Bagging Tree & Random Forest

구름

도시공학과 일반대학원

한양대학교

의사결정나무 알고리즘 확장

Dtree '84~'86

Bagging '94

Boosting '97

Random Forest '01

GBM '99

XGBoost '16
LightGBM

2
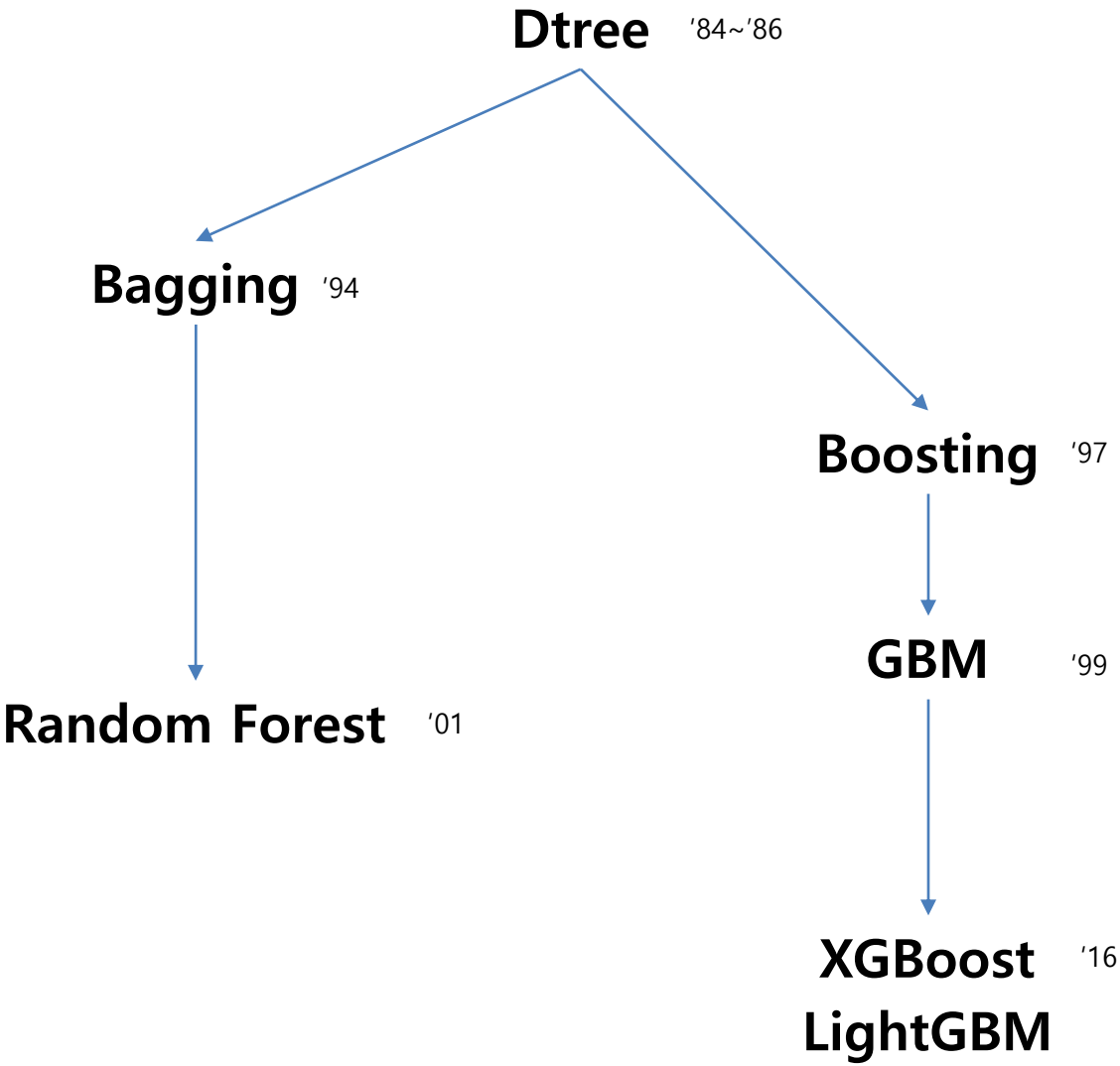
Ensemble Model   2개 이상의 학습 모델을 결합하는 메타 알고리즘

Voting

Data → Model A → Predict 0 (0:52%, 1:48%) → Hard Voting : 0

Data → Model B → Predict 0 (0:62%, 1:38%)

