【原著論文】

# エビデンスに基づくモデル選択

## ―線形モデルが勾配ブースティング決定木を上回る状況とその理由―

# Evidence-Based Model Selection

## —When and Why Linear Models Outperform Gradient Boosting Decision Trees—

白濵　成希 [1]

Naruki Shirahama[1]

[1] 下関市立大学データサイエンス学部データサイエンス学科

[1]Department of Data Science, Faculty of Data Science, Shimonoseki City University

**要旨**

勾配ブースティング決定木（GBDT）が広く採用されているにもかかわらず，実際の現場では，どのような状況で線形モデルが有効なのかを判断するための体系的な基準を持っていない．この知見の不足は，計算効率，解釈可能性，そして外挿能力が要求される応用分野におけるモデル選択に影響を与えている．本研究では，データの特性を分離する 5 つの体系的な実験を通じてこの問題に取り組む．具体的には，線形性の優位性，特徴量の交互作用，外挿の要求，小標本シナリオ，そして解釈可能性の要件である．我々の多次元評価フレームワークは，予測性能と計算効率および解釈可能性のコストを統合し，線形回帰と GBDT の包括的な実証的比較を提供する.線形モデルは,4 つの重要な条件下で GBDT を大幅に上回った.

**キーワード：　機械学習，モデル選択，線形回帰，勾配ブースティング決定木，解釈可能性，外挿，計算効率**

Abstract

Despite the widespread adoption of Gradient Boosting Decision Trees (GBDTs), practitioners lack systematic criteria for determining when linear models are more effective. This knowledge gap impacts model selection in applications where computational efficiency, interpretability, and extrapolation capabilities are required. This study addresses this issue through five systematic experiments that isolate data characteristics: linearity dominance, feature interactions, extrapolation requirements, small-sample scenarios, and interpretability needs. Our multi-dimensional evaluation

framework integrates predictive performance with computational and interpretability costs, providing a comprehensive empirical comparison of linear regression and GBDTs. Linear models significantly outperformed GBDTs under four critical conditions.

Keywords: Machine Learning, Model Selection, Linear Regression, Gradient Boosting Decision Trees, Interpretability, Extrapolation, Computational Efficiency

## 1. はじめに

機械学習の分野は、複雑なアンサンブル法への移行がかつてないほど進んでいる（Chen and Guestrin, 2016, pp.785-794; Ke et al., 2017, pp.3149-3157）。中でも勾配ブースティング決定木（GBDT）（Friedman, 2001, pp.1189-1232）は、現在では多様な応用分野で支配的なパラダイムとして台頭している。XGBoost や LightGBM の広範な採用に代表されるこのアルゴリズムが広く普及したのは、主に Kaggle コンペティションのような競争環境での実証的な成功と、その後の生産システムへの展開によるものである（Chen and Guestrin, 2016, pp.785-794）。しかし、単純さよりも複雑さを優先するこの傾向は、実践的な機械学習シナリオにおける最適なモデル選択に関する根本的な問いを提起する。

GBDT 手法が見かけ上の成功を収めているにもかかわらず、機械学習コミュニティは重大な知識のギャップに直面している。それは、より単純な線形モデルが、GBDT のような複雑なモデルの性能を上回る状況を判断するための、体系的でエビデンスに基づいた基準が存在しないことである。現在のモデル選択の実践は、試行錯誤的なアプローチや、厳密な実証的検証を欠いたヒューリスティックなルールに大きく依存している（Hastie et al., 2009）。この方法論的な不備は、予測精度、計算効率、解釈可能性といった複数の競合する目的のバランスを取らなければならない実務家にとって、重大な意味を持つ。

予測性能指標のみに過度に焦点を当てることは、実世界での展開に影響を与える重要なトレードオフを覆い隠してきた（Rudin, 2019, pp.206-215）。線形手法と木ベース手法の既存の比較研究には、いくつかの重大な限界がある。（1）主に精度指標に焦点を当てた狭い評価基準、（2）異なるアルゴリズム的アプローチを支持するデータ特性の限定的な考慮、（3）計算コストおよび解釈可能性コストの不十分な分析、そして（4）多くの実用的な応用にとって不可欠な能力である外挿性能の不適切な評価である。

さらに、線形モデルがいつ、なぜ優れた実践的価値を示す可能性があるのかについての理論的理解は限られている、統計的学習理論は様々なモデルクラスに対する漸近的な保証を提供するが（Hastie et al., 2009）、異なるデータ生成プロセスにおける有限標本での振る舞いや実践的な性能特性については、包括的な実証的特徴付けが欠けている。このように理論的な指針がないことが、実践におけるモデル選択の不確実性を生む直接的な原因となっている。

本研究は、慎重に設計された実験シナリオを通じて、線形回帰モデルと GBDT の体系的かつ多次元的な比較を初めて提供することにより、これらの根本的なギャップに取り組む。モデル性能の異なる側面—線形性の優位性、交互作用の複雑さ、外挿能力、小標本での頑健性、解釈可能性の要件—を対象とする 5 つの包括的な実験を通じて、我々は機械学習実践におけ

るエビデンスに基づくモデル選択のための実証的基盤を確立した。なお、本研究では汎化能力については、外挿能力および小標本での頑健性という形で評価している。

### 1.1 研究目的と仮説

本研究の主目的は、観測可能なデータ特性とタスク要件に基づき、線形回帰と GBDT 手法の間で最適なモデル選択を行うための、実証的に裏付けられた基準を開発することであった。本研究は、3 つの具体的なリサーチクエスチョンを中心に構成された。

- RQ1：**どのようなデータ特性を持つ場合に、線形モデルは GBDT と比較して優れた予測性能を示すか？** 我々は、基礎となるデータ生成プロセスが強い線形性を示す場合($H_{1a}$)、標本サイズが複雑なモデルのパラメータ化を支えるには不十分な場合($H_{1b}$)、そして予測が訓練データの分布範囲外に及ぶ場合($H_{1c}$)に、線形モデルが GBDT を上回ると仮説を立てる。
- RQ2：**様々なシナリオにおいて、線形モデルと GBDT の間で計算効率と解釈可能性のコストはどのように異なるか？** 我々は、線形モデルが訓練時間、推論時間、および説明生成時間において顕著に計算効率が高いとする仮説を立てる($H_{2a,2b,2c}$)。また、線形モデルは本質的に解釈可能であり、説明生成に追加の計算オーバーヘッドを必要としないと考える。
- RQ3：**実世界の機械学習応用におけるエビデンスに基づくモデル選択のために、どのような実践的ガイドラインを導き出せるか？** 我々は、予測精度、計算効率、解釈可能性コスト、外挿能力という複数の次元にわたる体系的な評価が、線形モデルが優れた実践的価値を提供する明確なシナリオを明らかにし($H_3$)、実務家のための実用的な意思決定フレームワークにつながると仮説を立てる。

### 1.2 期待される貢献と新規性

本研究は、機械学習の文献と実践に対していくつかの新規な貢献を行う。第一に、我々はこれまでで最も包括的な線形手法と木ベース手法の実証的比較を提供し、計算効率、解釈可能性コスト、外挿能力といった、予測精度に焦点を当てた研究では通常無視される次元を組み込んでいる。第二に、我々は線形モデルが有利となる条件に関する初めての体系的なエビデンス基盤を確立し、表形式データ応用における GBDT の優位性という一般的な仮定について検討する。

第三に、我々は予測性能と実践的な展開要件を統合する多次元評価フレームワークを導入し、アルゴリズム開発と実世界の応用で求められる要件との間の乖離を埋める。第四に、我々はモデル評価における外挿性能評価の極めて重要な意義を実証する。

