

# 電気情報工学セミナーⅡ

実験のためのPython & Git入門（機械学習による関数の近似実験）

～ 研究課題について ～

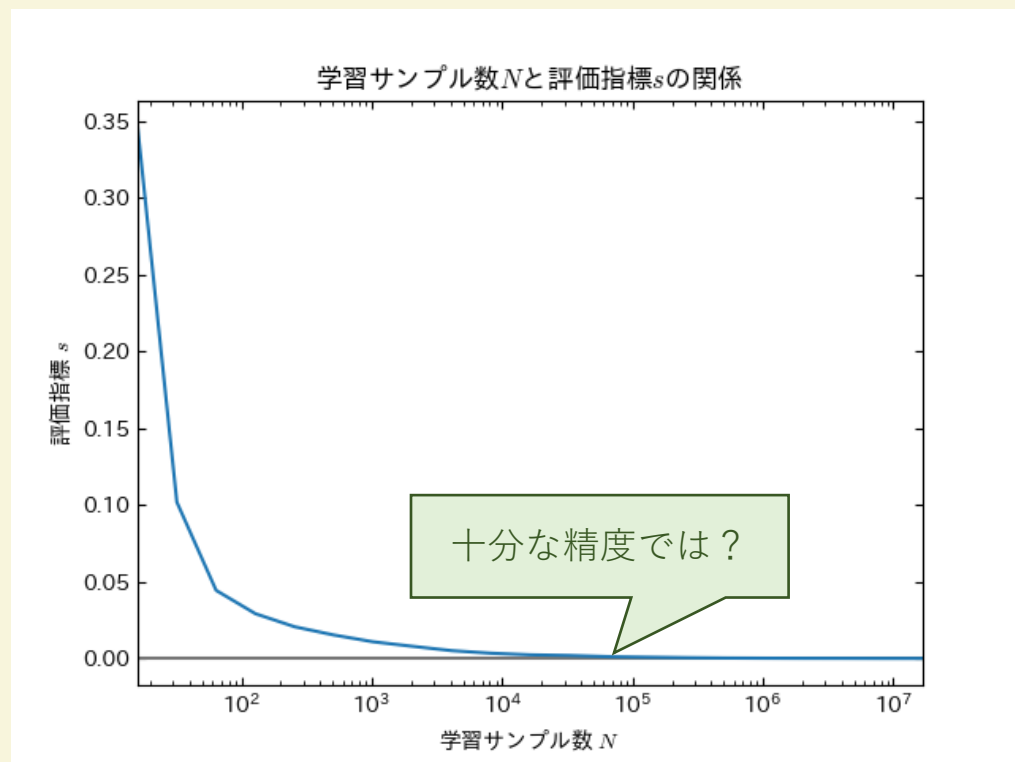
池原研究室

# 研究課題の例

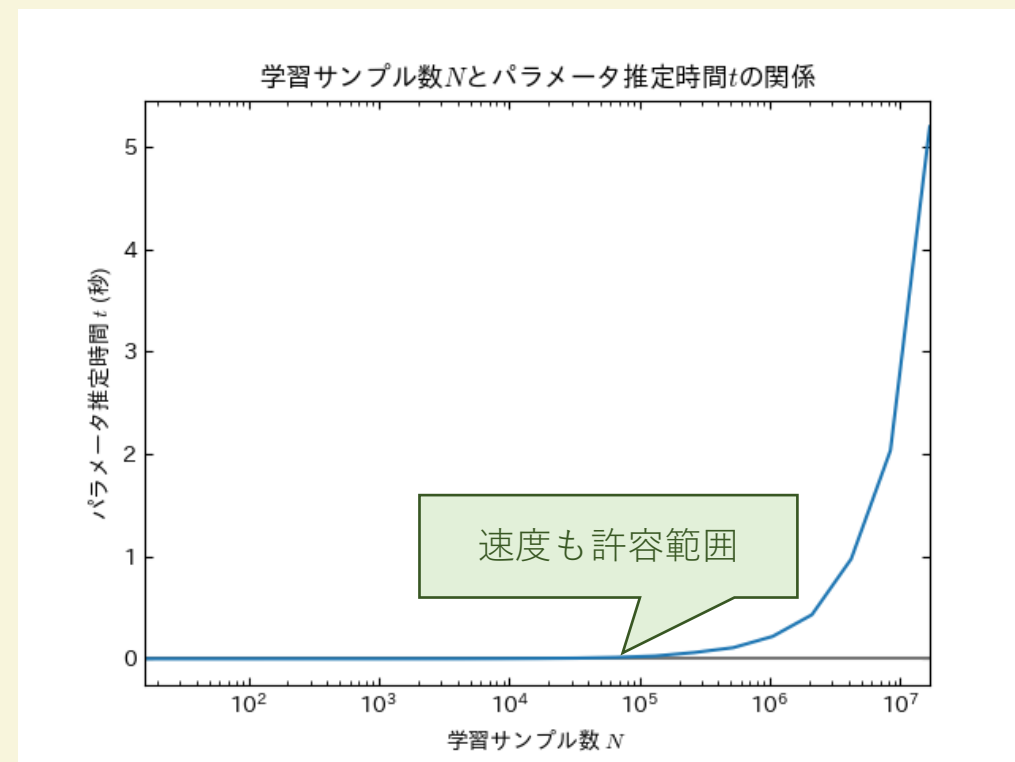
# 学習サンプルを増やしたときの効果・副作用

- $\sin(\pi x)$ を7次多項式でフィッティング
- 20回繰り返した際の平均をプロット

- $N = 2^n$  ( $n = 4, \dots, 24$ )で実験  
( $N = 16, \dots, 166777216$ )



