



# Foundation of Machine Learning 12주차



정재현, 우지수 / 2023.05.03



Computational Data Science LAB



# CONTENTS

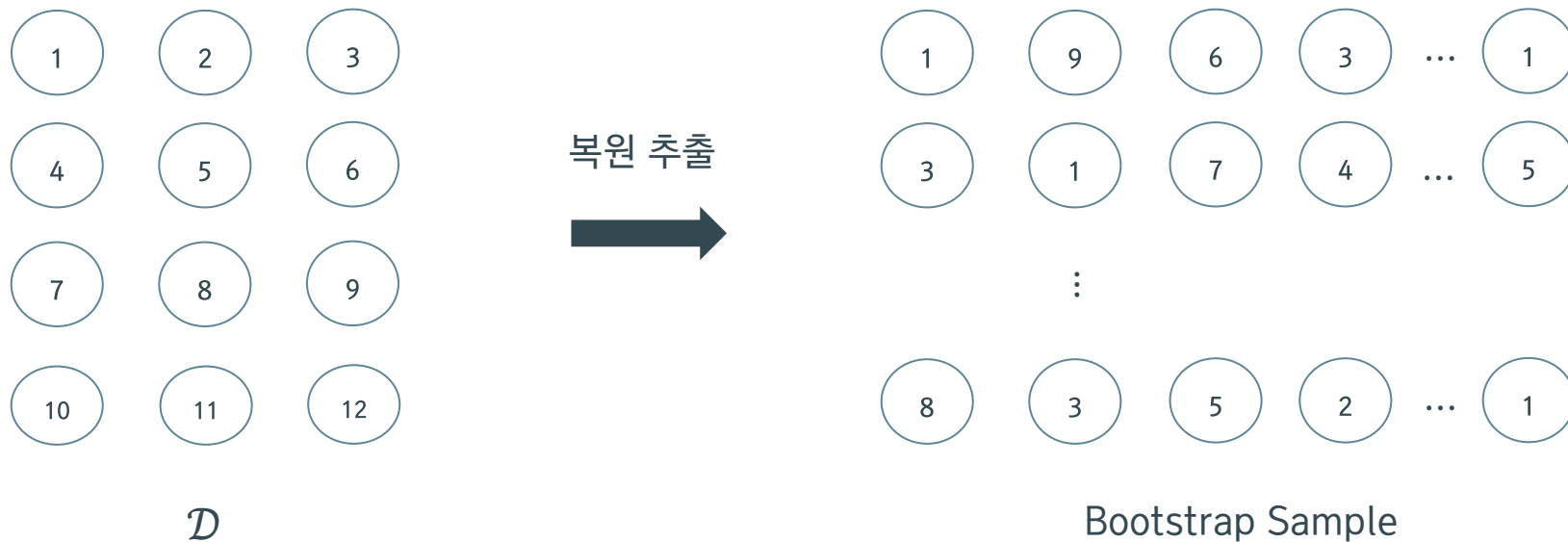
1. Introduction
  2. Bagging
  3. Random Forest
  4. Boosting
- 
- 

