1. Double-descent in random Feature regression

The result from using random Fourier features and the minimum-norm solution for regression on the Boston house price data set[1]. The experiment result, see Figure 1.1, shows what is known as the double descent phenomena [1].

The experiment included testing different kernel widths, γ values, $[0.001, 0.01, 0.1, 1]$. Interesting is that the smallest γ used, 0.001 gave the quickest descent in the training error and includes a (potential) global minima in the underparameterized regime. However, this seems to come at the price of the $||\theta||$ as it seems strictly larger than the other γ values tested.

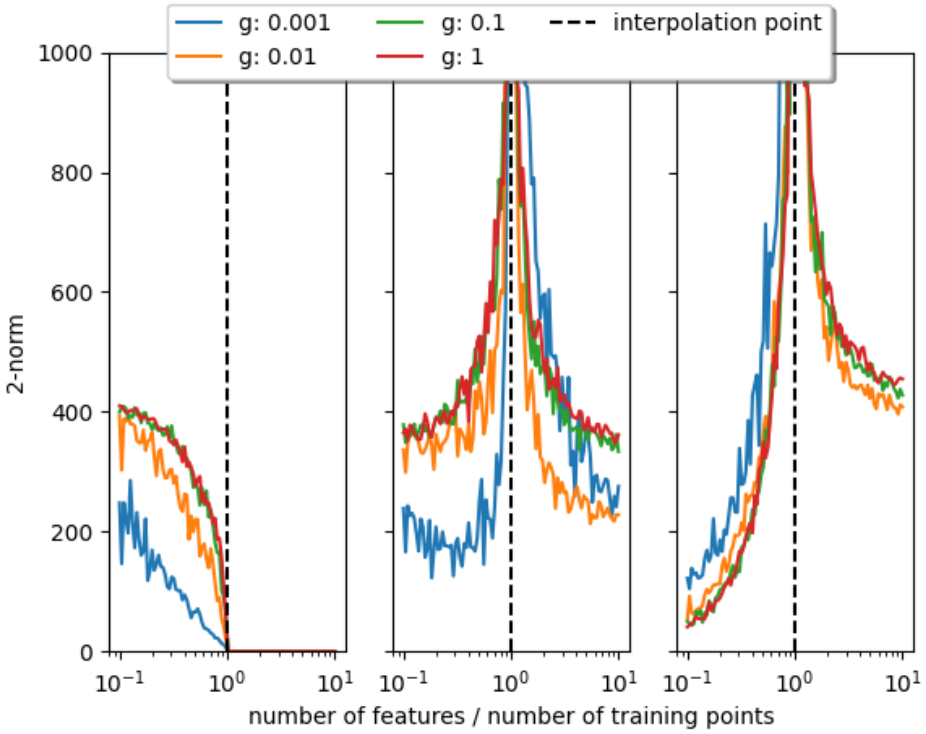

Figure 1.1. (left) training error (2-norm). (middle) testing error (2-norm). (right) θ norm. The colors represent the different γ values (g for short). The dashed line is the interpolation point. Training and test error is the 2-norm of the difference between estimated and true value.

References

[1] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

[1] can be retrieved from here: http://lib.stat.cmu.edu/datasets/boston