Data → Model C → Predict 1 (0:1%, 1:99%)

0 : 2개
1 : 1개

Soft Voting : 0

0 : 0.52 + 0.62 + 0.01 = 115
1 : 0.48 + 0.38 + 0.99 = 185

boosting

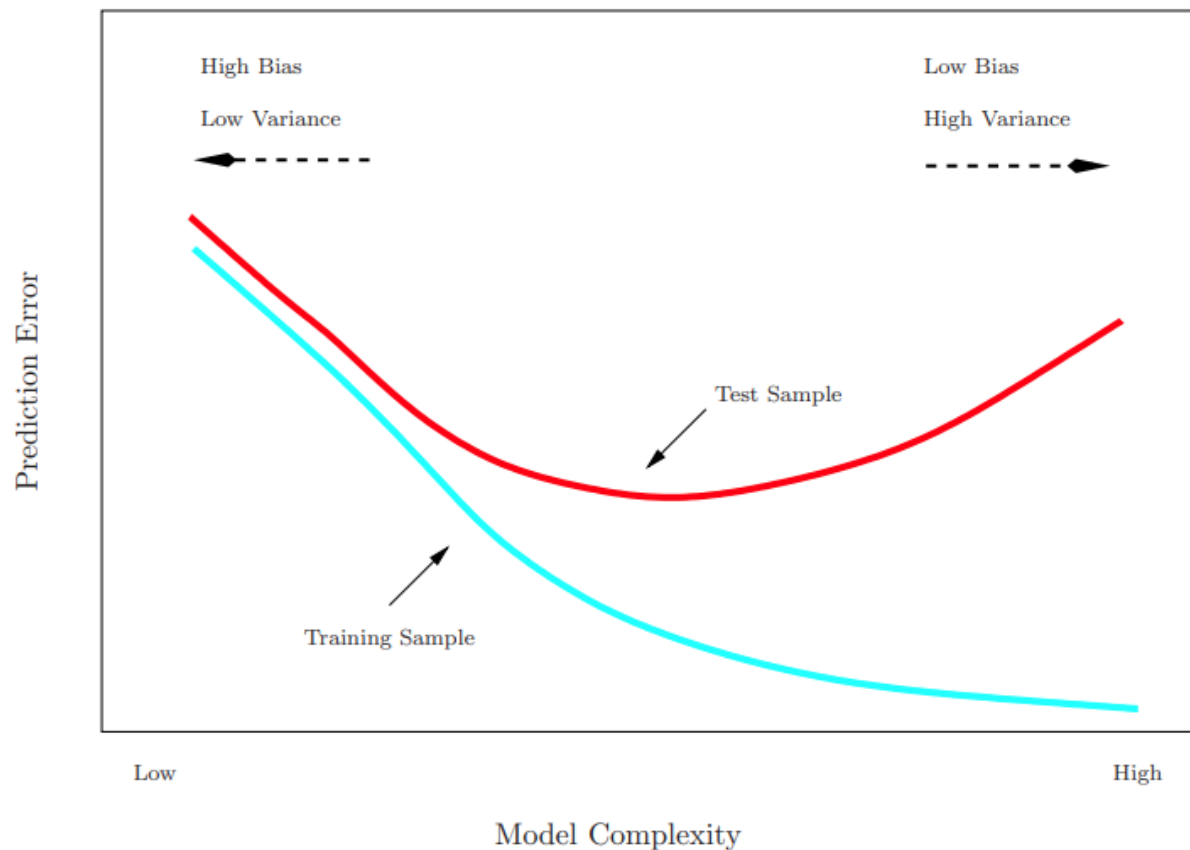Data → Model A → 잔차 → Model B → 잔차 → Model C → Predict
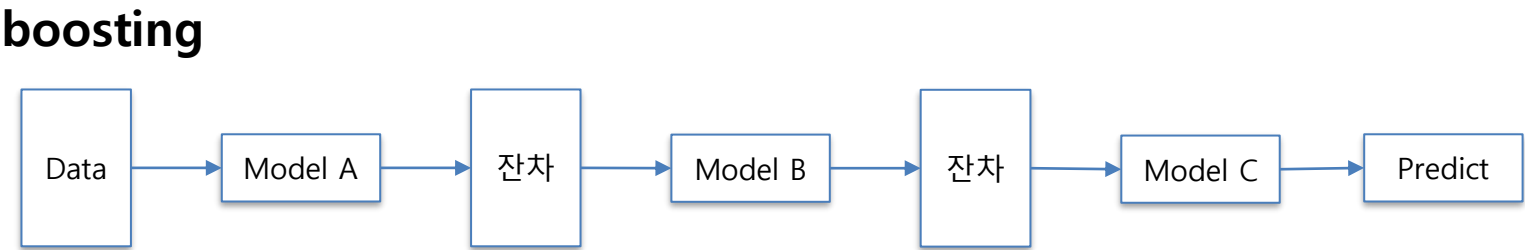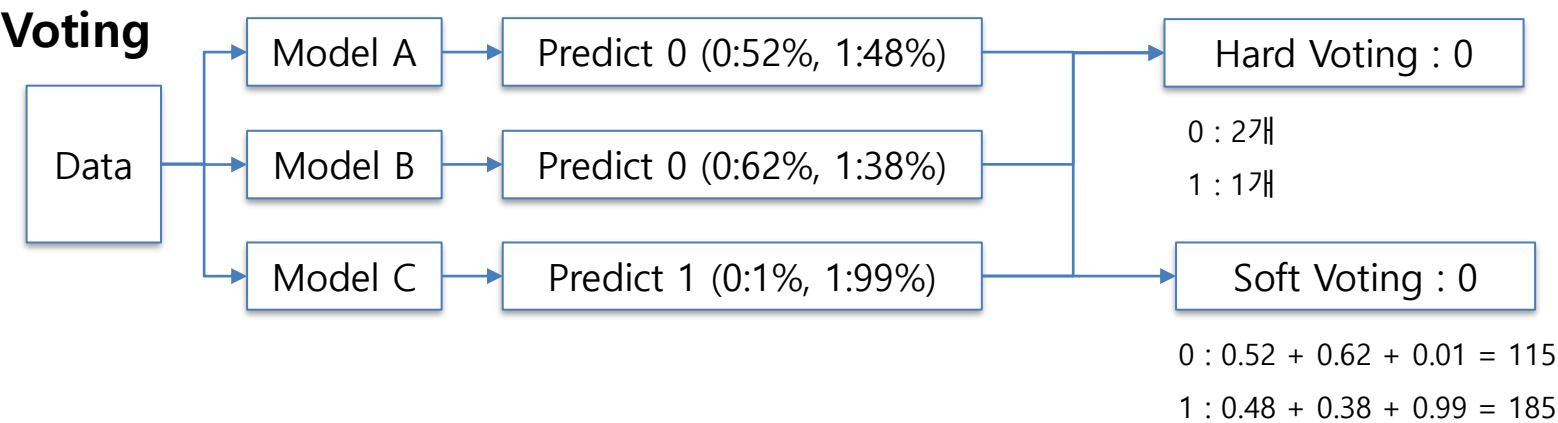