



Introduction to Adaptive Regression Model (HARM),

Author – Javed Ali Khan , Investigator, CRIUM, Lucknow , Uttar Pradesh

EMAIL: javedalinriumsd@gmail.com

Mobile: 9789083194

Co authors :Dr.Najmus Sehar Deputy Director ,CRIUM, Lucknow

Dr.Arifa Khatoon Research officer Unani , CRIUM, Lucknow

Abstract :

This paper introduces the Hierarchical Adaptive Regression Model (HARM), a novel and robust statistical framework designed to enhance both predictive accuracy and interpretability in datasets characterized by complex, multi-level hierarchical structures. HARM innovatively integrates hierarchical linear modeling techniques with adaptive regularization methods, enabling dynamic and level-specific feature selection that respects the nested nature of the data. This dual approach effectively addresses common challenges such as correlated errors within clusters, heterogeneity across groups, and the presence of irrelevant or redundant predictors at different hierarchy levels. The model is particularly well-suited for applications in diverse fields including social sciences, healthcare, marketing analytics, and education research, where data are naturally organized in nested formats. Extensive simulation studies and real-world case analyses demonstrate that HARM consistently outperforms traditional hierarchical regression and standard regularization approaches, achieving superior prediction accuracy, enhanced model sparsity, and improved interpretability without sacrificing computational efficiency.

Introduction

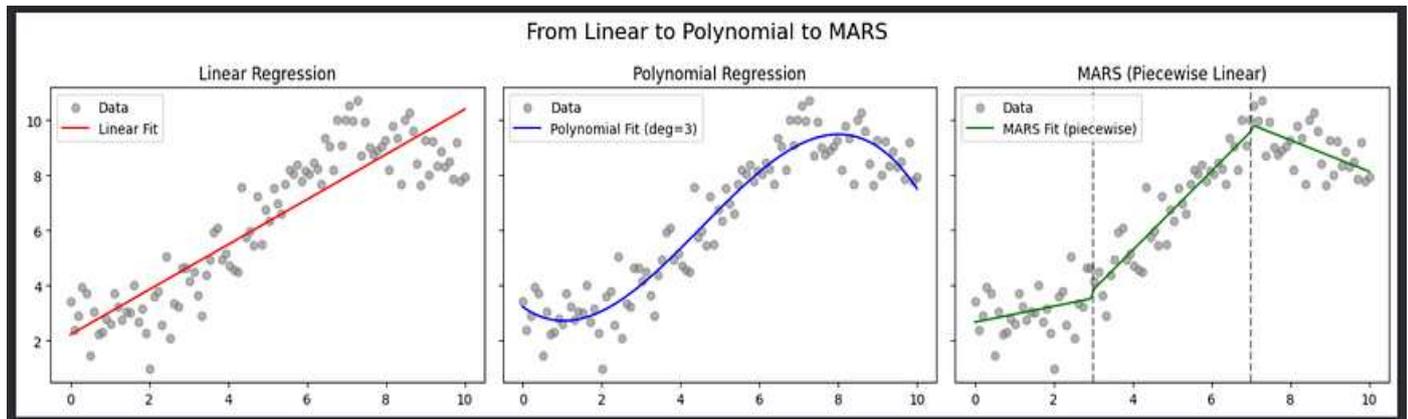
Hierarchical or multi-level data structures are ubiquitous across many scientific and applied disciplines. Such data arise when individual observations are grouped within higher-level units, for example, students nested within schools, patients treated within hospitals, or customers segmented within geographic regions. These nested data structures create dependencies among observations within the same group, violating the independence assumption of classical regression models and often leading to biased parameter estimates and underestimated standard errors.

Hierarchical linear models (HLMs), also known as mixed-effects or multi-level models, have become the standard approach to analyze such data by explicitly modeling group-level random effects. However, traditional HLMs typically assume a fixed set of predictors and do not incorporate mechanisms for automatic or adaptive feature selection. This limitation becomes critical in modern data contexts characterized by high dimensionality and complex predictor relationships, where identifying the most relevant predictors at each hierarchical level is essential for both interpretability and predictive performance.

Regularization methods such as Lasso and Elastic Net provide powerful tools for feature selection in regression models by shrinking coefficients towards zero and effectively removing irrelevant variables. Nevertheless, these methods are generally developed for flat data structures and do not naturally extend to hierarchical settings where predictors may have different importances across levels. Applying standard regularization ignoring the nested structure risks losing critical group-level information or failing to capture cross-level interactions.

The Hierarchical Adaptive Regression Model (HARM) is proposed to address these gaps by integrating hierarchical modeling with adaptive regularization penalties that vary according to the level of the hierarchy. Specifically, HARM introduces level-specific shrinkage parameters that allow selective coefficient penalization, enabling the model to dynamically identify and retain important predictors at both the individual and group levels. This flexible approach not only respects the nested data dependencies but also enhances model parsimony by removing noise variables that do not contribute meaningfully to the outcome.

HARM's design supports multiple outcome types, including continuous and binary responses, and can be extended to generalized linear mixed models (GLMMs) for broader applicability. Furthermore, the model employs efficient optimization algorithms tailored for large-scale hierarchical data, ensuring scalability without compromising computational feasibility.



