

住宅価格推定のための学習システムの構築と最適化

Optimization of Features and Hyperparameters for a Neural Network in House Price

小林晴, 清水恭介, 西山雄梨, 青柳美輝 (日本大学)

研究背景

課題:

住宅価格予測は、社会経済的に非常に重要な課題である。しかし無数の変数が複雑に絡み合うため解析が困難である。

従来手法の限界:

線形回帰: 非線形な関係性を捉えられない。

勾配ブースティング: 精度は向上したが、依然として専門家による手作業の

特徴量設計が必要である。

新たな手法と新たな問題:

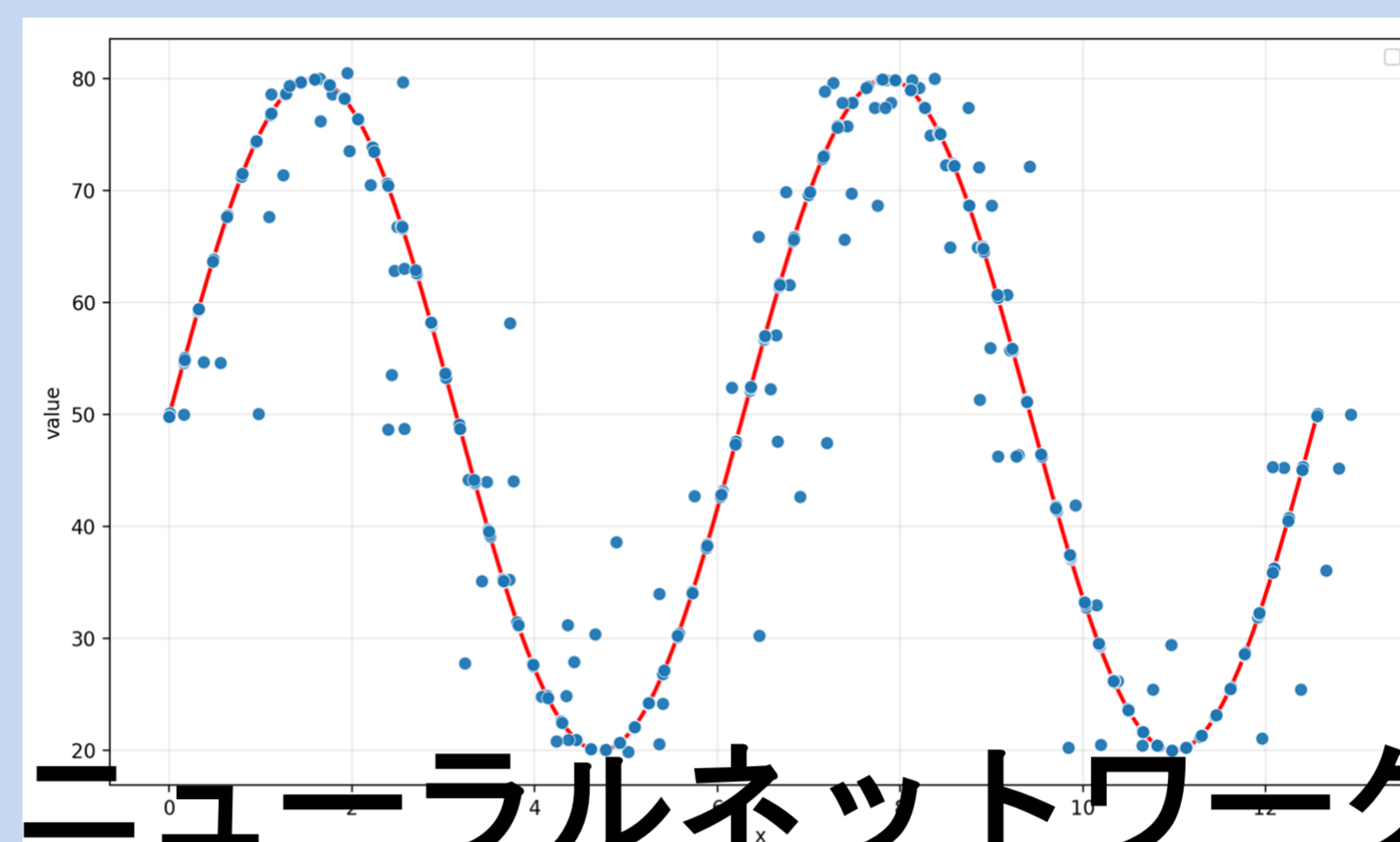
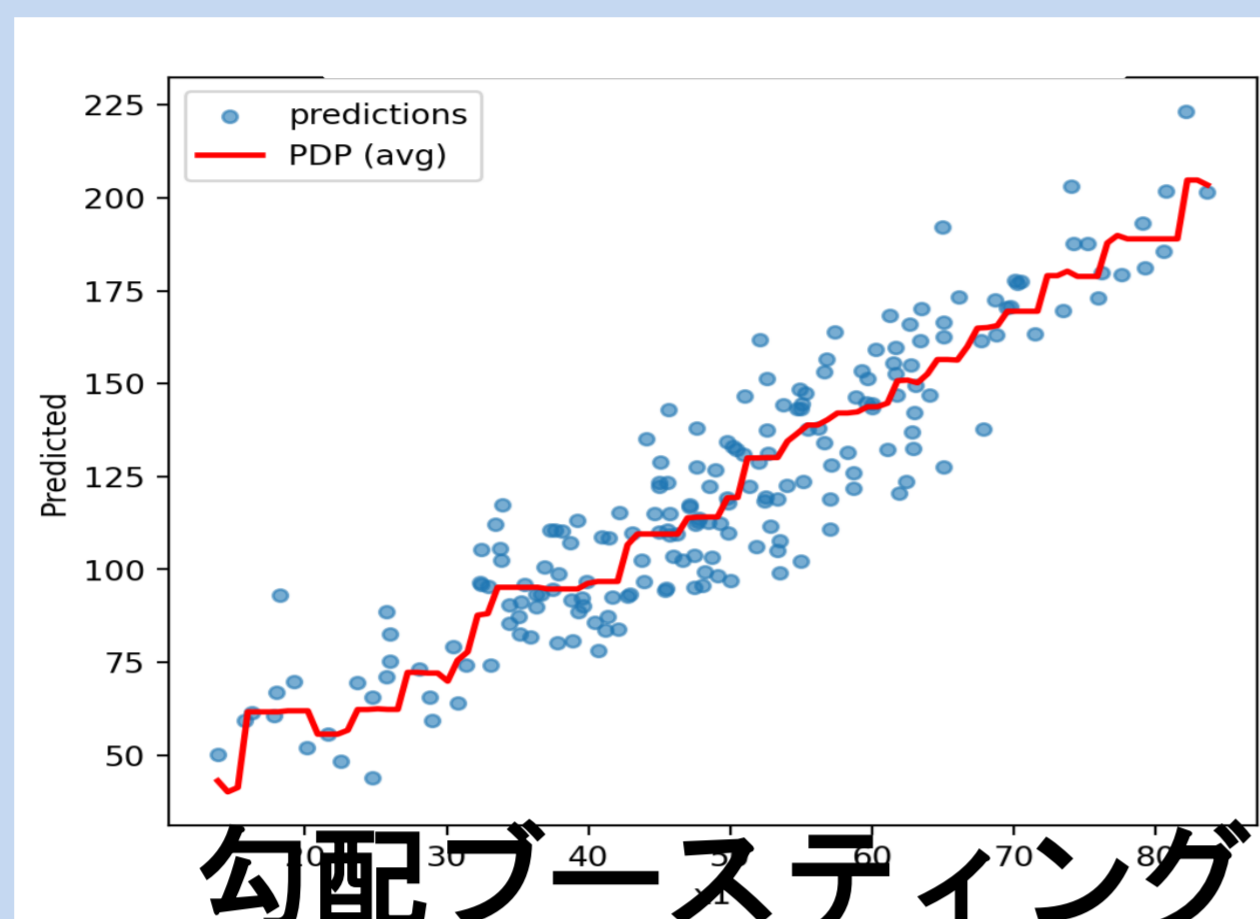
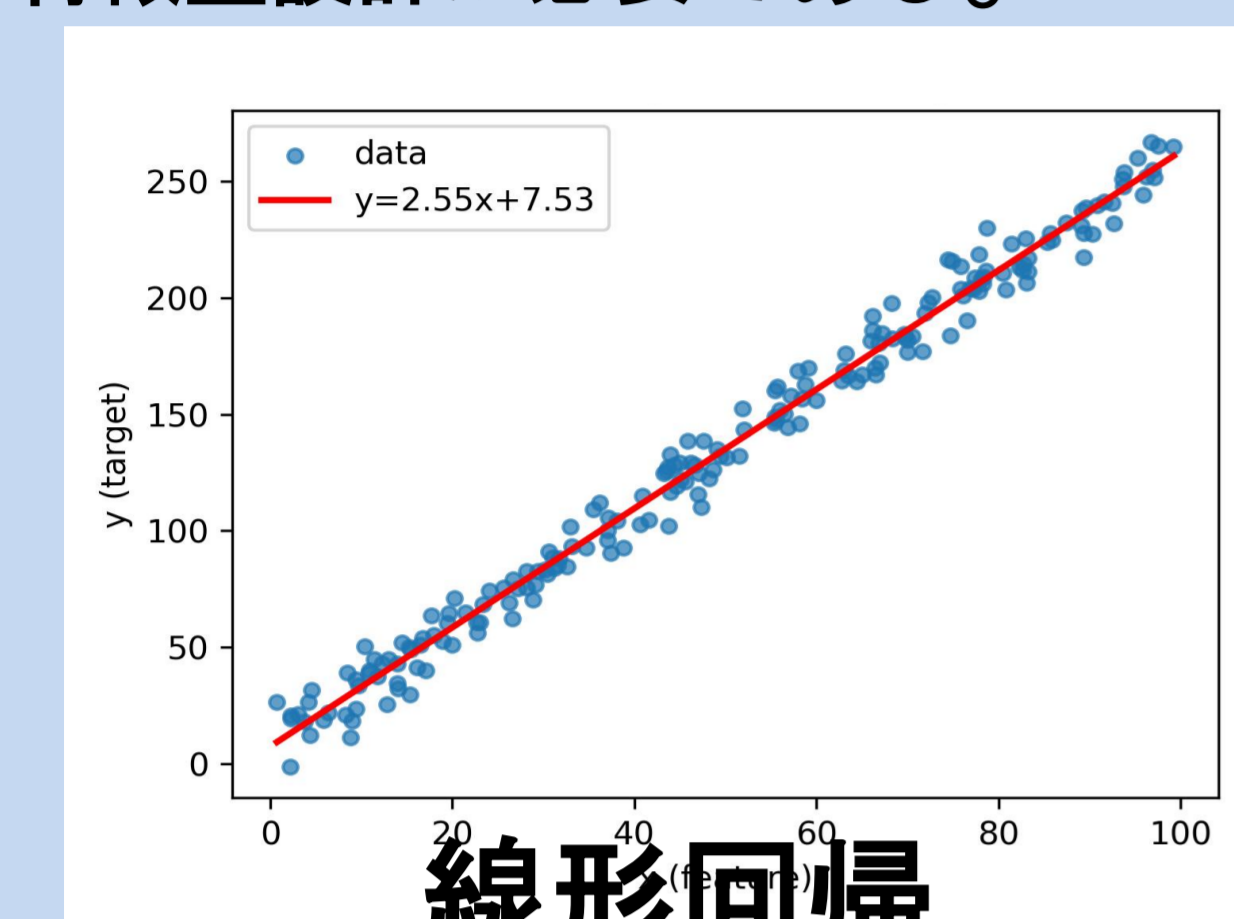
ニューラルネットワーク (NN): 複雑なパターンを自動学習できると期待されている。