# 01 | Introduction

## Bootstrap

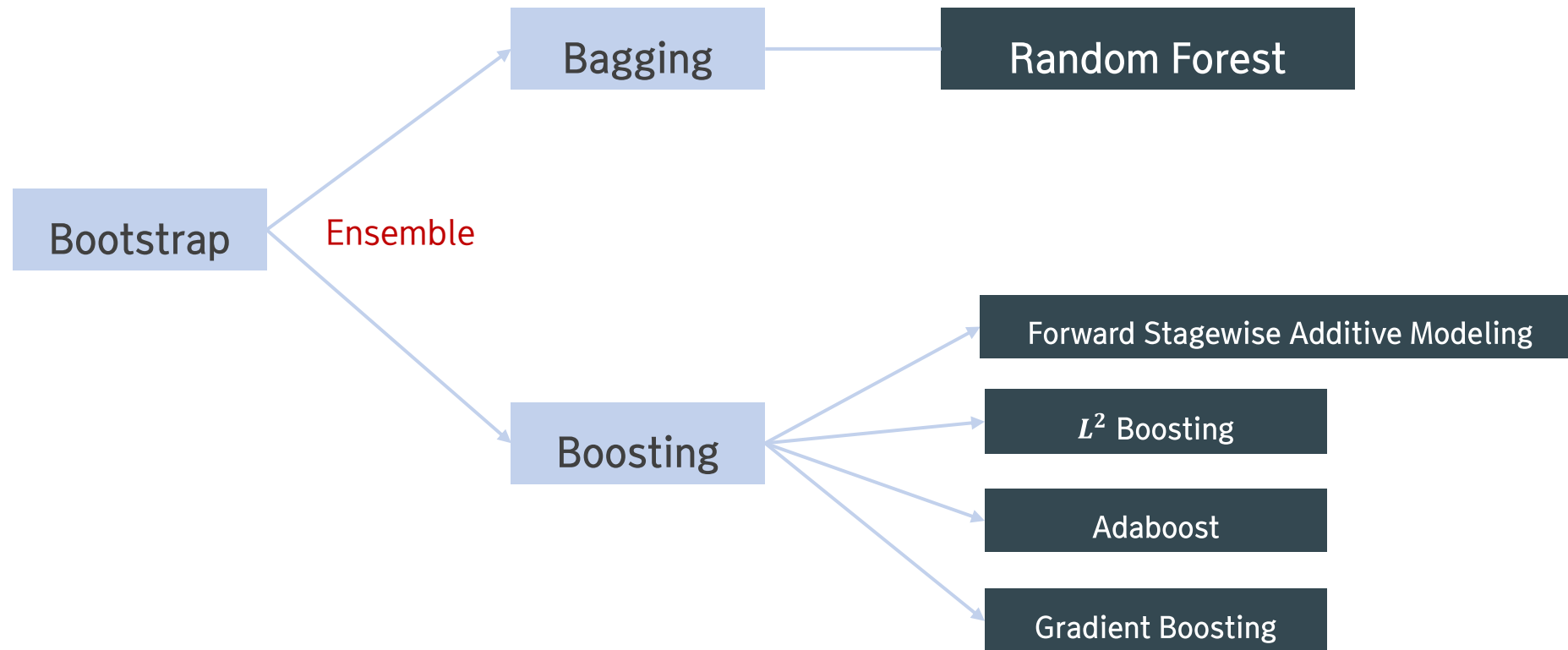
- Bootstrap

- ✓ 모집단( $\mathcal{D}$ )에서 무작위 복원 추출을 통해 여러 개의 학습 데이터 표본을 추출하는 것
- ✓ Bootstrap을 활용하여 Ensemble 사용 가능 → **Bagging, Boosting**



