Generalization Bounds

Theoretical Foundations of Deep Learning

Matteo Mazzarelli

December 17, 2024

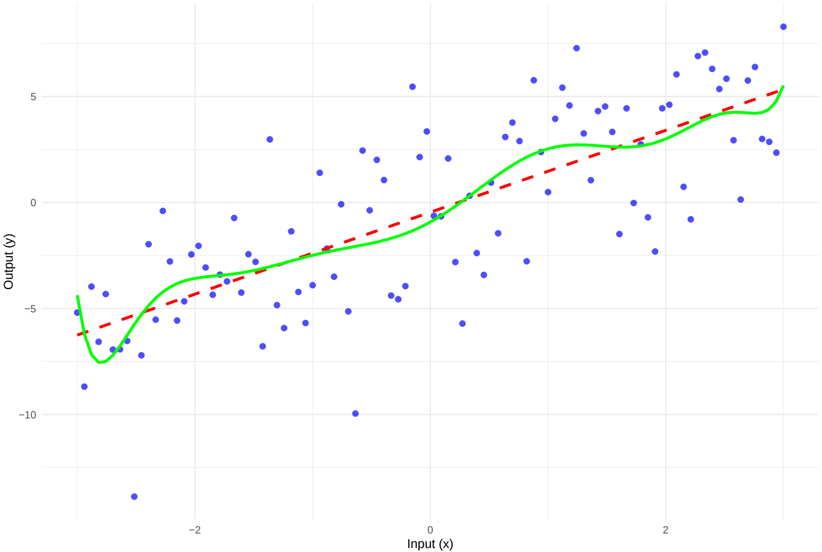
Motivation

- **Core Challenge**: How can a model learned from *limited training data* perform well on *unseen data*?
- Generalization lies at the heart of the machine learning process.
- A poorly generalized model risks:
  - **Overfitting**: Performing well on training data but poorly on unseen data.
  - **Underfitting**: Failing to capture the underlying patterns of the data.

The Perils of Overfitting: A Motivating Visualization

- **Overfitting in Action**:
  - A model can perfectly fit training data but fail to capture the true underlying pattern.
  - This often leads to poor performance on unseen data.
- **Demonstration**:
  - Dataset: A simple linear trend with noise.
  - Models:
    - Linear model: Captures the underlying trend.
    - High-degree polynomial: Overfits the noise in the data.

Overfitting Example: Linear vs. Polynomial Model



The Learning Problem

- **Supervised Learning**:
  - Goal: Learn a function $f : X \to Y$ mapping inputs X to outputs Y based on labeled training data.
- **Key Question**: Can the learned function perform well on unseen data?
- **Generalization**:
  - Ability of a model to extend its learning beyond the training data.
  - **Central Problem** in machine learning: balancing *empirical performance* with *future predictions*.

Why Theory Matters

- **Significance of Theory**:
  - Guides **algorithm design** by providing a foundation for developing new methods.
  - Allows **performance analysis** to identify the strengths and weaknesses of algorithms.
  - Reveals **limitations** of learning systems, helping us understand their boundaries.
- **Theoretical Understanding**:
  - Bridges the gap between empirical performance and guarantees on future behavior.

Introducing Generalization Bounds

- **What Are Generalization Bounds?**
  - Theoretical tools offering guarantees about a model's performance on unseen data.
  - Relate:
    - **Generalization Error**: How well the model generalizes.
    - **Empirical Risk**: Performance observed on training data.
    - **Model Complexity**: How expressive the model is.
- **Purpose**:
  - Provide insights into the trade-offs between model accuracy, complexity, and training data size.

Hoeffding's Inequality: A Starting Point

- **What is Hoeffding's Inequality?**
  - A fundamental result in probability theory used to bound the difference between the **empirical risk** and the **generalization error** for a fixed hypothesis.
  - Provides a way to measure how closely a model's performance on training data reflects its performance on unseen data.

Mathematical Formulation of Hoeffding's Inequality

- **Hoeffding's Inequality**:

$$P(|R(h) - R_{\text{emp}}(h)| > \varepsilon) \leq 2\exp(-2m\varepsilon^2)$$

  - $R(h)$: Generalization error (true performance on unseen data).
  - $R_{\text{emp}}(h)$: Empirical risk (error on training data).
  - ϵ: A small positive value (tolerance).
  - m: Size of the dataset.

Key Insights

- The probability that the generalization error $R(h)$ deviates significantly from the empirical risk $R_{\text{emp}}(h)$ decreases **exponentially** with:
  - Larger dataset size m.
  - Smaller tolerance ϵ.

Interpretation of Hoeffding's Inequality

- **What Does It Mean?**
  - As the dataset size (m) increases, the empirical risk becomes a more reliable indicator of the generalization error.
  - For a fixed hypothesis, we can be confident that the performance observed on training data is close to what can be expected on unseen data.
- **Why is it Important?**
  - Hoeffding's inequality gives a **quantitative guarantee** about the relationship between training performance and unseen data performance.

Limitations of Hoeffding's Inequality

- **The Challenge of Multiple Hypotheses**:
  - In practical machine learning, we often choose the best hypothesis from a large hypothesis class $\mathcal{H}$.
  - Hoeffding's inequality applies to a **single fixed hypothesis**, not to the case where multiple hypotheses are considered.
- **Implication**:
  - It doesn't directly address:
    - The **selection bias** introduced by choosing the hypothesis that minimizes the empirical risk.
    - The increased risk of overfitting when evaluating multiple hypotheses.
- **Motivation for Advanced Bounds**:
  - Hoeffding's inequality provides a foundation but highlights the need for bounds that account for hypothesis complexity, such as **VC dimension** or **Rademacher complexity**.