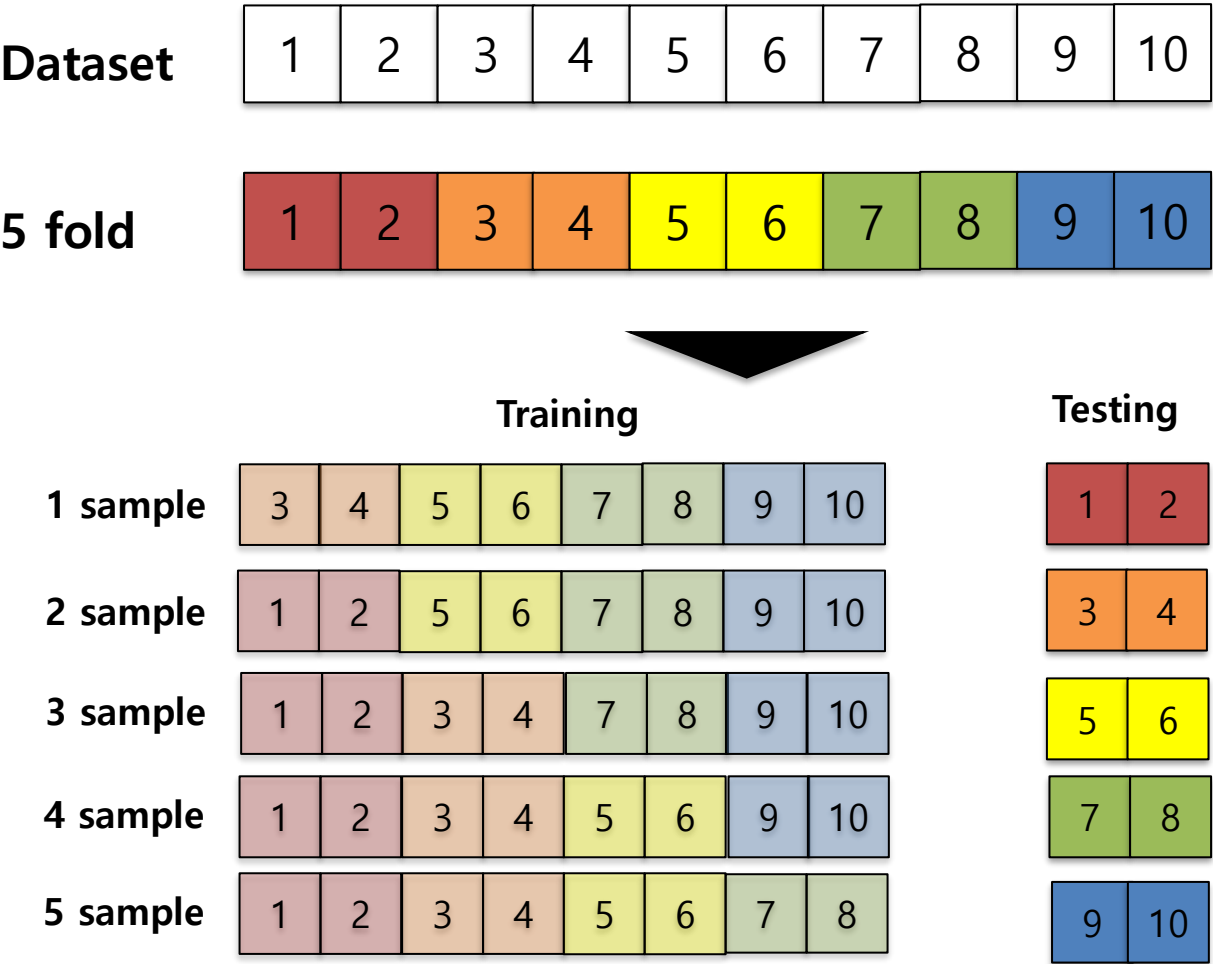
Bias-Variance Trade Off
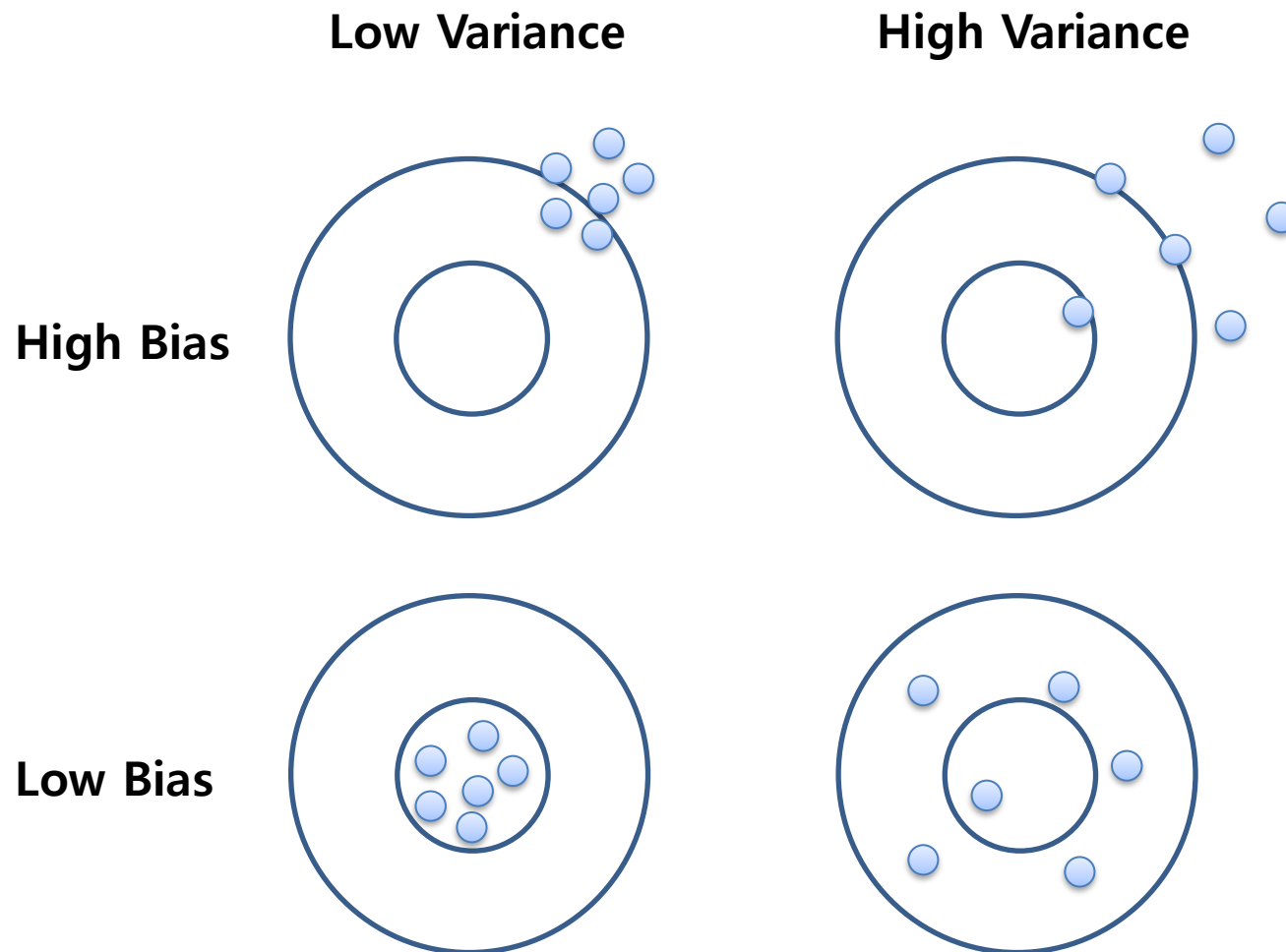
Error–complexity Trade Off

