

*KDD2018*読み会

複数のデータセットを横断して重要な超パラメータの分析

## Hyperparameter Importance Across Datasets

Jan N. van Rijn

Albert-Ludwigs-Universität Freiburg

Freiburg, Germany

vanrijn@cs.uni-freiburg.de

Frank Hutter

Albert-Ludwigs-Universität Freiburg

Freiburg, Germany

fh@cs.uni-freiburg.de

読み手：鹿島



# Frank Hutter

フォロー

Professor of Computer Science, [University of Freiburg](#), Germany

確認したメールアドレス: [cs.uni-freiburg.de](mailto:cs.uni-freiburg.de) - [ホームページ](#)

[Machine Learning](#) [Artificial Intelligence](#) [Automated Algorithm Design](#) [Combinatorial Optimization](#)

## タイトル

## 引用先

## 年

### Sequential model-based optimization for general algorithm configuration

F Hutter, HH Hoos, K Leyton-Brown  
International Conference on Learning and Intelligent Optimization, 507-523

767 2011

### ParamLS: an automatic algorithm configuration framework

F Hutter, HH Hoos, K Leyton-Brown, T Stützle  
Journal of Artificial Intelligence Research 36, 267-306

674 2009

### SATzilla: portfolio-based algorithm selection for SAT

L Xu, F Hutter, HH Hoos, K Leyton-Brown  
Journal of artificial intelligence research 32, 565-606

652 2008

### Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms

C Thornton, F Hutter, HH Hoos, K Leyton-Brown  
Proceedings of the 19th ACM SIGKDD international conference on Knowledge ...

447 2013

### Automatic algorithm configuration based on local search

F Hutter, HH Hoos, T Stützle  
Aaai 7, 1152-1157

247 2007

### Scaling and probabilistic smoothing: Efficient dynamic local search for SAT

F Hutter, D Tompkins, H Hoos  
Principles and Practice of Constraint Programming-CP 2002, 241-249

228 2002

### Algorithm runtime prediction: Methods & evaluation

F Hutter, L Xu, HH Hoos, K Leyton-Brown  
Artificial Intelligence 206, 79-111

218 2014

### Efficient and robust automated machine learning

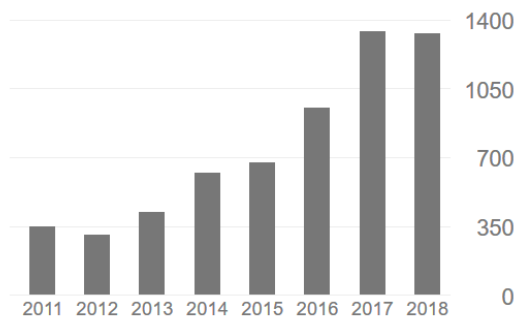
M Feurer, A Klein, K Eggensperger, J Springenberg, M Blum, F Hutter  
Advances in Neural Information Processing Systems, 2962-2970

193 2015

## 引用先

[すべて表示](#)

	すべて	2013 年以來
引用	6938	5402
h 指標	40	36
i10 指標	69	68



## 共著者

[すべて表示](#)

- Holger Hoos**  
Professor of Computer Science, ...
- Kevin Leyton-Brown**  
Department of Computer Scienc...
- Lin Xu**  
University of British Columbia
- Katharina Eggensperger**  
PhD Student, Machine Learning ...
- Aaron Klein**  
PhD Student, University of Freib...
- Marius Lindauer**  
University of Freiburg

# この論文の問いかけ： ある機械学習手法のどの超パラメータが重要？

- 機械学習手法にはいろいろな超パラメータがある：
  - SVM：（RBFカーネルの）幅や正則化パラメータ、縮小ヒューリスティクスの使用など
  - ランダムフォレスト：分岐の基準や、各決定木での特徴量の使用割合、葉頂点の最小データ数など
  - その他、手法とは独立に、欠損値の埋め方なども超パラメータとみなせる
- 超パラメータは「民間伝承」をもとに場当たりに決定される
- リサーチクエスチョン：
  - どの超パラメータが重要？
  - （さらに）それらの値域のどのあたりが重要？


## 関連研究との比較：

データ毎でなく学習手法レベルでの知見を得るのが目標

- 古くは： ランダムフォレストの重要パラメータの分析
- 最近では：
  - AutoML： 実験を繰り返しながら特定のデータセットに最も良い超パラメータの値を見つける)
  - メタ学習： データセットのメタデータから良い超パラメータ値を予測
  - これらはその場限りの低レベルな知識
- 本研究では： データセットの別を超えた、一般的な、より抽象度の高い知見（「SVMではカーネル幅が重要」）を見つける

