

앙상블

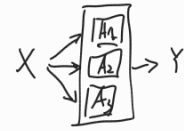
섞어 보겠다. (blending) 기존 model을 섞어(합쳐서) 성능을 리볼까?

앙상블 \Rightarrow 원딜 포지스트

집단리성. 한명의 전도가 VS 여러명의 일반인

\hookrightarrow 예측, 분류기 결합. best model 보다 더 좋은 예측 나올 수도.

투표 기반



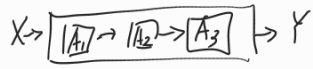
분류기 여러개

배깅/테스팅



데이터 여러개

부스팅



학습 방향

어디다
그래디언트
가중치 (다소 가중치)
오차 보정

- 투표 기반 분류기

정확도 80% 이상의 분류기 여러개 훈련. (ex) 로지스틱 회귀, 결정 트리, SVM

, 분류기의 다수결 투표. (가장 많이 선택된 클래스 예측)

장: 약한 학습기가 충분히 많고 다양 \Rightarrow 강한 학습기

단: 각 학습기가 독립적이고 문제의 상관관계가 없어야 한다. \Rightarrow 학습 데이터가 같으면 같은 종류 문제 생김, 잘못된 class 선택 가능.

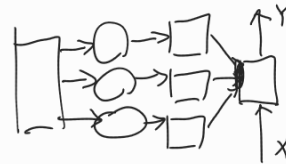
- 배깅, 테스트

훈련 데이터가 문제 서로 다른 훈련 알고리즘 \Rightarrow 무작위 훈련 데이터

배깅: 중복 허용 샘플링

테스팅: 중복 X 샘플링

\Rightarrow 모든 예측기의 예측을 다 써서 새로운 데이터에 대한 예측
CPU 코어 병렬 학습 가능

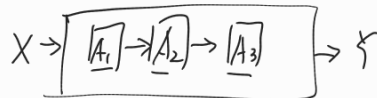


- 예측 수집 함수: 분류: 최빈값 \Rightarrow 편향/분산 조절 목적
회귀: 평균

- out of bag: 배깅을 사용할 때 사용되지 않는 데이터는 검증 데이터로 사용.

- 부스팅

약한 학습기 여러개 연결



어디다 부스트

- 과소적합 샘플의 가중치 더 높임 \Rightarrow 학습하기 어려운 샘플에 더 맞추어줌

- 경사하강법과 유사하나 경사하강법은 cost 함수 최소화, 어디다 부스트는 성능 향상 위해 예측기 추가 목적

• 그래디언트 부스트

- 오차 보정하도록 예측기 순차적 추가 (잔여 오차에 새로운 예측기 학습)

- 확률적 그래디언트 부스팅: 훈련 샘플 바뀔 주정

ex) XG Boost