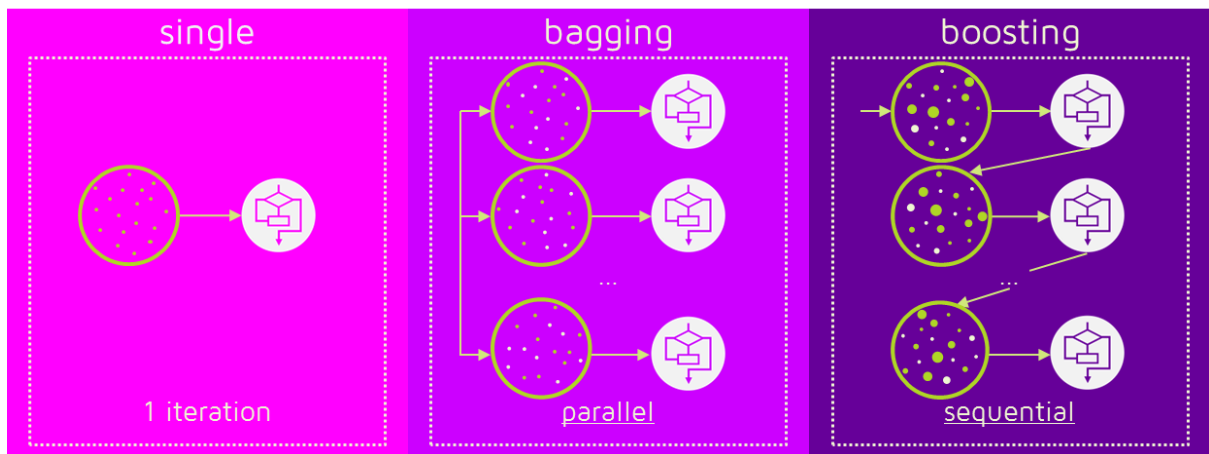


7장 앙상블 학습과 랜덤 포레스트

도입 앙상블(Ensemble)은 결정트리 뿐만 아니라, 다양한 알고리즘에 사용이 가능하며, 여러 알고리즘을 통해 더 좋은 예측을 할 수 있다. Bagging과 Boosting 등 다양한 방법이 있지만, 간단하게 생각하자면 하나가 아닌 여러 알고리즘을 통해 다수결 투표를 진행하여 더 나은 학습을 진행하는 것이라고 생각할 수 있다.

배깅과 부스팅



<https://quantdare.com/wp-content/uploads/2016/04/bb3.png>

부스팅은 배깅과 다르게 약한 학습기를 여러 개를 이용한다는 공통점이 있지만, 오답에 대해 가중치를 부여한다는 특징을 가지고 있다. 즉 앞서 알고리즘의 오차에 더 큰 가중치를 부여함으로써, 오차를 줄이는데 영향을 준다. 후에 이야기할 adaboost, xgboost 등이 있으며 이들은 높은 정확도를 보이지만, overfitting의 위험이 조금 더 클 수 있다.

랜덤 포레스트 – 랜덤포레스트는 배깅의 대표적인 방식으로, 말 그대로 랜덤하게 데이터를 추출하여 여러 결정트리를 통해 앙상블한다.

랜덤포레스트와 아다부스트 같은 앙상블 모델은 비선형 모델에 적합하며, SVM에 비하여 연산량이 적기 때문에 직관적이며 빠르다는 장점을 가지고 있다.