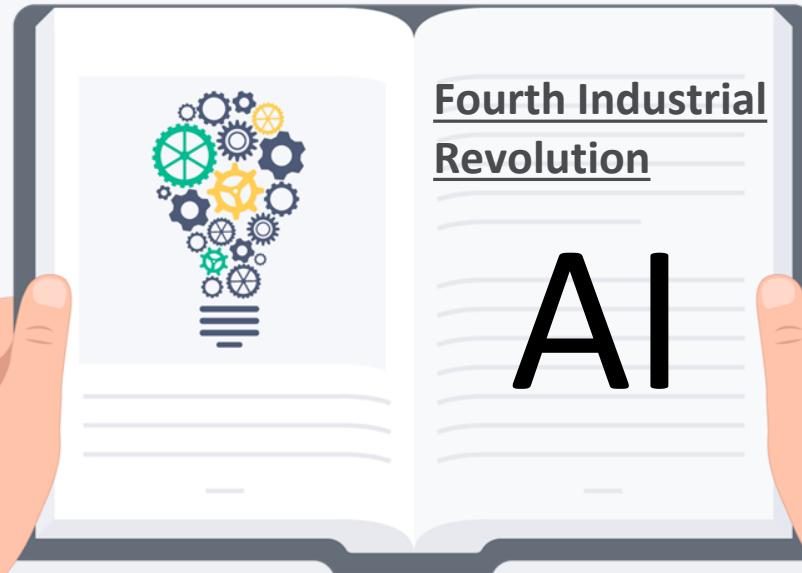


What is Artificial Intelligence(AI)