## 結果：

### 有効な超パラメータの特定と、その知見を使った精度向上

- OpenMLの100データセットを使ってfunctional ANOVA分析
  -  OpenML：様々なデータを様々なに分析した結果が集まっている
- 超パラメータ分析の結果：
  - SVMでは正則化パラメータとカーネル幅
  - ランダムフォレストでは各決定木で使用する特徴量数と、葉の最大データ数
  - AdaBoostでは学習率と各決定木の最大深さが重要であることがわかり、民間伝承に一致
- さらに、よく効きそうな領域を事前分布として入れると精度向上

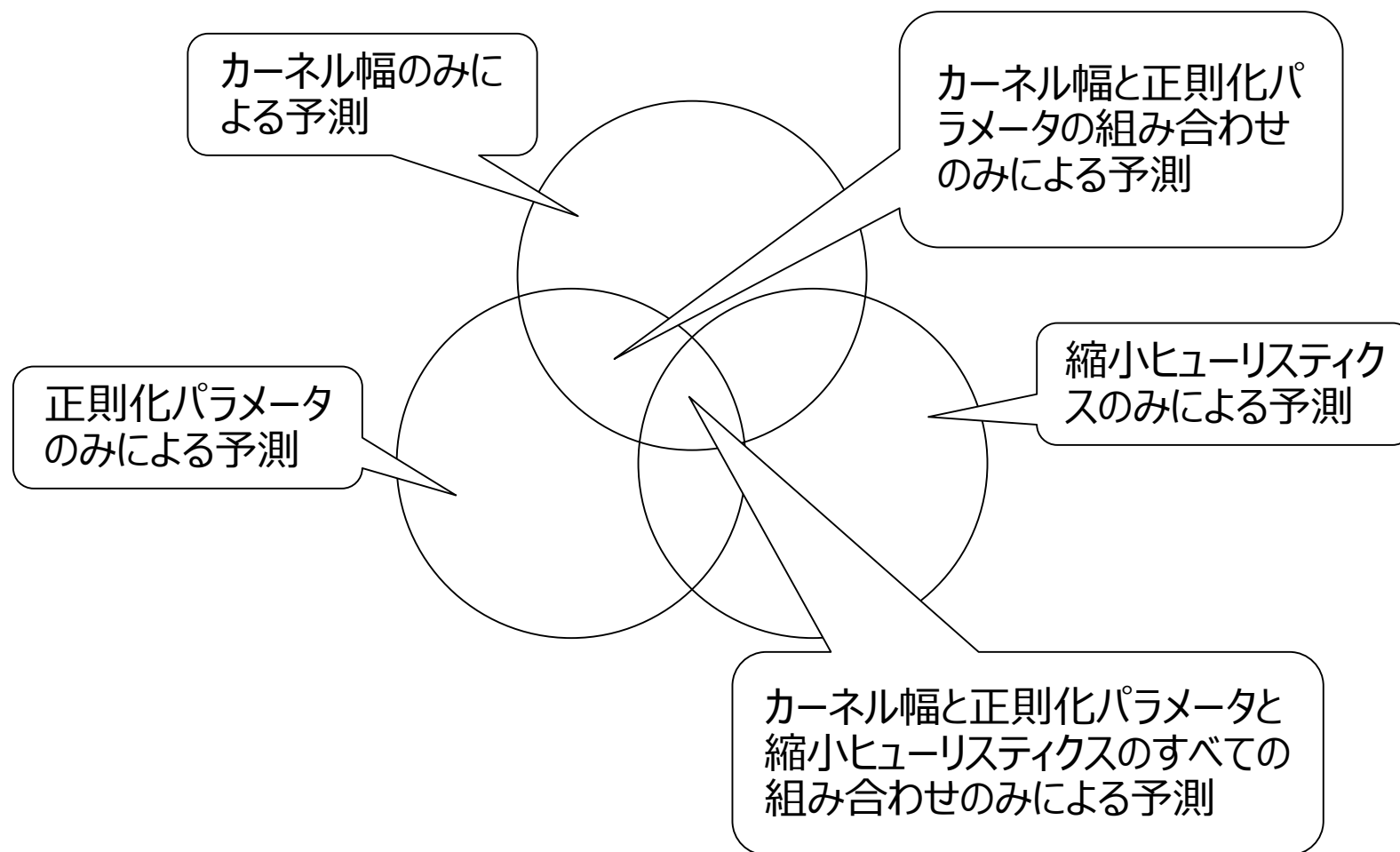
# Functional ANOVA :

