Gradient boosting – best function/classifier

Didn’t touch the first model.

Residual-MSE

-앙상블들의 앙상블(모델 스태킹)-진짜 중요함! 꼭 코드 익힐 것

분류기들 혹은 모델의 결과를 결합한다는 것

그 결과들로 하나의 모델을 만든다는 것

서로 다른 분류기 형태 간의 앙상블들의 앙상블

학습 시킨다음에 결과를 조합해서 또다른 모델에 그결과를 넣는 것

앙상블들의 앙상블은 스태킹이라고 함

블랜딩과 스태킹을 나누는 것은 의미 없음

* 블렌딩

훈련데이터를 반으로 조깬 다음

훈련 데이터 1/ 훈련데이터 2- 블랜딩- 각

레이어가 두개임: 블렌딩

첫번째 층은 여러 개의 모델을 이용해서 학습을 시킴/모델의 수만큼 결과가 나옴

첫번째 레이어와 두번째 레이어 에서 겹쳐 지는 훈련 데이터는 없어야 함.

* 스태킹 version 1

케이 폴드를 사용 함.

밸리데이션 체크’

학습 셋을 케이개로 나눠서 학습을 시켜

같은 크기의 차원만 다름

트레인 셋은 크로스 벨리데이션 기술을 이용해서 사이즈가 같게 나옴

* 블랜딩과 스태킹의 차이는 두개나 케이개 냐 임.
* 케이 폴드는 학습되지 않은 데이터를 가지고 다른 모델의 인풋데이터로 사용하기 위함\
* Stacking version2

테스트 데이터를 학습시키지 않음.

예측 결과 3개 + 테스트 셋 데이터와 트레인 데이터 전체를 이용해서 예측 결과를 만드는 것

예측 결과 전체를 이용해서 레이어 2 모델에 학습 시키고, 최종 예측 결과 모두를 가지고 성능을 평가하는 것

레이어 2 모델은 하나의 모델(메타 분류기 하나임)\_일반적으로

휴리스티한 모델

* 동일 형식 분류기와 부트 스트래핑을 이용한 앙상블의 앙상블

항상 예측 결과를 가져오거나 퍼포먼스를 측정할 때, 항상 학습 시키지 않은 데이터를 가지고 측정한다는 것이 중요함!

모델은 동일한데 데이터 셋이 다른 경우\_ 앙상블을 만들 수 있음

레이어 1의 모델이 전체 동일, 대신 모델에 적용하는 데이터가 다름(부트 스트랩)

학습 데이터에다가 훈련데이터 2를 넣어서 예측값을 뽑아내고

그 예측값을 모아서 다수결로 결정함

분류기(모델) 들을 결합하여 하나의 결과를 내는 것