



# HPO 1부) 통계와 머신러닝에서의 최적화

## 최적화 정의

: 주어진 성능 지수 함수를 고려하는 집합 안에서 제약 조건들을 만족시키며 최소화(또는 최대화)하는 해를 찾아내는 문제에 관한 학문 영역

$$\min_x J(x) \\ \text{subject to } x \in \Omega$$

- $J(x)$ : 최소화하고자하는 성능 지수 함수
  - 시스템 식별 문제에서는 추정 오차의 크기
  - 금융 공학에서는 포트폴리오의 위험도
  - 최적 제어 문제에서는 변수의 총 변화량
- $x$ : 최적화 변수 값이 변화함에 따라 성능함수 값이 변화  
변수의 수학적 특성(연속 변수 또는 이산 변수, 실수 또는 정수 변수 등에 따라 최적화 문제 특성 자체가 많이 변화)
- $\Omega$ : 최적화 변수가 속하는 집합 ; 등식 또는 부등식 제약 조건을 만족하는 최적화 변수의 집합

이론적 관점에서는 성능함수, 최적화 변수, 가능해 집합  $\Omega$ 의 특성에 따라 주어진 최적화 문제를 푸는 방법이 크게 달라짐

## ML 관점에서의 파라미터/하이퍼파라미터 최적화:

- 기계 학습에서 **파라미터 최적화**: 모델의 **예측 정확도**를 높이는 데 중요한 역할

- 학습 데이터를 사용하여 모델의 파라미터 값을 조정하고, 모델의 성능을 개선하는 것을 의미
- 파라미터는 모델의 가중치(weight) 및 편향(bias) 등과 같은 값들을 의미
- 기계학습에서 **하이퍼파라미터 최적화**: 모델의 **일반화 성능**을 높이는 데 중요한 역할
  - 모델 아키텍처와 관련된 하이퍼파라미터(hyperparameter) 값들을 조정하여 모델의 성능을 개선하는 것을 의미
  - 하이퍼파라미터는 모델의 학습률(learning rate), 배치 크기(batch size), 에포크 수(epochs), 규제 계수(regularization coefficient) 등과 같은 값을 의미

## 분포 관점에서의 파라미터/하이퍼파라미터 최적화:

- 분포에서 **파라미터 최적화**: 해당 분포에서 관찰된 데이터에 가장 적합한 파라미터 값을 찾는 것을 의미
  - 정규분포의 경우 평균과 표준편차가 해당 분포의 파라미터이
  - 파라미터를 최적화함으로써 모델의 성능을 향상
- 분포에서 **하이퍼파라미터 최적화**: 해당 분포를 생성하는 데 사용되는 하이퍼파라미터를 조정하여 분포의 성능을 향상시키는 것을 의미
  - 베타 분포의 경우, 하이퍼파라미터인 알파와 베타는 해당 분포의 형태를 결정하는데 사용
  - 하이퍼파라미터 값을 최적화함으로써 분포의 형태를 개선하고 모델의 성능을 향상
- 의의:
  - 파라미터와 하이퍼파라미터 최적화는 분포 모델링에서 중요한 개념
  - 이들 값을 올바르게 조정하면 모델의 예측력을 높일 수 있으며 잘못된 값을 선택하면 모델의 성능이 저하 가능
  - 즉, 분포 모델링에서는 최적화 과정을 통해 파라미터와 하이퍼파라미터를 적절하게 조정

## 최적화 기법의 예시

1. 최소자승법(OLS): 주어진 데이터와 모델 사이의 차이를 최소화하는 방법으로, 선형 회귀분석에서 많이 사용
2. 최대우도법(MLE): 주어진 데이터가 모델에서 나올 확률을 최대화하는 방법으로, 확률 모델링에서 많이 사용
3. 최대사후확률 : 베이저안 통계에서 사용되며, 사전 분포와 사후 분포를 이용하여 모델링 하고, 사후 분포의 평균 값을 최적화.