サンプル数を増やすと精度が上がる



パラメータ推定時間も増える

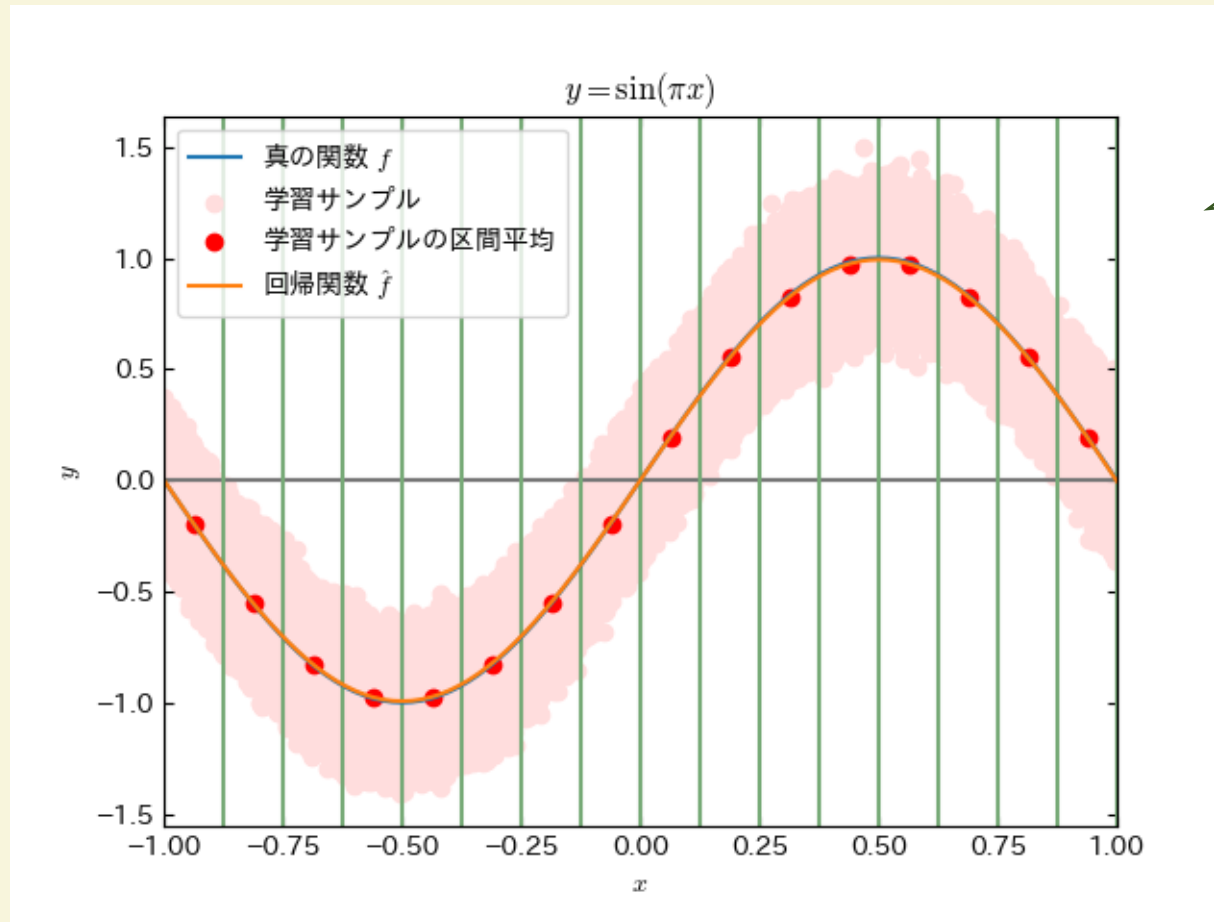
## 提案①：N=65536で打ち切る

$$a = \left( X^T \cdot X \right) \cdot X^T \cdot y$$

計算コストは $O(N(2d^2 + 1))$  + 逆行列の計算コスト

例)  $2^{24}$ 個のサンプルを $2^{16}$ 個だけ利用すれば計算コストは約1/256

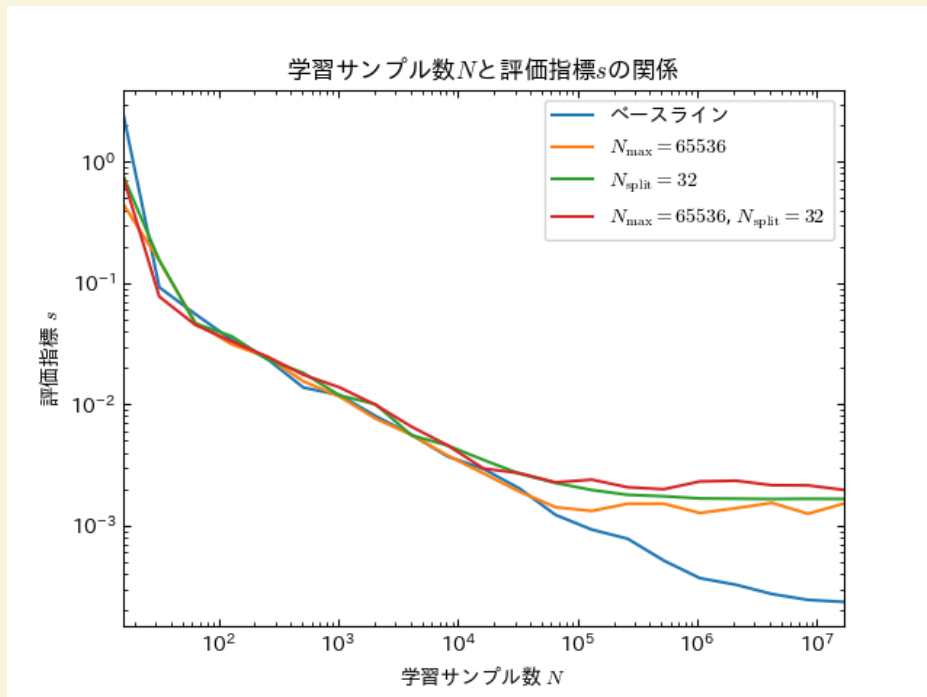
## 提案②：小区間に分割し平均をとる



$N_{\text{split}} = 16$ の例  
評価スコア0.07

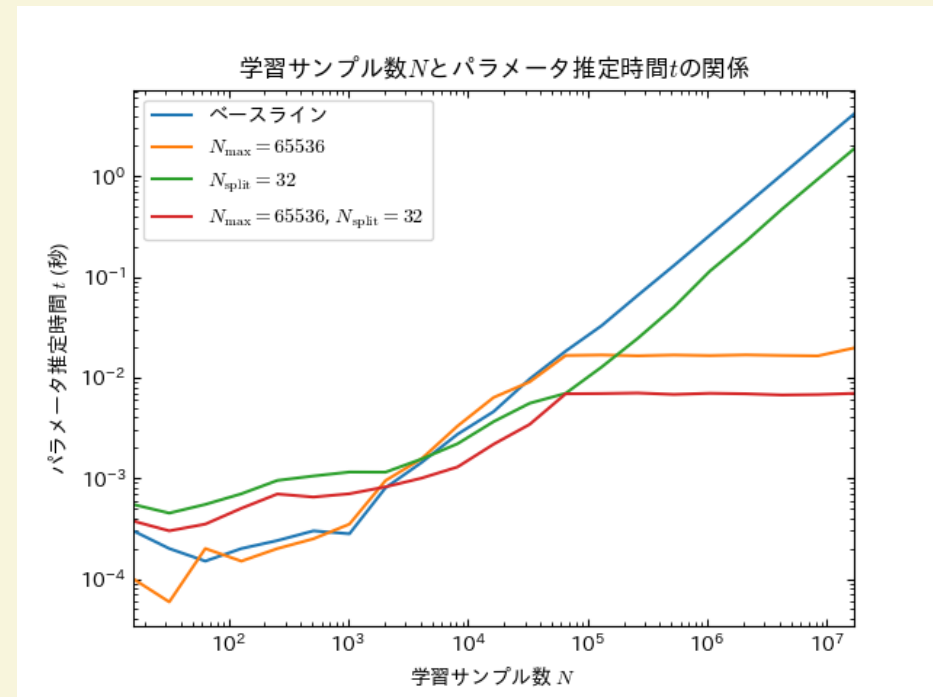
計算コストは  $O(N_{\text{split}}(2d^2 + 1)) + O(N) + \text{区間分割コスト} + \text{逆行列の計算コスト}$

