Complexity-reduced data models beyond the classical bias-variance trade-off

Jochen Merker and Gregor Schuldt
HTWK Leipzig University of Applied Sciences, Germany

In machine learning, a statistical model $F = F_T : \mathcal{U} \mapsto \mathcal{Y}$ is obtained from training data T. Such a model F estimates for inputs $u \in \mathcal{U}$ (distributed according to a random variable U) by Fu the output $y \in \mathcal{Y}$ (having the conditional distribution of a random variable Y given $U = u$). For a Hilbert space $\mathcal{Y}$, using the general definition $E(\langle FU - E(Y|U), Y - E(Y|U)\rangle_{\mathcal{Y}})$ of complexity of a statistical model F, and allowing randomness in the machine learning method like e.g. in stochastic gradient descent, we prove as an extension of [1] – under the assumption that the model F_T does not depend on the order of training data T – for a single randomly added data point (U, Y) with corresponding expected input-output map $F_{(U,Y)} := E_{\tilde{T}}(F_T)$, where $T = \{(U, Y), \tilde{T}\}$, the decomposition

$$\begin{aligned}&E_{T,F}\left(|F_T U - Y|_{\mathcal{Y}}^2\right) + 2E_{(U,Y),F}(\langle F_{(U,Y)}U - E(Y|U), Y - E(Y|U)\rangle_{\mathcal{Y}})\\ =&E_{T,F}\left(|F_T U - F_{(U,Y)}U|_{\mathcal{Y}}^2\right) + E_{(U,Y),F}(|F_{(U,Y)}U - E(Y|U)|_{\mathcal{Y}}^2) + E_{(U,Y)}(|Y - E(Y|U)|_{\mathcal{Y}}^2)\end{aligned}$$

of the expected squared error plus twice the complexity into variance w.r.t training data, squared bias and an unavoidable irreducible error inherent to the problem. This fine-grained decomposition generalizes the classical decomposition of mean squared error into squared bias and variance, and using this decomposition we provide a mathematical explanation for the double descent phenomenon, for which Belkin et al. [2] have first shown evidence.

References

[1] Merker, J., Schuldt, G., An attempt to explain double descent in modern machine learning, in *Tagungsband zur 26. Interdisziplinären Wissenschaftlichen Konferenz Mittweida (IWKM), 14.-15.04.2021, Scientific Reports* 2021 (2), 141–144, doi:10.48446/opus-12293

[2] Belkin, M., Hsu, D., Ma, S., Mandal, S., Reconciling modern machine-learning practice and the classical bias-variance trade-off, *Proceedings of the National Academy of Sciences* 116(32) (2019), 15849–15854.