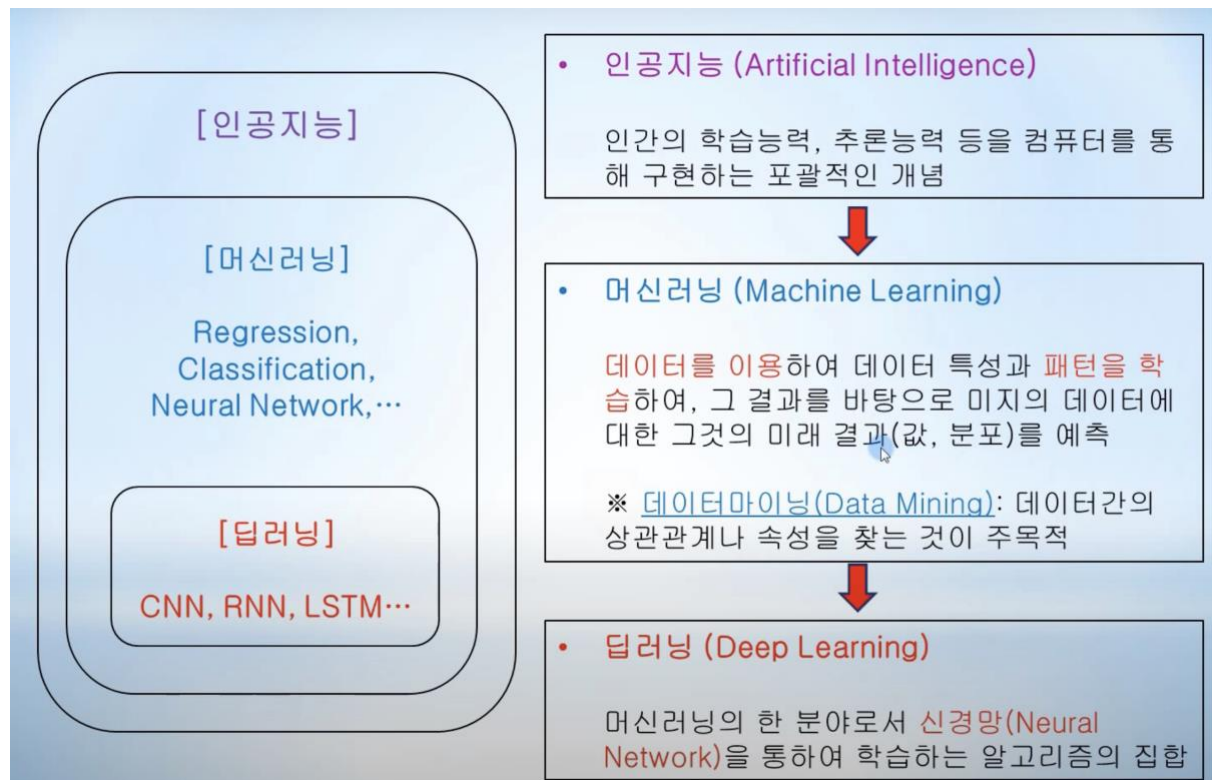
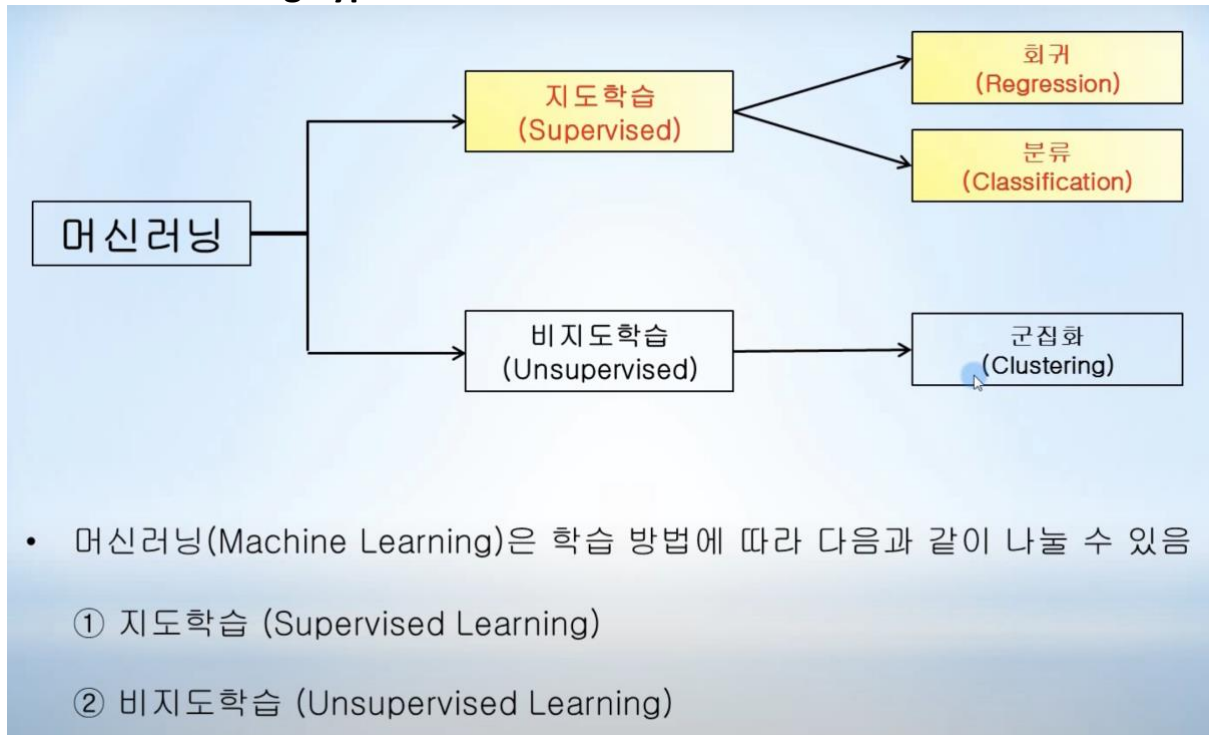