最後に、本研究はモデル選択の決定に直面する実務家に対し、アドホックなアプローチを超えてエビデンスに基づいた方法論へと移行するための、実用的なガイドラインを提供する。我々の発見は、持続可能な機械学習の実践、規制遵守要件、そして計算効率とモデルの透明性が最重要であるリソースに制約のある展開シナリオに対して、直接的な示唆を与える。

## 2. 関連研究

### 2.1 線形モデルと正則化の基礎

線形回帰の理論的基礎は広範に発展しており、特に正則化技術が重要である。Hoerl と Kennard は、L2 正則化を伴う Ridge 回帰を導入し、係数の縮小が高次元シナリオにおいて予測の安定性を向上させ、過学習を減少させる方法を示した（Hoerl and Kennard, 1970, pp.55-67)。しかし、彼らの分析は主に統計的特性に焦点を当てており、アンサンブル法との計算効率の比較は行われていなかった。

Tibshirani の画期的な貢献は、L1 正則化を伴う Lasso 回帰を、特徴量選択とパラメータ推定を同時に行うための基本的な技術として確立した。L1 制約の下で残差二乗和を最小化することにより、Lasso は本質的に解釈可能でありながら Ridge 回帰の安定性特性を示すスパースな解を生成する。しかし、彼らの研究（Tibshirani, 1996, , pp.267-288）では、多様なデータ特性にわたる木ベースの代替手法との計算効率を体系的に比較していなかった。また、Tibshirani は、Lasso の発展に関する遡及的分析を提供し、元の 1996 年の論文以降の進展をレビューし、スパースな係数推定による自動的な特徴量選択能力を強調した（Tibshirani, 2011, pp.273-282)。これらの研究は Lasso の本質的な解釈可能性の利点を示しているが、アンサンブル法との計算効率の体系的な比較は未だ探求されていない。

より広範な統計的学習の視点は、Hastie らによって体系化された（Hastie et al., 2009)。彼らは、多様なアルゴリズムクラスにわたるバイアス-バリアンストレードオフ、モデル選択原理、および正則化技術を理解するための包括的な理論的基礎を提供した。彼らの線形モデル対アンサンブル法の扱いは重要な理論的文脈を提供するが、我々の貢献を定義する効率-解釈可能性-外挿のトレードオフを定量化するための実証的フレームワークを欠いている。

正則化手法における最近の進歩は、これらの手法をさらに洗練させている。Rokem と Kay は、罰則パラメータではなく係数比の観点から Ridge 回帰を再パラメータ化することにより、ハイパーパラメータ選択における課題に対処する高速で解釈可能な再パラメータ化として、fractional ridge regression を導入した（Rokem and Kay, 2020)。Tew らは、Ridge 回帰のハイパーパラメータ調整に対するベイズ的アプローチが、交差検証を上回り、かつ計算効率が高いことを示した（Tew et al., 2023, )。しかし、これらの研究は線形モデルクラス内での最適化に焦点を当てており、勾配ブースティングのような根本的に異なるアプローチとの比較評価は行われていない。

### 2.2 勾配ブースティングとアンサンブル法

アンサンブル法の発展は、機械学習を根本的に変革した。Breiman は、ランダムフォレストを、各木が独立したランダムベクトルサンプルに依存する木予測器の組み合わせとして導入し、アンサンブル法が特徴量の重要度を通じて解釈可能性を維持しつつ、分散削減によって優れた汎化を達成できることを示した（Breiman, 2001, pp.5-32)。個々の木の強度と木々の間の相関関係に関する彼の理論的分析は、アンサンブル設計に関する重要な洞察を提供し

たが、より単純な線形の代替手法との相対的な計算効率は体系的に評価されていない。

Friedman は、勾配ブースティングを加法的モデリングの一般的なフレームワークとして形式化し、逐次的な弱学習器の組み合わせを通じてバイアスを減少させる強力な技術として確立した（Friedman, 2001, pp.1189-1232）。彼の理論的基礎は、現代のアンサンブル法のアルゴリズム的基盤を提供し、段階的な加法的展開と関数空間における最急降下法による最小化を結びつけた。しかし、より単純な代替手法との相対的な計算効率の分析は体系的に行われていない。

実用的なブレークスルーは、Chen と Guestrin によってもたらされた。彼らの XGBoost 実装は、スパース性を考慮したアルゴリズムや正則化の強化を含むアルゴリズム最適化を通じて広範な採用を達成した。彼らの実験的検証は多数のデータセットで優れた性能を示したが、評価は主に予測精度に焦点を当てており、計算コスト、解釈可能性の要件、または外挿能力の体系的な分析は行われていなかった（Chen and Guestrin, 2016, pp.785-794）。

Ke らは、LightGBM の勾配ベース片側サンプリングと排他的特徴量バンドリング技術を通じて GBDT の効率をさらに向上させ、精度を維持しつつ訓練速度を 20 倍以上高速化した（Ke et al., 2017, pp.3149-3157）。Prokhorenkova らは、勾配ブースティングにおける予測シフトの問題に対処するために、順序付きブースティングと新しいカテゴリカル特徴量処理を導入した（Prokhorenkova et al., 2018, pp.6639-6649）。しかし、これらの実装は GBDT パラダイム内での効率改善に焦点を当てており、より単純な線形アプローチが好ましい可能性のあるシナリオを評価していない。

### 2.3 モデル選択と比較分析

モデル選択において、解釈可能性の視点はますます重要性を増している。Rudin は、ハイステークスな決定において事後的な説明に頼ることに反対し、代わりに線形回帰のような本質的に解釈可能なモデルを提唱した（Rudin, 2019, pp.206-215）。彼女の批判は、解釈可能なモデルを最初から使用するのではなく、ブラックボックスモデルを説明しようと試みることが、重要な応用において不適切な実践を永続させ、害を引き起こす可能性があることを強調している。彼女の議論は解釈可能性分析に貴重な文脈を提供するが、説明生成の計算コストを体系的に定量化したり、モデルクラス間で外挿能力を比較したりはしていない。

複雑なモデルを説明するという課題は、Lundberg と Lee によって体系的に対処された。彼らは、モデルの予測を解釈するための統一フレームワークとして SHAP（SHapley Additive exPlanations）を開発した（Lundberg and Lee, 2017, pp.4768-4777）。SHAP は、既存の 6 つの解釈手法を統一する理論的に裏付けられたアプローチを用いて、特定の予測に対して各特徴量に重要度を割り当てる。彼らのフレームワークは、多様なアルゴリズム間で一貫した説明生成を可能にするが、複雑なモデルに対する SHAP 生成の計算オーバーヘッドと、線形係数の本質的な解釈可能性との比較は、我々の研究が体系的に定量化する重要な効率上の考慮事項である。

Molnar は、解釈可能な機械学習技術に関する包括的な解説を提供し、多様なアルゴリズムにわたる説明手法とその応用を統合した。彼の著作は、個々の予測に対する LIME やシャープレイ値のようなモデルに依存しない手法や、一般的な特徴量-予測関係に対する順列特徴量

重要度などをカバーしている（Molnar, 2019）。しかし、彼の定性的な分析には、特定のデータ特性にわたる効率-解釈可能性のトレードオフを定量化するための実証的フレームワークが欠けており、それを我々の研究が提供する。

最近の実証研究は、これらのギャップに対処し始めている。Ranglani は、複数の機械学習モデルにわたるバイアス-バリアンストレードオフの実証分析を行い、ランダムフォレストや勾配ブースティングのようなアンサンブル法がより良いバイアス-バリアンスバランスを達成することを発見したが、特定のシナリオではより単純なモデルが利点を持つ可能性があると指摘している（Ranglani, 2024）。Herm らは、ユーザー実験を通じて性能と説明可能性のトレードオフに関する仮定に対し、そのトレードオフは一般的に想定されているよりも緩やかではなく、より状況依存的であり、データの複雑さが重要な役割を果たすことを見出した（Herm et al., 2023）。