## 超パラメータの予測精度への貢献度を測れる分析法

- 目的：超パラメータ（の組み合わせ）の予測精度への影響を推定
  - それぞれの超パラメータの貢献度
  - 超パラメータ2つの組み合わせの貢献度
    - 単一超パラメータの貢献を差し引いた貢献度として定義
  - 超パラメータ3つの組み合わせの貢献度
    - 超パラメータの2つの組み合わせの貢献は差し引いて定義
  - ...
- 方法：Functional ANOVA によってそれぞれの貢献が取り出せる
  - 「予測精度の分散」への寄与大→動かすと精度が大きく変わる

# Functional ANOVA : モデル分解のイメージ

- モデルを超パラメータの組み合わせの貢献分に分解する



# 補足： 各手法のハイパーパラメータ

**Table 1: SVM Hyperparameters.**

hyperparameter	values	description
complexity (or: 'C')	$[2^{-5}, 2^{15}]$ (log-scale)	Soft-margin constant, controlling the trade-off between model simplicity and model fit.
coef0	$[-1, 1]$	Additional coefficient used by the kernel (sigmoid kernel only).
gamma	$[2^{-15}, 2^3]$ (log-scale)	Length-scale of the kernel function, determining its locality.
imputation	{mean, median, mode}	Strategy for imputing missing numeric variables.