인공지능 머신러닝 딥러닝 - review



Machine Learning Type



지도학습(Supervised Learning):

학습할 데이터에 입력과 그 입력에 대응되는 정답을 이용해서 데이터의 특성과 분포를 학습하고, 미래 결과를 예측하는 방법

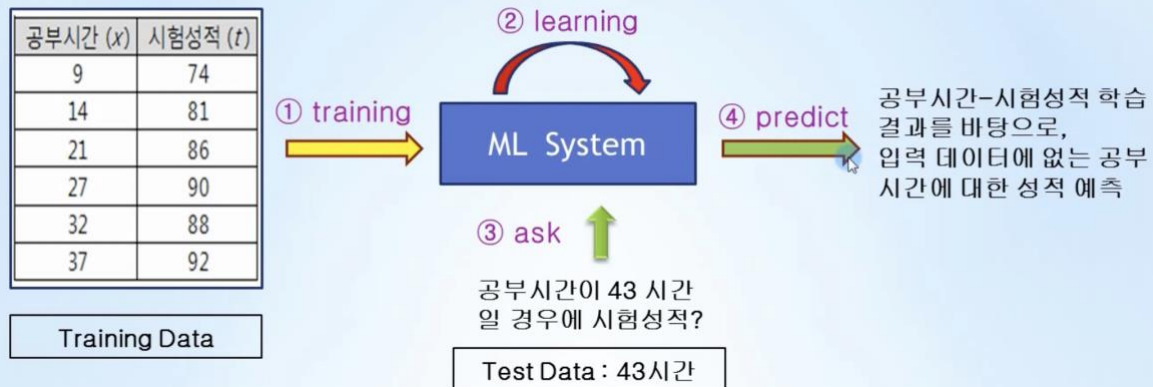
어떤 종류의 미래 값을 예측하는가에 따라 회귀(Regression)과 분류(Classification)로 나눔

비지도학습(Unsupervised Learning):

학습할 데이터에 정답은 없고 입력 값만 존재

정답이 없는 입력 데이터만을 이용하기 때문에 그러한 입력 값 자체의 특성과 분포만을 파악해서 Grouping하는 군집화에 주로 사용

Supervised Learning



지도학습(Supervised Learning)은 입력 값(x)과 정답(t , label)을 포함하는 Training Data를 이용하여 학습하고, 그 학습된 결과를 바탕으로 미지의 데이터(Test Data)에 대해 미래 값을 예측(predict) 하는 방법 \Rightarrow 대부분 머신러닝 문제는 지도학습에 해당됨

[예1] 시험공부 시간(입력)과 Pass/Fail (정답)을 이용하여 당락 여부 예측

[예2] 집 평수(입력)와 가격 데이터(정답) 이용하여 임의의 평수 가격 예측

Supervised Learning – Regression, Classification

- 지도학습은 학습결과를 바탕으로, 미래의 무엇을 예측하느냐에 따라 **회귀 (Regression)**, **분류(Classification)** 등으로 구분할 수 있음
 - 회귀(Regression)은 Training Data를 이용하여 **연속적인 (숫자) 값을 예측**하는 것을 말하며, 집평수와 가격 관계, 공부시간과 시험성적 등의 관계임
 - 분류(Classification)은 Training Data를 이용하여 **주어진 입력값이 어떤 종류의 값인지 구별**하는 것을 지칭함

Regression

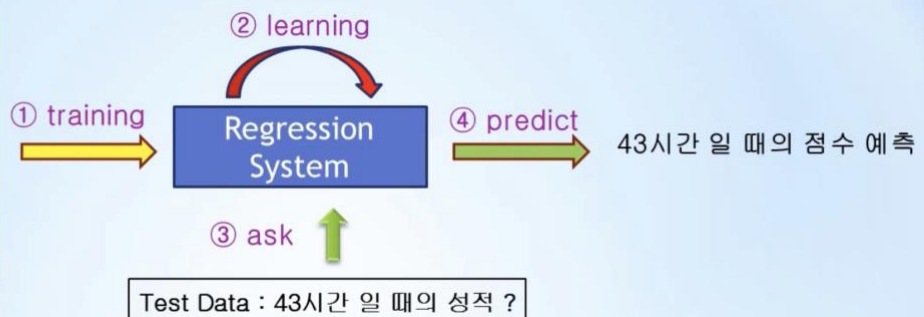
공부시간 (x)	시험성적 (t)	집평수 (x)	가격 (t)
9	74	20	98
14	81	25	119
21	86	30	131
27	90	40	133
32	88	50	140
37	92	55	196