affective.AI Lab

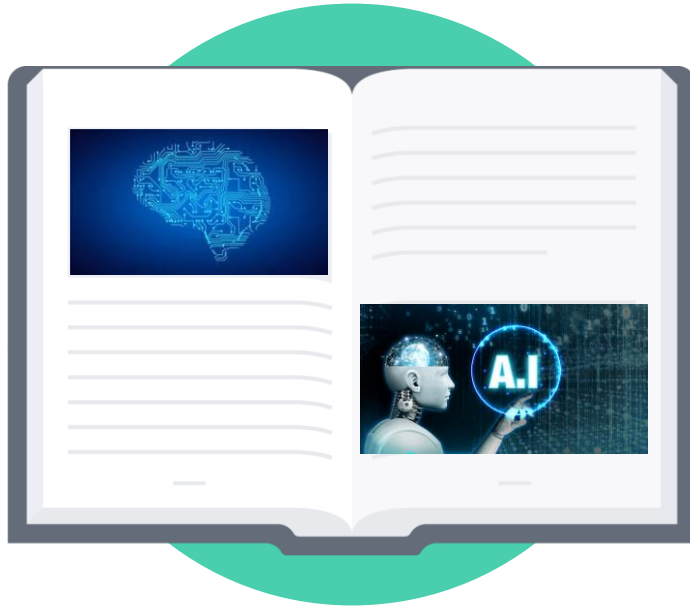


Department of Artificial Intelligence

Ji-Hun Yuk

Table of contents

목 차



1. Definition of AI

2. What is Machine Learning?

3. Weekly study

4. Summary

5. Q&A

★. Keyword

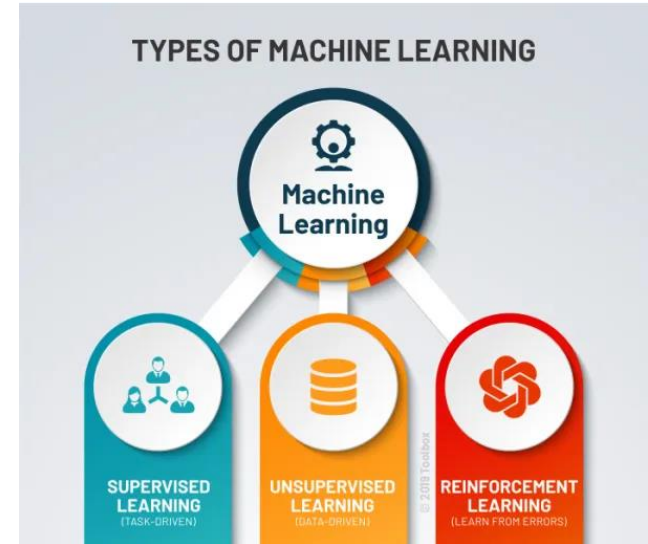
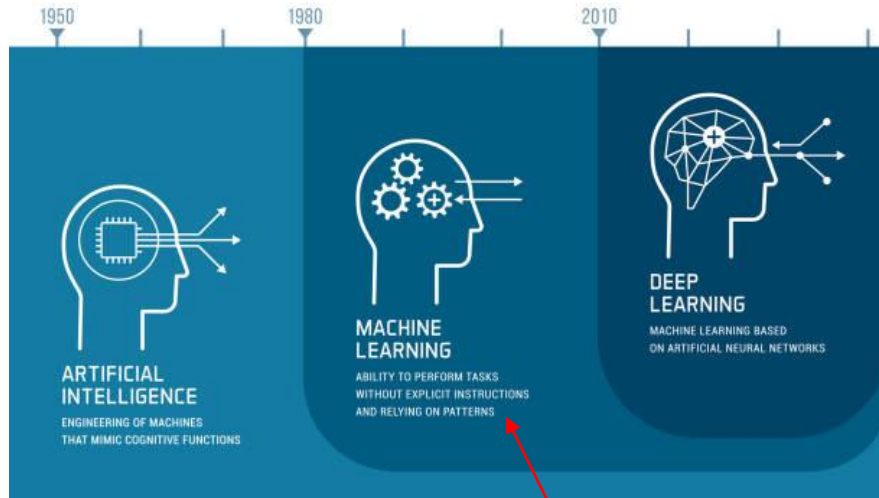
1.

What is AI?

Definition:

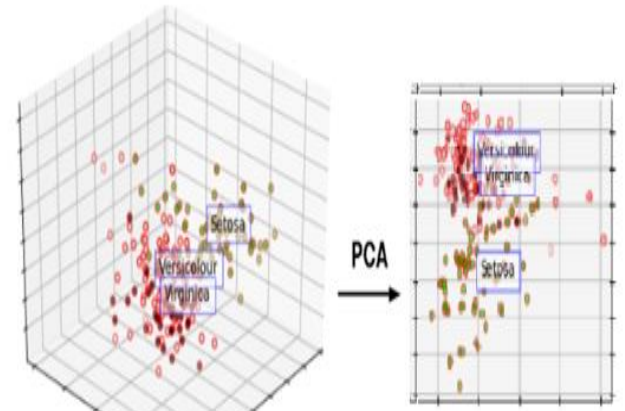
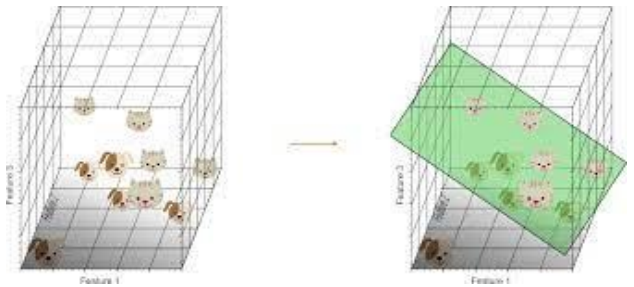
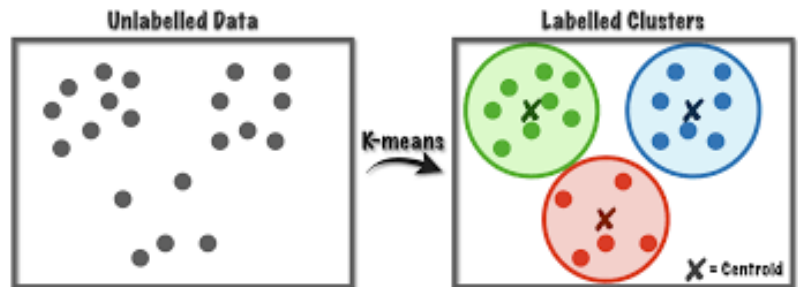
Artificial intelligence (AI) is intelligence—perceiving, synthesizing, and inferring information—demonstrated by machines, as opposed to intelligence displayed by humans or by other animals.