shrinking	{true, false}	Determines whether to use the shrinking heuristic (introduced in [24]).
tolerance	$[10^{-5}, 10^{-1}]$ (log-scale)	Determines the tolerance for the stopping criterion.

**Table 2: Random Forest Hyperparameters.**

hyperparameter	values	description
bootstrap	{true, false}	Whether to train on bootstrap samples or on the full train set.
max. features	$[0.1, 0.9]$	Fraction of random features sampled per node.
min. samples leaf	$[1, 20]$	The minimal number of data points required in order to create a leaf.
min. samples split	$[2, 20]$	The minimal number of data points required to split an internal node.
imputation	{mean, median, mode}	Strategy for imputing missing numeric variables.
split criterion	{entropy, gini}	Function to determine the quality of a possible split.

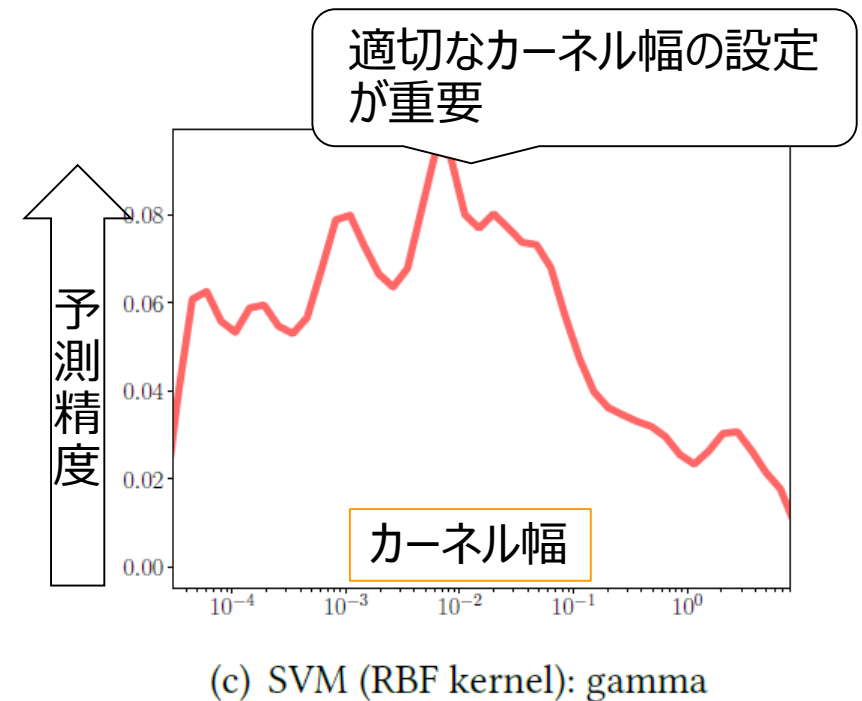
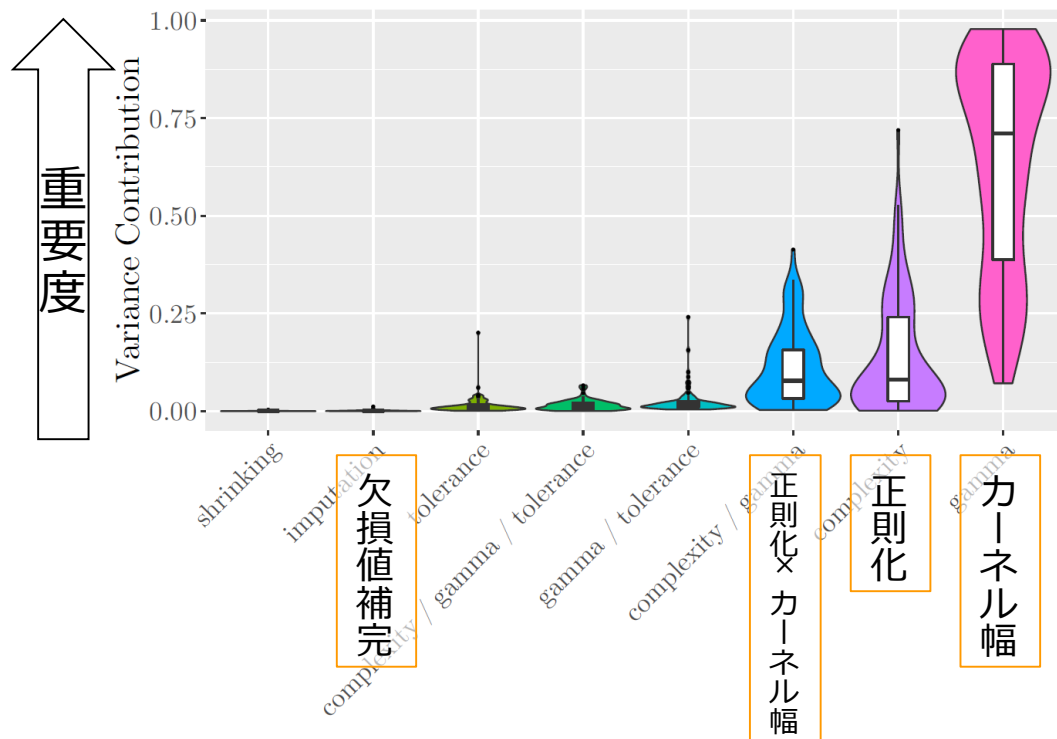
**Table 3: Adaboost Hyperparameters.**