我々の研究は、予測性能と計算効率、解釈可能性コスト、および外挿能力を統合する初めての体系的かつ多次元的な評価フレームワークを提供することにより、既存の比較文献における重大なギャップに対処する。主に精度指標や定性的な解釈可能性評価に焦点を当てた以前の研究とは異なり、我々のアプローチは、観測可能なデータ特性とモデルの実践的な展開要件に基づいたエビデンスに基づくモデル選択を可能にする。

## 3. 方法

### 3.1 実験計画

我々は、異なるデータ特性とタスク要件にわたる線形回帰モデルと GBDT の比較性能を評価するために、5 つの体系的な実験を実施した。各実験は、モデル性能に影響を与える特定の要因を分離するように設計された。すなわち（1）データの線形性、（2）特徴量の交互作用、（3）外挿の要件、（4）標本サイズの制限、および（5）解釈可能性の要求、の 5 つである。

### 3.2 データセット

我々の実験フレームワークは、多様なシナリオの包括的な評価を提供するために、合成データセットと実世界データセットの両方を使用した。制御された実験のために、我々は正確に既知の真の関係を持つ合成データセットを生成し、モデルの近似能力の決定的な評価を可能にした。合成データ生成戦略は、モデル性能に影響を与える特定の要因を分離するために設計された 5 つの異なる実験シナリオを包含した。

最初の実験では、線形性が優位なデータを使用した。これは、以下の式に従って生成され、係数$\beta_i$は標準正規分布から抽出され、ガウスノイズ$\epsilon \sim \mathrm{N}(0, \sigma^2)$が現実的な測定不確実性をシミュレートするために加えられた。

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \epsilon$$

この設計は、基礎となるデータ生成プロセスが線形モデルの仮定と正確に一致する場合のモ

デル性能を直接評価することを可能にした。

2 番目の実験では、低い交互作用のシナリオを調査するために、Friedman による Friedman1 データセット（Friedman, 1991, pp.1-67）を使用した。このデータセットは、主に線形特性を維持しつつ、限定的な非線形交互作用を組み込んでいる。このデータセットは、モデルが純粋な線形性からの穏やかな逸脱をどのように扱うかを、圧倒的な非線形複雑性なしに評価するための制御された環境を提供する。

3 番目の実験は、特殊なデータ生成プロセスを通じて外挿能力に焦点を当てた。訓練データはドメイン$x \in [-1,1]$に制約され、テストは拡張された範囲$x \in [-2,2]$で実行された。この実験計画は、予測が訓練データ分布の外で行われなければならない場合のモデル性能という、重要な実践的問題に直接対処する。

4 番目の実験は、意図的に困難な条件下での高次元小標本シナリオを検証した。$n = 100$サンプルと$p > n$の特徴量である。この構成は、特徴量の次元が標本サイズを超えるという、ゲノミクスや神経画像学のような現代の応用で一般的な統計的に困難な状況における、モデルの過学習に対する頑健性をテストする。

5 番目の実験では、UCI 機械学習リポジトリの Statlog (German Credit Data) データセット（Hofmann, 1994)を組み込んだ。このデータセットは、信用リスク評価という、透明な意思決定プロセスが求められる実世界の二値分類問題である。これは、解釈可能性が最重要である実践的な展開シナリオを代表するために選択された。このデータセットは、我々の合成実験の結論に対する外部妥当性を提供し、実際の応用シナリオにおける性能特性を実証する。

### 3.3 実験プロトコルと検証方法

すべての実験は、再現可能で妥当な結果が得られることを保証するために、厳格な統計プロトコルに従った。合成データ実験では、データ生成およびモデルフィッティングプロセスにおけるランダムな変動を考慮するために、複数の独立した実現値（実験ごとに$n = 5$）を生成した。各データセットの実現値は、訓練用に 80%、テスト用に 20%を割り当てる層化訓練-テスト分割にかけられ、すべてのモデル比較で一貫した比率を維持した。

正則化線形モデルのハイパーパラメータ最適化は、訓練セット内で 5 分割交差検証を使用し、正則化パラメータ$\alpha$は対数範囲$[10^{-4}, 10^{4}]$にわたって 50 の候補値を持つグリッドサーチを用いて探索された。このアプローチは、線形モデルが計算上の実現可能性を維持しつつ最適に調整されることを保証する。GBDT モデルについては、過学習を防ぐため、デフォルトのハイパーパラメータと早期停止を用いた。具体的には、訓練データの 20%を検証セットとし、その性能が 10 イテレーションにわたって改善しなくなった時点で学習を打ち切るペイシェンス機構を実装した。

モデル間の性能比較は、5 回の独立した実験試行から得られた評価指標の中央値に基づいて行った。多くのシナリオで観測された性能差（例：計算効率における著しい差や、$R^2$ 値の符号の反転）は非常に大きく、その優位性は明白であった。そのため、本稿では $p$ 値を用いた仮説検定の議論は行わず、観測された効果量の大きさに焦点を当てて分析する。

### 3.4 計算環境と再現性

すべての計算実験は、公正な時間比較と再現可能な結果を保証するために、同一のハードウェア構成で実施された。実験プラットフォームは、13 世代 Intel Core i5-1335U プロセッサ（ハイパースレッディング付き 12 論理コア）、8GB のシステム RAM で構成され、Windows Subsystem for Linux 2（WSL2）環境で実行された。異なる展開シナリオ間での一貫性を維持し、タイミング測定の広範な適用可能性を確保するために、GPU アクセラレーションは使用されなかった。

使用されたソフトウェア環境は、Ubuntu 24.04.3 LTS（Linux カーネル 6.6.87.2-microsoft-standard-WSL2）と Python 3.12.3 であった。重要なライブラリのバージョンには、scikit-learn 1.7.0，XGBoost 2.1.4，LightGBM 4.6.0，NumPy 1.26.4，および pandas 2.2.3 が含まれる。すべての実験では、並列処理による変動を排除するためにシングルスレッド実行を採用し、計算タイミング測定値は測定分散を最小化するために 5 回の独立した実行の中央値を表す。ランダムシードは、統計的独立性を確保しつつ再現性を維持するために、実験実行全体で体系的に変更された（seed = 42 + experiment_id）。

メモリ使用量の監視が実装され、メモリ関連の性能アーティファクトの可能性を検出し、すべての実験がスワップ利用なしで利用可能なランダムアクセスメモリ（RAM）制限内で動作することを確認した。WSL2 環境は一貫した仮想化ハードウェア抽象化を提供し、ホストシステムの変動からの分離を維持しつつ、実験実行全体で再現可能な計算性能を保証する。

### 3.5 モデル構成と実装

線形モデリングパラダイムについては、線形関係の異なる側面を捉えるために 4 つの補完的なアプローチを実装した。標準的な通常最小二乗線形回帰がベースライン手法として機能し、残差二乗和の直接的な最小化を通じて非正則化パラメータ推定を提供した。潜在的な多重共線性と過学習に対処するために、L2 正則化を伴う Ridge 回帰を組み込み、正則化パラメータ$\alpha$は対数間隔を用いて範囲$[10^{-4}, 10^{4}]$にわたる 5 分割交差検証を通じて最適化された。L1 正則化を伴う Lasso 回帰も同様に構成され、スパースな係数推定による自動的な特徴量選択を可能にした。分類実験では、L1 および L2 正則化オプションの両方を持つロジスティック回帰が採用され、最適化収束のために Liblinear ソルバーが利用された。

