“Error Goes as Dimension over Sample Size” Is a Far-Side Approximation; Near the Dimension the Error Diverges

— For least squares with Gaussian regressors the excess prediction error is not the familiar d/n but $d/(n-d-1)$ — the separator is whether the ratio of sample size to dimension is large or close to 1 — [Paper 743]

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

— *From thinking to syncing.* —

Abstract

The excess error of a regression is usually estimated as dimension over sample size. That is the form for samples far outnumbering the dimension.Set beside the exact form, the reach of the approximation and the manner of its failure are counted.No new theorem or law is claimed. First, the exact form is $d/(n-d-1)$——dimension 10 and 1000 samples give 0.010111, 1.011122 times the approximation (Section 2). **Second, fewer samples widen the gap——0.112360 at 100, 1.111111 at 20, 2.500000 at 15** (Section 2). **Third, and this is the core. As the sample size nears the dimension, the error diverges——10.000000 at 12 samples, 12.000000 times the approximation** (Section 2). **Fourth, holding the ratio removes the dependence on dimension——ten samples per dimension give 0.112360 at dimension 10 and 0.111235 at 100** (Section 3). **Fifth, simulation confirms it——dimension 5 gives 0.357921 at 20 samples (exact 0.357143) and 0.091716 at 60 (0.092593)** (Section 3). **Sixth, the separator is whether the ratio of sample size to dimension is large or close to 1——the approximation works only in the first case** (Section 4).
Keywords: least squares, prediction error, dimension and sample size, reach of an approximation, divergence, ratio of sample size to dimension.

Scope of this paper (scope note): **No new theorem or law is claimed——the least-squares prediction error $\sigma^2 d/(n-d-1)$ for Gaussian regressors is known (Seber 1977, Hastie 2009).Regressors are kept simple——independent Gaussian; correlation and outliers are not treated.Regularisation is not treated——shrinking coefficients and selecting variables are not treated.Only the near side of the boundary——for sample sizes at or below dimension plus one the expectation is not finite, and that region is not treated.No real data are claimed——particular data sets and learners are not treated.Relation to earlier papers**: **Paper 733 showed that the error has a minimum whose location moves with how much data one holds——there variance was placed in the form $\sigma^2 p/n$, with no claim about real learners. Here the exact version of that form is given, and the form itself is shown to fail where sample size nears dimension. The placed form, and the reach of that form.Paper 182 is also cited——as an instance, from the side of sample size, that the curse of dimensionality is not one curse.**What is added is **setting the exact form beside the approximation**, **showing the gap to reach $\mathbf{12.000000}$**, **showing independence from dimension at fixed ratio**, **confirming it by simulation**, and **placing the separator on the ratio of sample size to dimension**.

1. Notation and the Question

Setting: **take d independent Gaussian regressors and fit least squares on n samples. With noise variance σ^2, the expected excess prediction error at a new point is as follows.**

$$\mathbf{E}\big[\|\hat{\beta}-\beta\|^2\big]=\frac{\sigma^2 d}{n-d-1},\qquad \text{approximation } \frac{\sigma^2 d}{n}$$

Below, $\sigma^2=1$.

Question: **how far can "dimension over sample size" be used?**

2. The Approximation Fails, and the Error Diverges

Computation 2.1: **for varying dimension and sample size, the exact form, the approximation and their ratio were derived.**

dimension	samples	exact	approximation	ratio
10	12	10.000000	0.833333	12.000000
10	15	2.500000	0.666667	3.750000
10	20	1.111111	0.500000	2.222222
10	50	0.256410	0.200000	1.282051
10	100	0.112360	0.100000	1.123596
10	1000	0.010111	0.010000	1.011122
100	1000	0.111235	0.100000	1.112347

Observation 2.2

with many samples the approximation holds well.At dimension 10 and 1000 samples the ratio is 1.011122 ——“dimension over sample size” is the form for this side.

Observation 2.3

this is the core: as the sample size nears the dimension, the error diverges.At 12 samples it is 10.000000, 12.000000 times the approximation——the denominator $n - d - 1$ heads to 0, so what the approximation underestimates grows without bound.