しかし:

近年の研究は、モデル構造の複雑化 (深層化、CNN/RNNの利用) に焦点が当たりがちである。

その結果:

計算コストの増大、過学習、解釈性の低下といった新たな問題が発生している。



研究目的

本研究では、上記の問題意識から、あえてNNのアーキテクチャをシンプルな多層パーセプトロン (MLP) に固定する。

その上で、「特徴量エンジニアリング」と「ハイパーパラメータ最適化」という2点に徹底的に注力する。

この「徹底したデータと学習の最適化」というアプローチが、モデルの複雑化に頼るよりもいかに有効であるかを、定量的に明らかにすることを目的とする。

使用データ

本実験では、[アットホームデータセット] [1]から取得した不動産賃貸データを用いた。[アットホームデータセット]より以下のデータを入力特徴量として使用した。

物件情報: 「築年」「専有・建物面積 (平米)」「設備数」「建物構造」「間取りタイプ」

立地情報: 「徒歩(分)」「バス(分)」「所在階」

地理座標: 「緯度」「経度」「所在地名」

費用情報: 「管理費 (円)」「共益費 (円)」

また建物構造、所在地名に関してはOne-Hot エンコーディングを使用。

間取りタイプにはターゲットエンコーディングを使用しカテゴリの価格の中央値で数値に置き換えた。

その他の前処理上記の過程で生じた欠損値は全て0で補完

最後に学習の安定化を目的として、全特徴量にZスコア標準化を適用した

実験手法

本研究の目的 (シンプルなMLP+徹底した最適化の有効性を検証) を達成するため、以下の手順で実験・評価を行った。

1. 予測モデル: MLPに固定

目的の通り、実験を通じてモデルの構造はシンプルな多層パーセプトロン(MLP)に一貫して固定し、変更はなし。

構造: 入力層 → 中間層 (1層) → 出力層

2. 検証プロセス: 定量的評価

「徹底したデータと学習の最適化」アプローチの有効性を定量的に示すため、性能をステップバイステップで比較した。

基準点 (ベースライン) の設定

2つの軸での徹底的な最適化モデル構造は固定したまま、目的で掲げた2点に注力し、性能向上を図る。

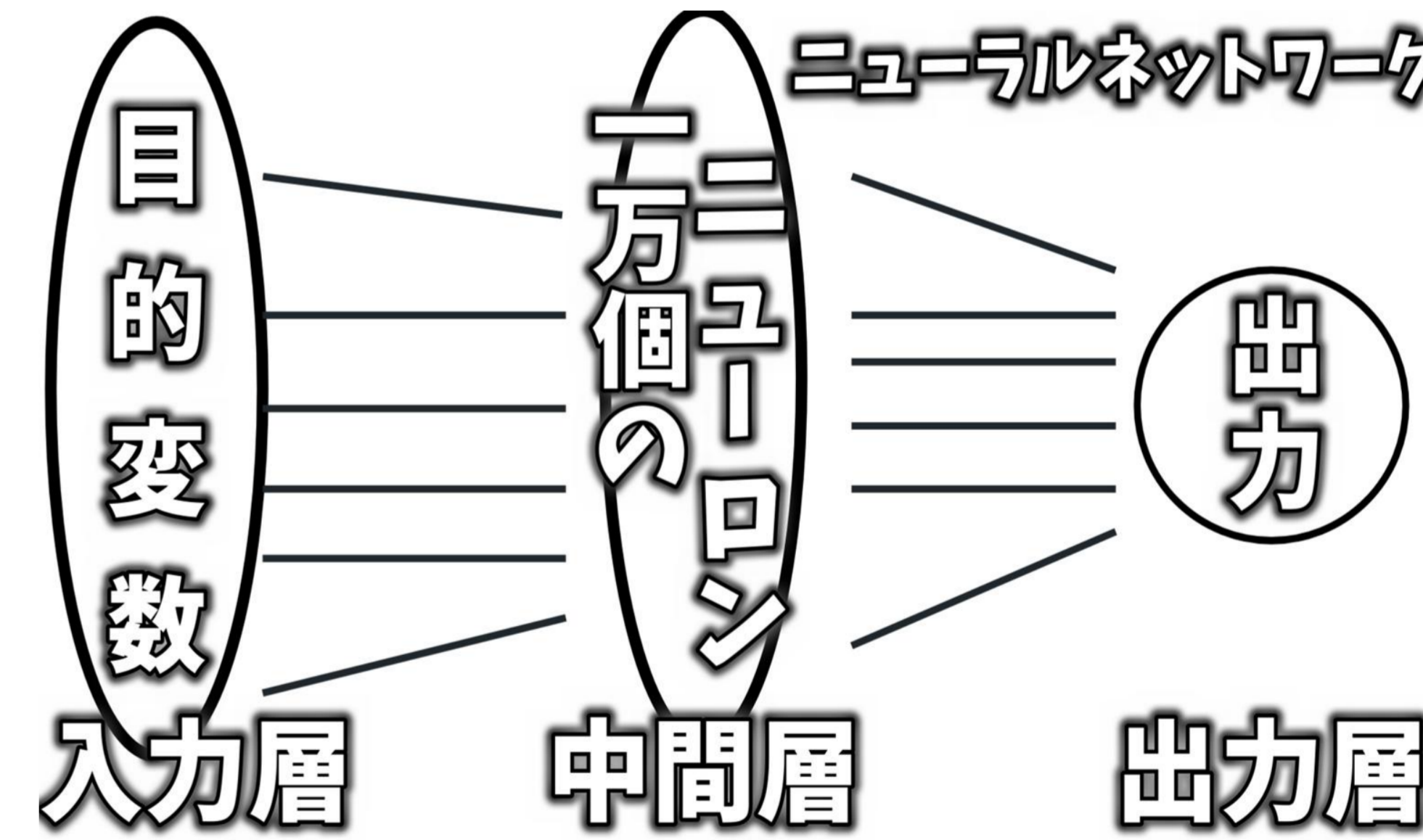
① 特徴量エンジニアリング (データの質を高める)