# 01 | Introduction

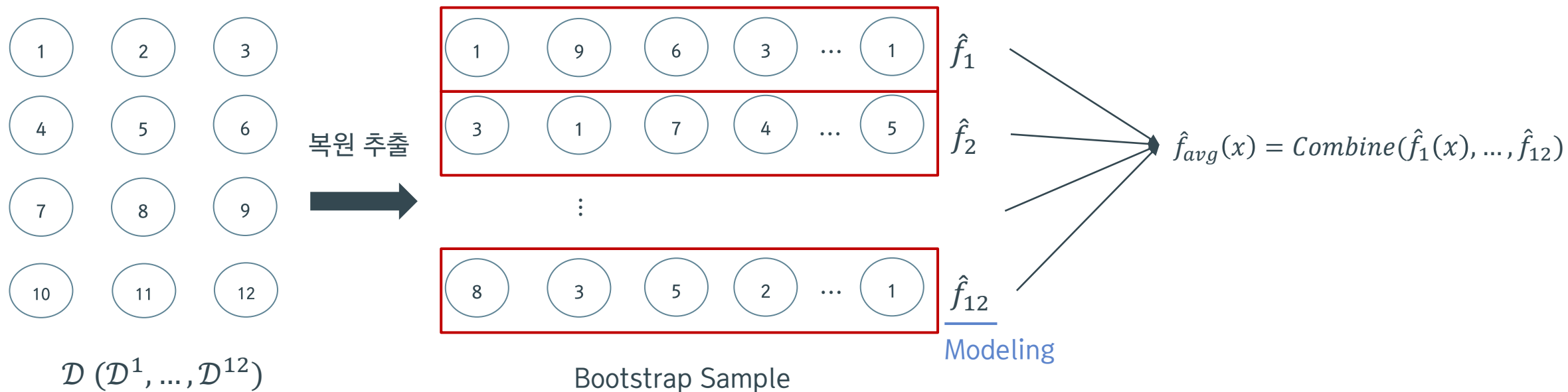
## Bootstrap



# 02 | Bagging

## Bootstrap Aggregating

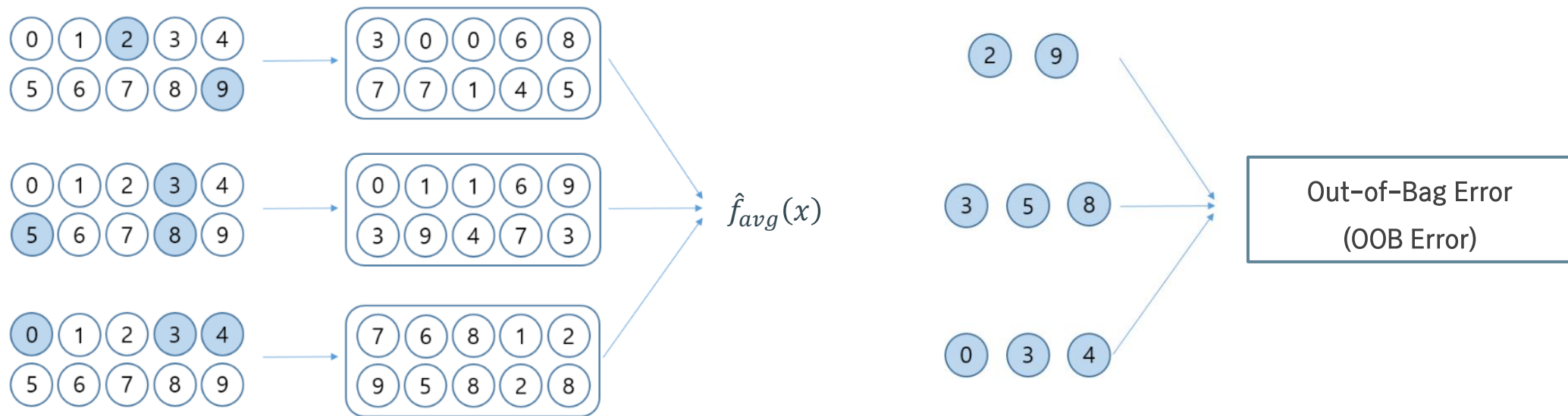
- Bagging (Bootstrap Aggregating)
  - ✓ 모델을 다양하게 만들기 위해 데이터를 재구성
  - ✓ Bootstrap을 통해 추출된 표본들로 다수의 Decision Tree를 만든 다음, 결합하여 최종 예측 수행