2. What is Machine Learning?



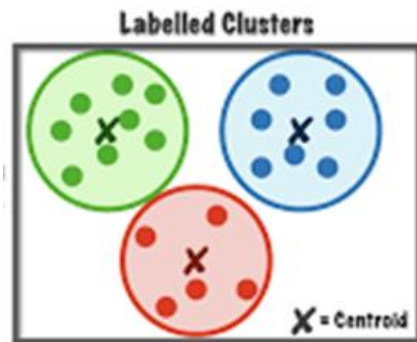
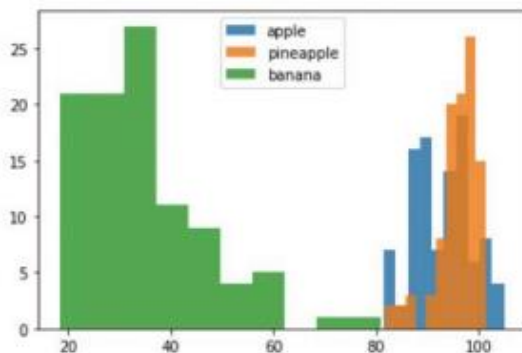
3-0. Weekly study

Ch6. 비지도 학습

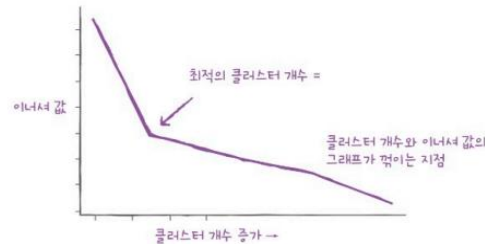


3-1. 비지도 학습

- 머신러닝의 한 종류로 훈련 데이터에 타깃이 없다. 타깃이 없기 때문에 외부의 도움 없이 스스로 유용한 무언가를 학습해야 한다. 대표적인 비지도 학습 작업은 군집, 차원 축소 등이 있다.
 - 히스토그램: 구간별로 값이 발생한 빈도를 그래프로 표시한 것이다. 보통 x축이 값의 구간(계급)이고 y축은 발생 빈도(도수)이다. (맷플롯립 함수, hist())
 - 군집: 비슷한 샘플끼리 하나의 그룹으로 모으는 대표적인 비지도 학습 작업이다. 군집 알고리즘으로 모은 샘플 그룹을 클러스터라고 부른다.



