Robin Eriksson; robin.eriksson@it.uu.se

1. Double-descent in Linear Regression with Random Covariates

The following Hand-in attempts reproducing the results given by Figure 2 in [1]. The experiment first examines the empirical MSE on the test-set for a minimum norm least squares model when looking at data with isotropic features. Then we add the asymptotics and see if they align.

Figure 1.1 tells a convincing story about the results: the empirics follow asymptotics with some noise. An interesting observation is that even though the SNR is increasing, the variance of the MSE is as well. This remark is counter-intuitive because a higher SNR should mean that the signal is better defined in the data and should lead to minor variance in error.

Further details about the experiment that generated the results. We try four different signal-to-noise ratios SNR $= r^2/\sigma^2$, for which we fix $\sigma^2 = 1$ and change $r^2 = \{1.0, 2.33, 3.66, 5.0\}$, the fixed l_2 norm of the data.

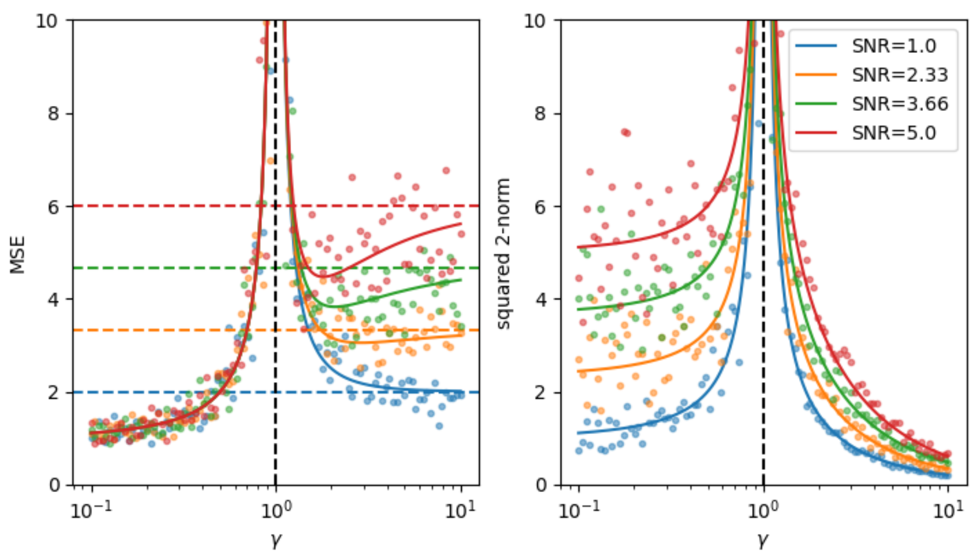

Figure 1.1. Empirical (dots) and asymptotics (lines) for four different combinations of SNR, when $\gamma = n/p$. The dashed line at $\gamma = 1$ indicates the interpolation point. (left) MSE on test set. The dashed lines is the MSE when $\gamma \to \infty$. (right) $||\theta||_2^2$.

References

[1] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *arXiv preprint arXiv:1903.08560*, 2019.