Stability and Sharper Risk Bounds with Convergence Rate $\tilde{O}(1/n^2)$

Bowei Zhu, Shaojie Li, Mingyang Yi, Yong Liu

Renmin University of China

Conference: NeurIPS 2025

中国人民大學
RENMIN UNIVERSITY OF CHINA

The Pursuit of Generalization

- **Core Question:** How well will a model trained on data perform on new, unseen data?
- **Key Concept:** Algorithmic Stability
  - A stable algorithm produces similar outputs given similar training datasets.
  - Stability provides a powerful framework for understanding generalization.
- **The Goal:** Derive high-probability, *excess risk* bounds that are both...
  - **Sharp:** Converge rapidly as sample size n increases.
  - **Dimension-free:** Hold even in very high-dimensional spaces.

Our Novel Framework: Gradient Stability

- **Traditional Approach:** Analyzes stability of the loss function $f(w; z)$.
- **Our Innovation:** Introduces and analyzes stability of the gradient$\nabla f(w; z)$.
- **Why This Is Powerful**:
  - Under common conditions (PL, smoothness), the loss is upper bounded by the squared gradient norm.
  - Controlling the gradient's stability provides a tighter handle on generalization error.
  - This leads to a direct and sharper bound on the excess risk.

Key Technical Results (I)

- **Generalization Bound for Gradients:** We prove a new, high-probability bound on $\|\nabla F(A(S)) - \nabla F_S(A(S))\|$.
- The Bound Features:
  - A fast $O(1/n)$ rate.
  - A leading constant that depends on the expected gradient norm $\mathbb{E}[\|\nabla f(A(S); Z)\|^2]$at the optimized model, which is typically small.
  - This is a key insight enabling sharper final rates.

Key Technical Results (II)

- **Excess Risk Bound:** Under Polyak-ojasiewicz (PL) and smoothness conditions, we derive:
  - $F(A(S)) - F(w^*) \lesssim \tilde{O}(1/n^2)$
- **This rate is achieved for:**
  - **Empirical Risk Minimization (ERM)**
  - **Projected Gradient Descent (PGD)**
  - **Stochastic Gradient Descent (SGD)** (with appropriate iterations)
- All bounds are **dimension-free**.

Why Our Results Are Significant

- **Finer-Grained Analysis:** Our gradient stability framework provides a more precise characterization of generalization.
- **Optimal Rates:** We achieve a quadratic improvement $\tilde{O}(1/n^2)$over the previous best stability-based bounds, which were linear $\tilde{O}(1/n)$.
- **Practical Relevance:** High-probability bounds are crucial in practice, where models are typically trained only once.
- **Broad Applicability:** The results hold for standard algorithms like ERM, PGD, and SGD.

Conclusion and Impact

- **Summary:** We introduced a novel framework based on gradient stability.
- **Outcome:** This enabled the derivation of the first dimension-free, high-probability excess risk bounds at a rate of Õ(1/nš)for standard algorithms.
- **Impact:** This represents a significant advance in learning theory, providing a sharper understanding of generalization through algorithmic stability.

Thanks!

Contact: {bowei.zhu, lishaojie95, yimingyang, liuyonggsai}
@ruc.edu.cn