わざと当てはまりを悪くすると予測が良くなり、どれだけ縮めるかは信号の強さが決める

— 係数を $n/(n+\lambda)$ 倍に縮めると、訓練側の外れは単調に増えるのに予測の誤差は $\lambda = 50$ で最小の **0.333333** をとり、最適な λ は次元×雑音÷信号である — 分離子は、信号が雑音×次元÷標本数より強いか弱いか — [論文744]

山岸 勇輝 (Yuuki YAMAGISHI)
Yamagishi Complexity Torus Lab, Japan｜山岸勇輝ラボ
2026年9月

— From thinking to syncing. —

要旨

手元のデータにいちばんよく合う係数が、新しいデータにいちばんよく当たるとは限らない。係数をわざと縮めると、予測は良くなりうる。どれだけ縮めればよいか、そしてどれだけ得をするかを、閉じた式とシミュレーションで数える。新しい定理も法則も主張しない。 第一に、縮めるほど訓練側では外れる——外れの項は $\lambda = 5 \cdot 50 \cdot 500$ で **0.002268**・**0.111111**・**0.694444** と単調に増える（第2節）。 第二に、それでも予測は良くなる——予測の誤差は **0.500000** から $\lambda = 50$ で **0.333333** まで下がり、500 では **0.708333** に戻る（第2節）。 第三に、これが本稿の芯である。最適な縮め方は、次元×雑音÷信号で決まる——信号 1 なら $\lambda = $ **50.000000**、信号 10 なら **5.000000**（第3節）。 第四に、信号が弱いほど得が大きい——誤差は信号 1 で **0.666667** 倍、信号 10 では **0.952381** 倍にしか下がらない（第3節）。 第五に、シミュレーションが裏を取る——20000 回で $\lambda = 0$ が **0.499461**、最適 λ が **0.332937**（第3節）。 第六に、分離子は、信号が雑音×次元÷標本数より強いか弱いかである——弱ければ大きく縮め、強ければほとんど縮めない（第4節）。 キーワード：正則化、リッジ回帰、係数の縮小、偏りと分散、信号と雑音の比、訓練と予測の逆向き。

本稿の射程（射程注記）：新しい定理も法則も主張しない——リッジ回帰、縮めることで偏りと分散を交換すること、直交した説明変数での最適な λ は、いずれも既知である（ホール＆ケナード 1970、ハスティー 2009）。直交した説明変数に限る——閉じた式を使うためで、相関のある説明変数は扱わない。λ の選び方を扱わない——信号の強さを知っているとして最適を出す。実際には知らないので、交差検証などで選ぶことになるが、それは扱わない。他の罰則を扱わない——ラッソや変数選択は扱わない。実在のデータを主張しない——個別のデータや学習器は扱わない。既刊との関係：論文733 は、誤差には最小が在り、その位置が手持ちのデータ数で動くことを示した——その射程注記は「交差検証、正則化、早期打ち切りの手順は扱わない」と明記していた。そこでは模型の自由度を選んで最小を作った。ここでは自由度を変えずに係数を縮め、同じ最小を連続なつまみで作る。自由度を選ぶことと、係数を縮めることになる。論文330 も引く——そこでは正規方程式で条件数が二乗になるのが恒等式だと示された。リッジの λ は、その行列の対角に足される量である。加えたのは、訓練側と予測側が逆向きに動くことを同じ表で示したこと、最適な λ を閉じた式と格子とシミュレーションで確かめたこと、得の大きさが信号の強さで **0.666667** 倍から **0.952381** 倍まで変わることを示したこと、分離子を「信号と雑音×次元÷標本数の比」に置いたことである。

1. 記法と問い

設定：直交した説明変数では、リッジ回帰は最小二乗の係数を一律に $n/(n+\lambda)$ 倍に縮める。予測の余分な誤差は、縮めたことによる偏りと、残った分散の和になる。

$$\hat{\beta}_\lambda = \frac{n}{n+\lambda}\,\hat{\beta}, \qquad \mathrm{E}(\lambda) = \Big(\frac{\lambda}{n+\lambda}\Big)^2 \|\beta\|^2 + \Big(\frac{n}{n+\lambda}\Big)^2 \frac{d\sigma^2}{n}, \qquad \lambda^* = \frac{d\sigma^2}{\|\beta\|^2}$$

次元 $d = 50$、標本 $n = 100$、雑音の分散 $\sigma^2 = 1$ とし、雑音×次元÷標本数は 0.500000 である。

問い：手元にいちばん合う係数は、いちばん当たる係数なのか。

2. 訓練側と予測側は、逆向きに動く

計算 2.1：信号 $\|\beta\|^2 = 1$ として、λ ごとに縮め係数、訓練側の外れの項、予測の誤差を出した。

λ	縮め係数	訓練側の外れの項	予測の誤差
0	1.000000	0.000000	0.500000
5	0.952381	0.002268	0.455782
50	0.666667	0.111111	0.333333
500	0.166667	0.694444	0.708333

> **観察 2.2**
>
> 縮めるほど、訓練側では外れる。外れの項は λ とともに単調に増える——手元のデータへの当てはまりだけを見れば、$\lambda = 0$ がいつも最善である。

> **観察 2.3**
>
> ところが予測の誤差は途中で最小になる。0.500000 から $\lambda = 50$ で 0.333333 まで下がり、$\lambda = 500$ では 0.708333 に戻る——わざと外すと当たるようになり、外しすぎるとまた当たらなくなる。

