

4. 머신러닝 모델은 어떻게 나눌 수 있는가?

1) Experience 종류에 따른 분류

-> 머신러닝은 dataset이라는 experience를 통해서 성능을 높여가는 알고리즘이다.

- Unsupervised Learning
 - > Unsupervised learning algorithms experience a dataset containing many features, then learn useful properties of the structure of this dataset.
 - > Dataset만 주어진다.
 - > $p(x)$ 를 배운다. (데이터셋의 분포를 학습한다)
- 예시) PCA, Kmeans Clustering,
- Supervised Learning
 - > Supervised learning algorithms experience a dataset containing features, but each example is associated with a label or target
 - > Dataset과 함께 라벨까지 주어진다. |
 - > $P(y | x)$ 를 배운다. (learning to predict y given x)

예시) knn, linear regression, logistic regression, support vector machine, decision tree

- Reinforcement Learning
 - > Some machine learning algorithms do not just experience a fixed dataset. For example, reinforcement learning algorithms interact with an environment. so there is a feedback loop between the learning system and its experience.
 - > Agent가 환경과 상호작용을 통해서 학습한다.
 - > Agent는 cumulative reward(미래)를 극대화 하기 위해서 더 좋은 정책을 학습한다.

2) Supervised Learning - 학습 방법에 따른 분류

-> 지도학습의 단계는 Inference와 decision으로 이루어진다.

-> 결국 Supervised Learning은 Posterior을 구해서 판단하게 된다.

-> 다음 세 가지 방법론이 있다.

- Generative Model
 - Generative Model은 다음 순서로 학습한다.
 - 1) Infer the class-conditional density $P(x | C_k)$ for each class C_k
 - 2) Infer the prior class probability $P(C_k)$
 - 3) User Bayes' theorem to find posterior probability $P(C_k | x)$
 - > Then apply decision theory to determine class membership.
- Discriminative Model
 - First solve the inference problem to determine posterior probabilities $p(C_k | x)$
 - > Then apply decision theory to determine class membership
- Discriminant Function
 - Find a function $f(x)$ that maps each input x directly onto a class label
 - > 확률에 대한 고려 없이 mapping 하는 Function이다.

-> 지도 학습은 추정하고자 하는 y 값이 categorical인지 numerical인지에 따라서 분류와 회귀로 나누어진다.

이제

- 1) 통계와 머신러닝에서는 보통 모수 theta를 도입해서 특정 모수를 가정하는 모델(parametic method)로 추정을 시도한다.
- 2) theta에 대해서는 점추정을 할 수도 있고 분포 추정을 할 수도 있다.
 - > Frequentist: 점추정 + 구간추정 / Bayesian: 분포추정
- 3) machine learning의 개괄적인 분류
 - > Unsupervised, Supervised, Reinforcement Learning

그렇다면 이렇게 많은 모델들 중에서 어떻게 모델을 선택해야 하는가?