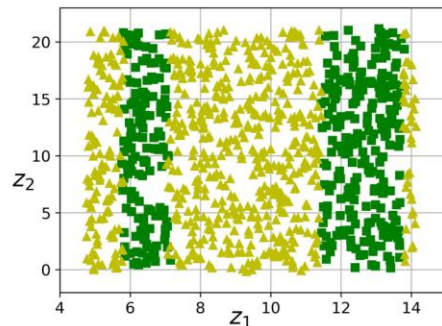
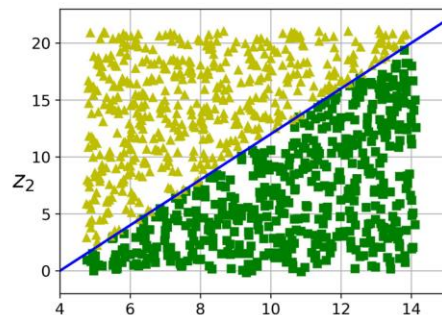
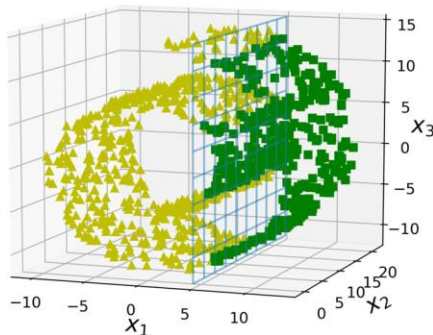
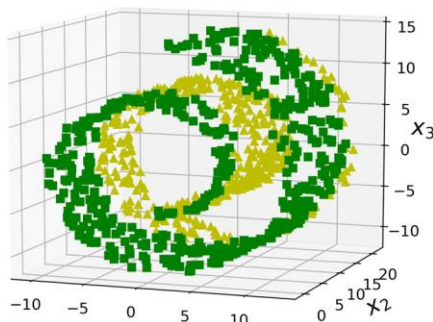
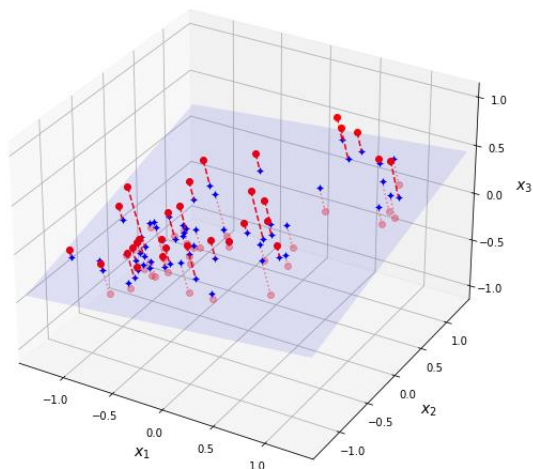
3-2.k-평균 알고리즘



- 처음에 랜덤하게 클러스터 중심을 정하고 클러스터를 만든다. 그 다음 클러스터의 중심을 이동하고 다시 클러스터를 만드는 식으로 반복해서(클러스터의 중심이 고정 될 때까지) 최적의 클러스터를 구성한다.
- 클러스터 중심: k-평균 알고리즘이 만든 클러스터에 속한 샘플의 특성 평균값이다. 센트로이드(centroid)라고도 부른다. 가장 가까운 클러스터 중심을 샘플의 또 다른 특성으로 사용하거나 새로운 샘플에 대한 예측으로 활용할 수 있다.
- 엘보우 방법: 최적의 클러스터 개수를 정하는 방법 중 하나다. 이너셔(inertia)는 클러스터 중심과 샘플 사이 거리의 제곱 합이다. 클러스터 개수에 따라 이너셔 감소가 꺾이는 지점이 적절한 클러스터 개수 k가 될 수 있다. 이 그래프의 모양을 따서 엘보우 방법이라고 부른다.