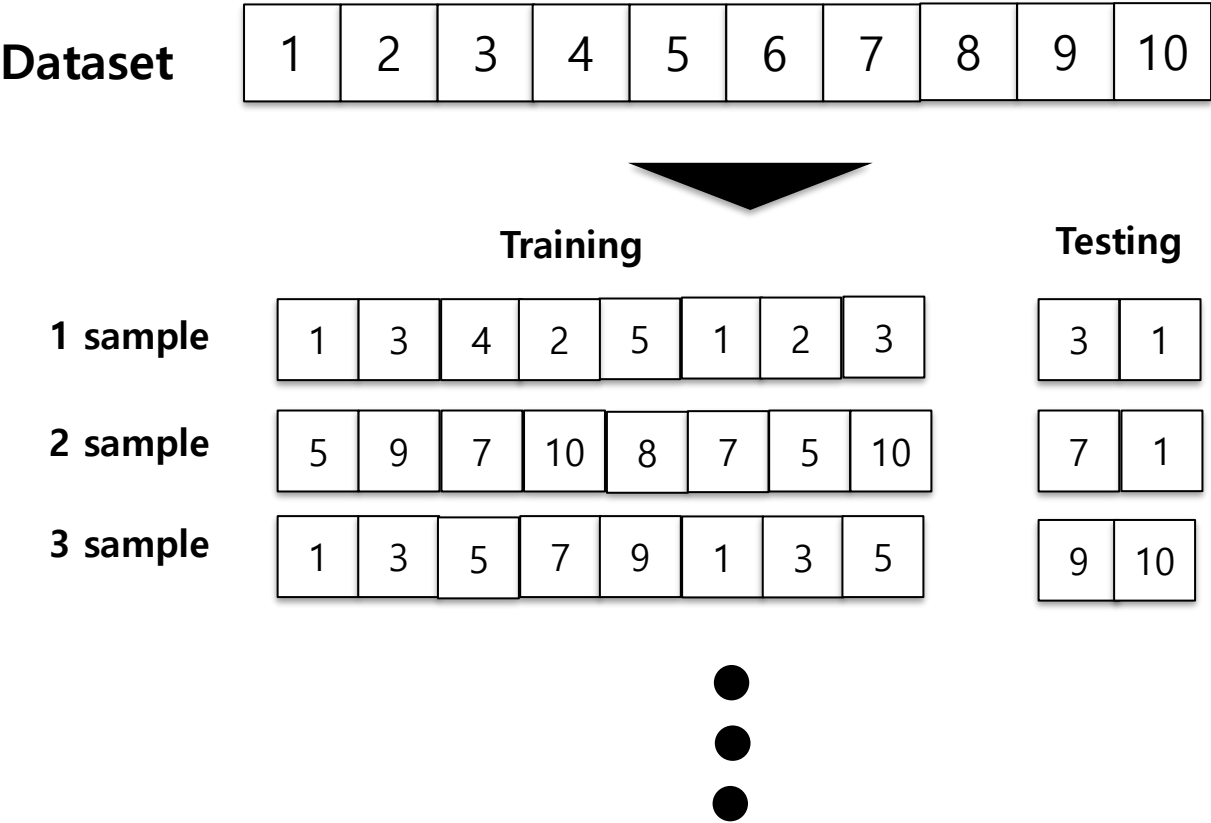## Training Error VS Test Error

Resampling : Cross Validation vs Bootstrap
