Team 22

1973099 김현지, 2076362 전재은

RESEARCH POSTER

XGBoost model에서 자동 최적화 알고리즘의 하이퍼파라미터 탐색 범위 및 자동 최적화 성능 분석

Definition of Hyperparameter

Hyperparameter은 ML 모델을 구성하거나
손실 함수를 최소화하는 알고리즘을 지정하는 데 사용되는 parameter
모델의 성능에 큰 영향을 미치는 요소 중 하나

Definition of Hyperparameter tuning

모델을 최적화하기 위해 hyperparameter을 조정하는 과정

• Hyperparameter Tuning

보이지 않는 데이터에서 잘 작동하는 견고한 model을 구축하기 위해 필요하다.

Hyperparameter 설정에 따라 model의 성능이 상이하므로 Machine learning 또는 deep learning model을 효과적으로 활용하려면 적절한 hyperparameter의 선택이 필요하다.

• Automated Hyperparameter Tuning을 해야 하는 이유

'rules of thumb'라고도 불리는 Manual Hyperparameter Tuning은 경험 또는 직감으로 hyperparameter 값을 설정하는 방식을 말한다.

Manual Hyperparameter Tuning은 다수의 hyperparameter, 복잡한 model, 시간이 오래 걸리는 model 평가, 비선형 hyperparamer 상호 작용 등 특정 요인으로 인해 발생하는 많은 문제에 대해 비효과적이다.

인간의 판단에 편향되어 있기 때문에 Automated Hyperparameter Tuning이 더 효과적이다.

• Automated Hyperparameter Optimization에서 Hyperparameter 탐색 범위의 영향

실험에서 사용되는 hyperparameter의 탐색 범위는 model의 성능에 직접적인 영향을 미친다.

범위가 너무 좁으면 최적의 hyperparameter을 놓칠 수 있고, 반대로 범위가 너무 넓으면 계산 비용이 늘어날 뿐만 아니라 부적절한 설정이 발생할 수 있다.

• 실험 목적

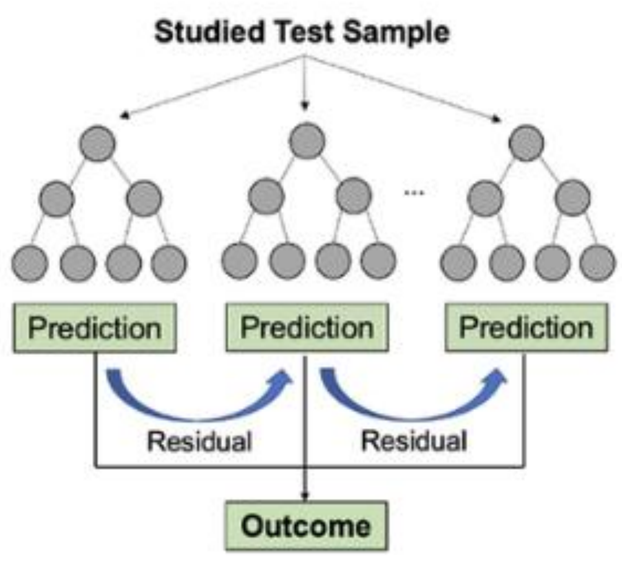
XGBoost model을 이용해서 Automated Optimization Methods (Grid Search, Random Search, Hyperopt, Optuna)의 효과적인 hyperparameter 탐색 범위를 실험해보고 최적화 성능을 분석해보고자 한다.

• 실험의 배경지식

1. XGBoost model

여러 개의 decision tree가 보여 boosting 앙상블로 구현된 ML 모델

이전 모델의 negative gradient를 줄이는 방향으로 학습한다.



2. Automated Optimization Methods

1) Grid Search

가장 기본적인 Hyperparameter Optimization Method

가능한 모든 조합의 hyperparameter로 훈련시켜서 최적의 조합을 찾는다.

2) Random Search

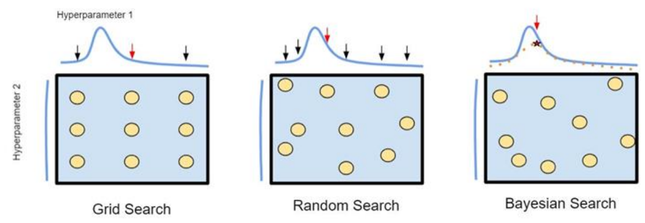
사용자가 지정한 hyperparameter 탐색 범위 내에서 임의의 조합을 추출하여 최적의 조합을 찾는 방법

Grid Search와 마찬가지로 최적의 hyperparameter를 찾기 위해 많은 경우의 수가 존재한다.

3) Hyperopt

Bayesian Optimization Model을 기반으로 한 Automated Hyperparameter Tuning Framework

Bayesian Optimization은 목적함수를 최대 또는 최소로 하는 최적해를 찾는 방법을 말한다.



4) Optuna

Hyperparameter Optimization Task를 자동화해주는 Framework

• 실험 방법

1. 사용한 Dataset

주택별 건축연도, 설비, 넓이, 지역 등 설명변수가 79개이며, 목적변수는 주택가격이다. 총 1460채에 대한 주택 정보가 학습 데이터로 주어지는 데이터의 크기가 큰 dataset이다.

2. 사용한 예측 model과 Automated Optimization Algorithm

주택 가격을 예측하는 모델은 XGBoost 모델이고, 실험에 사용되는 자동 최적화 알고리즘은 Grid Search, Random search, Hyperopt, Optuna이다.