② ハイパーパラメータ最適化 (学習の質を高める)

定量的比較

「初期モデル」と、2軸の最適化を経た「最終モデル」を比較。

「精度 (Loss)」と「効率 (Epochs)」がどれだけ改善したかを定量的に評価し、本アプローチの有効性を明らかにする。

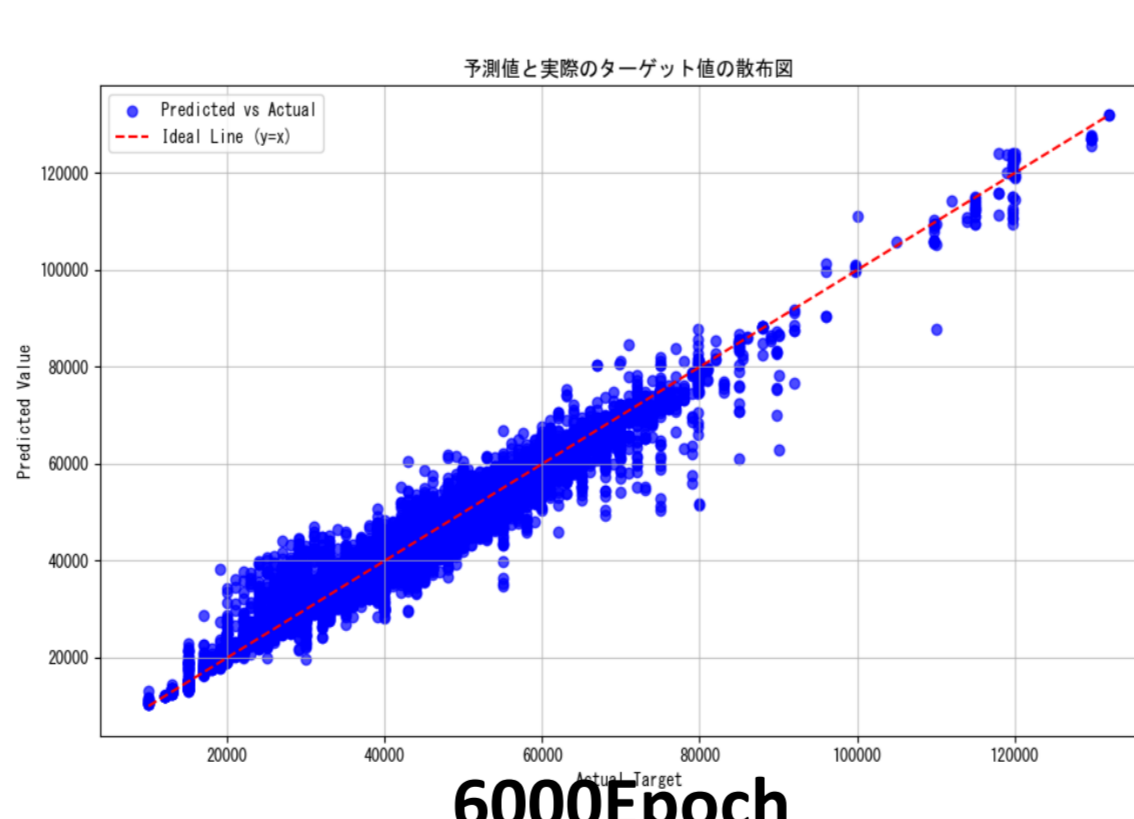
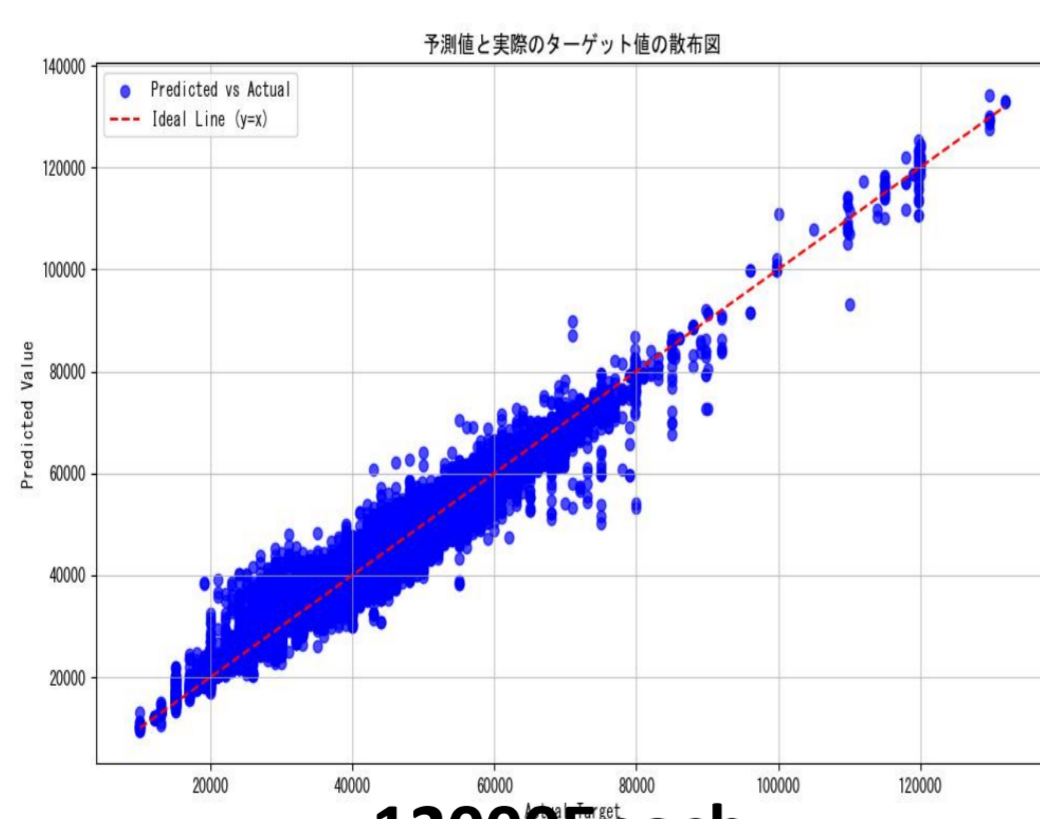