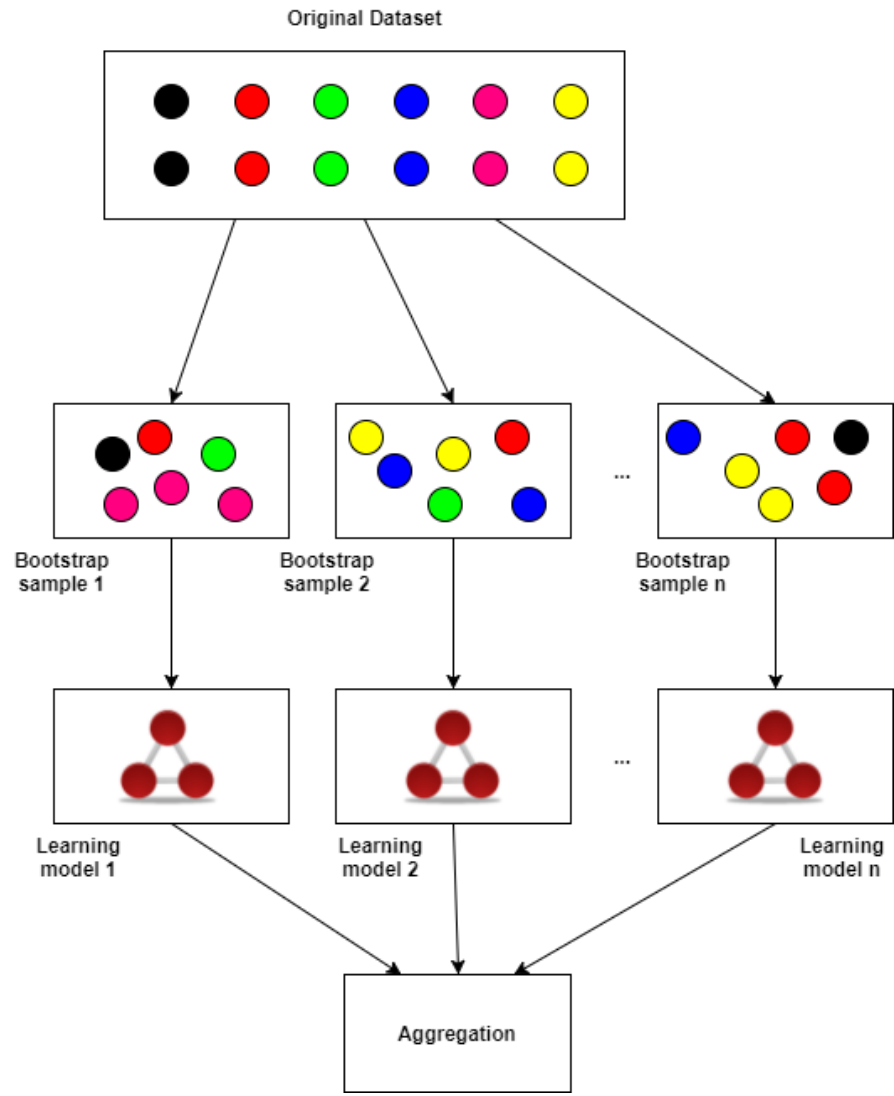
K-fold Cross Validation

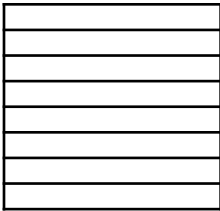# Resampling : Cross Validation vs Bootstrap