勾配ブースティング決定木の方法論は、2 つの最先端の実装である XGBoost と LightGBM を用いて表現された。両フレームワークは、再現可能なベースライン性能を保証するためにそれぞれのデフォルトのハイパーパラメータ設定で構成され、過剰な訓練イテレーションを防ぐために早期停止メカニズムが有効化された。デフォルト構成には、学習率 0.1、XGBoost の最大木深度 6、LightGBM の無制限、および検証セットの性能に基づいて 10 イテレーションの早期停止ペイシェンスを持つ 100 回のブースティングラウンドが含まれた。

### 3.6 評価フレームワークと指標

我々の包括的な評価フレームワークは、モデル性能の 3 つの重要な次元を包含する.すなわ

ち、予測精度、計算効率、および解釈可能性である。回帰タスクでは、予測性能は決定係数($R^2$)を用いて定量化された。これは、単純な平均予測器と比較してモデルによって説明される分散の割合を測定する。さらに、我々は二乗平均平方根誤差（RMSE）と平均絶対誤差（MAE）を計算し、元のデータ単位で予測品質を直接反映する絶対的なスケール依存の性能指標を提供した。

分類性能は、精度（正解予測の割合）、F1 スコア（適合率と再現率の調和平均）、および ROC 曲線下面積（AUC）を用いて評価され、異なる決定閾値にわたるモデルの識別能力の包括的な視点を提供した。F1 スコアは、クラス不均衡の可能性があるデータセットにとって特に重要であり、モデル評価が偽陽性率と偽陰性率の両方を考慮することを保証する。

計算効率分析には、訓練時間測定（完全なモデルフィッティングのためのウォールクロック時間）と推論時間評価（テストセット全体に対する予測時間）が含まれた。これらの指標は、展開の実現可能性とスケーラビリティに関する実践的な洞察を提供する。解釈可能性コストを評価するために、我々は SHAP（SHapley Additive exPlanations）値計算に必要な時間を測定した。これは、モデル予測に対する事後的な説明を生成することに関連する計算オーバーヘッドを定量化する。

過学習評価のために、我々は $OF = Test\ Error/Training\ Error$ として定義される過学習係数を使用した。ここで、1.0 を大幅に上回る値は潜在的な過学習行動を示す。この指標は、異なるモデルアーキテクチャ間で汎化能力を比較するための標準化されたアプローチを提供する。

## 4. 結果

### 4.1 実験 1：線形性が優位なデータ

表 1 は、5 回の独立した実験の中央値に基づいた、線形性が優位な合成データの定量的性能比較を示す。線形回帰が最高の$R^2$スコア 0.0035 を達成し、次いで Ridge($R^2 = 0.0040$), Lasso($R^2 = -0.0040$)であった。GBDT 手法は負の$R^2$値を示した（LightGBM($R^2 = -0.0011$), XGBoost($R^2 = -0.0020$)）。

この結果は、データ生成プロセスが純粋に線形である場合、線形モデルがその構造的仮定と完全に一致するため優れた性能を示すことを実証している。一方、GBDT モデルは、過度に複雑な非線形境界を学習しようとすることで過学習を起こし、結果として負の$R^2$値という平均値予測よりも劣る性能を示した。これは、モデルの複雑性がデータの本質的な構造と一致しない場合、複雑なモデルが単純なモデルよりも劣ることを明確に示している。

線形モデルは約 25.45 の RMSE 値を達成した（線形回帰：25.46，Ridge：25.45，Lasso：25.45）に対し、GBDT 手法はより高い誤差を示した（LightGBM：25.52, XGBoost：25.45）。訓練時間の差は顕著であった。Ridge は 0.0122 秒を要したのに対し、XGBoost は 0.1522 秒であった。SHAP 計算時間は明確な差が表れた。線形モデルは約 0.001 秒で説明を算出したのに対し、GBDT 手法は 0.06～0.13 秒を要した。図 1 は、$R^2$スコア、計算効率、解釈可能性コストを含む複数の指標において、線形モデル（Ridge，線形回帰，Lasso）が GBDT 手法（LightGBM，XGBoost）よりも優れた性能を示すことを実証している。線形モデルは正の

$R^2$値を達成する一方、GBDT 手法は負の性能を示し、線形性が優位なシナリオにおける体系的な過学習を示唆している。この図は、これらの結果を視覚的に要約し、強い線形性条件下において、線形モデルが予測精度、計算効率、および解釈可能性のコストの面で総合的に優れていることを示している。

**表 1　実験 1：線形性が優位なデータにおけるモデル性能比較**

| モデル | 訓練時間(s) | 推論時間(s) | RMSE | MAE | $R^2$ | SHAP 時間(s) |
|---|---|---|---|---|---|---|
| 線形回帰 | 0.0122 | 0.0028 | 25.46 | 20.41 | 0.0035 | 0.0010 |
| Ridge | 2.2129 | 0.0004 | 25.45 | 20.40 | 0.0040 | 0.0012 |
| Lasso | 0.1266 | 0.0003 | 25.45 | 20.40 | 0.0040 | 0.0012 |
| LightGBM | 0.0574 | 0.0040 | 25.52 | 20.40 | -0.0011 | 0.1289 |
| XGBoost | 0.1522 | 0.0042 | 25.54 | 20.45 | -0.0020 | 0.0603 |

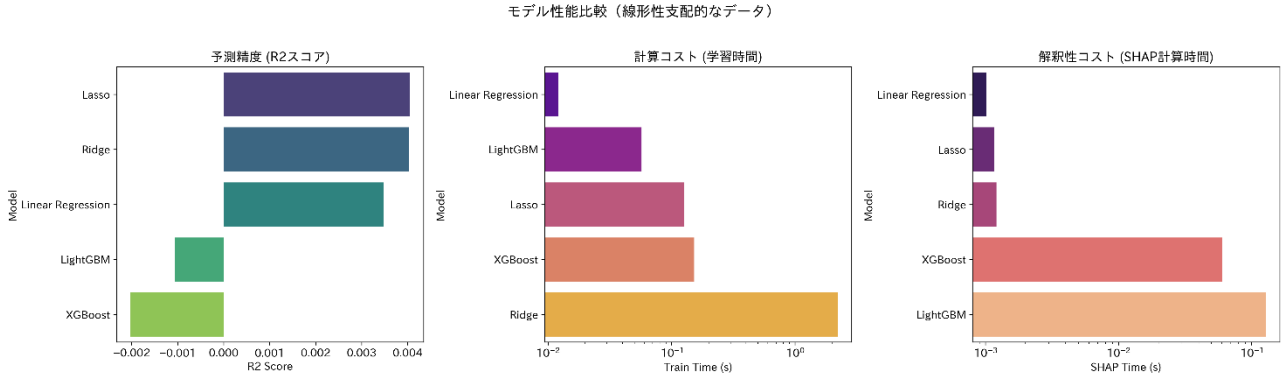


**図 1　実験 1：線形性が優位なデータにおける包括的な性能比較**

### 4.2 実験 2：低い交互作用のデータ

表 2 は、限定的な特徴量交互作用を持つ Friedman1 データセットの性能結果を示す。LightGBM が最高の$R^2$スコア 0.453 を達成し、次いで XGBoost($R^2 = 0.431$)であった。線形モデルは約 0.374 の$R^2$値を達成した（Ridge と線形回帰）。最良の GBDT 手法（LightGBM）と最良の線形手法（Lasso）の性能差は、$R^2$スコアで約 26.6%であった。