3. 실험 과정

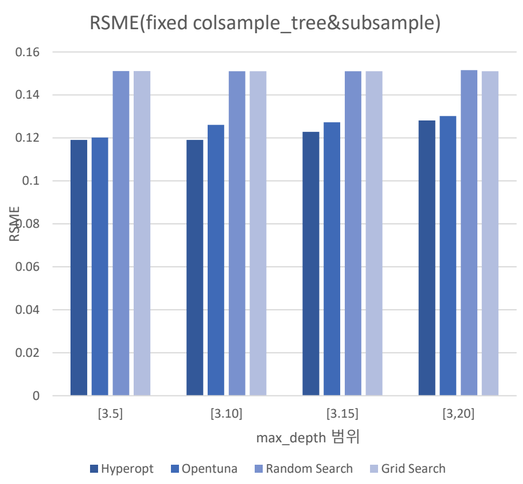
자동 최적화를 위한 목적함수를 정의하는 과정에서 하이퍼파라미터 공간을 정의할 때, 모델 트리의 깊이 (max_depth), 각 트리를 구성하는 데 사용할 feature의 비율 (colsample_bytree) 변수를 조정하여, 자동 최적화 알고리즘 별 RSME(Root Mean Squared Error)를 비교하였다.

데이터 중 샘플링을 하는 변수인 sub_sample은 0.8, 학습 속도를 조절하는 변수인 learning rate은 0.05로 고정하였다.

a) 첫번째 실험에서 colsample_bytree의 값은 0.8, max_depth의 최소 tree depth 값은 3으로 고정하고, 최대 tree depth 값을 20까지 5씩 늘려가며 hyperparameter의 탐색 범위를 조정하였다.
b) 두번째 실험에서는 max_depth의 값을 5로 고정하였다. Colsample_bytree의 범위는 0.5로 최소로 하고 최대 허용범위를 0.7에서 1.0까지 조정하였다.
c) 세번째 실험에서 colsample_bytree는 0.2에서 0.9 사이의 어떤 값이든 선택될 수 있도록 하였다. Max_depth의 값은 첫번째 실험과 동일하게 조정하였다.

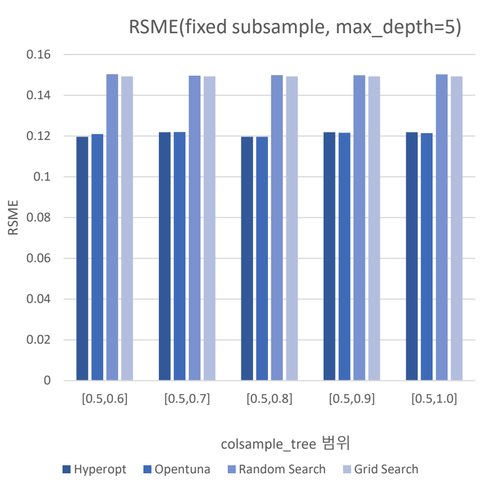
• 실험 결과

a) 첫번째 실험



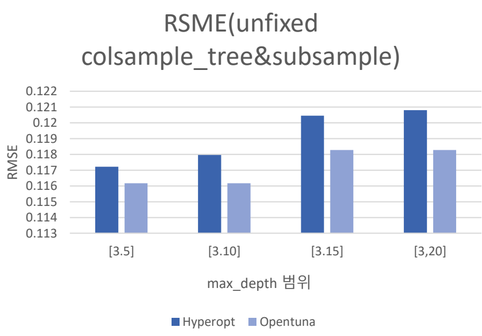
자동 최적화를 통한 최적의 model depth는 알고리즘별로 각각 3, 4, 5, 5로 나왔다. RSME는 Hyperopt, Optuna 순으로 높았고, Random Search와 Grid Search의 값은 동일하게 나왔다.

b) 두번째 실험



Grid Search 알고리즘의 RSME는 colsample_bytree의 범위를 조정해도 줄어들지 않았다. Hyperopt에서 RSME 값이 적었던 colsample_bytree의 범위는 [0.5,0.6], [0.5,0.8]이었고, Optuna에서는 [0.5,0.8], Random Search에서는 [0.5,0.7]이었다.

c) 세번째 실험



자동 최적화를 통한 최적의 model depth는 Hyperopt, Optuna 모두 3이었으며, Optuna의 RSME가 더 작았다.

• 결론

XGBoost 모델의 자동 최적화 알고리즘으로 사용할 feature의 비율을 고정하고 모델 트리 깊이가 고정되지 않았을 때는 Hyperopt, Optuna 순으로 적절했다.

모델 트리 깊이를 고정하고, 사용할 feature의 비율을 달리하였을 때는 알고리즘 별로 RSME가 적은 범위가 달랐다.

하지만, 이때도 Hyperopt, Optuna 순으로 RSME가 작았고, 적절한 알고리즘이었다.

Hyperopt와 Optuna에서 사용할 feature의 비율을 유연하게 바꾸고 모델 트리 깊이를 최적화했을 때 Optuna의 성능이 더 좋았다.

• 참조

Li Yang, Abdallah Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," Neurocomputing, vol. 415, 2020, pp. 295-316, https://doi.org/10.1016/j.neucom.2020.07.061.
Zachary Warnes, "Hyperparameter Tuning — Always Tune your Models", Medium, July 7, 2021, https://towardsdatascience.com/hyperparameter-tuning-always-tune-your-models-7db7aeaf47e9.
Petru Potrimba, "What is Hyperparameter Tuning? A Deep Dive.", roboflow, June 16, 2023, https://blog.roboflow.com/what-is-hyperparameter-tuning/.
Jakub Czakon, "Optuna vs Hyperopt: Which Hyperparameter Optimization Library Should You Choose?", neptune.ai, October 20, 2023, https://neptune.ai/blog/optuna-vs-hyperopt.