

앙상블 방법

오류 결합의 기초

- 오류 결합 방법은 앙상블 방법론(ensemble methods)이라고 한다.
- 특정한 하나의 예측 방법이 아니라, 복수의 예측 모델을 결합하여 더 나은 성능의 예측을 하려는 시도이다.
- 오류 결합 방법은 크게 취합(aggregation) 방법론과 부스팅(boosting) 방법론이 있다.
 - 취합 방법론(aggregation): 사물론 오류의 결합이 이치적합리가 있음.
 - 부스팅 방법론(Boosting): 사물론 오류를 점진적으로 줄여간다.
- 각 방법론의 대표적인 방법들은 아래와 같다.
 - 취합 방법론(Aggregation)
 - 1) 다수결 (Majority Voting) < "hard voting"
 - 2) 배깅 (Bagging)
 - 3) 랜덤 포레스트 (Random Forests)
 - 부스팅 방법론(Boosting)
 - 1) 아다 부스트 (Ada Boost)
 - 2) 그레디언트 부스트 (Gradient Boost)
- 오류 결합 방법을 사용하면, 일반적으로 계산량은 증가하지만 다음과 같은 효과가 있다.
 - 단일 모델을 사용할 때보다 성능 분산이 감소하고, 즉 과적합화를 방지한다.
 - 개별 모델이 성능이 낮은 경우에는 결합 모델의 성능이 더 향상된다.

취합 방법론

- ① 다수결 방법 (Majority Voting) < "Hard Voting": 관순 투표, 개별 모델의 결과가
- < "Soft Voting": 가중치 투표, 개별 모델의 조건부 확률의 합 기준.

오류 결합을 사용한 성능 향상

- 다수결 모델이 개별 모델보다 더 나은 성능을 보이는 이유는 다음과 같다.

↳ 만약 어떤 개별 모델이 정확을 예측할 확률이 p 인 경우, 이러한 모델을 N 개 모아서 다수결 모델을 만들면 정확을 예측할 확률이 다음과 높아진다.

$$\Rightarrow \sum_{k \geq \frac{N}{2}} \binom{N}{k} p^k (1-p)^{N-k}$$

