



#2 - Categories of ML Problem


주제 : Categories of ML Problem

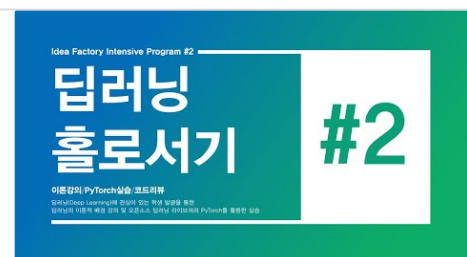
링크 :

[#2.Lec] ML Basic - 딥러닝 홀로서기 (Eng Sub)

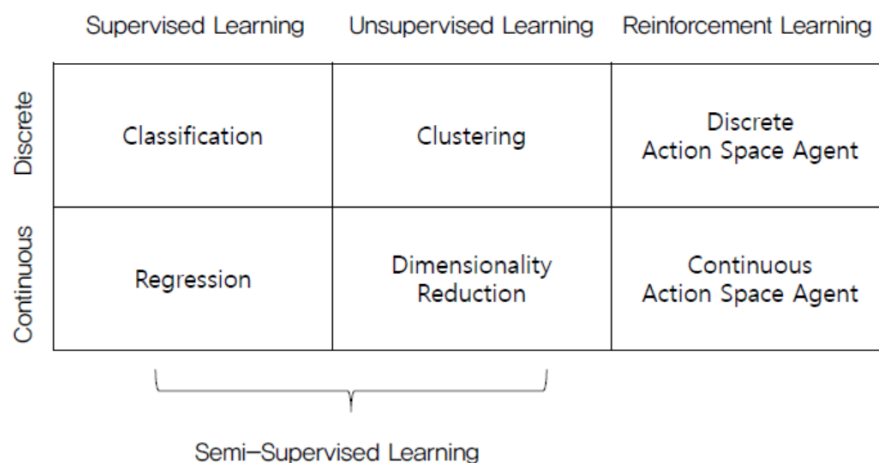
Slides Link : <https://github.com/heartcored98/Standalone-DeepLearning/blob/master/Lec1/Lec1-B.pdf>

Code Link : <https://github.com/heartcored98/Standalone-DeepL...>

 <https://www.youtube.com/watch?v=hPXeVHdldmw&feature=youtu.be>



1. Categories of ML Problems



- Reinforcement Learning

Input의 상태에 따라 Action을 취하고 상태를 확인한 후 Improve되었다면 보상 후 다시금 Action 반복하며 최적화된 Action을 찾아가는 것

- Semi-Supervised Learning

Supervised와 Unsupervised Learning의 섞은 방법론

Ex) 노래를 분류할 경우, 미리 라벨링을 준 100개의 노래셋을($X \rightarrow Y$) 이용하여 학습을 한 후 나머지 X 를 자동 라벨링하는 방식

2. Regression Problem

- Continuous Output을 도출
- 함수 $F(x)$ 가 가장 쉬운 예시 (X 를 통해 Y 를 찾아내는 과정)
- Linear Regression / Nonlinear regression으로 나눌 수 있다

3. Classification

- Discrete Output을 도출
- X 를 통해서 Y 를 분류

4. Clustering

- 여러가지 Instance(=Sample)간의 Distance를 정의하여 비슷한 Instance끼리 그룹화
 - ※ Feature Selection : 유의미한 피쳐들을 선택하는 것
 - ※ Feature Extraction : 유의미한 피쳐들을 선택하여 모델에 데이터로 주입하는 것

5. Dimensionality Reduction

- 데이터의 손실을 최소화한 상태로 고차원의 데이터를 저차원으로 변형하는 과정
- PCA 기법 : 데이터의 분산이 가장 커지는 축을 사용하여 차원을 축소하는 기법
- Dimensionality Reduction은 다른 다양한 알고리즘을 활용한다
 - ex) Case1 : 3차원의 X 피쳐들을 2차원의 Y 피쳐로 축소
 - ex) Case2 : 숫자 이미지의 차원 축소
 - 28 pixel x 28 pixel 단위의(255개의 값 per pixel) 숫자 이미지가 있다
(즉, 784개의 정보를 담은 이미지 = 784차원짜리 X 벡터)

- X벡터를 0~9까지의 숫자(10개의 클래스)에 해당하는 Discrete Space에 맵핑하고 싶다고 하자
- 기본적으로 거리를 잴 때 유클리디안 Distance 방식을 활용하는데, 이때, 각 숫자이미지들의 거리를 Direct로 계산하다보면 다른 숫자이미지끼리의 거리가 더 가깝게 측정되는 오류가 발생할 수 있다
- 따라서 Dimentionality Reduction 알고리즘을 사용하여 784차원의 X벡터를 3차원으로 변경할 수 있다면 ($x \rightarrow h$), 머신러닝 알고리즘으로 활용 가능하다