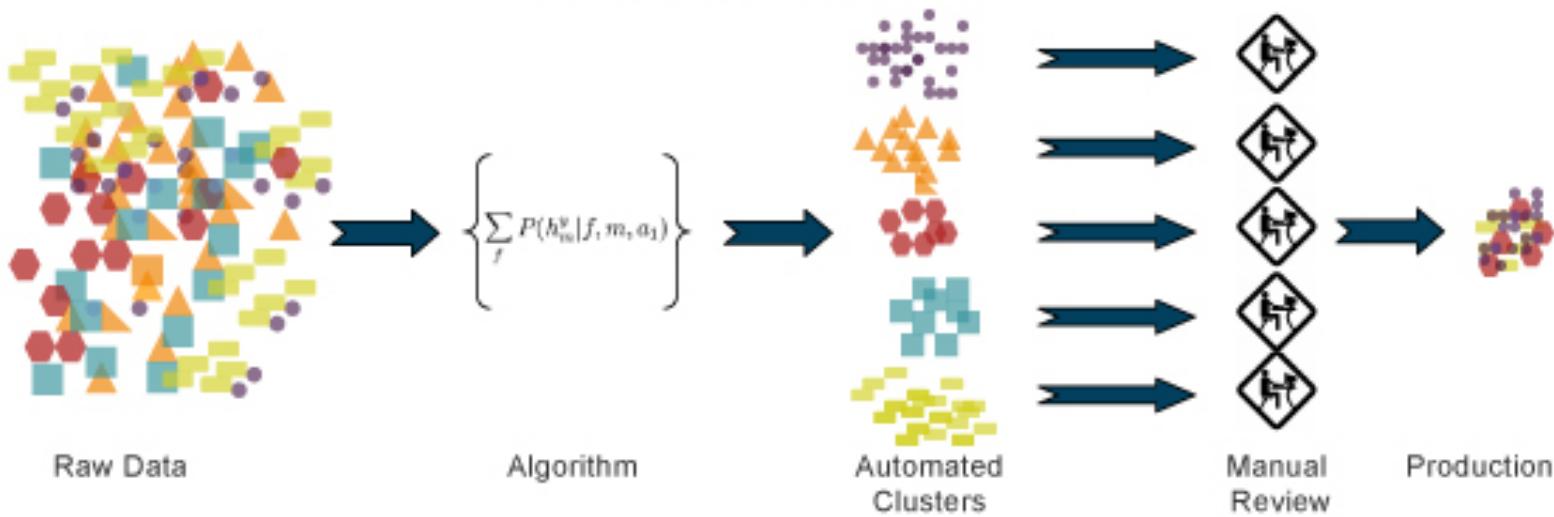
- Logistic Regression/Classification
- K-Nearest Neighbours
- Support Vector Machines
- Kernel Support Vector Machines
- Naive Bayes
- Decision Tree Classification
- Random Forest Classification



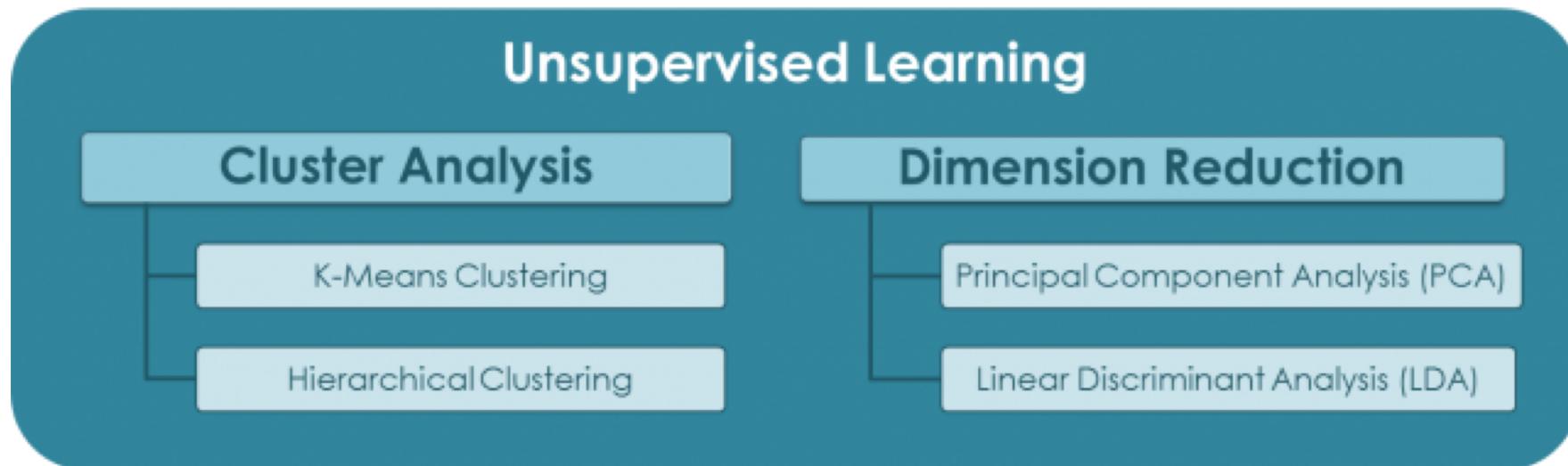
Multi-label Classification

비지도학습 (Unsupervised Learning)