# 02 | Bagging

## Out-of-Bag Error

- Out-of-Bag Error



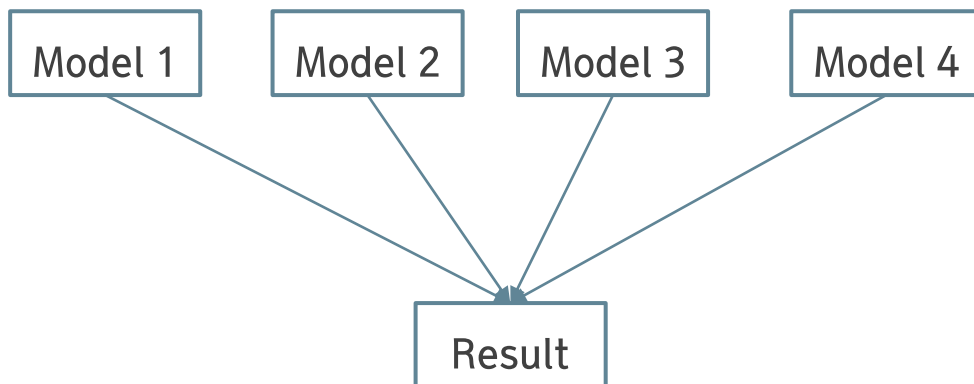
- ✓ Bootstrap을 통해 Sampling을 하다 보면 전체 데이터  $\mathcal{D}$ 에서 약 63%만 추출
- ✓ 남은 37%는 OOB(Out-of-bag)이라고 정의한 후, 모델 검정 용도로 사용 → 모든 트리의 OOB 오차를 평균한 값으로 성능 추정
- ✓ 별도의 검증 데이터 없이 OOB 데이터를 이용하면 모델의 성능 평가나 Hyper Parameter 최적화 가능

# 02 | Bagging

## Tree vs Bagging

- Tree vs Bagging

✓ 깊이 성장한 Tree : 분산 증가, 편향 감소 → **Overfitting**



✓ Bagging : 샘플을 무작위적으로 추출하는 Bootstrap을 활용하여 다양한 모델을 만들고 결합

Bootstrap을 활용하면서 전체 데이터에서 일부만을 이용해 학습하므로 편향은 유지되고 전체 분산은 감소

# 03 | Random Forest

## Definition

- Random Forest

### [Bagging과 Random Forest]

- ✓ 여러 개의 Tree들의 조합이기 때문에 Bagging model의 분산은 각각의 트리들의 분산과 그들의 공분산으로 구성

$$\underline{Var(X + Y)} = Var(X) + Var(Y) + \underline{2Cov(X, Y)}$$

$X + Y$ 의 분산이  $X$ 와  $Y$ 의 분산을 더한 것 보다 큰 형태  
= Bagging model의 분산이 더 커질 수 있다는 의미

모델들이 대부분 중복되는 데이터를 가지고 있기 때문에  
비슷한 Tree가 만들어질 확률이 높으므로  $2Cov(X, Y) \neq 0$

∴ 각 공분산을 줄일 수 있는 방법이 필요



# 03 | Random Forest

## Definition

- Random Forest

- ✓ Bootstrap을 한 후, 모델을 다양하게 만들기 위해 데이터 뿐만 아니라 변수도 재구성한 후 결과를 취합
- ✓ Bagged Decision Tree를 사용할 때, Tree들 간의 상관관계(공분산)를 줄이는 것이 목적

Ex) 건강 위험도 예측



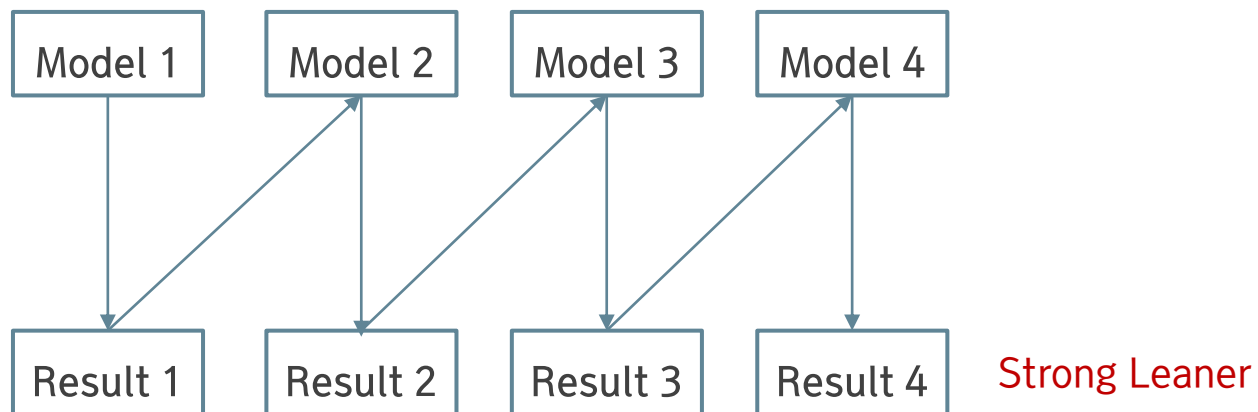
- ✓ 수많은 요소(Feature)를 기반으로 건강의 위험도를 예측한다면 Overfitting 발생
  - ✓ 여러 개의 Feature 중 랜덤으로 n개의 Feature만 선택해서 하나의 Tree를 만드는 과정을 반복하여 여러 개의 Tree 만들기
- Feature의 총 개수의 제곱근 정도로 설정
- ✓ 여러 Decision Tree들이 내린 예측 값들 중 가장 많이 나온 값을 최종 예측 값으로 설정