hyperparameter	values	description
algorithm	{SAMME, SAMME.R}	Determines which boosting algorithm to use.
imputation	{mean, median, mode}	Strategy for imputing missing numeric variables.
iterations	$[50, 500]$	Number of estimators to build.
learning rate	$[0.01, 2.0]$ (log-scale)	Learning rate shrinks the contribution of each classifier.
max. depth	$[1, 10]$	The maximal depth of the decision trees.



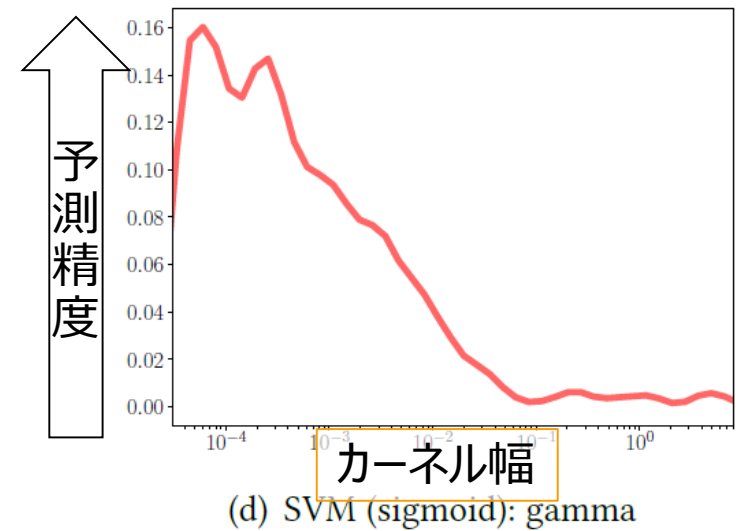
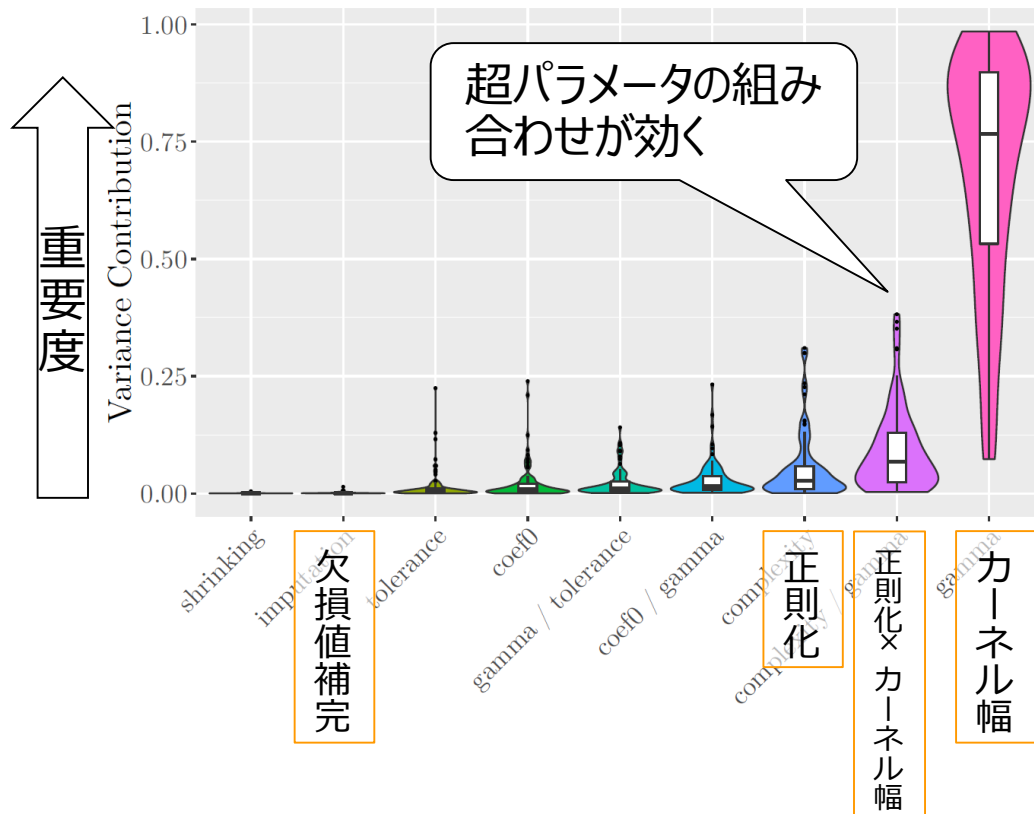
# SVM (RBFカーネル) の超パラメータ分析結果： (言い伝えのとおり) カーネル幅や正則化が重要