## 통계적 최적화

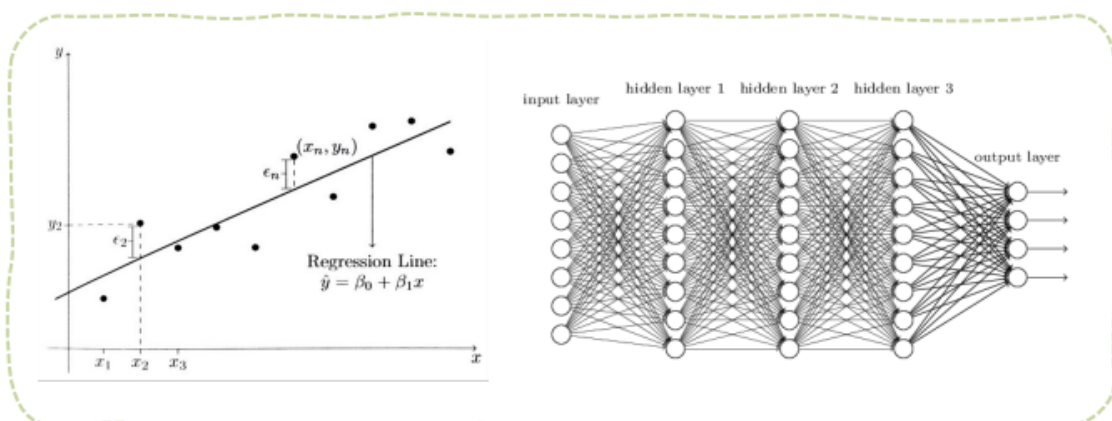
통계적 최적화는 데이터 분석에서 매우 중요한 역할 함

주어진 조건 하에서 목적 함수를 최소화하거나 최대화하기 때문

- 선형 회귀분석에서 OLS를 사용하여 데이터와 모델 사이의 차이를 최소화하여 모델 향상 가능
- MLE를 사용하여 파라미터 값을 추정 → 데이터의 생성 확률을 최대화하여 모델 구축 가능

## Parameter vs Hyper parameter:

### Parameter란?



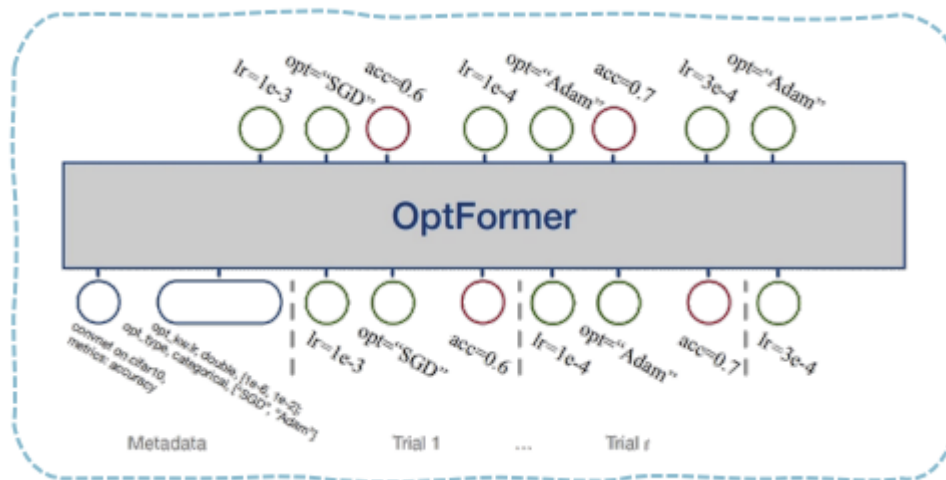
- 학습 과정에서 생성되는 변수

- 모델을 학습할 때 데이터로부터 추정되어 최적화되는 값 (모델 내부에서 학습) → 모델의 학습 과정 중에 조정되어 가장 좋은 성능을 내는 값을 구하며, 새로운 입력에 대해 예측을 수행할 때 사용

ex) 랜덤 포레스트 모델과 딥러닝 모델의 학습 가중치

ex) 선형 회귀 모델에서의 기울기와 절편

## Hyper Parameter란?



- 모델을 구성할 때, 학습을 제어하고 모델의 성능을 조절하는 데 사용되는 변수
- 이들은 모델의 파라미터와는 달리 직접적으로 학습되지 않음 → 모델 구성 단계에서 사용자가 지정하며, 학습을 수행할 때 모델 파라미터를 최적화하기 위해 사용
- 하이퍼 파라미터는 모델 자체로부터 최적화되지 않음 → 여러가지 기법을 통해 하이퍼 파라미터 최적화 가능

ex) 랜덤 포레스트 모델

- 트리의 개수
- 트리의 깊이

ex) 딥러닝(신경망) 모델

- layer의 개수
- 에폭(학습횟수)의 수
- 은닉층의 수
- 뉴런의 수

- 학습률