# 04 | Boosting

## Definition

- Boosting
  - ✓ 모델이 맞추기 어려워하는 문제를 맞힐 수 있도록 학습하는 것
  - ✓ 오분류된 데이터에 초점을 맞추어 더 많은 가중치를 주는 방식
  - ✓ Weak learner 여러 개를 합쳐 Strong learner를 만드는 방법

Weak Learner

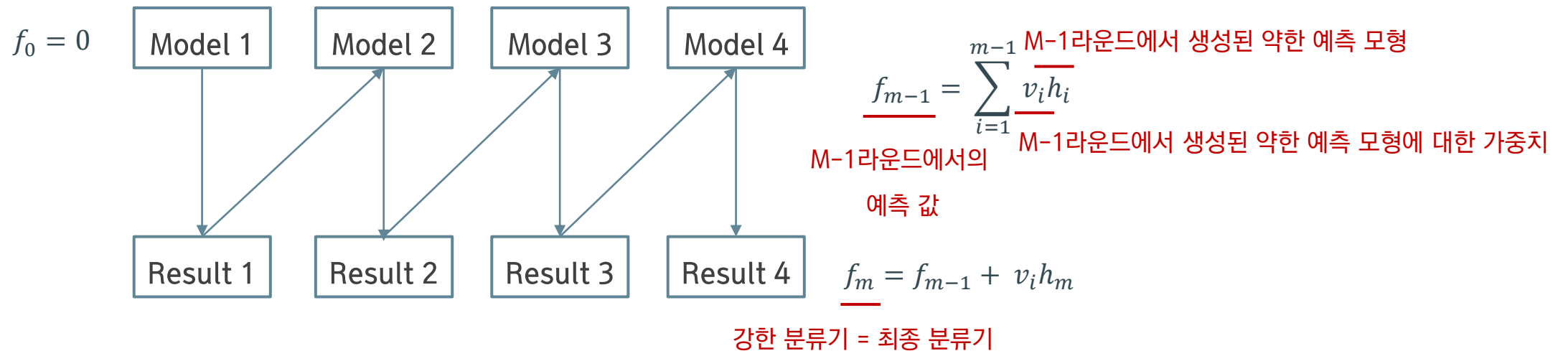


# 04 | Boosting

## Example

- Forward Stagewise Additive Modeling (FSAM)

- ✓ M-1라운드까지 반복적으로 예측 값을 갱신하며 최종적으로 강한 예측 모형인  $f_m$ 을 구하는 방식
- ✓ 오차를 줄이는 방향으로 새로운 변수를 추가해 모델의 예측 값을 업데이트 → 선형 회귀 모델에 적용