- SVM (RBFカーネル) ではカーネル幅、次いで、正則化パラメータが重要
- こういう知見は知られていたが、初めての定量的な証拠だと



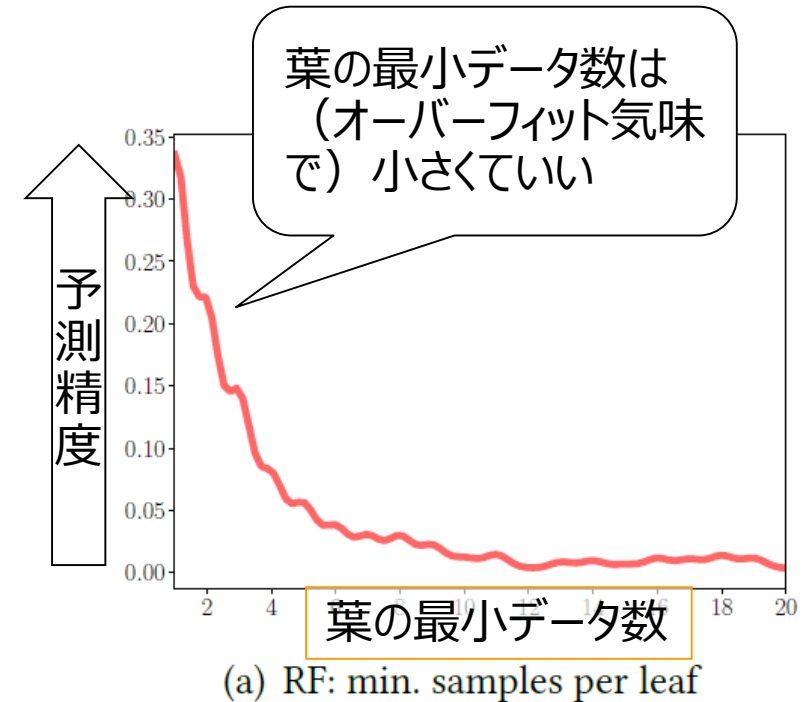
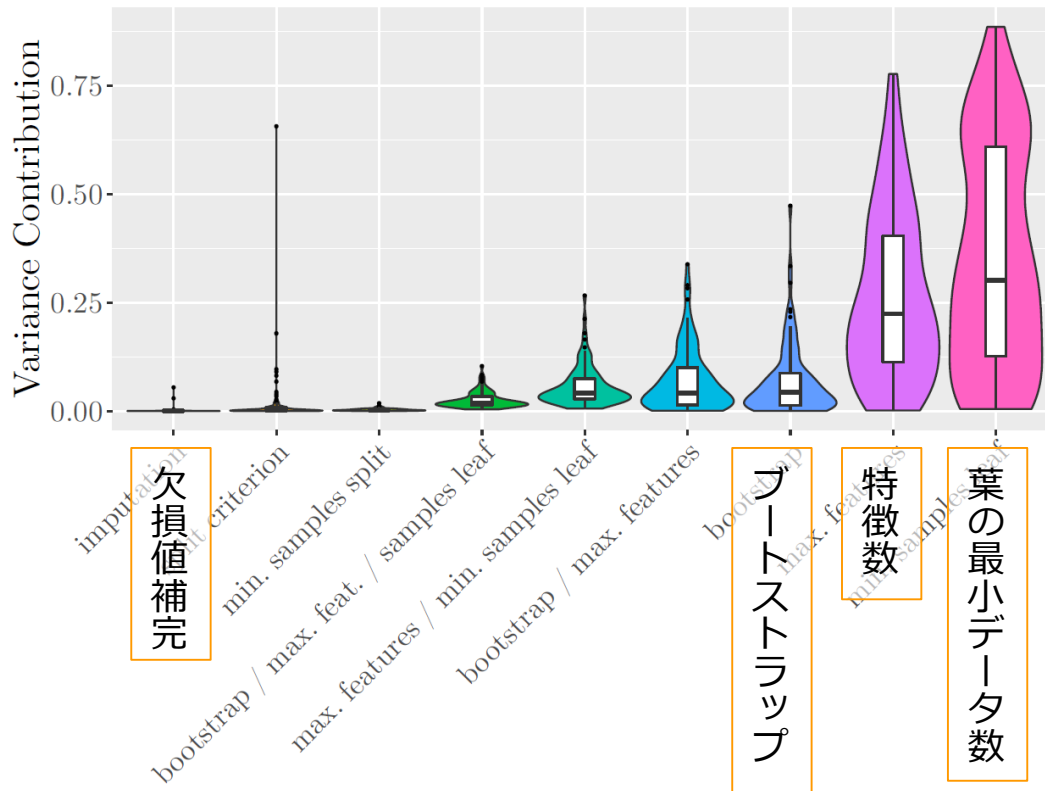
# SVM (シクモイドカーネル) の超パラメータ分析結果： RBFカーネルと同様の結果だが、こちらは組み合わせが効く