Bootstrap

Dataset

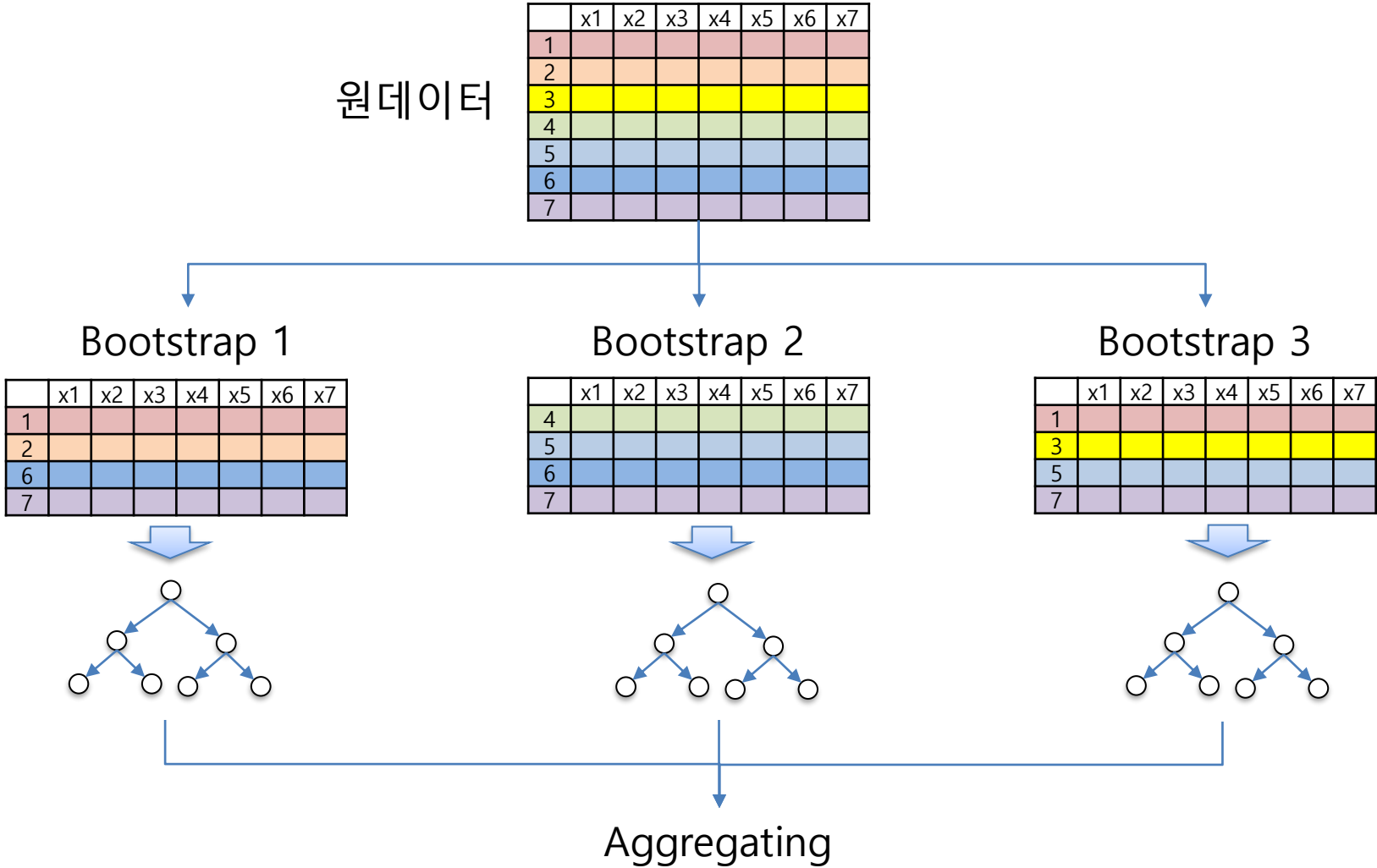| 1	2	3	4	5	6	7	8	9	10

Training **Testing**

1 sample

1	3	4	2	5	1	2	3

3	1

2 sample

5	9	7	10	8	7	5	10

7	1

3 sample

1	3	5	7	9	1	3	5

9	10

●
●
●

<u>B</u>ootstrap <u>Agg</u>regat<u>ing</u> : Bagging

원데이터

Bootstrap **Agg**regat**ing** : Bagging

Aggregating

Random Forest