The Error Has a Minimum, but Where That Minimum Sits Is Not a Property of the Model: It Moves with How Much Data You Hold

— at **20** points it sits at five degrees of freedom and at **2500** it sits at twenty-seven, so **125** times the data moves it by **5.400000** — the separator is whether the error counted is on the data fitted or on data not fitted — [Paper 733]

Yuuki YAMAGISHI (山岸 勇輝)
Yamagishi Complexity Torus Lab, Japan
September 2026

— From thinking to syncing. —

Abstract

A model too complex and a model too simple both err, and a minimum lies between. What is counted here is that where the minimum lies is not a property of the model.Its location is shown to move with the amount of data.No new theorem or law is claimed. First, the two terms move in opposite directions——at one degree of freedom, bias 4.000000 and variance 0.010000; at thirty-two, 0.003906 and 0.320000 (Section 2). **Second, so a minimum lies inside——at 100 points it is 1.139383 at nine degrees of freedom, below the 5.010000 and 1.323906 at the ends** (Section 2). **Third, and this is the core. The location of the minimum moves with the data——five degrees of freedom at 20 points and twenty-seven at 2500, a factor of 5.400000** (Section 3). **Fourth, so "the right complexity" is not a property of the model——within one family of models, the answer changes with how much data one holds** (Section 3). **Fifth, the noise term does not go away——even at 10^6 points the error cannot fall below the noise at 1.0** (Section 3). **Sixth, the separator is whether the error counted is on the data fitted or on data not fitted——on the first it falls with complexity and there is no minimum at all** (Section 4). **Keywords**: bias and variance, error decomposition, optimal complexity, sample size, irreducible noise, which data the error is counted on.

Scope of this paper (scope note): **No new theorem or law is claimed——the decomposition into bias, variance and noise, the opposing directions of the first two, and the growth of the optimal complexity with sample size are all standard (Geman 1992, Hastie 2009).The forms are assumed——squared bias is taken as $4/p^2$ and variance as $\sigma^2 p/n$. This claims nothing about the behaviour of any real learner; only the structure, that the two oppose and the minimum moves, is read.Procedures are not treated——cross-validation, regularisation and early stopping are not treated.Double descent is not treated——the re-descent of error at still higher complexity does not arise in this form.No real model is claimed——particular data sets and learners are not treated.Relation to earlier papers**: **Paper 701 showed that the more data there is, the more a penalised criterion prefers the more complicated model——there what the criterion selects was the subject, with the crossing given numerically. Here the error itself is split and the complexity minimising it grows with the data. Selection, and the error of what is selected: both point the same way, toward complexity as data grows.**What is added is **tabulating the three terms together**, **showing the minimum to lie inside by comparison with both ends**, **putting the movement of the minimum at 5.400000**, **showing the noise floor to remain**, and **placing the separator on which data the error is counted upon**.

1. Notation and the Question

Setting: **the error on data not fitted splits into three: squared bias, variance and noise.**

$$\text{error} = \text{bias}^2(p) + \text{variance}(p, n) + \sigma^2, \qquad \text{bias}^2 = \frac{4}{p^2}, \qquad \text{variance} = \frac{\sigma^2 p}{n}$$

p is the degrees of freedom and n the number of data points. The two forms are assumed, with $\sigma^2 = 1.0$.

Question: **the right complexity is a property of what?**

2. Two Terms Oppose, and a Minimum Lies Inside

Computation 2.1: **at 100 data points, the three terms were derived against the degrees of freedom.**

degrees of freedom	squared bias	variance	total
1	4.000000	0.010000	5.010000
2	1.000000	0.020000	2.020000
4	0.250000	0.040000	1.290000
8	0.062500	0.080000	1.142500
16	0.015625	0.160000	1.175625
32	0.003906	0.320000	1.323906

Observation 2.2

the directions are opposite.Bias falls with the degrees of freedom and variance rises——so both ends are bad and a minimum lies inside.

Observation 2.3

the minimum is 1.139383 at nine degrees of freedom.Below the 5.010000 at one and the 1.323906 at thirty-two——the badness of the simple end comes from bias, that of the complex end from variance.

