

## 앙상블(Ensemble)

목적

: 여러 분류기를 하나로 연결하여 개별 분류기보다 더 좋은 일반화 성능을 달성

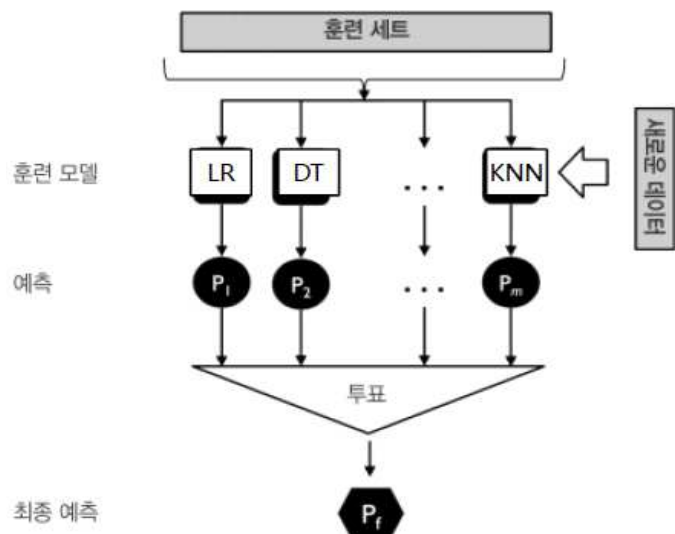
방법

1. 여러 분류 알고리즘 사용: 다수결 투표(Voting)
2. 하나의 분류 알고리즘을 여러 번 사용: 배깅(Bagging), 부스팅(Boosting)

### 다수결 투표(Majority Voting)

동일한 학습 데이터로 모델을 구축한다

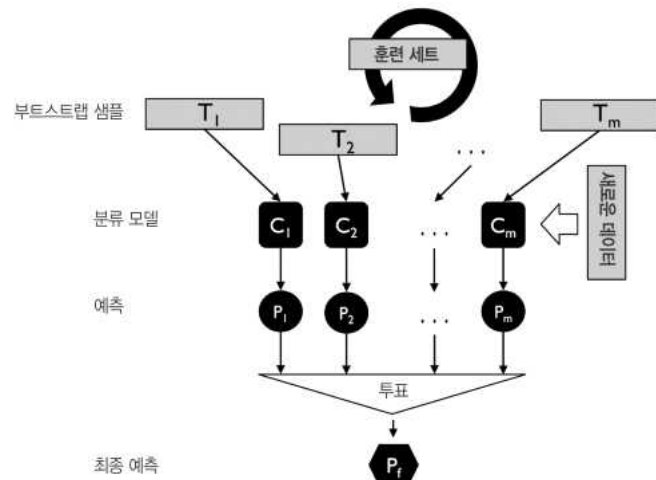
샘플을 뽑을 때 중복은 없다



### 배깅(Bagging)

알고리즘마다 서로 다른 학습 데이터를 추출(샘플링)하여 모델 구축에 사용한다

부트스트랩(Bootstrap): 사용하면 학습 데이터 샘플링 시 복원 추출(중복) 허용



## 랜덤 포레스트(Random Forest)

배깅의 일종

단일 분류 알고리즘(Decision Tree) 사용

Forest 구축: 무작위로 예측변수 선택하여 모델 구축

결합 방식: 투표(분류), 평균(예측)

## 부스팅(Boosting)

샘플을 뽑을 때 잘못 분류된 데이터의 50%를 재학습에 사용 혹은 가중치 적용

AdaBoost: 전체 학습 데이터를 사용하고 잘못 분류된 데이터에 가중치 적용한다

