勾配ブースティング決定木における ハイパーパラメータの自動決定

大阪教育大学附属高校池田校舎 1年　谷澤一蕗

■要旨

勾配ブースティング決定木（GBDT）のハイパーパラメータ min_gain_to_split（γ）は, 現状グリッドサーチ等の総当たり探索で決められ, 「なぜその値か」を説明することができない。本研究では, ノード内の勾配をランダムに並べ替えて得られる「偶然による最大利得」の帰無分布を用い, その上側95%点をγとして採用する導出法を提案する。実験の結果, 実データ（Diabetes）では探索と同等の精度を約1/4の葉数で達成し, 説明可能な過学習制御としての有効性を示した。

■導入　ハイパーパラメータの抱える問題

複数の決定木を組み合わせ, 次々残差を補正していく勾配ブースティング決定木(GBDT)モデル。その分割の可否を判断する評価式は, LightGBM（Ke et al.,2017）では次の式である。

$$Gain = \frac{1}{2}\left[\frac{(\sum G_L)^2}{\sum H_L + \lambda} + \frac{(\sum G_R)^2}{\sum H_R + \lambda} - \frac{(\sum G_P)^2}{\sum H_P + \lambda}\right] - \gamma$$

ここで, 各記号は次の通り定義される。

- g_i：サンプル i における損失関数 L(y_i, ŷ_i) の予測値 ŷ_i に関する一階微分
- h_i：同じく二階微分（ヘッセ）
- ΣG, ΣH：ノードに属する全サンプルについての g_i, h_i の総和
- P, L, R：分割前の親ノード（Parent）, 分割後の左（Left）・右（Right）の子ノード
- λ：葉の重みに対する L_2 正則化係数（過学習防止）
- γ：分割を許可するための最小利得の閾値。

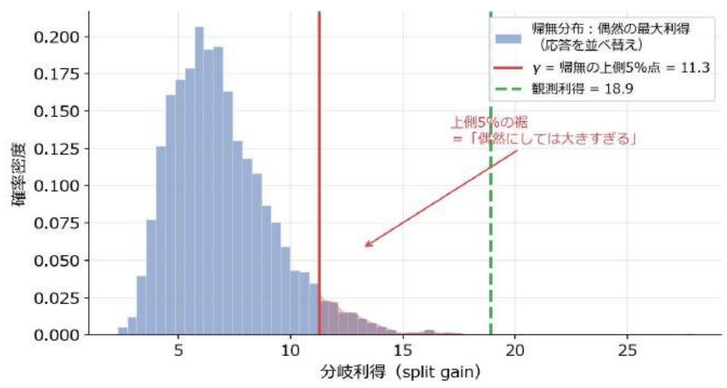
本研究の対象（LightGBM では min_gain_to_split）Gainは「分割によって損失がどれだけ減るか」の近似値であり、Gain > 0 のときのみ分岐が実際に行われる。

この際, ハイパーパラメータとして用いられるパラメータγ（lightGBMで“min_gain_to_split”）はどれくらいの利得が得られるならさらなる分岐を許すかを司る閾値であり, 過度な分岐が行われピーキーなモデルとなってしまう過学習現象が発生する際にはこのパラメータを上げることによって抑え込むことができる（Ke et al.,2017）。**現状, そのパラメータを定める方法としては総当たり的な手法を取るグリッドサーチ, optuna（ベイズ最適化により効率化された自動探索）やランダムサーチによるランダムな探索などが取られている。しかし, そのハイパーパラメータに設定したのかという理由がブラックボックス化してしまいがちであるため, そのハイパーパラメータを数理的な方法で示すことができないかと考えた。**なお, 統計的検定をGBDTに応用する近年の試みとしてScoreStop（Hines & Hines, 2026）があるが, これはブースティングの反復回数を検定する手法であり, 本研究が対象とする分岐を司る閾値であるmin_gain_to_splitへの適用とは異なる問題を扱っている。

■手法　帰無分布に基づき閾値を決定する

本手法はGBDTの学習前に, 以下の手順でγを導出する。導出したγは通常のハイパーパラメータと同様にLightGBMに渡して学習し学習アルゴリズム自体は変更しない。

