Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Optimistic Estimate Uncovers the Potential of Nonlinear Models

Yaoyu Zhang * [1,2], Zhongwang Zhang † [1], Leyang Zhang ‡ [3], Zhiwei Bai § [1], Tao Luo ¶ [1,4,5], and Zhi-Qin John Xu || [1,2]

[1]School of Mathematical Sciences, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China.
[2]School of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China.
[3]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, United States.
[4]CMA-Shanghai, Shanghai Jiao Tong University, Shanghai 200240, China.
[5]Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Abstract. We propose an optimistic estimate framework to evaluate the potential of nonlinear models in fitting target functions at overparameterization. In our framework, such potential is quantified by estimating the smallest possible sample size needed for a model to recover a target function, referred to as an optimistic sample size. Following the framework, we derive the optimistic sample sizes for matrix factorization models, deep models, and two-layer neural networks (NNs) with fully-connected or convolutional architectures. For each nonlinear model, we confirm via experiments that the target functions can be fitted at overparameterization as predicted by our analysis. Our results suggest a hierarchical inductive bias of nonlinear models towards simple functions with smaller optimistic sample sizes intrinsic to their architecture. The dynamical realization of the suggested inductive bias remains an open problem for the further study.

Keywords:
Nonlinear regression,
Optimistic sample size estimate,
Matrix factorization model,
Two-layer neural network,
Hierarchical inductive bias.

Article Info.:
Volume: 4
Number: 3
Pages: 192 - 222
Date: September/2025
doi.org/10.4208/jml.231109

Article History:
Received: 09/11/2023
Accepted: 18/06/2025

Communicated by:
Qianxiao Li

1 Introduction

In recent years, large nonlinear models like deep neural networks (DNNs) have achieved unprecedented success, powering amazing applications as AlphaGo [17], AlphaFold [10], and GPT [5,15]. However, for many decades, the community had severely underestimated their potential. One major factor is that traditional theoretical frameworks for generalization excessively focus on the worst-case scenarios in which overfitting easily happens in absence of regularization for overparameterized models. To circumvent the limitation of the worst-case based estimates, we propose an opposite approach of optimistic estimate

*Corresponding author. zhyy.sjtu@sjtu.edu.cn
†0123zzw666@sjtu.edu.cn
‡leyangz_hawk@outlook.com
§bai299@sjtu.edu.cn
¶luotao41@sjtu.edu.cn
||Corresponding author. xuzhiqin@sjtu.edu.cn

to evaluate the best-possible fitting performance of nonlinear models at overparameterization. In this work, fitting means recovering the target function exactly.

We demonstrate the idea of the optimistic estimate using a chemical reaction analogy. By applying the law of element conservation, we derive the following optimistic estimates for chemical reactions: graphite has the potential to be converted to diamond, while gold cannot be obtained from copper. These optimistic estimates inform the potential of a chemical reaction, leading to the abandonment of alchemy and subsequently the invention of synthetic diamonds. Likewise, our optimistic estimate for nonlinear regression aims to uncover the fitting potential of nonlinear models in terms of a minimal sample size for recovering a target. Our quantitative estimate will motivate later studies comparing the empirical sample size requirement to our theoretical estimate.

To assess how well a model fits a target function, the sample size necessary for fitting serves as a natural quantity. In linear regression, the sample size is in general determined by the model size, i.e. the size of model parameters. Therefore, linear models with fewer parameters are preferred in practice. In the case of nonlinear regression, although theoretical understanding is limited, it has been empirically demonstrated that DNN models can effectively fit target functions even at heavy overparameterization [22]. To gain deeper understanding of this widely observed phenomenon, we propose an optimistic estimate framework for nonlinear models to evaluate the smallest possible sample size – which we term the optimistic sample size – necessary for fitting each target function they can express. Unlike traditional sample complexity analyses that focus on worst-case scenarios for achieving specified performance levels [16], our optimistic sample size quantifies the minimum number of training samples needed to recover a target function under ideal conditions. This approach provides new insights beyond existing work, such as Zhong *et al.*'s analysis [25] of two-layer neural network recovery based on local strong convexity assumptions, which break down in overparameterized settings.

Our results in the following show that the optimistic sample sizes of certain simple targets indeed can be much smaller than the model size for many widely considered nonlinear model families like deep models [21], matrix factorization models [2, 8, 18] and two-layer NNs. Furthermore, the optimistic sample size as a functional over the model function space exhibits rich structures, especially for the NN family.

Given a target function f^* with optimistic sample size R^* for a model $f_{\boldsymbol{\theta}}$, our results indicate that R^* training samples are necessary for fitting f^*. However, it remains an open problem to estimate how many samples are sufficient for fitting f^* in practice. Our experiments in matrix factorization models and two-layer tanh-NNs show that these model families can achieve near-optimism fitting performance with proper hyperparameter tuning (specifically, initialization scale and learning rate). In this context, "near-optimism" means the empirical sample size necessary for fitting f^* in experiments can approach or even match R^*. In particular, when R^* is much smaller than the model size M, the empirical sample size is generally much closer to R^* than to M. The potential mechanism for this phenomenon is discussed in Section 6.2.

Our optimistic estimates suggest quantitative answers to some long-standing problems in nonlinear regression discussed in Section 7. These problems include determining which model can fit at overparameterization, the effective size of parameters of a model, the