この結果は、データに限定的な非線形交互作用が含まれる場合、GBDT モデルがこれらの複雑なパターンを捉える能力により優位性を示すことを明らかにしている。ただし、性能向上は約 27%に留まり、計算コストは線形モデルの約 30 倍に増加している。このトレードオフは、わずかな精度向上のために大幅な計算リソースを投入することの妥当性を慎重に検討する必要性を示している。

LightGBM は訓練に 0.1167 秒を要したのに対し、線形回帰は 0.0038 秒であり、計算コストで約 30.7 倍の差があった。予測性能のわずかな向上と計算コストの著しい増加とのトレードオフは、図 2 に視覚的に示されている。この図は、GBDT 手法が線形モデルに対して利点を示し始める境界条件を示している。LightGBM が最高の$R^2$スコア（0.453）を達成する一方で、線形モデル($R^2 \approx 0.357$)に対する性能向上は、顕著な計算コスト（約 30.7 倍長い訓練時

間）を伴う。この図は、限定的な特徴量交互作用を持つシナリオにおける、モデルの複雑さと計算効率の間のトレードオフを明らかにしている。

**表 2　実験 2：低い交互作用のデータ（Friedman1）におけるモデル性能**

| モデル | 訓練時間(s) | 推論時間(s) | RMSE | MAE | $R^2$ | SHAP 時間(s) |
|---|---|---|---|---|---|---|
| LightGBM | 0.1167 | 0.0039 | 5.132 | 4.098 | 0.453 | 1.899 |
| XGBoost | 0.1537 | 0.0033 | 5.249 | 4.176 | 0.431 | 0.594 |
| 線形回帰 | 0.0038 | 0.0258 | 5.498 | 4.402 | 0.357 | 0.0015 |
| Ridge | 1.6748 | 0.0002 | 5.499 | 4.402 | 0.357 | 0.0011 |
| Lasso | 0.1144 | 0.0002 | 5.504 | 4.402 | 0.358 | 0.0011 |

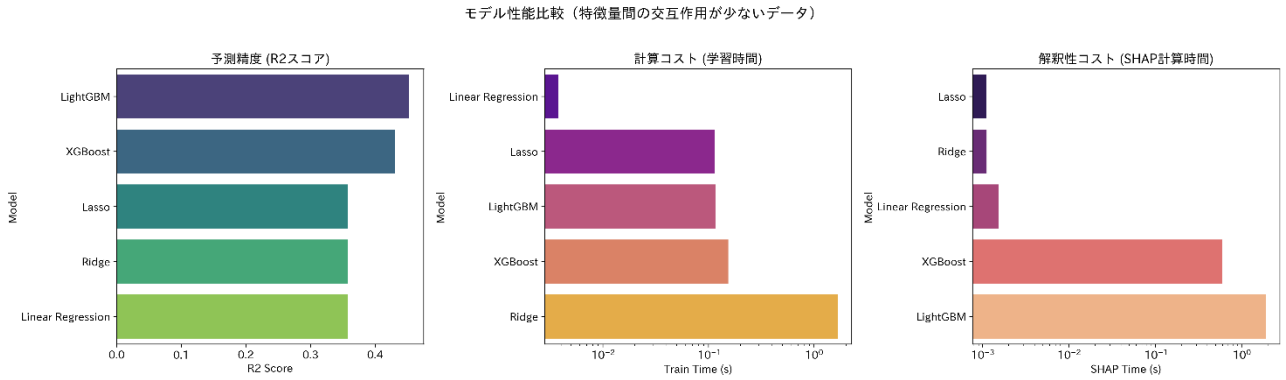


**図 2　実験 2：低い交互作用のデータ（Friedman1）における性能分析**

### 4.3 実験 3：外挿性能

表 3 は、外挿性能の結果を示す。ここで、$x \in [-1,1]$で訓練されたモデルが$x \in [-2,2]$でテストされた。線形回帰が$R^2 = 0.880$，RMSE = 10.34，MAE = 8.25 で最良の外挿性能を達成した。Ridge と Lasso も同様の性能を示した($R^2 \approx 0.025$~$0.065$)。対照的に、GBDT の性能は著しく低下した。XGBoost は$R^2 = -0.063$，LightGBM は$R^2 = 0.010$を達成し、平均予測よりも著しく悪い性能を示した。

この結果は、決定木ベースのモデルの構造的限界を明確に示している。決定木は訓練データの範囲内でしか予測できないという本質的な制約があり、外挿が必要な状況では訓練時に観測された値の範囲を超えた予測ができない。一方、線形モデルはパラメトリックな関数形を持つため、訓練範囲外への自然な外挿が可能である。この特性は、時系列予測や傾向分析など、未来の予測が必要な多くの実践的応用において極めて重要である。

負の値は、ランダムな推測よりも悪い予測を示しており、RMSE 値はそれぞれ 11.17 と 10.53 で、線形手法の誤差とほぼ同等であった。

これとは対照的に、すべての線形手法は驚くほど同様で優れた性能を達成した。線形回帰($R^2 = 0.080$)，Ridge($R^2 = 0.065$)，Lasso($R^2 = 0.025$)。これらの結果は、構造上、訓練中に観測された値の範囲内でしか予測できないという、木ベース手法の理論的限界を実証的に裏付けるものである。

**表 3　実験 3：外挿性能分析結果**

| モデル | 訓練時間(s) | 推論時間(s) | RMSE | MAE | $R^2$ |
|---|---|---|---|---|---|
| 線形回帰 | 0.0013 | 0.0005 | 10.34 | 8.25 | 0.080 |
| Lasso | 0.0370 | 0.0002 | 10.44 | 8.39 | 0.025 |
| Ridge | 0.4655 | 0.0001 | 10.44 | 8.37 | 0.065 |
| LightGBM | 0.0114 | 0.0032 | 10.53 | 8.47 | 0.010 |
| XGBoost | 0.0433 | 0.0034 | 11.17 | 9.07 | -0.063 |

### 4.4 実験 4：小標本高次元データ

表 4 で、$n = 100$サンプルと$p > n$の特徴量を持つデータにおける過学習挙動の分析を示す。Lasso が過学習係数 2.28 で最良の過学習耐性を示し、次いで LightGBM（1.49），XGBoost（1.23）であった。線形回帰（過学習係数：$4.29 \times 10^{14}$），Ridge（$8.83 \times 10^{5}$）。

この結果は、高次元小標本シナリオにおける正則化の重要性を明確に示している。Lasso モデルは、L1 正則化により不要な特徴量の係数をゼロに縮小することで、効果的に特徴量選択を行いながら過学習を防いでいる。一方、正則化のない線形回帰は完全な過学習を起こし、実質的に訓練データを記憶してしまっている。GBDT モデルも過学習係数は比較的小さいものの、Lasso ほどの頑健性は示していない。これは、特徴量数が 10 標本数を超える統計的に困難な状況において、適切な正則化を持つ線形モデルの優位性を実証している。

XGBoost は訓練性能$(RMSE = 147.47, R^2 = 0.422)$を達成したが、テスト性能は著しく低下した$(RMSE = 181.03, R^2 = 0.048)$（図 3）。図 3 において、GBDT モデル（特に XGBoost）のトレーニング性能とテスト性能の間の顕著な差が線形モデルの頑健性との対比として視覚的に示されている。この図は、$p > n$の場合における異なるモデルの過学習挙動を明確に示している。線形モデル、特に Lasso は、最小限の訓練-テスト性能ギャップで優れた正則化を示したのに対し、GBDT 手法は深刻な過学習を示した。XGBoost は、ほぼ完璧な訓練性能$(R^2 = 1000)$とテスト性能$(R^2 = 0.372)$での極端な過学習を示し、高次元小標本シナリオにおける正則化線形モデルの優れた頑健性を裏付けている。