Observation 2.4

the failure is not uniform.The ratio is 1.282051 at 50 samples, 2.222222 at 20 and 3.750000 at 15, rising steeply near the boundary——treating the approximation’s error as “about ten percent” is right only on the far side.

3. Holding the Ratio Removes the Dependence on Dimension

Observation 3.1

with samples held at ten times the dimension, changing the dimension changes little.0.112360 at dimension 10 with 100 samples, 0.111235 at dimension 100 with 1000——what acts is not dimension or sample size but their ratio.

Computation 3.2: **at dimension 5, coefficients, regressors and noise were drawn at random, least squares was solved, and the excess error averaged.**

dimension, samples (runs)	simulation	exact	relative difference
5, 20 (6000)	0.357921	0.357143	0.002179
5, 60 (4000)	0.091716	0.092593	0.009468

Observation 3.3

the simulation sits on the exact form.Even at 20 samples the relative difference is 0.002179——the actual value is 0.357143, not the approximation's $5/20 = 0.25$.

Observation 3.4

more dimensions demand samples in proportion.Held ratio, nearly unchanged error; ratio near 1, divergence——a worsening ratio matters more than dimension itself.

4. The Separator

Observation 4.1

the separator is whether the ratio of sample size to dimension is large or close to 1.The approximation works only in the first case; in the second the error estimate itself collapses——using “dimension over sample size” requires first saying which side the ratio is on.

Observation 4.2

beside Paper 733.Paper 733 showed the minimum of error to move with the amount of data, placing variance as $\sigma^2 p/n$——here the exact version of that form is given and its reach cut. It is also an instance, from the side of sample size, of Paper 182’s point that the curse is not one curse.

Observation 4.3

what is not said.Correlated regressors, outliers, regularisation and the region beyond the boundary are not treated——what is shown is the reach of the approximation and the divergence outside it.

5. Roll Call and Closing

1. **0.010111 at 1000 samples** (Computation 2.1): **the approximation holds.**
2. **ratio 1.011122** (Computation 2.1): **the far side.**
3. **0.112360 at 100** (Computation 2.1): **beginning to fail.**
4. **1.111111 at 20** (Computation 2.1): **ratio 2.222222.**
5. **10.000000 at 12** (Observation 2.3): **the core.**
6. **12.000000 times** (Observation 2.3): **diverging.**
7. **0.111235 at dimension 100** (Observation 3.1): **same at fixed ratio.**
8. **simulation 0.357921** (Computation 3.2): **exact 0.357143.**
9. **0.091716 at 60 samples** (Computation 3.2): **exact 0.092593.**
10. **ratio large or near 1** (Observation 4.1): **the separator; Paper 733 and Paper 182.**

Closing: **“Error goes as dimension over sample size” is the far-side approximation; as the sample size nears the dimension, the error diverges.For Gaussian regressors the excess error of least squares is $d/(n-d-1)$: at dimension 10 it is 0.010111 with 1000 samples, 1.011122 times the approximation, but 0.112360 with 100, 1.111111 with 20 and 2.500000 with 15——and with 12 samples 10.000000, 12.000000 times the approximation.Holding samples at ten times the dimension gives 0.112360 at dimension 10 and 0.111235 at 100: what acts is the ratio——more dimensions demand samples in proportion, and as the ratio nears 1 the estimate itself collapses.Simulation at dimension 5 with 20 samples gave 0.357921, on the exact 0.357143 and not the approximation’s 0.25.The separator is whether the ratio of sample size to dimension is large or close to 1.Placed among the earlier papers——Paper 733 placed a form for variance and read the moving minimum; this cuts the reach of that form.To be honest——correlated regressors, outliers and regularisation are not treated, nor the region beyond the boundary.**

References

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
2. Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley.
3. Yamagishi, Y. (2026). *“The Curse of Dimensionality” Is Not One Curse* [Paper 182]. Zenodo. 10.5281/zenodo.22104170
4. Yamagishi, Y. (2026). *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* [Paper 733]. Zenodo. 10.5281/zenodo.23015055

— 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)