- 훈련 데이터에 결과가 주어지지 않은 상태에서 자율적으로 학습함.
- 훈련 데이터의 패턴을 관찰하여 스스로 의미를 찾아냄. 서로 유사한 패턴끼리 묶거나 (**Clustering**), 연관 규칙 (**Association**) 등을 찾아냄.
- 예측이나 추정 문제보다는 데이터가 어떻게 구성되어 있는지 밝히는데 주로 사용함
 - 데이터 요약, 정리



비 지도학습 (Unsupervised Learning)의 유형

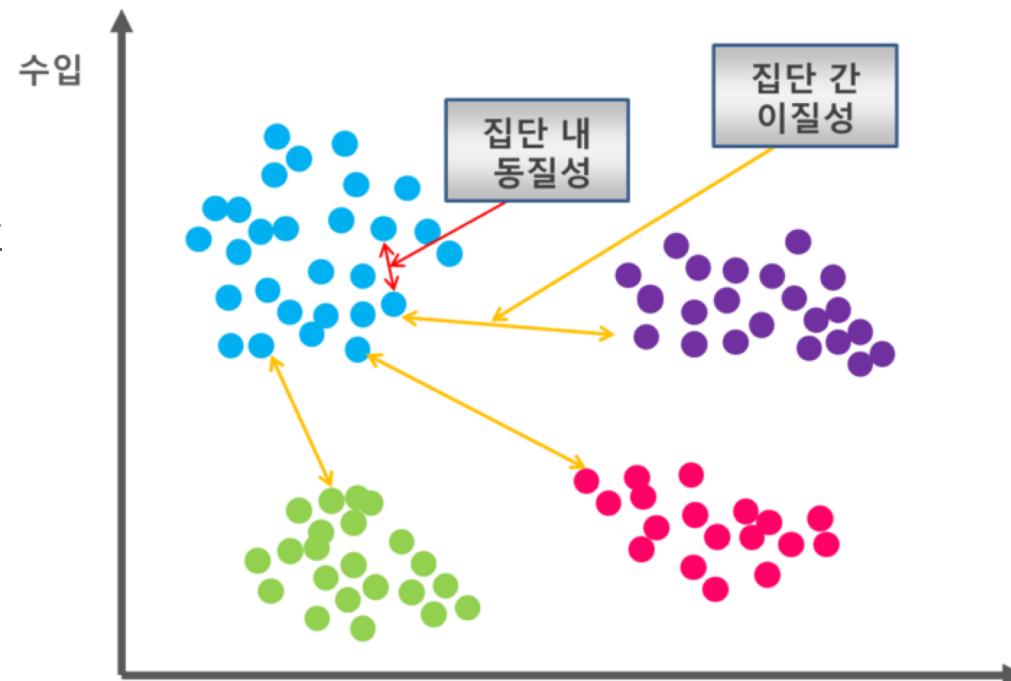


비 지도학습 (Unsupervised Learning)의 유형

1) Clustering

- 유사한 엔티티를 함께 그룹화하는 프로세스
- 그룹화 된 데이터는 클러스터를 만드는데 사용됨
- 데이터 포인트에서 유사점을 찾아 비슷한 데이터 포인트를 그룹화하고 새로운 데이터가 어떤 클러스터에 속해야 하는지를 파악