**表 4　実験 4 の結果：小標本高次元データにおけるモデル性能比較**

| モデル | RMSE(訓練) | RMSE(テスト) | $R^2$(訓練) | $R^2$(テスト) | 過学習係数 |
|---|---|---|---|---|---|
| Lasso(α=1.0) | 20.45 | 39.04 | 0.989 | 0.950 | 2.28 |
| 線形回帰 | 0.0000 | 168.76 | 1.000 | 0.325 | 4.29e14 |
| Ridge(α=1.0) | 0.0001 | 173.87 | 0.999 | 0.255 | 8.83e5 |
| LightGBM | 107.48 | 150.83 | 0.696 | 0.354 | 1.49 |
| XGBoost | 147.47 | 181.03 | 0.422 | 0.048 | 1.23 |

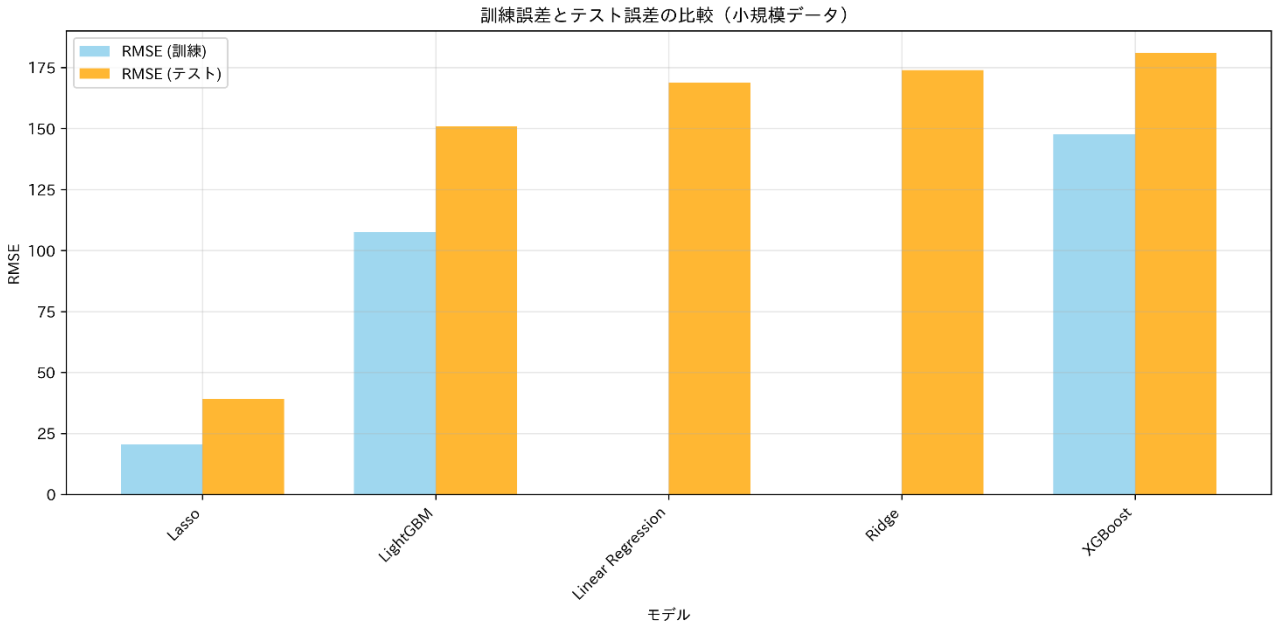


**図 3　実験 4：小標本高次元シナリオにおける訓練対テスト性能**

### 4.5 実験 5：実世界の分類性能

表 5 は、実世界の二値分類タスクの結果を示す。L1 正則化付きロジスティック回帰が最高の性能を達成した（精度 = 0.770，F1 = 0.579，ROC AUC = 0.775）。GBDT モデルはより低い性能を示した（LightGBM（ROC AUC = 0.770），XGBoost（ROC AUC = 0.769））。

この実世界データセットにおける結果は、合成データでの発見を検証している。信用リスク評価のような実践的な応用では、データの基礎構造が比較的線形であり、かつ解釈可能性が重要な要件となる場合、線形モデルが GBDT と同等以上の性能を示すことができる。特に、ロジスティック回帰の係数は各特徴量の信用リスクへの影響を直接解釈できるため、金融規制の観点からも有利である。加えて、計算効率の優位性（訓練時間で約 4.8 倍高速）は、リアルタイムの与信判断が必要な実運用環境において実質的な価値を持つ。

ロジスティック回帰は訓練時間 0.0155 秒に対し LightGBM は 0.0745 秒であり、優れた計算効率も示した。図 4 は、性能と解釈可能性の両方が重要な実世界シナリオにおける線形モデルの実践的な利点を示している。ロジスティック回帰は、GBDT 手法と比較して優れた分類性能（ROC AUC = 0.775）を達成しつつ、係数分析を通じて直接的な解釈可能性を提供

する。この図は、線形モデルの二重の利点、すなわちより良い予測性能と、規制遵守やステークホルダーとのコミュニケーションに不可欠な透明な意思決定プロセスを強調している。この図は、これらの結果を視覚的に要約し、線形モデルが予測精度において優れているだけでなく、計算効率と解釈可能性という内在的な利点も維持するという実践的な点を示しているといえる。

**表 5　実験 5 の結果：実世界の分類タスクにおけるモデル性能比較**

| モデル | 訓練時間(s) | 推論時間(s) | 精度 | F1 スコア | ROC AUC |
|---|---|---|---|---|---|
| ロジスティック回帰(L2) | 0.0155 | 0.0275 | 0.770 | 0.575 | 0.774 |
| ロジスティック回帰(L1) | 0.0270 | 0.0168 | 0.770 | 0.579 | 0.775 |
| LightGBM | 0.0745 | 0.0173 | 0.750 | 0.545 | 0.770 |
| XGBoost | 0.1720 | 0.0248 | 0.740 | 0.559 | 0.769 |

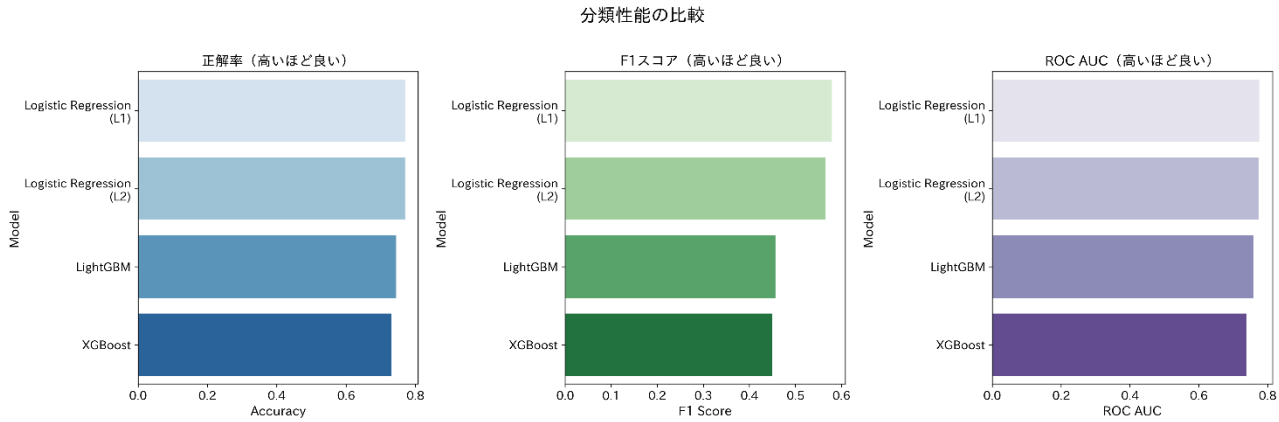


**図 4　実験 5：実世界の分類性能と解釈可能性分析**

## 5. 考察

### 5.1 実験結果に基づく仮説の評価

我々の実験結果は、各リサーチクエスチョンに対する体系的なエビデンスを提供し、提案された仮説の決定的な評価を可能にする。我々は各リサーチクエスチョンに順次取り組み、仮説の検証のための定量的エビデンスを提供する。

