

ドッキングシミュレーションおよび 機械学習を活用した医薬品化学構造の設計

June. 10, 2016
ディスカッション

酒井研究室 宮崎 大輝

学習性能評価

-実装したこと

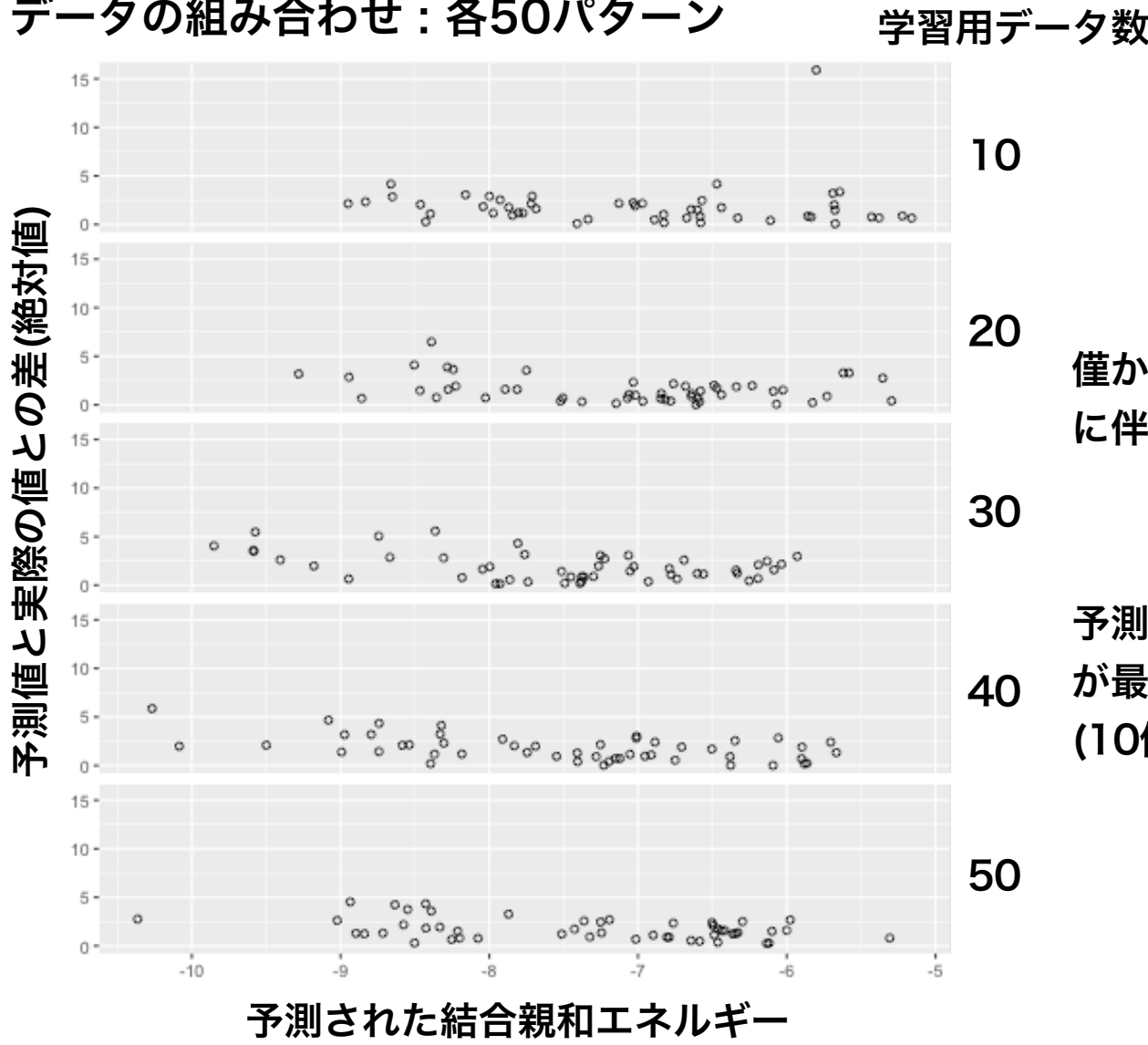
- ・ 学習データ数とデータの組み合わせを変えた学習用データとテストデータの作成
- ・ 候補探索計算を学習データ数とデータの組み合わせを変えてのループ化視化
- ・ 計算により得られた候補の
 - 予測目的変数
 - 予測値と実際の値との差
 - CAS No. を保存 → csvファイル
- ・ 各試行における予測値と、予測誤差を可視化

-問題点

- ・ ドッキングシミュレーションから得られる結合親和エネルギーの有効桁数が少なく、学習用データのy値が被る → エラー
- ・ CAS No. ではなく構造式として可視化したい

