

4-1

1 訓練誤差 : $2 \cdot 10^{-15}, 2 \cdot 10^{-14}, 2 \cdot 10^{-13}, 2 \cdot 10^{-12}, 2 \cdot 10^{-11}, 2 \cdot 10^{-10}, 2 \cdot 10^{-9}, 2 \cdot 10^{-8}, 2 \cdot 10^{-7}$
 $(MSE = 0.4071530532)$

テスト誤差 : $2 \cdot 10^{-6}$
 $(MSE = 0.4583527600)$

2 訓練誤差 : $2 \cdot 10^{-16}$
 $(MSE = 0.4071530599)$

テスト誤差 : $2 \cdot 10^{-7}$
 $(MSE = 0.4560211245)$

3 リッジ回帰 : $2 \cdot 10^{-6}$
 λ : $2 \cdot 10^{-7}$

4 テスト誤差が最小になる正規化係数 λ を選択した際、リッジ回帰のテスト誤差が、ランソンのテスト誤差がわずかに小さくなっている。
 一般には、リッジ回帰は全ての変数を利用して予測をするため、予測能力が高くなると考えられる。

今回の問題では、4特徴選択が必要のないデータであるため、ランソンの全ての特数量を利用している。しかし、4特徴の決定係数が高い場合、ランソンのリッジ回帰に比べ特徴選択をすることから、

- ・ 計算効率の向上
- ・ 目標変数の予測精度向上
- ・ 学習結果の解釈性の向上

などのメリットが期待できる

ランソン

リッジ回帰

```
alpha= 64.00000, Train MSE = 0.4076564423, Test MSE = 0.4583527600
Feature Coefficients
1 volatile acidity -0.177736
6 total sulfur dioxide -0.120092
7 density -0.077231
4 chlorides -0.074966
8 pH -0.032287
2 citric acid -0.029137
5 free sulfur dioxide 0.030532
3 residual sugar 0.047255
0 fixed acidity 0.073536
9 sulphates 0.126044
10 alcohol 0.271555
```

```
alpha= 0.00781, Train MSE = 0.4089173915, Test MSE = 0.4560211245
Feature Coefficients
1 volatile acidity -0.180921
6 total sulfur dioxide -0.109501
4 chlorides -0.073494
8 pH -0.044440
7 density -0.018053
2 citric acid -0.012326
0 fixed acidity 0.016785
5 free sulfur dioxide 0.017682
3 residual sugar 0.018687
9 sulphates 0.116373
10 alcohol 0.308392
```