- ここでもカーネル幅 (1位)、正則化パラメータ (3位) は重要
- 特に、これらの「組み合わせ」(2位) が重要



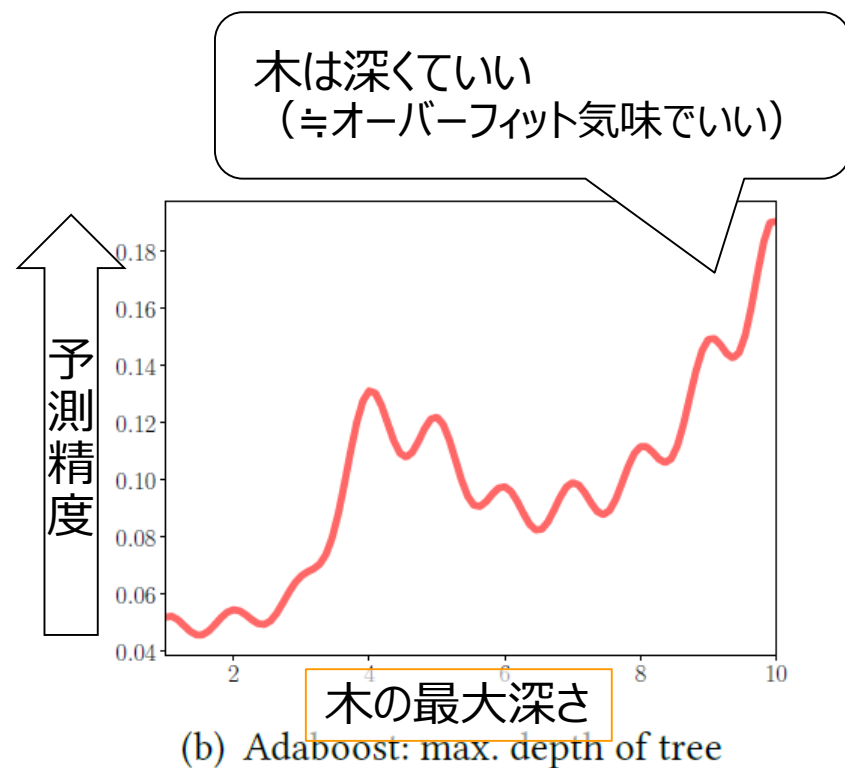
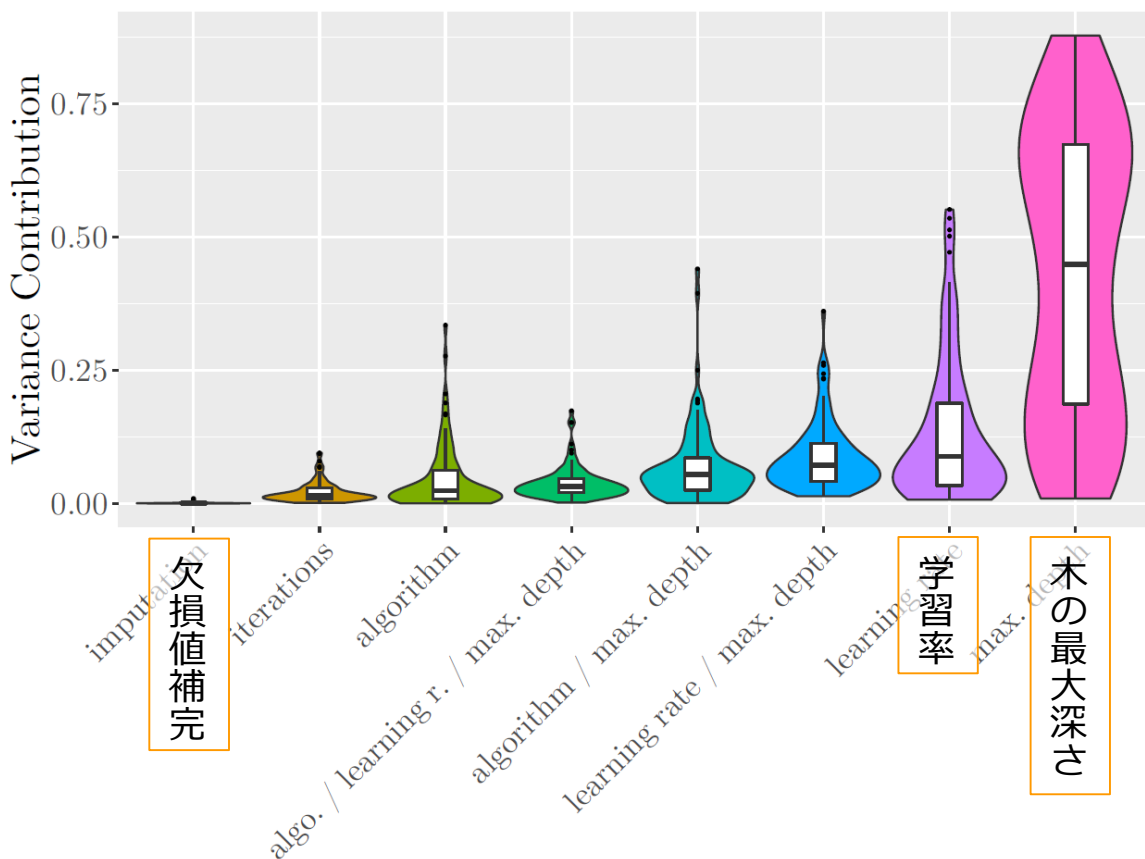
# ランダムフォレストの超パラメータ分析結果： 葉の最小データ数、使用特徴数、ブートストラップが効く