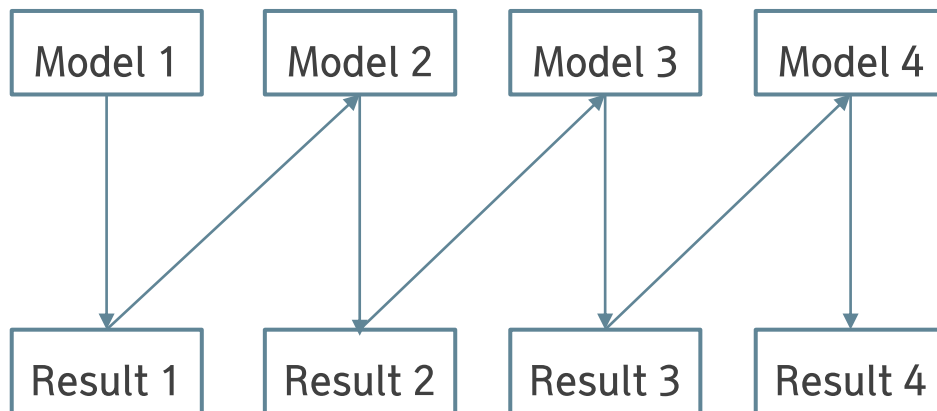


# 04 | Boosting

## Example

- $L^2$  Boosting

- ✓ 이전 라운드에서의 예측 값과 새로운 약한 분류기를 결합한 후 실제 값 간의 차이를 제공하여 오차 계산
- ✓ 평균 제곱 오차(Mean Squared Error, MSE)를 최소화하는 방향으로 학습



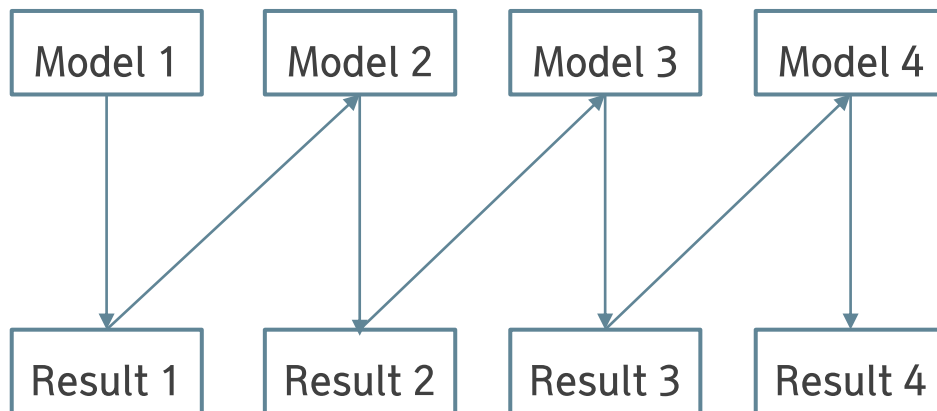
$$J(v, h) = \frac{1}{n} \sum_{i=1}^n \left( \overbrace{y_i}^{\text{실제 값}} - \left[ \underbrace{f_{m-1}(x_i)}_{\text{M-1 라운드의 예측 값}} + \underbrace{vh(x_i)}_{\text{약한 분류기의 가중치} \times \text{약한 분류기}} \right] \right)^2$$

# 04 | Boosting

## Example

- AdaBoost (Adaptive Boost)

- ✓ 초기에 모든 샘플의 가중치를 균등하게 초기화한 후 학습을 진행
- ✓ 이전 Learner에서 잘못 분류된 샘플에 가중치를 부여해 다음 Learner에서 높은 가중치를 가진 샘플에 더 많은 비중을 두는 방식



$$\begin{aligned} \text{Strong Learner } H(x) &= \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots + \alpha_t h_t(x) \\ &= \sum_{t=1}^T \alpha_t \text{Weak Learner } h_t(x) \end{aligned}$$

# 04 | Boosting

## Example

- Gradient Boosting

- ✓ 이전 라운드의 최종적인 모델이 예측한 데이터의 오차를 예측하는 Weak Learner를 학습
- ✓ 오차에 대한 Gradient를 계산해 이를 예측 값에 더해주는 방식으로 오차를 보완

$$y = h_0(x) + error$$

$$error = h_1(x) + error2$$

$$error2 = h_2(x) + error3$$

$$y = \underbrace{h_0(x)}_{\text{Strong Learner}} + \underbrace{h_1(x) + h_2(x) + \dots}_{\text{Weak Learner}} + small\ error$$

- ✓ 해당 과정을 반복하여 여러 개의 분류기를 결합하여 강력한 분류기 만들어내는 방법

# Q&A

---

감사합니다.