これは, 条件付き推論木(ctree)に用いられる**目的変数と各説明変数の間に「関連がない」という帰無仮説を検定し, 棄却されない場合分岐を止める手法**(hothorn et al.,2006)**をGBDTに移植・応用したものである。**

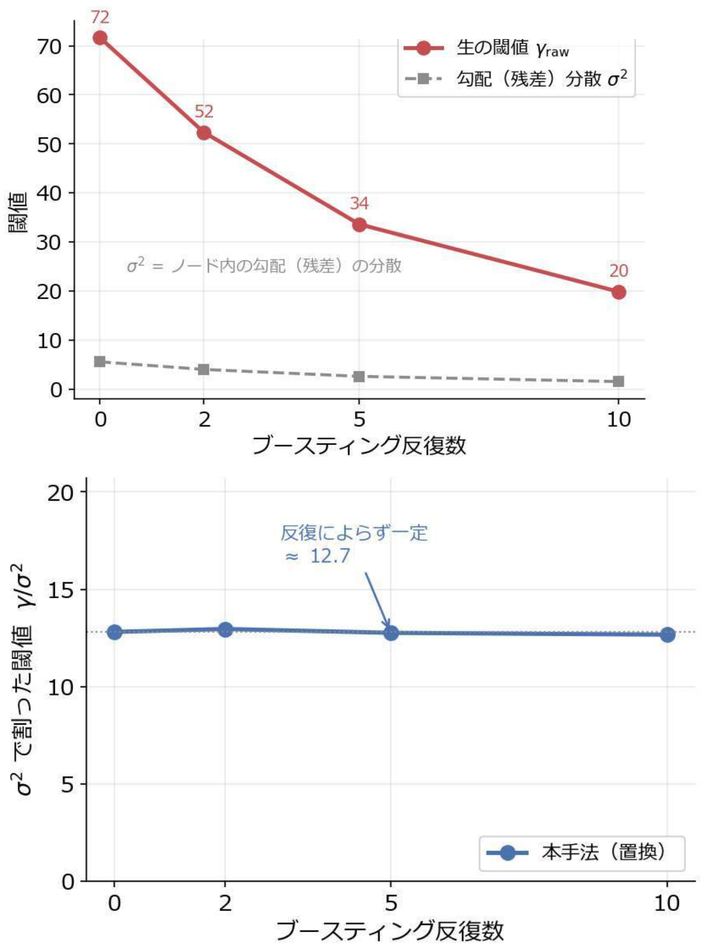


1. **あるノードで, 現在の勾配をペアで並べ替える。**(g_i, h_i) を組のまま並べ替えるのは, 勾配とヘシアンの対応関係を保ち, 特徴量との関連のみを破壊するためである。また, GBDTの利得は生の目的変数ではなく現在の勾配上で計算される（Ke et al.,2017）ため, 並べ替えの対象も勾配とする。
2. 並べ替えデータ上で, 候補全体（特徴×分割点）にわたる**最大の利得を計算する。**
3. 2を多数回繰り返して偶然得ることができた最大の利得を用い帰無分布を計算する。(帰無仮説 H_0：ノード内で特徴量と残差の間に関連がない。)
4. その95%点をハイパーパラメータとすることで, 観測した利得がパラメータを超えた時, それは偶然によるものではない可能性が高いとわかる。

各並べ替えで単一候補ではなく全候補の最大を記録するのは, 実際の木も全候補から最大利得の分岐を選ぶためであり（Ke et al.,2017）, この設計が多重性の補正に相当する（実験②で検証を行った）。

