

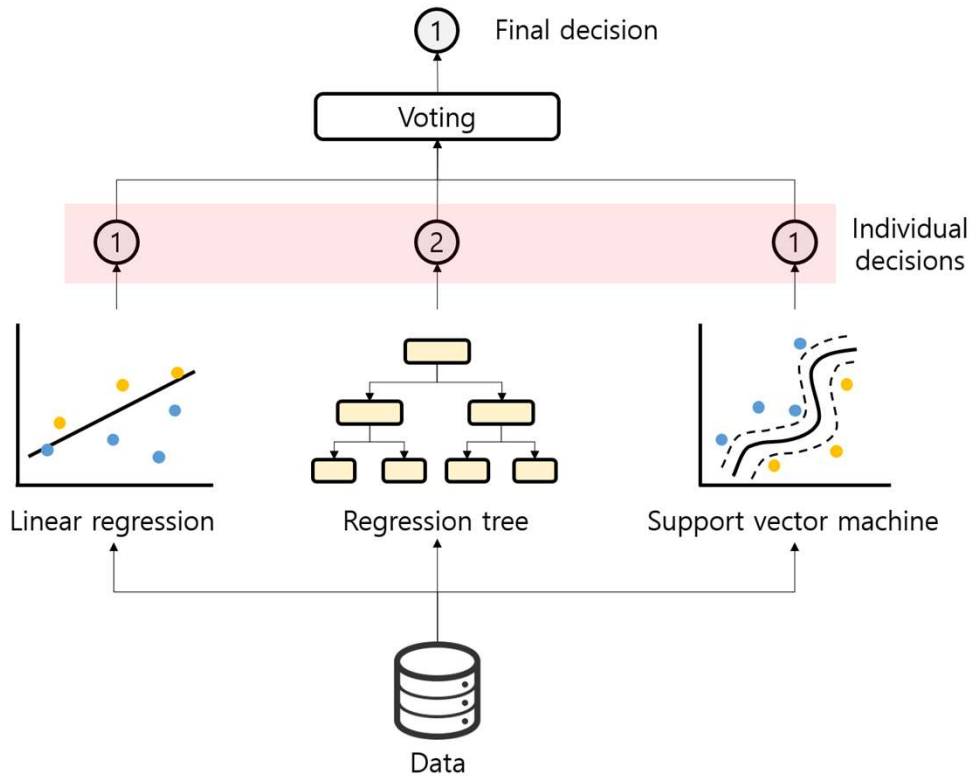
# 앙상블 (Ensemble)

데이크루 1기 jihyeheo

# 1. 앙상블(Ensemble)

- 앙상블(Ensemble)의 사전적 의미는 ‘함께, 조화’입니다. 머신러닝에서 이야기하는 앙상블은 여러 개의 모델을 활용하여 더 나은 결과를 도출하는 것을 의미합니다.
- 이 모형은 여러 개의 weak learner(약한)를 결합한다면 하나의 single learner(단일)보다 더 나은 성능을 얻을 수 있다는 생각에서 나온 방법입니다.
- 앙상블의 종류는 다음과 같습니다.
  - 1) Voting : 동일한 train set을 가지고 여러 모델을 훈련
  - 2) Bagging : train set에서 각 알고리즘마다 별도의 train set을 만들어 훈련
  - 3) Boosting : 배깅과 유사한 진행이지만 분류하는 중 틀린 데이터에 가중치를 부여

