

✓ 앙상블과 랜덤포레스트

8강. 앙상블과 랜덤포레스트

- Voting / Bagging / Boosting / Stacking
- 랜덤포레스트(Random Forest)

■ 지도학습의 종류



■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법

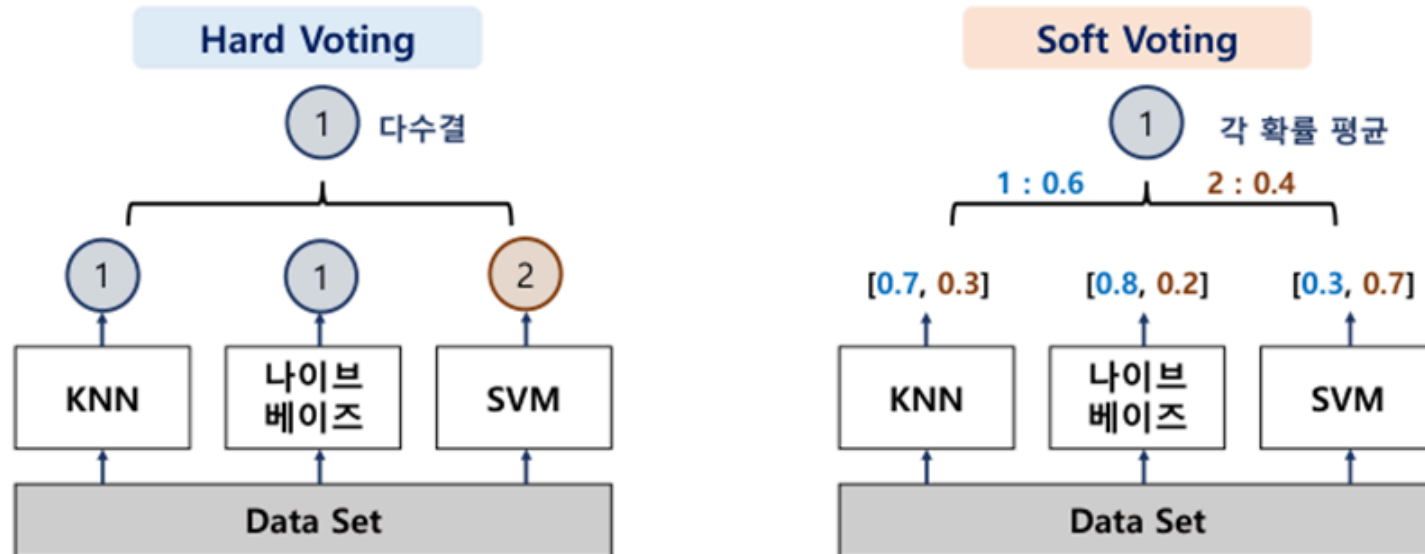


■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법