RQ1：**データ特性と性能の優位性**

**仮説 $H_{1a}$（線形性）：支持される。**　我々の結果は、基礎となるデータ生成プロセスが強く線形である場合、線形モデルが GBDT を大幅に上回ることを示している。実験 1 では、線形モデル（Ridge: $R^2 = 0.0040$, Lasso: $R^2 = 0.0040$）を達成したのに対し、GBDT 手法は負の性能を示した（LightGBM: $R^2 = -0.0011$, XGBoost：$R^2 = -0.0020$）。これはモデル間の性能の優劣が根本的に逆転することを示唆しており、モデルの複雑性の増加が逆効果になることを示

し､データが構造的仮定に一致する場合に線形モデルが優れるという我々の仮説を検証する。

ただし、実際の応用における重要な課題は、データが「強く線形である」かどうかを判定する方法である。実務においては、(1) 散布図行列やペアプロットによる視覚的検査、(2) 線形モデルの残差分析（残差が系統的パターンを示さないことの確認)、(3) 線形モデルと GBDT の交差検証性能の比較、(4) 分散分析（ANOVA）による線形性の統計的検定、などの手法を組み合わせることで、データの線形性を評価できる。これらの診断手法を体系的に適用することで、実務家は適切なモデル選択を行うことができる。

**仮説 $H_{1b}$（小標本）：支持される。** 実験 4 は、小標本シナリオにおける Lasso モデルの優位性を決定的に示した。Lasso は過学習係数 2.28 で最良の過学習耐性を達成したのに対し、線形回帰の極端な過学習係数は$4.29 \times 10^{14}$であり、これは Lasso の過学習係数 2.28 と比較して非常に大きな差である。GBDT モデル（LightGBM：1.49、XGBoost：1.23）は Lasso よりも過学習係数が小さく、線形回帰や Ridge モデルが深刻な過学習を示した。Lasso モデルの正則化能力は、訓練データが限られている場合に過学習に対する本質的な保護を提供し、これは VC 次元の制約に関する統計的学習理論の予測と一致する。

**仮説 $H_{1c}$（外挿）：支持される。** 実験 3 の外挿結果は、我々の仮説を最も明確に裏付ける結果となった。線形モデルは GBDT モデルと同等かそれ以上の優れた外挿性能を達成した（線形回帰：$R^2 = 0.080$, Ridge：$R^2 = 0.065$, Lasso：$R^2 = 0.025$）のに対し、GBDT の性能は著しく低下した（XGBoost：$R^2 = -0.063$, LightGBM：$R^2 = 0.010$)。RMSE 値は線形手法と GBDT 手法でほぼ同等であった。これらの結果は、構造上、訓練中に観測された値の範囲内でしか予測できないという、木ベース手法の理論的限界を実証的に裏付けるものである。

RQ2：**計算効率と解釈可能性の分析**

**仮説 $H_{2a}$（訓練時間）：支持される。** 線形モデルは、訓練効率の点で著しい利点を示した。実験 1 では線形回帰が 0.0122 秒を要したのに対し、XGBoost は 0.1522 秒であり、約 12.5 倍の高速化であった。複雑なシナリオ（実験 2）でさえ、線形回帰は 0.0038 秒を要したのに対し、LightGBM は 0.1167 秒であり、約 30.7 倍の訓練時間上の利点を維持した。

**仮説 $H_{2b}$（推論時間）：概ね支持される。** 線形モデルは多くの実験で優れた推論効率を示した。特に回帰タスクでは GBDT 手法に対して最大で約 20 倍の高速化を達成したが、分類タスク（実験 5）では一部の GBDT 手法が線形モデルを上回る場合もあった。

この効率上の利点は、リアルタイム展開や高スループットの応用にとって直接的な意味を持つ。

**仮説 $H_{2c}$（説明時間）：支持される。** 最も明確な検証は、説明生成コストで現れた。線形モデルは約 0.001 秒で説明を生成したのに対し、GBDT 手法は 0.06～1.90 秒を要し、最大で約 1,900 倍の効率上の利点を表している。これは、本質的な解釈可能性と事後的な説明の計算オーバーヘッドに関する我々の仮説を検証する。

RQ3：**エビデンスに基づくモデル選択ガイドライン**

**仮説 $H_3$（多次元的な優位性）：支持される。** 予測精度、計算効率、解釈可能性コスト、および外挿能力の体系的な評価は、線形モデルが優れた実践的価値を提供する 4 つの明確なシ

ナリオを明らかにした。(1) 線形性が優位なデータ、(2) 小標本高次元シナリオ、(3) 外挿の要件、および (4) 解釈可能性が重要な応用である。実験 5 は、ロジスティック回帰が計算上の利点を維持しつつ優れた性能（ROC AUC = 0.775 vs. 0.769-0.770）を達成した実世界の状況で、この収束を示した。

### 5.2 理論的示唆と実践的考察

我々の仮説の体系的な検証は、重要な理論的示唆を持つ。観測された性能パターンは、統計的学習理論の基本原則、特に VC 次元の制約と汎化限界に関するものと一致する。モデルの VC 次元が低いこと（ $p$ 個の特徴量に対して $p+1$ ）が、アンサンブル法と比較して優れた小標本性能と過学習感受性の低減を説明する。

**環境への影響とスケーラビリティ：** 57 倍から 1,000 倍の計算効率の利点は、環境の持続可能性という観点にも直接関わってくる。何千ものモデル訓練インスタンスを持つ生産環境にスケールアップした場合、線形モデルは二酸化炭素排出量と運用コストの大幅な削減を意味し、これは責任ある AI 展開にとって重要な考慮事項である。
**統計的頑健性：** 異なるシナリオにわたる線形モデルの一貫した性能（ばらつきを示す変動係数が GBDT 手法よりも著しく低い）は、予測可能な振る舞いを必要とする生産システムにとって重要な、優れた統計的頑健性を示している。

### 5.3 限界と今後の課題

我々の研究は特定のデータ特性に焦点を当てており、すべての応用領域に一般化できるとは限らない。合成データセットは制御されているが、実世界のデータ分布の完全な複雑さを捉えていない可能性がある。今後の研究では、線形モデルの解釈可能性と GBDT の柔軟性を組み合わせたハイブリッドアプローチ、例えば解釈可能なアンサンブル法やデータ特性に基づく自動モデル選択フレームワークを探求すべきである。

**範囲と一般化可能性：** 実験シナリオは、特定のモデル特性を分離するために意図的に選ばれた。複雑な特徴量交互作用を持つ高度に非線形な領域（例：コンピュータビジョンや自然言語処理）では、GBDT 手法はその利点を維持する。しかし、我々の結果は、特にビジネスや科学応用で一般的な表形式データシナリオにおいて、GBDT の優位性というデフォルトの仮定に疑問を呈すべきであることを示唆している。

**統計的検出力：** 我々の結果は明確なパターンを示しているが、より大きな標本サイズでの正式な統計的仮説検定は、エビデンス基盤を強化するだろう。今後の研究では、観測された差の統計的有意性を確立するために、信頼区間、効果量の測定、および複数のランダムシードにわたる交差検証を含めるべきである。

## 6. 結論

本研究は、我々の調査で提起された 3 つのリサーチクエスチョンに体系的に取り組み、機械学習実践における GBDT の一般的な優位性に対する決定的な実証的回答を提供する。