## 2. 앙상블(Ensemble) - Voting



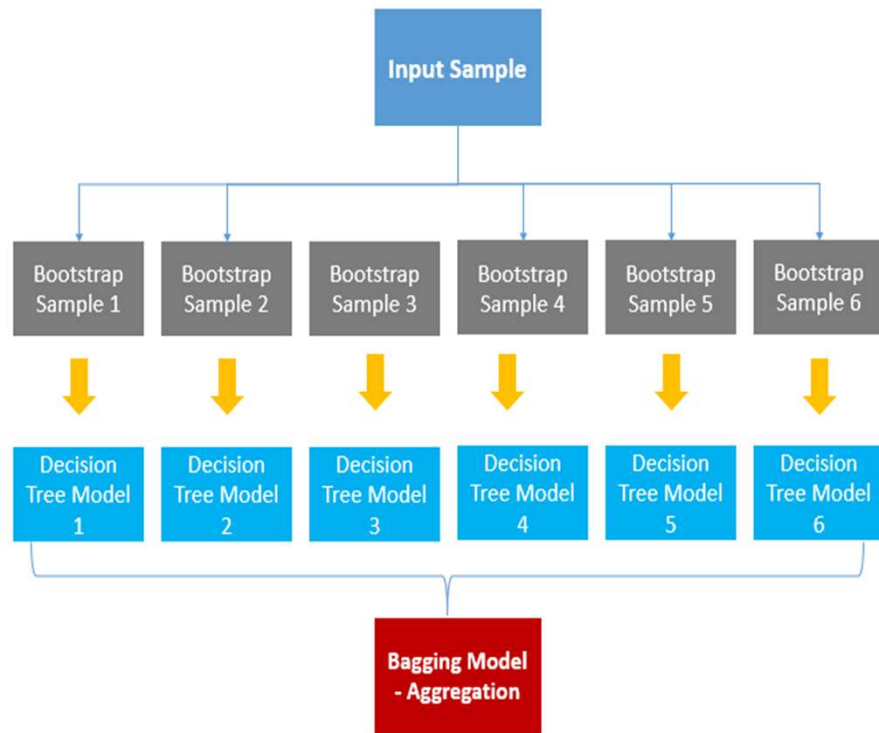
- Voting은 투표를 통해서 결정한다는 의미입니다. 사 진처럼 똑같은 데이터를 다른 모델을 이용하여 결과를 낸 후 투표를 통해서 최종적으로 값을 도출하는 것입니다.

- Voting의 종류는 다음과 같습니다.

- 1) Hard Voting : 결과물에 대한 최종 값을 투표하여 결정한다.
- 2) Soft Voting : 결과물이 나올 확률값을 다 더해서 각각의 확률을 구한 후 최종값을 도출한다.

Soft Voting 방식이 더 합리적이라고 하여 보통 설정에서 'Soft'로 지정하여 한다고 합니다.

## 2. 앙상블(Ensemble) - Bagging



- Bagging은 Bootstrap Aggregation의 약자로, 여러 번 뽑아 각 모델을 학습시켜 결과물을 집계하는 방법입니다.

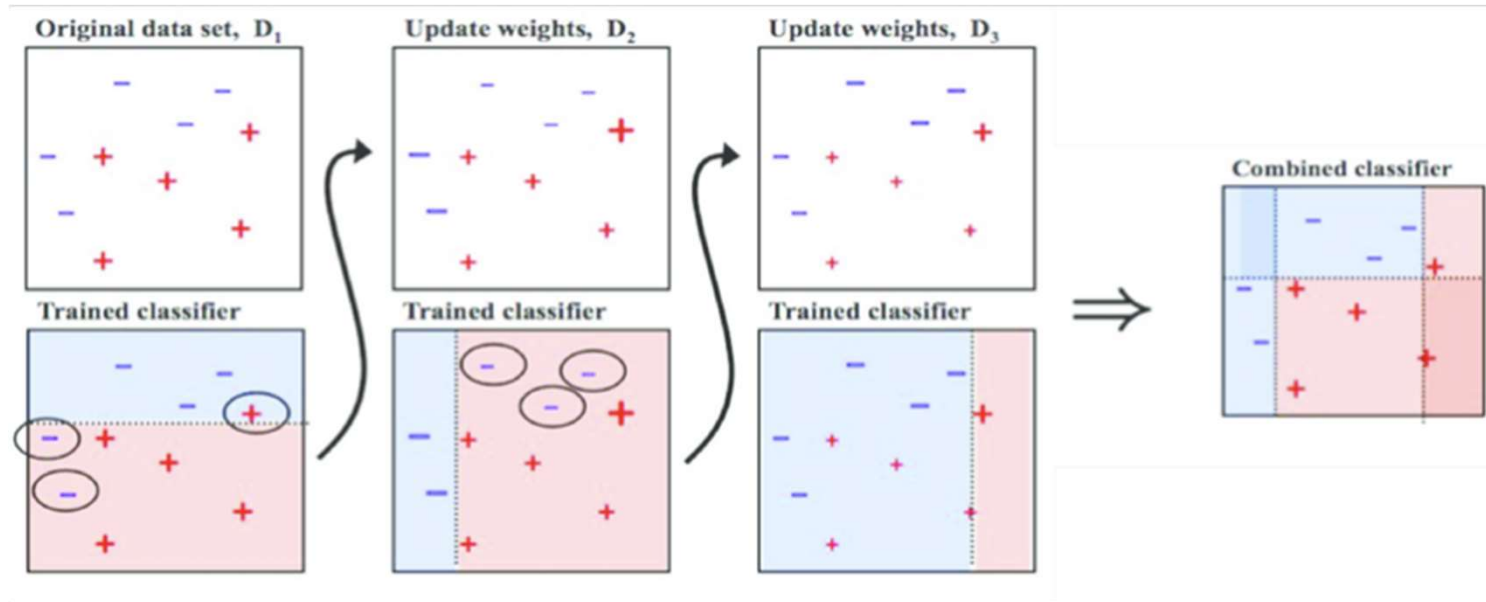
- 복원 랜덤 샘플링을 통한 데이터 추출 후 학습된 모형의 결과를 집계하는데,

Categorical Data -> Voting으로 집계

Continuous Data -> 평균으로 집계합니다.