서로 다른 모델을 활용하여 예측한 값을 다수결 혹은 평균으로 최종결과 선정



일반적으로 Soft Voting의 성능이 좋음

■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법

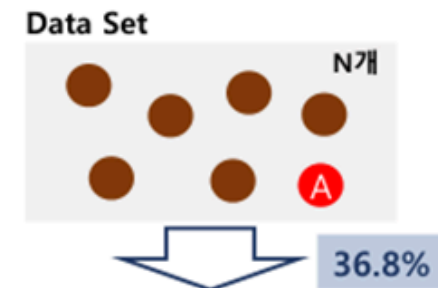
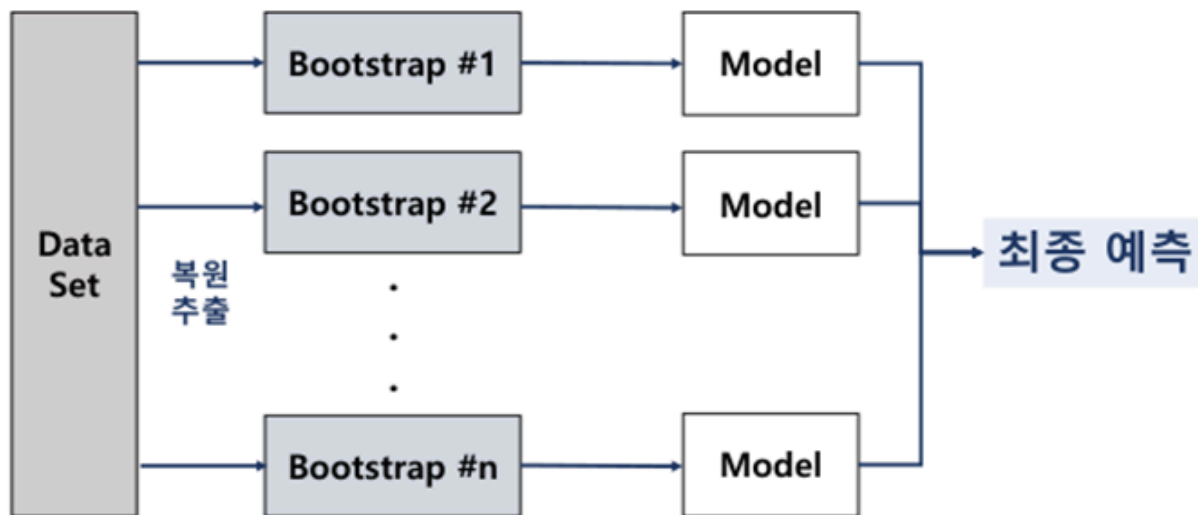
Voting

