

예습과제3 DA박보영

GBM

약한약습기에 가중치 부여를 통해 오류를 개선해 나감, 가중치 업데이트를 경사하강법 이용.

오류식=실제값-예측값

장) 예측 성능 뛰어남, 과적합에도 강함.

단) 수행시간 오래걸림.

XGBoost (\in GBM)

장) 예측 성능 뛰어남, 빠른 수행시간, 과적합 규제, tree pruning, 자체내장된 교차검증, 결손값 자체 처리

파라미터

booster, silent, nthread, eta, num_boost_rounds, min_child_weight, gamma, max_depth, sub_sample, colsample_bytree, lambda, alpha, scale_pos_weight, 조기중단 학습, 데이터셋용 DMatrix, data, label 입력

피쳐명을 f0,1,....

sklearn프레임워크와 연동

LightGBM(\in GBM)

장) 학습시간 단축, 적은 메모리 사용량, 예측 성능 뛰어남,

단) 쉬운 과적합 발생

리프중심트리불할(최대손실값을 가지는 리프노드를 지속적으로 분할하면서 트리가 깊어짐)

튜닝:num_leaves, min_data_in_leaf, max_depth

스태킹앙상블

개별알고리즘으로 예측한 데이터를 기반으로 다시 예측을 수행.

두 개의 모델 필요. 성능 높힐 때 사용

과적합방지, 개별 모델들이 각 교차검증으로 메타모델을 위한 학습,테스트용으로 생성. 이를 기반으로 메타 모델이 학습,예측 수행.