# 結果



- サンプルが増えるほど精度は上がる
- 提案①②ともに精度は0.002程度で頭打ちになる

➤ 提案②は提案①より精度が低く、時間もかかるため単体での使用は向かない



- サンプルが増えるほど推定時間も長くなる
- 提案①によって推定時間に上限ができる
- 提案②は推定時間を半分程度に短縮する

# 研究課題

# 要件

- ✓ 1人1つ以上自分でテーマを決めて取り組む
  - テーマ例から選んでも、完全にオリジナルでもOK
  - 好きなだけ取り組んでよいが1つで十分
- ✓ 1人5～30分程度の発表時間
  - 提案内容はシンプルに伝える
  - 質疑応答の時間も別途設ける
  - 動機付け→（提案→）実験の順で発表するとまとめやすい
- ✓ 自分で実験をおこなう
  - どういう比較・調査をするか実験をデザインする
  - プログラミングして実測する
  - 結果はグラフや表で示す
- ✓ 提案型の場合、提案内容がうまくいかなくてもOK
  - 「バグでプログラムが動かなかった」というのはNG
  - うまくいかなかった理由を考える



# テーマ例（多項式フィッティング）

- $\sin(n\pi x)$ を近似するには何次多項式を使うのが良い？
- さまざまな関数を多項式フィッティングしてみた（うまくいく例、いかない例）
- 高次の多項式を使えばうまくいくのか？
- $x^p$ 以外を基底関数としてフィッティング

# テーマ例（ガウス過程回帰）

- $G^{-1}$ を用いず、 $\mathbf{g}^T(x)$ を合計が1になるように正規化するとオリジナルと比べどうなる？
- 最適な $\sigma_x$ や $\sigma_y$ のを見つけ方はある？（特に少ないサンプル数で）
- 多項式フィッティングと比較してメリット・デメリットは？
- 多項式フィッティングの最適な次数 $d$ をガウス過程回帰で予測してみた