学習性能評価

学習用データ数：10 – 50 個
データの組み合わせ：各50パターン

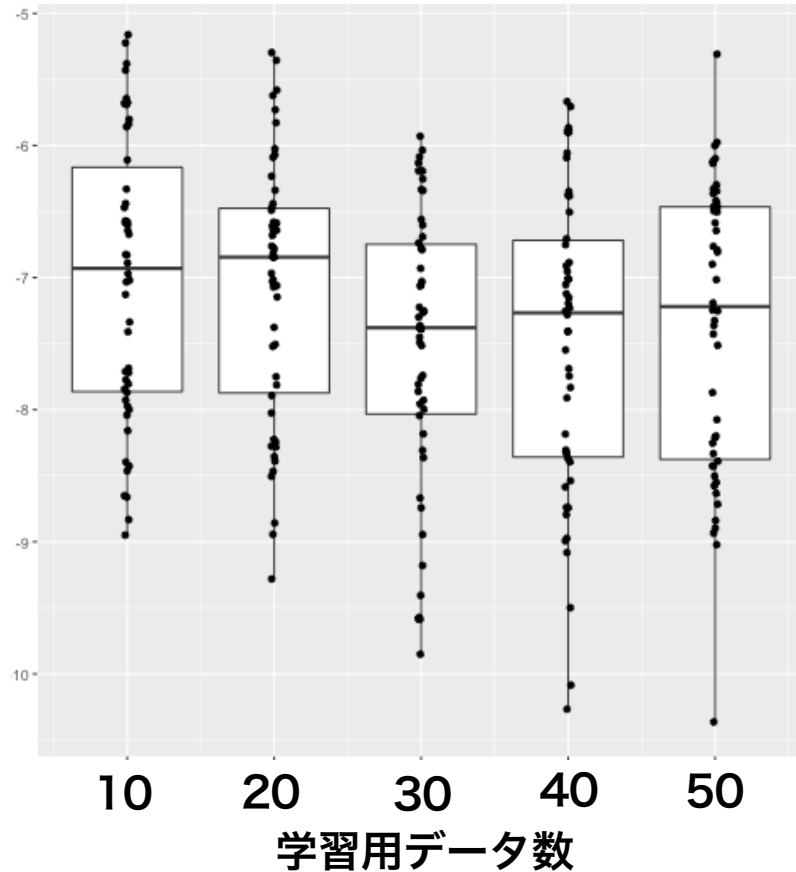


僅かだが、学習用データ数増加に伴い予測値は良くなっている？

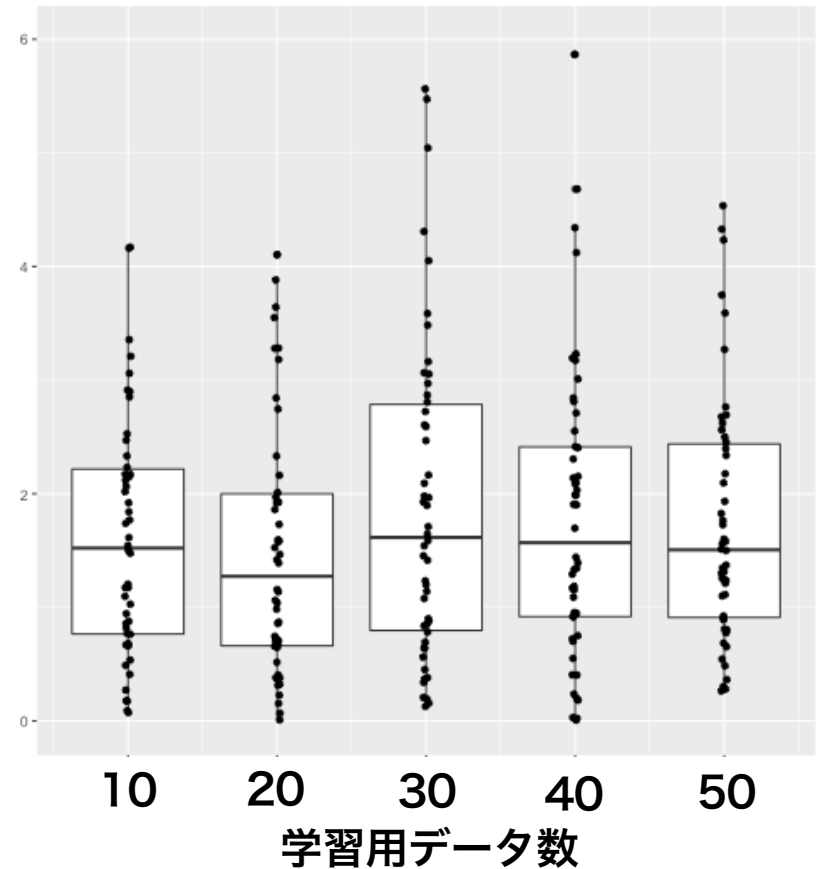
予測誤差は学習用データ50個の 때가最も小さい？
(10個の時も小さい)

学習性能評価

予測値の箱ひげ図



予測誤差値の箱ひげ図



予測値 : 学習データの増加に伴いより良い予測値が得られる傾向?

予測誤差 : 全体的に誤差は小さく、性能は良い?

試行回数が少なく、有意な結果が得られていない可能性