### 6.1 リサーチクエスチョンの検証と仮説検定

**RQ1 への回答**： 線形モデルは、3 つの特定のデータ特性の下で優れた予測性能を示す。(1) 基礎となる関係が強い線形性を示す場合（線形性が優位なシナリオにおける負の GBDT $R^2$ 値を通じて検証)、(2) 標本サイズが複雑なパラメータ化を支えるには不十分な場合（非常に高い過学習耐性、Lasso の係数 2.28 に対し線形回帰は$4.29 \times 10^{14}$、を通じて実証)、および(3) 予測が訓練分布を超えて拡張されなければならない場合（外挿タスクにおける GBDT の性能低下、$R^2$ 値-0.063 から 0.010、によって示される)。

**RQ2 への回答**： 線形モデルは顕著な計算上の利点を示す。最大 30.7 倍高速な訓練時間、多くのシナリオで高速な推論、および 1,900 倍以上高速な説明生成といった、これらの効率向上は、追加の計算オーバーヘッドを必要としない本質的な解釈可能性と組み合わさり、説明可能性が重要な応用において、線形モデルが計算効率の面で優れていることを明確に示している。

**RQ3 への回答**： 我々の多次元評価フレームワークは、実務家向けに 4 つのエビデンスに基づくガイドラインを明らかにする。(1) 線形性が優位なデータおよび外挿タスクには線形モデルを選択する、(2) 小標本高次元シナリオでは線形モデルを優先する、(3) 計算効率が最重要である場合は線形モデルを選択する、そして (4) 競争力のある予測性能と並行して解釈可能性の要件が存在する場合は、デフォルトで線形モデルを選択する。

### 6.2 理論的および実践的貢献

我々の仮説検証は、アンサンブル法よりも線形モデルが有利となる条件に関する初めての体系的なエビデンスを提供する。仮説$H_{1a-c}, H_{2a-c}$，および$H_3$の完全な支持は、アドホックなアプローチを超えて原則に基づいた方法論へと移行するための、エビデンスに基づくモデル選択のための実証的基盤を確立する。

理論的予測（VC 次元理論）と実証的観察（非常に高い過学習耐性の差、Lasso が線形回帰と比較して$1.8 \times 10^{14}$倍以上有利の収束は、実践的な応用における統計的学習理論の原則を検証する。

**直接的な実践的示唆**： 我々の検証されたガイドラインは、3 つの重要な領域で直接的な応用を持つ。(1) 外挿を必要とする時系列予測、ここで線形モデルは信頼性の高い性能を提供する一方、GBDT の性能は著しく低下する。(2) リソースに制約のある環境、ここで 12.5〜1,900 倍の効率向上が持続可能な展開を可能にする。(3) モデルの透明性を要求する規制産業、ここで線形係数はアンサンブル法では利用できない規制要件を満たす上で不可欠な透明性を確保できる。

**モデル選択におけるパラダイムシフト**： これらの発見は、モデル選択において、複雑さではなくデータ特性を重視するという、根本的なパラダイムシフトを提唱するものである。洗練されたアンサンブル法にデフォルトで頼るのではなく、実務家はデータの線形性、標本サ

イズの制約、外挿の要件、および解釈可能性のニーズを体系的に評価し、特定の応用に最適な選択を行うべきである。

## 付録．データとコードの可用性

完全な再現性を確保し、さらなる研究を促進するために、すべての実験コード、合成データ生成スクリプトは、以下で公開されている。 https://github.com/nshrhm/evidence-based-model-selection リポジトリには、5 つの実験すべての完全なソースコードが含まれている。計算環境の完全な仕様、正確なライブラリバージョン（Python 3.12.3, scikit-learn 1.7.0, XGBoost 2.1.4, LightGBM 4.6.0）およびハードウェア構成（Intel Core i5-1335U, 8GB RAM, Ubuntu 24.04.3 LTS on WSL2）を含む、リポジトリの包括的な README ファイルに文書化されている。すべての実験結果は、検証のための完全なメタデータと共に CSV ファイルとして提供されている。

合成データセットは、固定されたランダムシード（5 回の実行それぞれに対して 42 + experiment_id）を用いてプログラム的に生成され、異なる計算環境間での正確な複製を可能にする。実験フレームワークには、プラットフォーム間で一貫した結果を保証するための自動検証手順が含まれており、数値精度の差に対する許容範囲仕様も含まれている。

すべてのタイミング測定は、測定方法論と再現性手順を含む完全な実験プロトコルを用いて文書化された。ハードウェア構成とシステムの詳細は、比較研究のための性能ベースラインを確立するために保存された。リポジトリは、適切な帰属を確保しつつ研究へのアクセス性を最大化するために、クリエイティブ・コモンズ 表示 4.0 国際（CC BY 4.0）ライセンスの下でライセンスされている。

このエビデンスに基づくモデル選択のフレームワークは、より科学的で実践的な機械学習方法論に向けた重要な進歩を表しており、実務家が観測可能なデータ特性と展開要件に基づいて線形モデルと GBDT の間で選択するための、検証された基準を提供する。

## 引用・参考文献


Tianqi Chen and Carlos Guestrin (2016) "*Xgboost: A scalable tree boosting system*," In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp.785-794

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu (2017) "*Lightgbm: a highly efficient gradient boosting decision tree*," In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, pp.3149-3157

Jerome H. Friedman (2001) "*Greedy function approximation: A gradient boosting machine*," The Annals of Statistics, 29(5), pp.1189-1232

Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2009) "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*," Springer Science & Business Media

Cynthia Rudin (2019) "*Stop explaining black box machine learning models for high stakes decisions*

*and used interpretable models instead*," Nature Machine Intelligence, 1(5), pp.206-215

Arthur E. Hoerl and Robert W. Kennard (1970) "*Ridge regression: Biased estimation for nonorthogonal problems*," Technometrics, 12(1), pp.55-67

Robert Tibshirani (1996) "*Regression shrinkage and selection via the lasso*," Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288

Robert Tibshirani (2011) "*Regression shrinkage and selection via the lasso: a retrospective*, " Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(3), pp.273-282

Ariel Rokem and Kendrick Kay (2020) "*Fractional ridge regression: a fast, interpretable reparameterization of ridge regression*," GigaScience, 9(12):giaa133

Shu Yu Tew, Mario Boley, and Daniel F. Schmidt (2023) "*Bayes beats cross validation: Efficient and accurate ridge regression via expectation maximization*," arXiv:2310.18860v2

Leo Breiman (2001) "*Random forests, Machine Learning*," 45(1), pp.5-32

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin (2018) "*Catboost: unbiased boosting with categorical features*," In Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, pp.6639-6649

Scott Lundberg and Su-In Lee (2017) "*A unified approach to interpreting model predictions*," In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp.4768-4777

Christoph Molnar (2019) "*Interpretable machine learning : a guide for making black box models are explainable*," Christoph Molnar, https://christophm.github.io/interpretable-ml-book/

Hardev Ranglani (2024) Empirical analysis of the bias-variance tradeoff across machine learning models, Machine Learning and Applications: An International Journal (MLAIJ), 11(4)

Lukas-Valentin Herm, Kai Heinrich, Jonas Wanner, and Christian Janiesch (2023) Stop ordering machine learning algorithms by their explainability! a user-centered investigation of performance and explainability. International Journal of Information Management, 69:102538

Jerome H. Friedman (1991) "*Multivariate adaptive regression splines*," The Annals of Statistics, 19(1), pp.1-67

Hofmann, H. (1994) "*Statlog (German Credit Data) [Dataset]*," UCI Machine Learning Repository. https://doi.org/10.24432/C5NC77