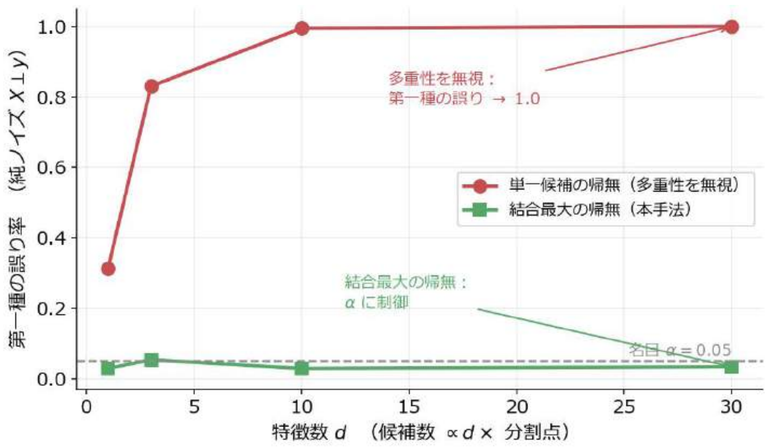
■実験①　γの変化の様子

- 目的：導出されるγがGBDTで行われる決定木の反復とともにどのように変化するのかを調べる。
- 設定：合成データ（NumPy, n=2000, 特徴量6, シード20260703）を用いる。反復m=0, 2, 5, 10のルートノードでそれぞれ1000回の置換を行ってγを導出した。
- 結果：導出されるγは反復とともに72→52→34→20と単調に減少した。なお, γを残差の分散σ^2で割った値は全反復でおよそ12.7と一定になった。
- 考察：γの減少はブースティングが残差を縮めることによってσ^2が減少したからであると考えられる。**また, γをσ^2で割った値は一定になったため, σ^2を測って12.7を掛ければ各反復のγに近い値が得られ, 計算コストを大幅に削減することができる。**



■実験②　多重性の考慮

- 目的：GBDTでは非常に多数の分割点候補があるので, それを全て試行して最も利得の大きかった1個を選ぶ場合, 全くそれらが目的関数と関係ない場合も偶然大きな利得を出す恐れがある。そのため, γの導出においてそれらの多重性を考慮する必要があるかどうかを, 全候補の最大を取る今回の手法と, 多重性を考慮せず単一の候補から帰無分布をなす方法でそれぞれ比較することにより検証する。
- 設定：NumPyで生成した純粋なノイズ（n=500, 特徴量 d=1, 3, 10, 30, シード20260703）を用いる。単一候補と今回提唱する手法である全候補の最大を取る帰無で, ノイズなのに分岐を許してしまう第一種の誤りが発生する確率を比較する。
- 結果：単一候補の帰無では誤り率がd=1で0.31, d=30で1.0まで膨れ上がってしまった。しかし、**全候補の最大を取る帰無では閾値が候補数に応じて自動的に上がり, 誤り率は全条件でα=0.05付近（0.03–0.05）にとどまった。**
- 考察：木をどこで分割するか, それとも分割しないかは非常に多数の候補があるため, **多重性を無視してしまうと全くそれらが目的関数と関係ない場合でも偶然大きな利得を出してしまう。**そのため, 今回の手法で用いたように全候補の最大を取る方法を用いるなどの工夫が必須である。

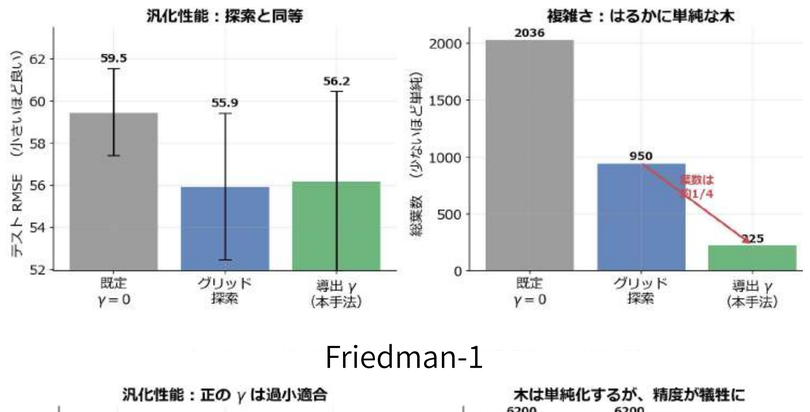


※これらの検証用データの作成, 当手法の実装, データ解析等に使用したPythonスクリプトはClaude Sonnet5 (2026,Anthropic)を用いて作成した。その後, 生成されたコードを精査した。

