
SIGNATE Student Cup 2021春

1st place solution

鈴木 克弥(Ka28)

目次

- 自己紹介
- モデル構成
- 特徴量
- 最終サブには選べなかったが最もスコアの良かったモデル
- 初心者がこの解法をサブミットするまでのロードマップ

自己紹介

氏名：鈴木 克弥 (Ka28)

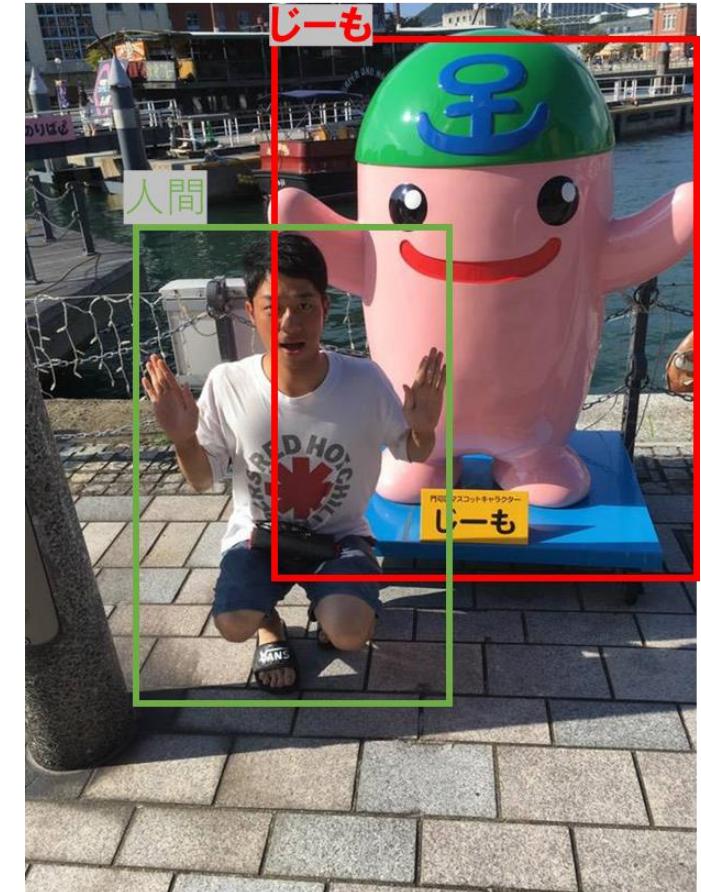
所属：宮崎大学大学院 工学研究科

工学専攻 機械・情報系コース

計測制御学研究室 M2

インターン：Datumix株式会社

研究：物体検出



* 注意 *

- 中級者以上の方々は今回の解法から学ぶことはないかもしれませんが怒らないでいただきたい。



モデル構成

Stratified Kfold(n=5)

- lightGBM(cv:0.6133)
- kNN (0.6530)
- kNN+lightGBM (0.6450)
- kNN+MLP(4層) (0.6030)

