

Online Realizable Regression and Applications for ReLU Networks

Ilan Doron-Arad *Massachusetts Institute of Technology, Cambridge, MA, USA*

Idan Mehalel *The Hebrew University of Jerusalem, Jerusalem, Israel*

Elchanan Mossel *Massachusetts Institute of Technology, Cambridge, MA, USA*

Editors: Steve Hanneke and Tor Lattimore

Abstract

Realizable online regression can behave very differently from online classification.¹ Even without any margin or stochastic assumptions, realizability may enforce *horizon-free* (finite) cumulative loss under metric-like losses, even when the analogous classification problem has an infinite mistake bound. We study realizable online regression in the adversarial model under losses that satisfy an approximate triangle inequality (approximate pseudo-metrics). Recent work of [Attias et al. \(2023\)](#) shows that the minimax realizable cumulative loss is characterized by the scaled Littlestone/online dimension \mathbb{D}_{onl} , but this quantity can be difficult to analyze.

Our main contribution is a generic potential method that upper bounds \mathbb{D}_{onl} by a concrete Dudley-type entropy integral [Dudley \(1967\)](#); [Rakhlin et al. \(2015\)](#) that depends only on covering numbers of the hypothesis class under the induced sup pseudo-metric. For an hypothesis class \mathcal{H} , we define an *entropy potential* $\Phi(\mathcal{H}) = \int_0^{\text{diam}(\mathcal{H})} \log N(\mathcal{H}, \varepsilon) d\varepsilon$, where $N(\mathcal{H}, \varepsilon)$ is the ε -covering number of \mathcal{H} under the sup pseudo metric $\sup_x \ell(f(x), g(x))$, and show that for every c -approximate pseudo-metric loss it holds that $\mathbb{D}_{\text{onl}}(\mathcal{H}) \leq O(c \cdot \Phi(\mathcal{H}))$. In particular, polynomial metric entropy implies $\Phi(\mathcal{H}) < \infty$ and hence a horizon-free realizable cumulative-loss bound with transparent dependence on effective dimension.

We illustrate the method on two families. For the class \mathcal{H}_L of all L -Lipschitz functions on $[-1, 1]^d$ under the loss $\ell_q(y, y') = |y - y'|^q$, we establish a sharp phase transition: if $q > d$ then $\mathbb{D}_{\text{onl}}(\mathcal{H}_L) = \Theta_{d,q}(L^d)$, and the bound is achievable efficiently, whereas if $q \leq d$ then $\mathbb{D}_{\text{onl}}(\mathcal{H}_L) = \infty$. Complementing these metric-specific results, for any continuous loss with $\ell(y, y) = 0$, the loss along realizable sequences in \mathcal{H}_L satisfies $\ell(\hat{y}_t, y_t) \rightarrow 0$ as $t \rightarrow \infty$.

As a second application, we study bounded-norm k -ReLU networks [Goel et al. \(2021\)](#) over $[-1, 1]^d$ with squared loss and highlight a regression–classification separation: realizable online classification is impossible already for $k = 2, d = 1$ under 0/1 loss, yet realizable regression admits finite total loss, including a $\tilde{O}(k^2)$ cumulative-loss upper bound, a matching lower bound up to polylogarithmic factors, and an efficient $O(1)$ guarantee for a single ReLU ($k = 1$) independent of the input dimension d . Assuming Gap-ETH [Dinur \(2016\)](#); [Manurangsi and Raghavendra \(2016\)](#), we also rule out efficient proper online learners that achieve realizable accumulated loss $\tilde{o}(d)$ for any constant $k \geq 2$.

1. Extended abstract. Full version appears as arXiv:2602.19172v2

References

- Idan Attias, Steve Hanneke, Alkis Kalavasis, Amin Karbasi, and Grigoris Velegkas. Optimal learners for realizable regression: PAC learning and online learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- Irit Dinur. Mildly exponential reduction from gap 3sat to polynomial-gap label-cover. *Electron. Colloquium Comput. Complex.*, TR16-128, 2016. URL <https://eccc.weizmann.ac.il/report/2016/128>.
- Richard M Dudley. The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis*, 1(3):290–330, 1967.
- Surbhi Goel, Adam R. Klivans, Pasin Manurangsi, and Daniel Reichman. Tight hardness results for training depth-2 relu networks. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, Virtual Conference, January 6-8, 2021*, volume 185 of *LIPICs*, pages 22:1–22:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPICs.ITCS.2021.22. URL <https://doi.org/10.4230/LIPICs.ITCS.2021.22>.
- Pasin Manurangsi and Prasad Raghavendra. A birthday repetition theorem and complexity of approximating dense csps. *CoRR*, abs/1607.02986, 2016. URL <http://arxiv.org/abs/1607.02986>.
- Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *J. Mach. Learn. Res.*, 16(1):155–186, 2015.