- 葉の最小データ数、各決定木で使う特徴数が効く
- ブートストラップ（サンプリングしたデータで訓練）も効く



# AdaBoostの超パラメータ分析結果： 決定木の深さは深いほどいい

- 決定木の最大深さが劇的に効く（かつ、深いほどいい）
- 学習率も効く



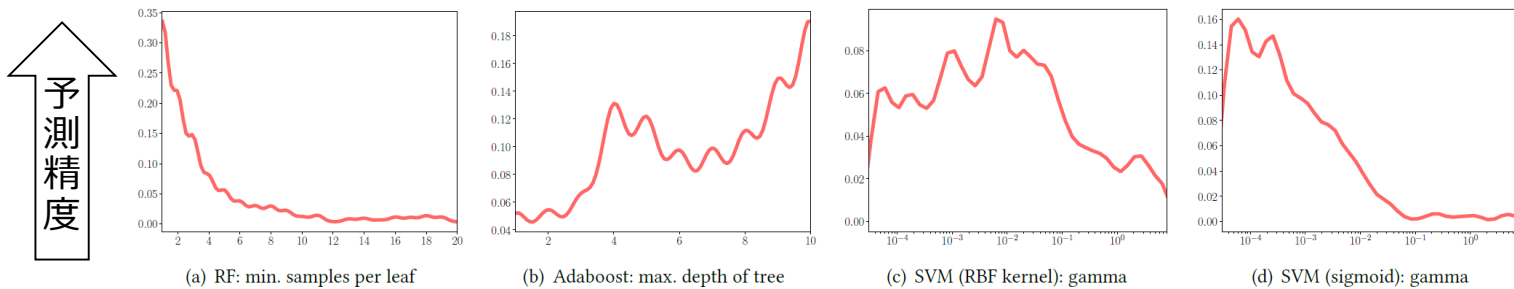
# 全手法を通して得られた知見：

## 少数の超パラメータが独立に有効なケースが多い

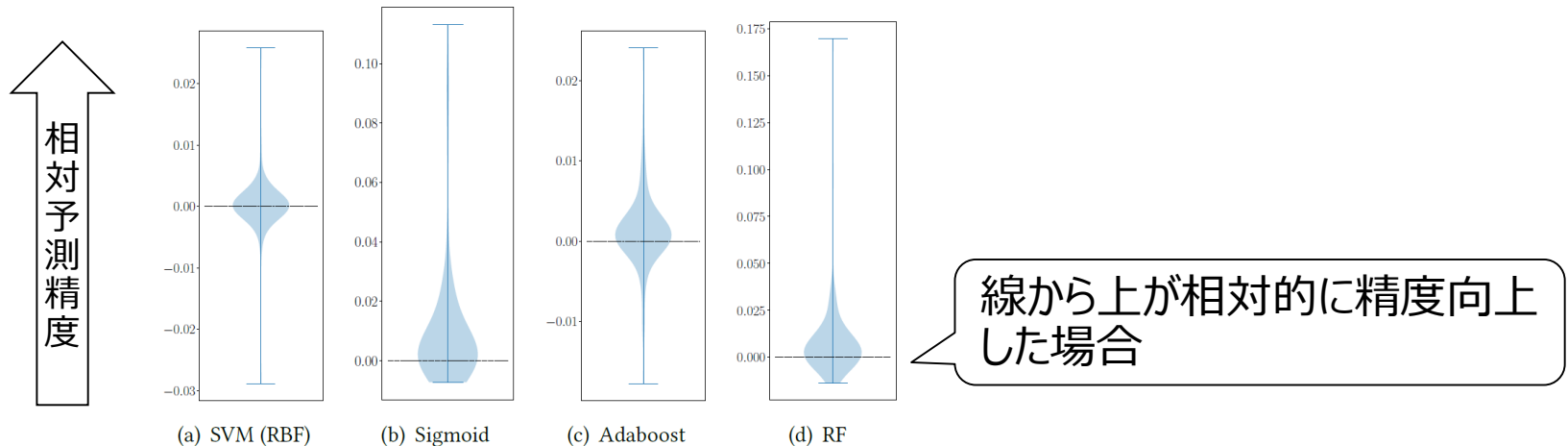
- 多くのデータセットにおいて、少数の共通の超パラメータの影響が大きい
- 超パラメータの組み合わせは意外に効かない
  - 結構独立に効く
- あと、欠損値の補完法の選択は割とどうでもいい
  - どの手法でも順位は低い
  - 補完が必要ないという意味ではない

# 超パラメータ最適化への応用： 得られた知見を事前分布として使って予測精度向上

- 効く超パラメータを動かしたときの予測精度を、事前分布として AutoML手法 (hyperband) に与える



- 一様な事前分布よりも精度向上



## まとめと所感：

# 機械学習法の超パラメータ重要度を実データで定量分析

- それぞれの機械学習法の超パラメータのうちどれが重要かを調べた
  - Functional ANOVA, OpenMLの100データセット
- さらに、超パラメータ最適化の事前分布に使ってみた
- 手法的には著者らの過去の結果をつかった「だけ」ともいえる

F. Hutter, H. H. Hoos, and K. Leyton-Brown. An efficient approach for assessing hyperparameter importance. In *Proc. of ICML*. 754–762 (2014).

- 本論文の貢献は、OpenMLデータの分析の部分
  - これまで皆がなんとなく思っていた「民間伝承」を定量的に分析
- 「この方法はどういうときにうまくいく（いかない）のか」を調べたい