Classification

공부시간 (x)	시험성적 (t)	집평수 (x)	가격 (t)
9	Fail	20	Low
14	Fail	25	Low
21	Pass	30	Medium
27	Pass	40	Medium
32	Pass	50	Medium
37	Pass	55	High

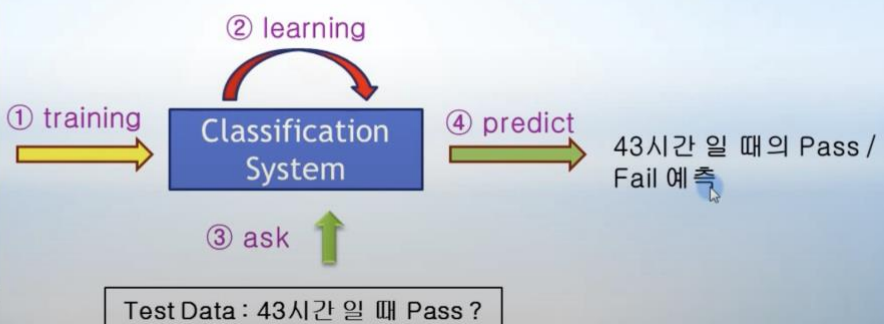
공부시간 (x)	시험성적 (t)
9	74
14	81
21	86
27	90
32	88
37	92

Training Data



공부시간 (x)	시험성적 (t)
9	Fail
14	Fail
21	Pass
27	Pass
32	Pass
37	Pass

Training Data

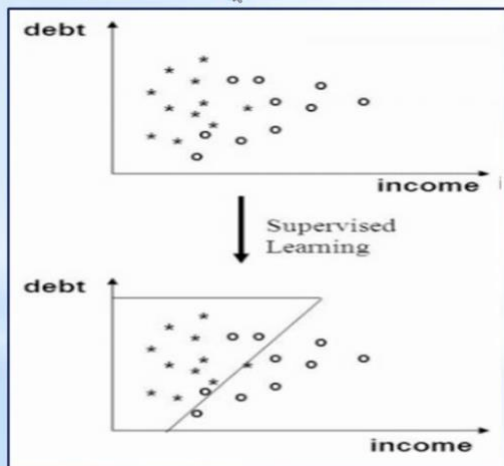


Unsupervised Learning

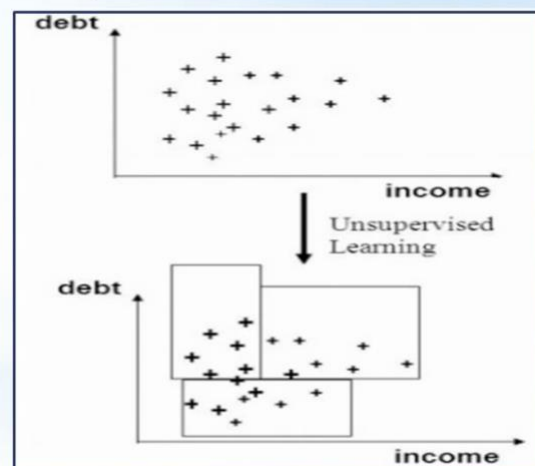
- 비지도학습(Unsupervised Learning)은 트레이닝 데이터에 정답은 없고 입력 데이터만 있기 때문에, 입력에 대한 정답을 찾는 것이 아닌 입력데이터의 패턴, 특성 등을 학습을 통해 발견하는 방법을 말함

[예] 군집화(Clustering) 알고리즘을 이용한 뉴스 그룹핑, 백화점의 상품 추천시스템 등

지도학습 분류 (Classification)



비지도학습 군집화 (Clustering)



- 지도학습에서의 분류는 Training Data의 입력에 대응되는 정답을 알고 있기 때문에, 이러한 정답을 가장 잘 나타낼 수 있는 임의의 직선을 찾아서 그 직선을 기준으로 아래와 위, 또는 좌, 우 등으로 분류가 가능
 - 비지도학습의 Training Data에는 정답이 없이 오로지 입력 데이터만이 존재. 즉, 위의 Training Data를 봤을 때, 입력 데이터에 해당하는 income에 대한 가격만 나타남을 알 수 있음
- 이처럼 비지도학습에서는 정답이 없이 입력 데이터만을 가지고 있기 때문에 이러한 입력 데이터의 특성과 분포를 위와 같이 주제별로 묶거나, 유사한 특성을 갖는 데이터들끼리 묶는 '군집화 알고리즘'을 사용
- 입력 데이터의 분포와 패턴을 이용해서 Grouping이나 추천시스템에 사용