Bagging(Bootstrap Aggregating) Boosting

Stacking

Bootstrap된 데이터를 활용하여 동일한 모델로 데이터를 학습하여 예측

* Bootstrap : 모집단으로부터 데이터를 복원 추출하는 샘플링 방식



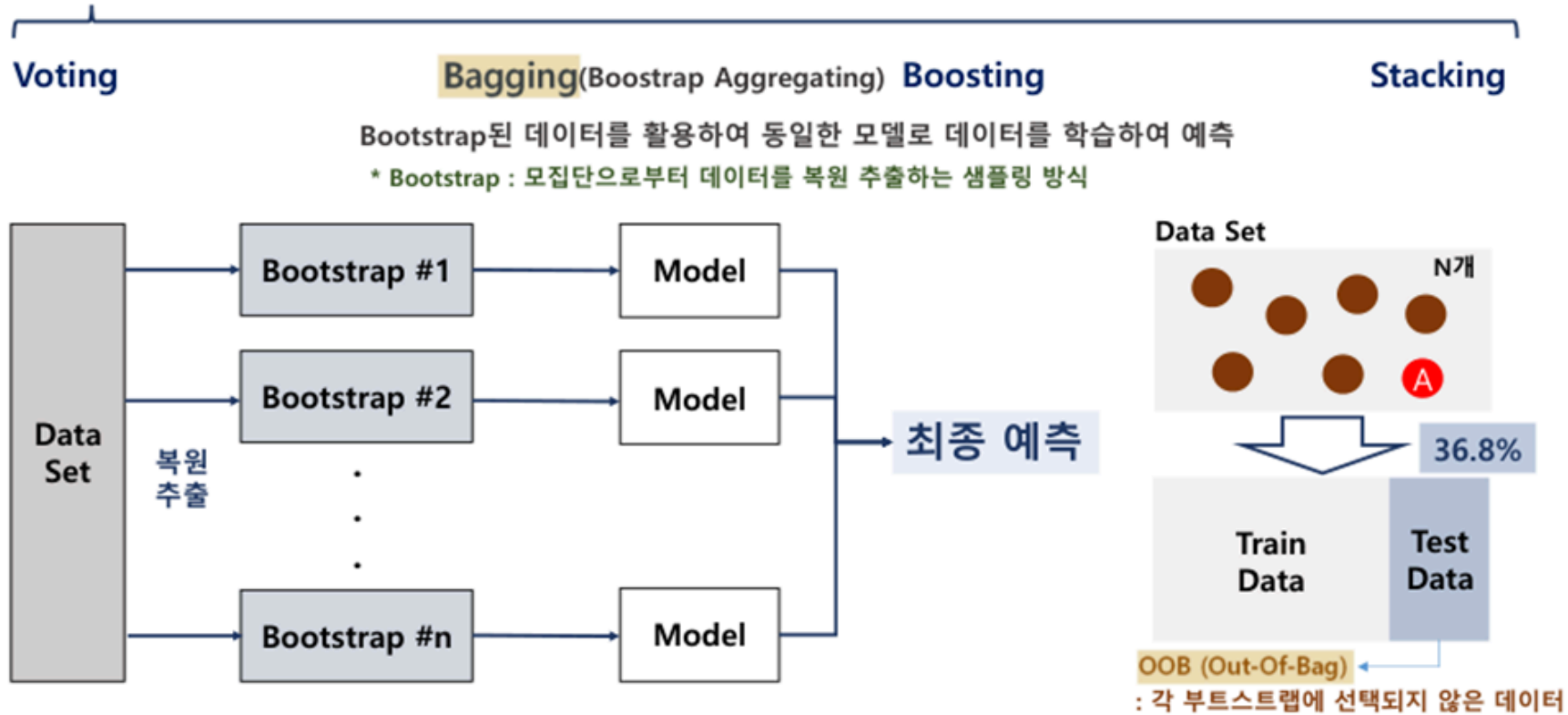
N번 복원 추출 시 특정 A 데이터 미선택될 확률

$$P = \left(1 - \frac{1}{N}\right)^N$$

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = 0.368$$

■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법



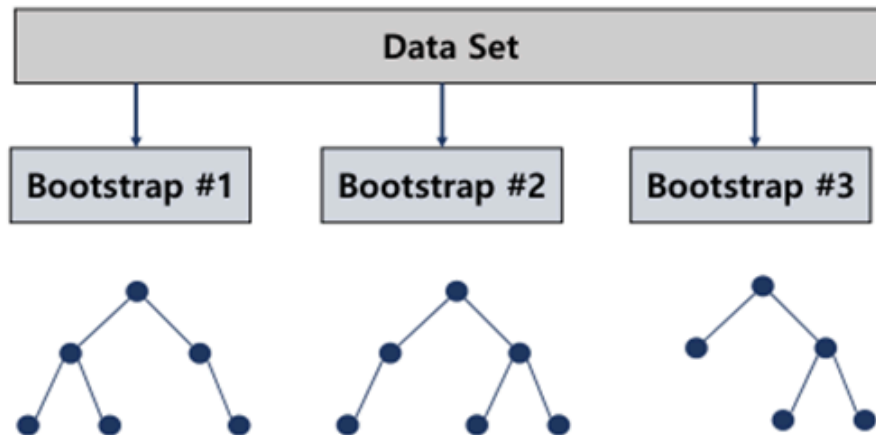
■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법



랜덤 포레스트(Random Forest)