結果

本研究のアプローチ (シンプルなMLP構造 + 最適化) により、予測精度と学習効率の両面で、著しい性能向上が確認された。

- 予測精度の向上 (Lossの低減) 予測誤差を示すLoss値は、初期モデル (ベースライン) の 0.0473 に対し、最終モデル (最適化後) では 0.0396 へと明確に減少した
- 学習効率の向上 (Epochの削減) 学習回数 (Epoch) は、初期モデルの 12000 回から、最終モデルでは 6000 回へと半減した

比較項目	初期モデル (ベースライン)	最終モデル (最適化後)
予測精度 (loss)	0.0473	0.0396
学習回数 (epoch)	12000	6000



考察

・精度の向上: 特徴量エンジニアリングと学習率の最適化により、予測誤差 (Loss) は 0.0473 から 0.0396 へと約16%改善した。

・効率の飛躍的向上: より少ない学習 (12000 → 6000 Epochs) で、より高い精度を達成した。これは、最適化によって学習効率が約2倍に高速化したことを示している。

この結果から、モデル構造を複雑化させることなく、データと学習プロセスを徹底的に最適化するアプローチが、性能向上に極めて有効であったと結論付けられる。

謝辞

本研究の実施にあたり、国立情報学研究所のIDRデータセット提供サービスによりアットホーム株式会社から提供を受けた「アットホームデータセット」を使用いたしました。貴重なデータをご提供くださいました関係各位に、心より感謝申し上げます。

参考文献

[1]アットホーム株式会社 (2024): アットホームデータセット. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.13.1>

[2] 斎藤 康毅 (2016): 『ゼロから作る Deep Learning — Python で学ぶディープラーニングの理論と実装』, オライリー・ジャパン