**부스팅(GBM, XGBoost, LGBM)**

**부스팅 알고리즘**

**Theory**

부스팅 알고리즘은 여러 개의 약한 학습기(Weak Learner)를 순차적으로 학습 예측하면서 잘못 예측한 데이터에 가중치 부여를 통해 오류를 개선해 나가면서 학습하는 방식입니다.

부스팅의 대표적인 구현은 AdaBoost(Adaptive boosting)와 그레디언트 부스트가 있습니다. AdaBoost는 오류 데이터에 가중치를 부여하면서 부스팅을 수행하는 대표적인 알고리즘입니다.

**GBM(Gradient Boosting Machine)**

**Theory**

GBM도 AdaBoost와 유사하나, 가중치 업데이트를 경사 하강법(Gradient Descent)을 이용하는 것이 큰 차이입니다. 간단히 말해, 반복 수행을 통해 오류를 최소화할 수 있는 방향으로 가중치를 업데이트하며 학습을 이어가는 방법입니다.

일반적으로 GBM이 RF보다는 예측 성능이 뛰어난 경우가 많습니다. 하지만, 수행 시간이 오래걸리고, 하이퍼 파라미터 튜닝 노력도 많이 필요합니다. 특히 수행 시간 문제는 GBM이 극복해야 할 중요한 과제입니다.

**하이퍼파라미터**

Loss : [‘log\_loss’, ‘deviance’, ‘exponential’] default = ‘log\_loss’

**Learning\_rate : 0~1, default = 0.1**

**n\_estimators : default = 100**

subsample = 0~1, default = 1

max\_depth, min\_samples\_leaf, min\_samples\_split

**XGBoost(eXtra Gradient Boost)**

**Theory**

XGBoost는 트리 기반의 앙상블 학습에서 가장 각광받고 있는 알고리즘 중 하나로 GBM의 단점인 느린 수행 시간 및 과적합 규제(Regularization) 부재 등의 문제를 해결하여 다른 머신러닝 모델보다 우수한 성능을 나타냅니다. **( XGB = GBM + regularization + early\_stopping )**

**하이퍼파라미터**

**Learning\_rate : 0~1, default = 0.1**

**n\_estimators : default = 100**

subsample : 0~1, default = 1

max\_depth : default = -1

**조기중단 관련 파라미터**

Early\_stopping\_rounds : default = 100

Eval\_metric : [‘log\_loss’, ‘deviance’, ‘exponential’] default = ‘log\_loss’

Eval\_set = [ (X\_test, y\_test) ]

**LGBM(Light Gradient Boosting Machine)**

**Theory**

LGBM은 XGBoost의 장점은 계승하고 단점은 보완하는 방식으로 개발됐기 때문에 예측 성능은 비슷하지만, 훨씬 다양성 있는 기능을 제공합니다. LGBM의 가장 큰 장점은 XGBoost보다 학습에 걸리는 시간도 적을 뿐만 아니라, 메모리 사용량도 상대적으로 적다는 점입니다.

또한, LGBM은 일반 GBM 계열의 트리 분할 방법과 다르게 리프 중심 트리 분할 방식을 사용함으로, 비대칭적인 규칙 트리가 생성됩니다. 하지만 이렇게 최대 손실값을 가지는 리프 노드를 지속적으로 분할해 생성된 규칙 트리는 학습을 반복할수록 결국 균형 트리 분할 보다 예측 오류 손실을 최소화 할 수 있다는 것이 LGBM의 기본 사상입니다.

**하이퍼파라미터**

**Learning\_rate : 0~1, default = 0.1**

**n\_estimators : default = 100**

subsample : 0~1, default = 1

num\_leaves : default=31

min\_child\_samples : default=20

max\_depth : default= -1

**조기중단 관련 파라미터**

Early\_stopping\_rounds : default = 100

Eval\_metric : [‘log\_loss’, ‘deviance’, ‘exponential’] default = ‘log\_loss’

Eval\_set = [ (X\_test, y\_test) ]

텍스트이(가) 표시된 사진

자동 생성된 설명