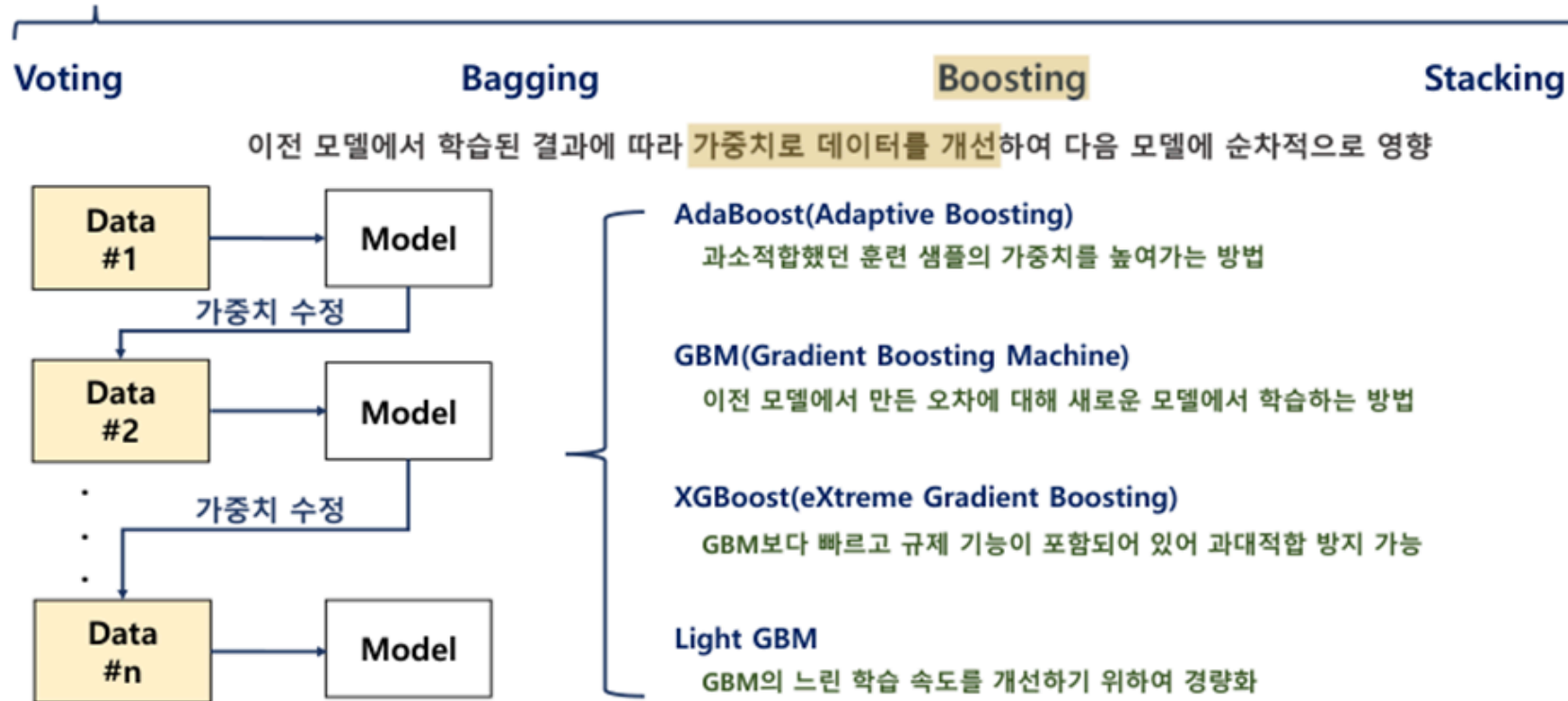
여러 개의 무작위로 생성된 결정트리를 활용한 배깅 방식의 알고리즘



- 트리 노드 분할 시 무작위성 주입
: 극단적 무작위 랜덤 포레스트인 경우
→ 엑스트라 트리
- 내부 구조를 알 수 없는 블랙박스 모델
- 편향을 손해 보는 대신 분산을 낮춤
: Overfitting에 강력한 모델
- 많은 메모리를 소모하지만 빠른 속도