3. The Minimum Moves with the Data

Computation 3.1: for each amount of data, the minimising degrees of freedom and the three terms there were derived.

data	minimising d.o.f.	squared bias	variance	error
20	5	0.160000	0.250000	1.410000
100	9	0.049383	0.090000	1.139383
500	16	0.015625	0.032000	1.047625
2500	27	0.005487	0.010800	1.016287

Observation 3.2

this is the core: 125 times the data moves the minimising degrees of freedom from five to twenty-seven, a factor of 5.400000.Not one thing about the family of models has changed——"this model is about right at this complexity" silently assumes how much data one holds.

Observation 3.3

the direction comes from the variance side.Variance is inversely proportional to the data, so more data makes complexity cheaper——the bias side does not depend on the data, so one may spend the savings on complexity.

Observation 3.4

the noise still does not go.Even at 10^{6} points the error cannot fall below 1.0——aiming at zero error does not stand inside this decomposition.

Observation 3.5

the three terms share a unit and not a nature.Bias answers to the choice of model, variance to the amount of data, noise to neither——they tabulate together and have different knobs.

4. The Separator

Observation 4.1

the separator is whether the error counted is on the data fitted or on data not fitted.On the first it falls with complexity and no minimum exists——that a minimum exists at all is a consequence of looking at data from outside.

Observation 4.2

beside Paper 701.Paper 701 showed that the more data there is, the more a penalised criterion prefers the more complicated model——there selection was the subject. Here the decomposition of error moves the same way. The selecting side, and the error of what is selected.

Observation 4.3

what is not said.Cross-validation, regularisation and double descent are not treated, and the two forms are assumed——what is shown is that the minimum moves and what moves it.

5. Roll Call and Closing

1. **5.010000 at one d.o.f.** (Computation 2.1): **bias dominates.**
2. **1.323906 at thirty-two** (Computation 2.1): **variance dominates.**
3. **minimum 1.139383** (Computation 2.1): **at nine.**
4. **five d.o.f. at 20 points** (Computation 3.1): **where the minimum sits.**
5. **twenty-seven at 2500** (Computation 3.1): **it has moved.**
6. **a factor of 5.400000** (Observation 3.2): **the core.**
7. **error down to 1.016287** (Computation 3.1): **at 2500 points.**
8. **the noise 1.0 remains** (Observation 3.4): **zero error is unavailable.**
9. **variance goes as $1/n$** (Observation 3.3): **complexity gets cheaper.**
10. **which data is counted** (Observation 4.1): **the separator; Paper 701.**

Closing: **The error has a minimum, but where that minimum sits is not a property of the model: it moves with how much data you hold.Splitting the error into squared bias, variance and noise, the first two oppose: at 100 points, one degree of freedom gives 4.000000 plus 0.010000, hence 5.010000; thirty-two give 0.003906 plus 0.320000, hence 1.323906; and nine give the minimum 1.139383——the badness of the simple end comes from bias, that of the complex end from variance.Yet the location moves: five degrees of freedom at 20 points and twenty-seven at 2500, a factor of 5.400000 for 125 times the data——nothing about the family of models has changed, so "this model is about right at this complexity" silently assumes how much data one holds.The direction comes from the variance side——variance goes inversely with the data, so more data makes complexity cheaper, while bias does not depend on the data at all, leaving the savings to be spent on complexity.The noise still does not go: even at 10^{6} points the**

error cannot fall below 1.0——aiming at zero error does not stand inside this decomposition.The separator is whether the error counted is on the data fitted or on data not fitted——on the first it falls with complexity, so that a minimum exists at all is a consequence of looking at data from outside.Placed among the earlier papers——Paper 701 treated selection; here it is the error of what is selected, moving the same way.To be honest——the two forms are assumed, and cross-validation and double descent are not treated.

References

1. Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
3. Yamagishi, Y. (2026). *The More Data There Is, the More a Penalised Criterion Prefers the More Complicated Model* [Paper 701]. Zenodo. 10.5281/zenodo.23014898

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

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