

CatBoost

앙상블의 종류

보팅 스택킹 배깅 부스팅

보팅

말그대로 같은 train데이터로 훈련시킨모델 다른 알고리즘 기반

소프트 보팅 - 클래스당 예측확률!

하드 보팅 - 마지막 predict로

배깅

랜덤포레스트.

데이터 복원추출이나 비복원추출해서 의사결정나무 돌리는것

스택킹

train data에 한해서 정확도가 높다 다만 과적합의 위험성..

이걸 보완하기 위해서 나온게 cv세트기반 스택킹..

이걸 cv로 과적합을 막는데

그 cv도 데이터 불균형이 있을수있으므로

계층별 k겹교차검증을 한다...

부스팅..

xgboost lightGBM Catboost

GB는 손실함수의 미분값 기울기...

Gradient + boosting

정확도가 높아진다는것은 과적합이 된다는거다?

그러므로 부스팅 계열은 정규화 알고리즘이 꼭필요하다

학습률을 예측된 값에 곱해가면서 기존 answer값과 빼가면서 잔차를 줄여가는 식으로 예측을한다.

catboost 는 시계열 데이터를 처리를 잘한다 왜지?

과거데이터로 response ending한다고