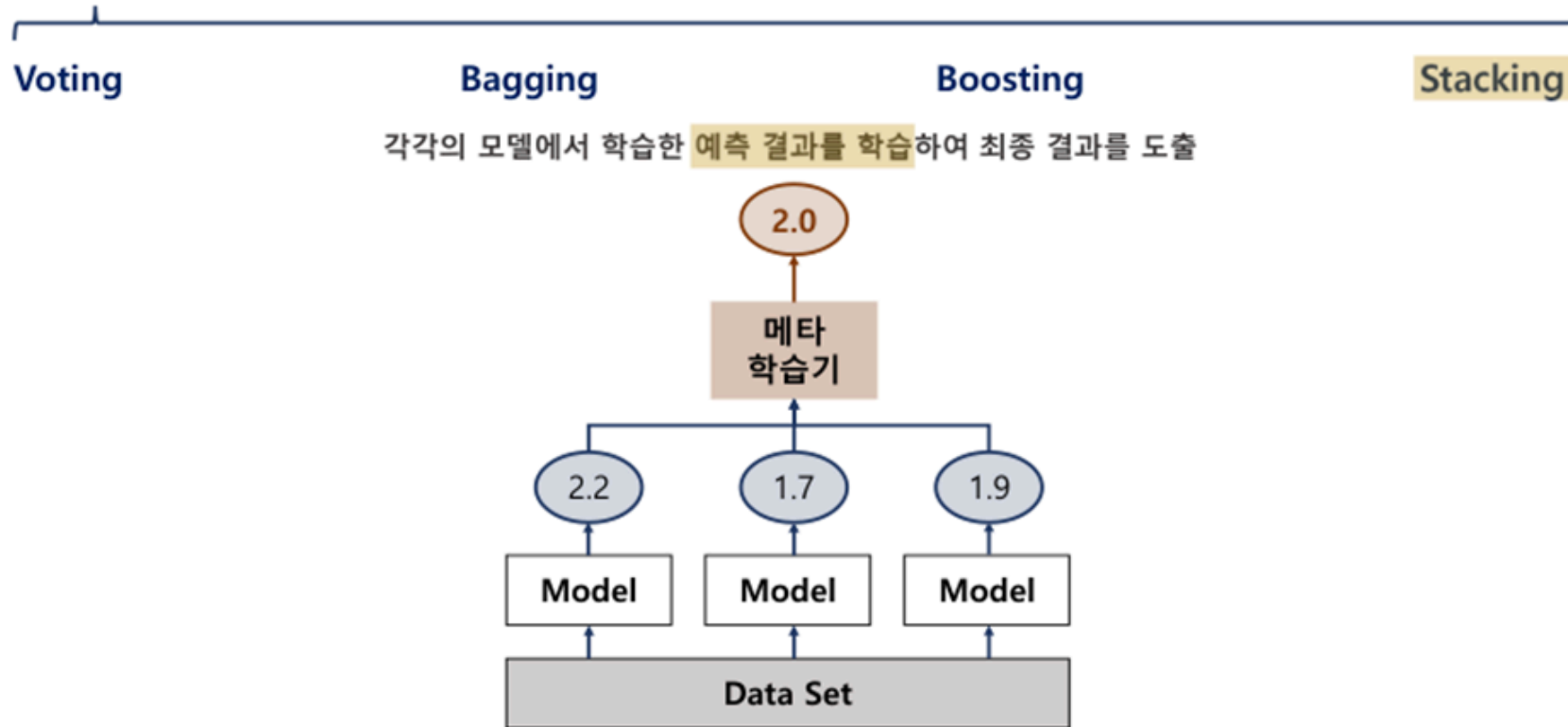
■ 앙상블(Ensemble)

여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법



■ 앙상블(Ensemble)

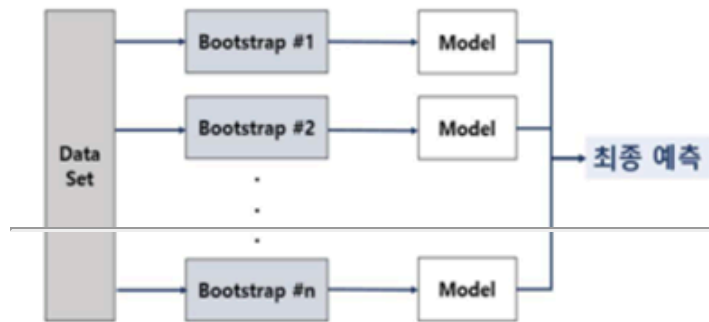
여러 모델의 결과를 조합하여 결과를 향상시키는 학습 방법



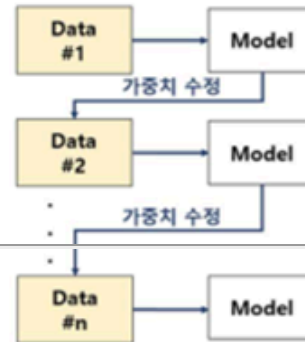
■ 앙상블(Ensemble)

분산 컴퓨팅 활용 가능 여부

Bagging



Boosting



Stacking



✓ 앙상블(Ensemble)
각 모델이 독립적으로 학습

모델이 순차적으로 학습

메타 학습 전 독립적 학습