- K-Means Clustering
- Hierarchical Clustering



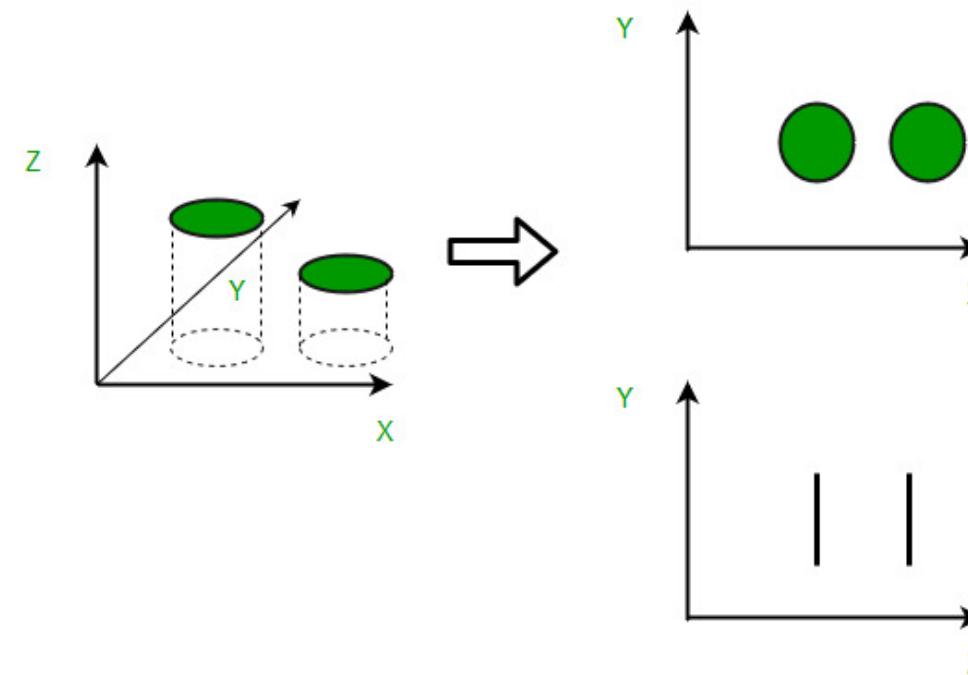
비 지도학습 (Unsupervised Learning)의 유형

2) Dimensionality Reduction

- 입력 데이터에서 원하지 않는 데이터를 제거하기 위해 데이터의 크기를 줄임
- 데이터의 바람직하지 않은 features을 제거하는 데 사용
- 큰 차수를 갖는 데이터 세트를 동일한 데이터 및 작은 크기를 갖는 데이터로 변환하는 프로세스

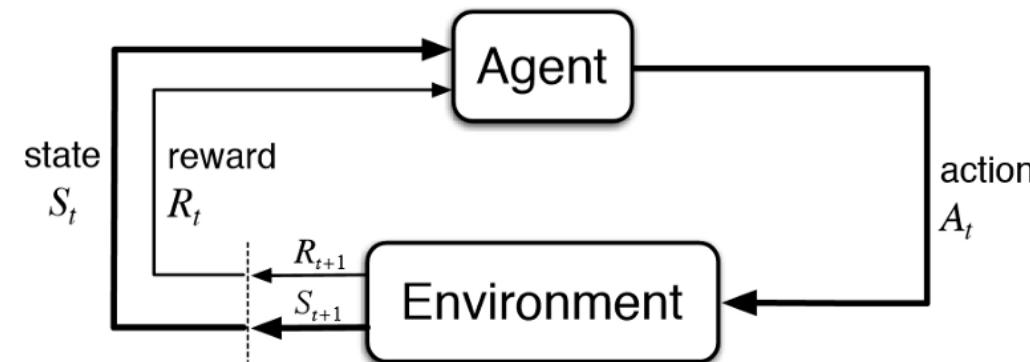
- Principal Component Analysis
- Linear Discriminant Analysis
- Kernel Principal Component Analysis

Dimensionality Reduction



강화학습(Reinforcement Learning)

- 과거 데이터를 학습하는 것이 아니라 **현재 상태에서 시행착오 혹은 시뮬레이션에 의해 최적의 행동을 결정함.** 시행착오 경험에 의해 행동을 점차 개선시킴.
- 현재의 데이터 (혹은 Experience)를 학습하여 어떤 Action을 취하고 결과에 대해 보상 (Reward)을 받음.
- 보상은 결과에 따라 양수 (+ : Reward) 일 수도 있고 음수 (- : Punishment)일 수도 있음.
- 강화 학습은 **보상이 최대가 되는 Action을 찾는 방법**임.
 - 아이가 걷는 것을 배우는 것처럼 어떻게 행동할 줄 모르지만 환경과 상호작용하면서 걷는 법을 알아가는 것과 같은 학습 방법



강화학습(Reinforcement Learning)의 특징

1. Trial and Error

- 해보지 않고 예측하고 움직이는 것이 아니고 해보면서 자신을 조정해 나가는 것

2. Delayed Reward

- 강화학습은 시간의 순서가 있는 문제를 풀기 때문에 지금 한 행동으로 인한 환경의 반응이 늦어질 수가 있다. (혹은 다른 행동과 합해져서 더 좋은 환경의 반응을 받아낼 수도 있음)
- 이럴 경우에 환경이 반응할 때까지 여러 가지 다른 행동들을 시간의 순서대로 했기 때문에 어떤 행동이 좋은 행동이었는지 판단하기 어려운 점이 있다

강화학습(Reinforcement Learning)의 유형

1. Q-Learning
2. SARSA [State Action Reward State Action]
3. Deep Q-Network
4. Markov Decision Processes
5. DDPG[Deep Deterministic Policy Gradient]

반 지도 학습(Semi-Supervised Learning)

- 두 가지 유형의 raw data를 사용하는 네 번째 유형의 기계 학습
- 지도 및 비 지도 학습의 혼합
- 분류, 회귀 및 예측과 같은 방법과 함께 사용할 수 있음
- 지도 학습을 위한 엄청난 양의 데이터에 라벨을 지정하는 과정은 종종 시간이 많이 걸리고 비용이 많이 들게 되므로 이 반 지도 학습은 최종 모델의 정확성을 높이고 건물을 만드는 데 소요되는 시간과 비용을 줄이는 점에서 유용함