3. 縮め方は、信号の強さが決める

計算 3.1：信号の強さを二通りにして、最適な λ と、そのときの誤差を出した（格子で探した最小とも照らした）。

信号 $\|\beta\|^2$	最適な λ	最適での誤差	$\lambda = 0$ に対する比
1	50.000000	0.333333	0.666667
10	5.000000	0.476190	0.952381

> **観察 3.2**
>
> これが本稿の芯である。最適な縮め方は、次元×雑音÷信号で決まる。信号が 10 倍になると最適な λ は 10 分の 1 になり、格子で探した最小もそれぞれ 50.00 と 5.00 に一致する——手元への当てはまりではなく、信号と雑音の比が縮め方を決めている。

> **観察 3.3**
>
> 信号が弱いほど、得が大きい。信号 1 で誤差は 0.666667 倍、信号 10 では 0.952381 倍にしか下がらない——最適での誤差は雑音×次元÷標本数に信号の割合を掛けた形で、信号が雑音側に比べて弱いときに縮める値打ちが出る。

計算 3.4：信号 1 で、係数に雑音を足して縮める操作を 20000 回繰り返した。

λ	シミュレーション	閉じた式
0	0.499461	0.500000
50	0.332937	0.333333

観察 3.5

二つとも 2 % 以内で閉じた式に乗る。最適での誤差が $\lambda = 0$ の 3 分の 2 になることを、別の道で確かめた——式の都合ではない。

4. 分離子

観察 4.1

分離子は、信号が雑音×次元÷標本数より強いか弱いかである。弱ければ大きく縮め、強ければほとんど縮めない——「正則化は効くか」という問いには、この比を言うまで答が無い。

観察 4.2

論文733 と並べる。733 は、誤差の最小の位置が手持ちのデータ数で動くことを示し、正則化の手順は扱わないと明記した——そこでは模型の自由度を選んで最小を作った。ここでは自由度を変えずに係数を縮め、同じ最小を連続なつまみで作る。

観察 4.3

論文330 とも並べる。330 は、正規方程式で条件数が二乗になるのが近似ではなく恒等式であることを示した——リッジの λ はその行列の対角に足される量で、本稿はそれを予測の誤差の側から数えた。

観察 4.4

言わないこと。相関のある説明変数も、λ の選び方も、他の罰則も扱わず、信号の強さを知っているとして最適を出した——示したのは、訓練と予測が逆向きに動くことと、縮め方を決めている比である。

5. 点呼と結び

1. $\lambda = 0$ で予測 **0.500000**（計算 2.1）：手元に最もよく合う。
2. $\lambda = 50$ で **0.333333**（計算 2.1）：わざと外すと当たる。
3. $\lambda = 500$ で **0.708333**（計算 2.1）：外しすぎ。
4. 訓練側の外れ **0.111111**（計算 2.1）：単調に増える。
5. 信号 **1** で最適 **50.000000**（計算 3.1）：本稿の芯。
6. 信号 **10** で **5.000000**（計算 3.1）：**10** 分の **1**。
7. 得は **0.666667** 倍（観察 3.3）：信号が弱い。
8. 得は **0.952381** 倍（観察 3.3）：信号が強い。
9. シミュレーション **0.332937**（計算 3.4）：閉じた式に乗る。
10. 信号と雑音×次元÷標本数の比（観察 4.1）：分離子・論文**733**・**330**。

結び：わざと当てはまりを悪くすると予測が良くなり、どれだけ縮めるかは信号の強さが決める。直交した説明変数・次元 **50**・標本 **100** で係数を縮めると、訓練側の外れは $\lambda = 5$・**50**・**500** で **0.002268**・**0.111111**・**0.694444** と単調に増えるのに、予測の誤差は **0.500000** から **0.455782**・**0.333333** と下がって **0.708333** に戻る——手元への当てはまりだけを見れば $\lambda = 0$ が最善で、それは予測では最善でない。最適な λ は次元×雑音÷信号で、信号 **1** なら **50.000000**、信号 **10** なら **5.000000** になり、誤差はそれぞれ **0.666667** 倍と **0.952381** 倍に下がる——信号が雑音×次元÷標本数に比べて弱いときほど、縮める値打ちが出る。**20000** 回のシミュレーションでは **0.499461** と **0.332937** で、閉じた式と **2 %** 以内だった。分離子は、信号が雑音×次元÷標本数より強いか弱いかである——「正則化は効くか」には、この比を言うまで答が無い。既刊との位置——論文**733** は自由度を選んで最小を作り、正則化は扱わないと明記した。ここは係数を縮めて同じ最小を連続なつまみで作る。正直に言えば——直交した説明変数に限り、λ の選び方も他の罰則も扱わず、信号の強さを知っているとした。

参考文献

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
2. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*.
3. 山岸勇輝 (2026).『条件数が二乗になるのは、近似ではなく恒等式である』[論文330]. Zenodo. 10.5281/zenodo.22681423
4. 山岸勇輝 (2026).『誤差には最小が在るが、その最小の位置は模型の性質ではなく、手持ちのデータ数で動く』[論文733]. Zenodo. 10.5281/zenodo.23015055

山岸 勇輝 (Yuuki YAMAGISHI)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)