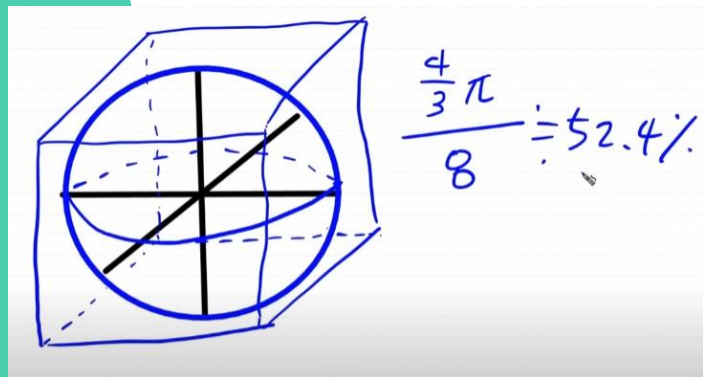
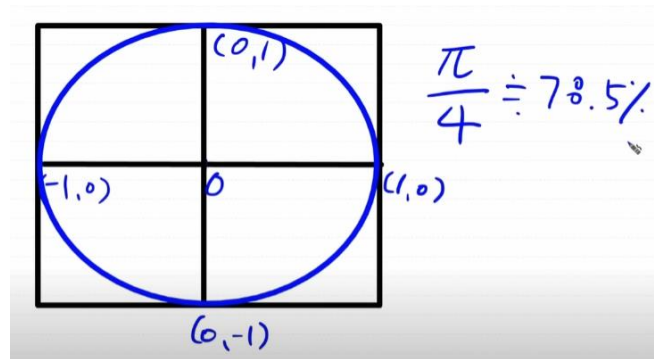
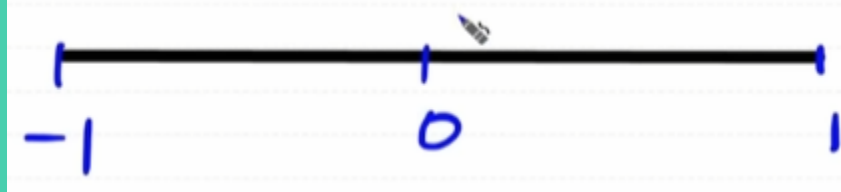
3-3. 차원 축소

- 원본 데이터의 특성을 적은 수의 새로운 특성으로 변환하는 비지도 학습의 한 종류. 차원 축소는 저장공간을 줄이고 시각화하기 쉽다. 또한 다른 알고리즘의 성능을 높일 수도 있다.



3-3. 차원의 저주

- 차원이 높을 때의 위험?



4차원?

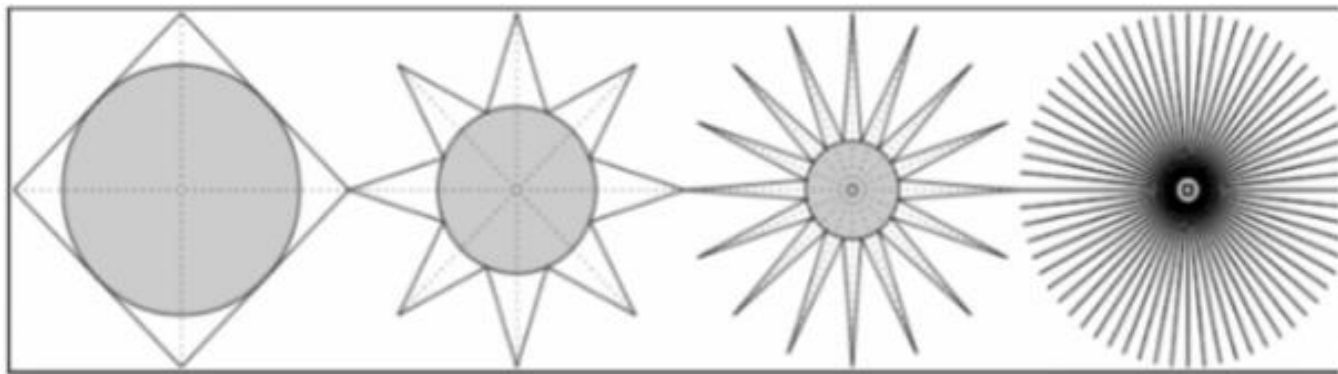


$$\left\{ (a,b,c,d) \mid \sqrt{a^2+b^2+c^2+d^2} \leq 1 \right\}$$

$$\frac{\pi^2}{2} \quad \frac{\pi^2}{16} \approx 30.8\%$$

3-3. 차원의 저주

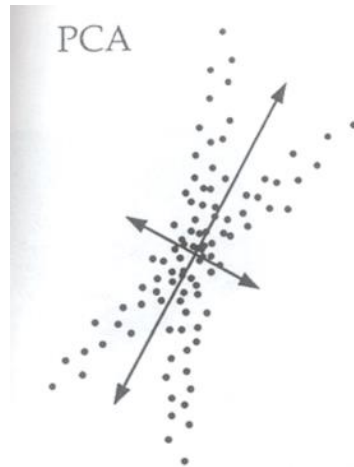
- n 이 무한히 커짐에 따라 해당 비율이 0으로 수렴.
- 거리 기반 유사도를 대표하기 어려움.
- 차원 축소의 필요성.