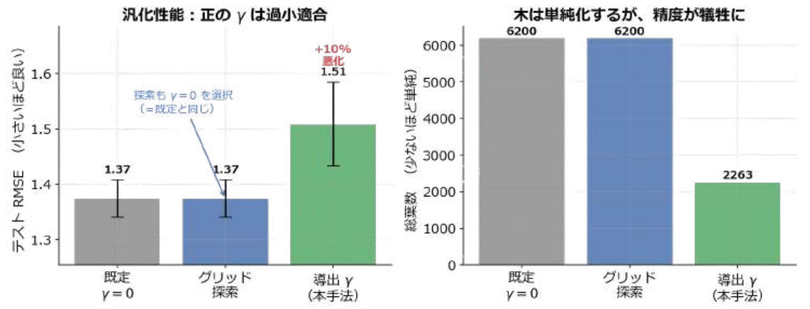
■実験③　実用性検証

- 目的：導出したγと探索（グリッドサーチ）によって求めた値を実データを用いて比較することで, 手法の実用性を検証する。
- 設定：糖尿病患者の診療データセットであるScikit-Learn DiabetesとFriedman-1の二つの回帰モデル用のデータセットで, γ=0（既定）/グリッド探索/本手法のγをRMSE（二乗平均平方根誤差）, 葉数（モデルの複雑さ）で比較する。
- 結果：Diabetes では本手法のRMSEは56.2±4.3となり, 探索の55.9±3.5と同等に保ったままで葉数は225対950とおよそ1/4（既定 γ=0 比では約1/9）まで削減することができた。しかし, Friedman-1では探索自体がγ=0を選び, 正のγはいずれもRMSEを約10%悪化させた。
- 考察：本手法を用いることにより, 偶然によって行われてしまう分岐を防ぐことには成功し, **過学習の抑制に対しては一定の効果を発揮することができた。**しかしながら, Friedman-1のような個々には利得が小さいが本物である分岐をひとつひとつ有意であると判定しなければならないようなモデルでは, 正のγを入れることによって, **本当は有効だが弱い分岐まで一部が道連れとなって弾かれることで性能が悪化してしまう。**

Scikit-Learn Diabetes



Friedman-1



■結論・今後の展望

本研究では, ノード内の勾配をランダムに並べ替えて得られる「偶然による最大利得」の帰無分布から, min_gain_to_split（γ）を「なぜその値か」を説明できるような形で導出する手法を提案した。また, それを糖尿病の予測データセットであるScikit-Learn Diabetesに投入した際には探索と同等の精度を約1/4の葉数で達成し, **説明可能な過学習制御としては有効であることがわかった。**

しかし, 小さな利得しかない分岐を残すことが求められるFriedman #1では, 正のγはどれも過小適合している。そのため, **本手法は過学習で無駄な分岐が増えやすいモデルの制御として機能するのであって, 現時点では探索を完全に置き換えることができるものではないといえる。**今後はさらに汎化性能を高め, 今回は95%の点で固定とした有意水準を操作可能にするなどの方法で様々なデータセットに対応できるように研究を進めていきたい。

■参考文献

- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu (2017) . 「LightGBM: a highly efficient gradient boosting decision tree」 . ACM Digital Library . https://dl.acm.org/doi/10.5555/3294996.3295074 , (最終閲覧 2026-07-21) .
- Oliver J. Hines, Christian L. Hines(2026) . 「ScoreStop: Gradient-based early stopping using functional score tests」 . ArXiv . https://arxiv.org/abs/2606.02740 , (最終閲覧 2026-07-21) .
- Torsten Hothorn, Kurt Hornik, Achim Zeileis (2006) . 「Unbiased Recursive Partitioning: A Conditional Inference Framework」 . Research homepage of Achim Zeileis . https://www.zeileis.org/papers/Hothorn+Hornik+Zeileis-2006.pdf , (最終閲覧2026-07-21) .