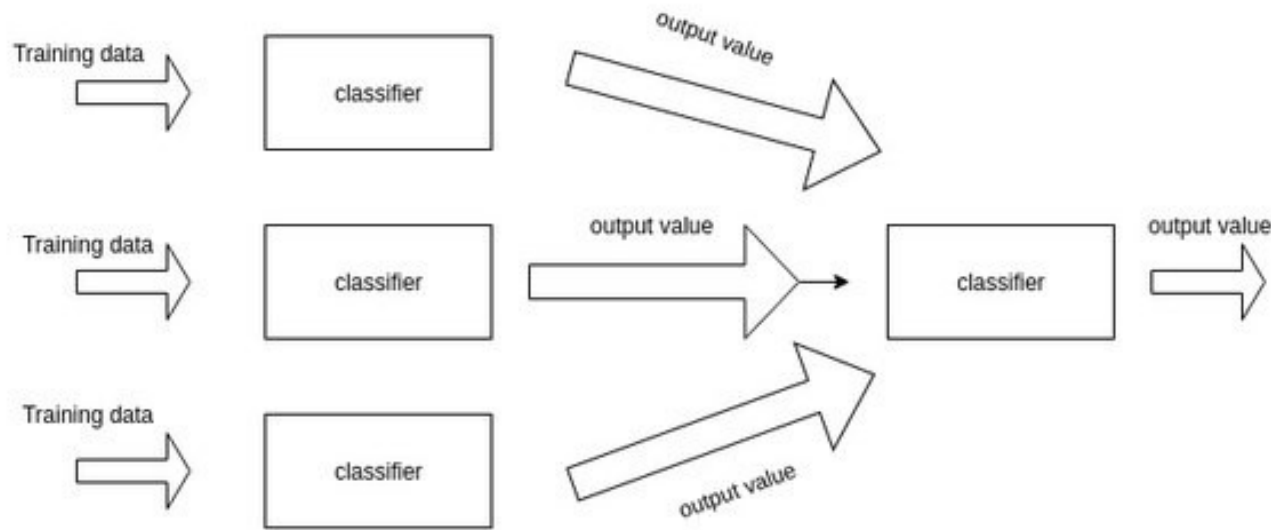
- 배깅 기법을 활용한 대표적인 모형으로 랜덤 포레스트를 들 수 있습니다.

## 2. 앙상블(Ensemble) - Boosting



- Boosting은 가중치를 활용하여 weak 분류기를 strong 분류기로 만드는 방법입니다.
- 부스팅은 다른 기법과 다르게 모형간 팀워크를 이루어 가중치를 부여한다. 결과를 예측한 후 예측 결과에 따라 데이터에 가중치가 부여되고 또 다음 모형에 영향을 줍니다.
- 잘못 분류된 데이터에 집중하여 새로운 분류 규칙을 만드는 단계를 반복합니다.

## 2. 앙상블(Ensemble) - Stacking



- Stacking은 Meta Modeling이라고 불리기도 하는데 서로 다른 모델을 조합해서 최고의 성능을 내는 모델을 생성합니다. 조합을 통해 장점을 취하고 약점을 보완하는 역할을 합니다.
- 따라서 어마무시한 연산량이 일어납니다.
- 단계를 나누자면 각각의 모델으로 계산 후 나온 결과를 합치는 단계, 결과를 바탕으로 모델링을 하여 결과를 내는 단계로 구성할 수 있습니다.

# Reference

1. Voting 사진 : <https://untitledblog.tistory.com/156>
2. Voting 개념 : <https://velog.io/@guns/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-%EC%8A%A4%ED%84%B0%EB%94%94-%EC%95%99%EC%83%81%EB%B8%94-Ensemble-Voting>
3. Bagging, Boosting, stacking 개념 : <https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-11-%EC%95%99%EC%83%81%EB%B8%94-%ED%95%99%EC%8A%B5-Ensemble-Learning-%EB%B0%B0%EA%B9%85Bagging%EA%B3%BC-%EB%B6%80%EC%8A%A4%ED%8C%85Boosting>