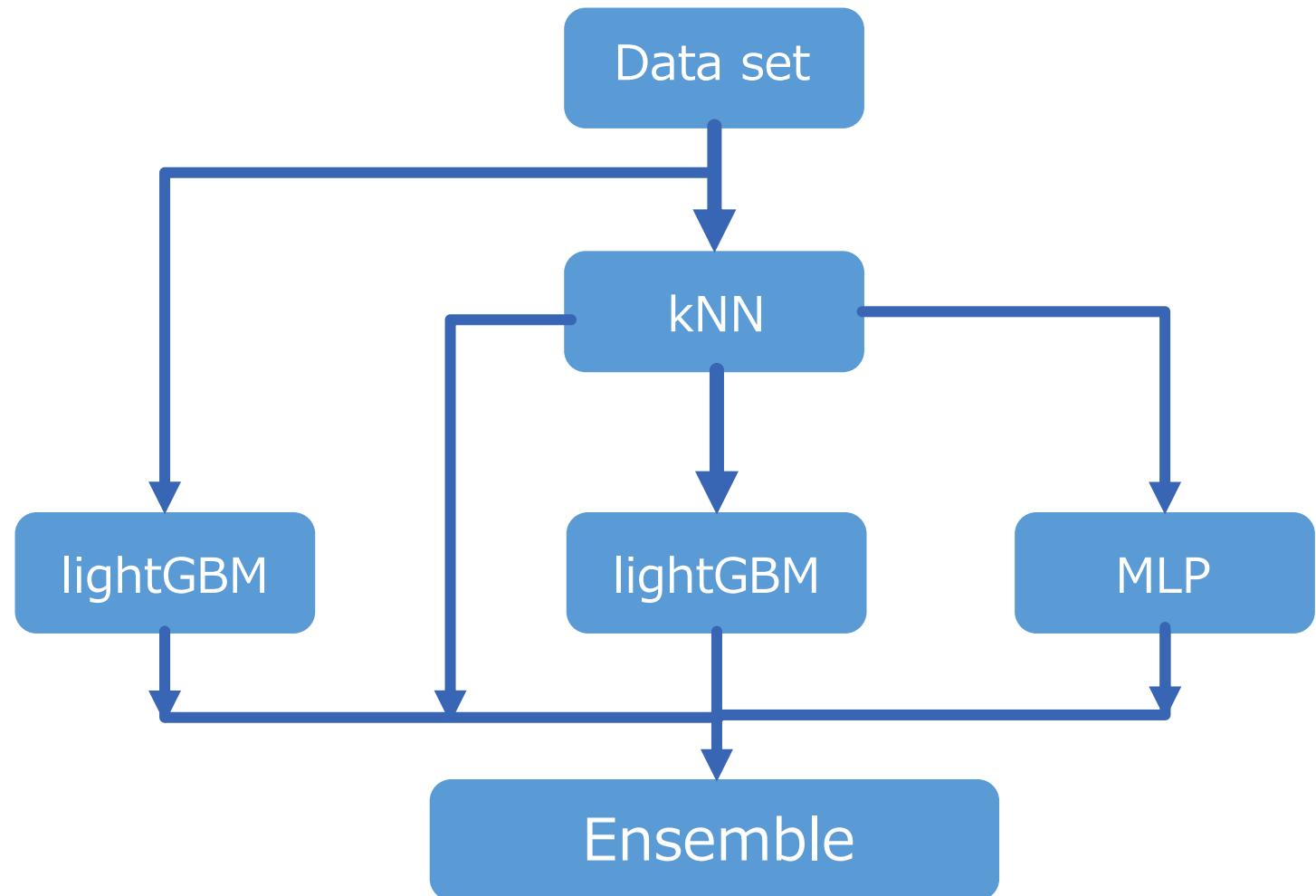
Ensemble:多数決

最終スコア

CV : 0.6567

Public : 0.6721

Private : 0.7098



特徴量

- lightGBM (<https://signate.jp/competitions/449/discussions/lgbm-baseline-lb06240>)
masato8823さんのフォーラム (LGBM Baseline)
- kNN (<https://signate.jp/competitions/449/discussions/knn-baseline-cv06256-lb06417>)
nagissさんのフォーラム (kNN Baseline)
- kNN+lightGBM (<https://signate.jp/competitions/449/discussions/lgbmknn-lb06567-1>)
nagissさんのフォーラム (LGBM+kNN)
- kNN+MLP
同上

最終サブでは選べなかったが効果的だったもの

- K-meansでの各クラスタの中心点からの距離特徴量

K-means特徴量	CV	Public	Private
あり	0.6622	0.6389	0.6980
なし	0.6133	0.6306	0.6500

- Catboostもアンサンブルに含める

CatBoost	CV	Public	Private
あり	0.6743	0.6561	0.7129
なし	0.6567	0.6721	0.7098

完全初心者がこの解法を実装するまでの ロードマップ

- kaggleスタートブック ← 流れをつかむ
- kaggleで勝つデータ分析の技術 ← ざっと目を通してコンペ参加しながら読み返す（評価指標、特徴量作成、モデルの評価）
- Pytorch公式チュートリアル ← 簡単なモデルを組めるようになる

https://yutaroogawa.github.io/pytorch_tutorials_jp/
- コンペに参加しフォーラムやDiscussionを読む
- Twitterでコンペ強そうな人をフォローする

お金をかけたくない!!

- SIGNATE Campusに登録
(<https://campus.signate.jp/campaigns/referral/1000/7E506040C13218qt>)
- 学生なら講座が無料で受けられる
- Pythonデータ分析コースLv.1をとりあえずやってみる
- 何かしらのコンペに参加



運営の皆さん、 参加者の皆さん
ありがとうございました！