Abundant regression and benign overfitting: when minimum-norm prediction works for $p > n$

Liliana Forzani
FACULTAD DE INGENIERA QUIMICA - UNL, United States
liliana.forzani@gmail.com

We study high-dimensional linear regression driven by a low-dimensional sufficient reduction $X = UZ + \varepsilon$, where the response depends on X only through the index Z. The regression is *abundant* when the loadings accumulate signal with the dimension, $U^T U/\eta(p) \to M > 0$ with $\eta(p) \to \infty$: most predictors carry information about the response. We propose a definition of abundance, relate them to existing notions in partial least squares, envelopes and sufficient dimension reduction, and contrast them with the benign-overfitting literature, whose bounded-spectrum regime is precisely where the minimum-norm (ridgeless) interpolator is not consistent. Under abundance the covariance is spiked, and we give explicit non-asymptotic bounds on the prediction risk, valid for every n, p with $p > n+1$, showing that the second descent of the double-descent curve reaches the irreducible error.

Trabajo en conjunto con R. Dennis Cook (University of Minnesota), Miguel Marcos (UNL) y Pedro Morin (UNL).

Referencias

[1] Cook, R.D., Forzani, L. and Rothman, A. (2013). Prediction in abundant high-dimensional linear regression. *Electronic Journal of Statistics* **7**, 3059–3088

[2] Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Annals of Statistics* **50**, 949–986.