

05

GBM

: weak learner ← 가중치 부여해 오류 개선

• AdaBoost : 오류데이터에 가중치 → \sum
대해 다음 분류마다 잘 학습

• GradientBoost : Gradient Descent
손실함수최소화 ($h(x) = y - F(x)$)
 Δ 수렴시간↑ (병목X)

→ params

loss = 'deviance' : cost func.

learning-rate = 0.1

n_estimators = 100

subsample = 1 : train data 사용

06

XGBoost : 트리 기반 앙상블, GBM 기반

Δ 병목X → split

교차할 규제

Tree pruning : 중요도 높은 분할 out

교차검증 + 조기중단 (정확도, 과피해와 같은 것)
적은값 자체처리

Δ C/C++ 기반의 코어 lib

eval-set
eval-matrix

파라미터명 | → 파라미터 pack : 미리 호환X

치이 | → skl pack : X, XGBClassifier
XGBRegressor

• 파이썬 wrapper params : 일반 / 부속 / 학습테스트

↳ DMatrix

• sklearn Wrapper XGB

* params 명 차이

n_estimators = num_boost_round
↓ skl ↓ ps

조금만 → fit() 안에 작성

07

LightGBM : time ↓ mem ↓ 기증 ↑

* 리프노드심도제한 → depth ↑ 조절됨.

마관까지, skl 호환 → LGBMClassifier

LGBM Reg

08

평가) ROC-AUC 사용 : 불균형

09

under sampling : 많음 data → 적음 data

* over sampling ←

↳ SMOTE : 적은 data 많게 생성 (원본 data 수치 조절)

: imbalanced-learn, fit_sample로 증식

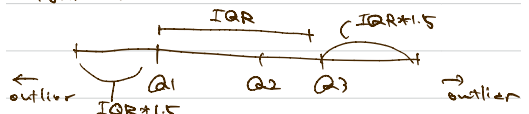
Δ 개체 ↑ 정도 ↓

불균형 LGBM 생성시

boost-from-average = False 설정
↓ X

scaling → (log4p) 사용

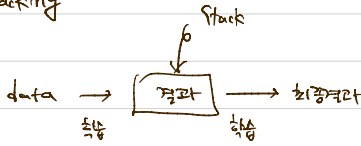
• 이상치데이터



Q1, Q3 값 ← np.percentile()

10

Stacking



CV set 개수

폴드별 학습 \rightarrow 회귀 meta model's test/train set