3-3. 주성분 분석(PCA)

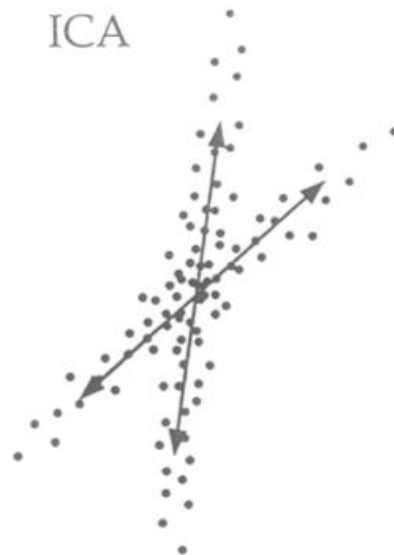
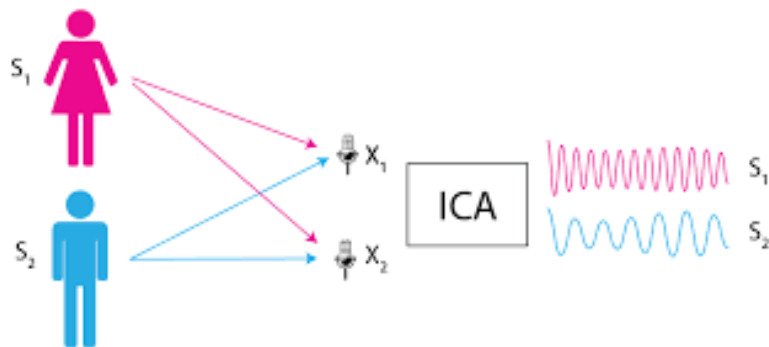
차원 축소 알고리즘의 하나로 데이터에서 가장 분산이 큰 방향을 찾는 방법이다. 이런 방향을 주성분이라 부르며, 원본 데이터를 주성분에 투영하여 새로운 특성을 만들 수 있다. 일반적으로 주성분은 원본 데이터에 있는 특성 개수보다 작다.

- 설명된 분산: 주성분 분석에서 주성분이 얼마나 원본 데이터의 분산을 잘 나타내는지 기록한 것. 사이킷런의 PCA 클래스는 주성분 개수나 설명된 분산의 비율을 지정하여 주성분 분석을 수행할 수 있다.



3-3. 독립 성분 분석(ICA)

차원 축소 알고리즘의 하나로 데이터에서 주성분을 이용한다는 점에서 PCA와 비슷하지만 데이터를 가장 잘 설명하는 축을 찾는 PCA와는 달리 가장 독립적인 축을 찾는 알고리즘이다.

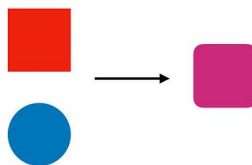


3-3. PCA vs ICA

- PCA는 고차원 데이터의 성질을 효율적으로 나타내는 목적으로 이용.
- ICA는 혼합 데이터를 독립한 성분으로 분리하는 목적으로 이용.
- <https://wordbe.tistory.com/entry/Linear-Factor-Model-PCA-ICA-Sparse-coding-%EC%84%A4%EB%AA%85>

PCA

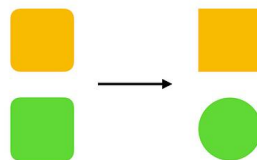
Compresses information



Requires preprocessing: autoscaling

ICA

Separates information



Requires preprocessing: autoscaling

Often benefits from first applying PCA

4.

Summary



5.

Q&A



Thanks for listening

