

モデル解説

実装しているモデル

- RandomForestClassifier
 - ✓ 分類木
- RandomForestRegressor
 - ✓ 回帰木
- Xgboost
 - ✓ ハイパーパラメータにて分類回帰を選択

RandomForestのパラメータ

パラメータ名	説明
n_estimators	実装する木の数(決定木をいくつ作るか?)
criterion	分割する際の基準(分岐する際にこういった基準で分割するか)
max_depth	木の深さ(決定木がどれくらい条件分岐するか?)
min_samples_split	分岐する際のサンプル数(いくつのデータまで分岐するか?)
max_leaf_nodes	葉の数の上限(決定木のたどり着く結論の数)

※ max_depthは数値を入力しない場合、深さは無制限になります

※ max_leaf_nodesは数値を入力しない場合、葉の数は無制限になります

Xgboostのパラメータ

パラメータ名	説明
max_depth	木の深さ(決定木がどれくらい条件分岐するか?)
learning_rate	学習率の割合(小さいほうが過学習しにくい)
objective	決定木の種類(Regressorで回帰木、Classifierで分類木)
num_round	実装する木の数(決定木をいくつ作るか?)
subsample	決定木にて使用するデータの割合 (あえて使えるデータを制限することで過学習を防止できます)

引用、参考資料

ランダムフォレストで分類するときの過学習対策の検討

<https://www.haya-programming.com/entry/2019/06/22/235058>

決定木分析についてざっくりまとめ_理論編

https://dev.classmethod.jp/articles/2017ad_20171211_dt-2/

XGBoostのパラメータチューニングまとめ

<https://qiita.com/c60evaporator/items/a9a049c3469f6b4872c6>

XGBoostパラメータのまとめとランダムサーチ実装

<https://qiita.com/FJyusk56/items/0649f4362587261bd57a>