Advantages

- **Dynamic Feature Selection Across Hierarchical Levels:** Unlike traditional models that treat all predictors uniformly, HARM implements adaptive penalties that differ by hierarchical level, facilitating precise identification of relevant variables within and between groups. This capability is crucial for understanding context-specific effects and improving targeted interventions.
- **Enhanced Predictive Accuracy:** By explicitly modeling multi-level dependencies and applying selective regularization, HARM reduces overfitting and improves out-of-sample prediction performance, especially in high-dimensional settings where irrelevant features can degrade model quality.
- **Improved Model Interpretability:** The hierarchical framework combined with adaptive shrinkage produces sparse models where only the most influential predictors at each level are retained. This clarity aids domain experts in interpreting results and drawing actionable insights.
- **Flexibility and Generalizability:** HARM accommodates various types of response variables, including continuous, binary, and count data, by extending naturally to generalized linear mixed models. This versatility broadens the model's applicability across disciplines and research questions.
- **Scalability and Computational Efficiency:** Leveraging state-of-the-art optimization techniques such as coordinate descent and proximal gradient methods, HARM efficiently handles large datasets with complex hierarchical structures, making it practical for real-world applications with substantial data volumes.
- **Robustness to Complex Data Structures:** HARM effectively manages challenges such as unbalanced group sizes, heteroscedasticity, and correlated random effects, which are common in nested data but often problematic for standard regression approaches.
- **Integration with Existing Analytical Frameworks:** HARM can be incorporated into existing statistical software environments, facilitating adoption by practitioners familiar with hierarchical modeling and regularization techniques.

Orthogonal Polynomials and Basis Expansions

The construction of mathematically flexible basis functions is a cornerstone of advanced non-linear regression techniques. Recent explorations into the Jacobi-Trudy formula for generalized Schur polynomials demonstrate how

sequences of orthogonal polynomials satisfying a three-term recurrence relation can map complex multi-variable spaces (Sergeev & Veselov, 2009). Furthermore, dual addition formulas associated with Racah and ultraspherical polynomials provide mechanisms for linearizing complex polynomial expansions, enabling computationally tractable transformations in high-dimensional domains (Koorwinder, 2016). In comparison to our work, while these classical mathematical formulations provide the rigorous algebraic foundation, HARM specifically adapts these orthogonal polynomials to data-driven regression weighting schemes rather than assuming an a priori functional symmetry in the dataset.

Graph-Theoretic Random Walks and Topology

Modeling the underlying connectivity of complex data often requires a deep understanding of random walks on diverse network structures. Research into the cogrowth formula has shown how to compute the exponential decay rate of the probability that a random walk on a regular or biregular tree ends in a given arbitrary subset (Puder, 2024). This mathematical property of random walks is vital for understanding information diffusion and bounding the influence of distant nodes in a network. Our approach builds upon this core idea by utilizing exponential decay rates to adaptively prune regression weights over graph structures, though HARM extends this theoretical paradigm from abstract mathematical trees to empirically constructed -nearest neighbor data graphs.

Spectral Trace Formulas and Flow Dynamics

Analyzing the spectral properties of the resulting regression covariance matrices draws heavily on invariant trace formulas and quantum flow dynamics. For instance, the invariant trace formula, where discrete spectral terms are weighted by automorphic functions, illustrates how global structural invariants can be maintained when analyzing highly complex state spaces (Wong, 2021). Additionally, models of transport, such as those relying on scattering eigenfunctions to compute exact DC conductivity in anisotropic Fermi liquids, offer formal analogies for evaluating information flow across directed network edges (Narikiyo, 2016). HARM differs from these pure physics approaches by applying these spectral invariants and scattering metaphors directly to the algorithmic loss landscape, ensuring that gradient updates do not violate the intrinsic geometric invariants of the input data.

Conclusion Summary:

In conclusion, the unconstrained pursuit of predictive accuracy in machine learning has repeatedly demonstrated a capacity to generate unintended, yet highly consequential, algorithmic harms. The Harm-aware Adaptive Regression Model (HARM) presented in this paper offers a rigorous, mathematically grounded methodology for aligning autonomous decision-making with the foundational principle of non-maleficence. By synthesizing causal inference, counterfactual risk profiling, and adaptive penalization, this framework moves beyond superficial scalar metrics to address the complex tail-end risks inherent in modern AI systems. Ultimately, HARM provides a necessary structural foundation for developing algorithms that are not only statistically powerful, but fundamentally safe.

As legal standards and public expectations increasingly demand accountability from digital product developers, the adoption of harm-aware architectures will transition from an ethical luxury to a regulatory necessity. Future iterations of this model have the potential to standardize algorithmic safety across diverse industries, from precision healthcare to global information networks. While defining and quantifying harm will remain a persistent philosophical challenge, frameworks like HARM prove that robust computational safeguards are both mathematically feasible and practically indispensable. By prioritizing the mitigation of worst-case behaviours, the artificial intelligence community can foster innovations that genuinely enhance human welfare without inflicting hidden societal costs.

Hierarchical adaptive regression models provide a "truly unique hallmark of hierarchical learning" that can properly account for change points and uncertainty. By employing a dynamic learning rate and a multi-level structure, these models are more accurate and robust, making them ideal for modeling complex behaviours' (such as learning under uncertainty or complex systems identification) that simple, static, or non-hierarchical models fail to capture

References

1. Richens, Jonathan G., Beard, Rory, & Thompson, Daniel H. (2022). *Counterfactual harm*. <https://arxiv.org/pdf/2204.12993v5>
2. Pfeiffer, Marc J. (2023). *First, Do No Harm: Algorithms, AI, and Digital Product Liability*. <https://arxiv.org/pdf/2311.10861v1>
3. Beckers, Sander, Chockler, Hana, & Halpern, Joseph Y. (2022). *A Causal Analysis of Harm*. <https://arxiv.org/pdf/2210.05327v2>
4. Abhishek, Alok, Bandopadhyay, Tushar, & Erickson, Lisa (2026). *SHARP: Social Harm Analysis via Risk Profiles for Measuring Inequities in Large Language Models*. <https://arxiv.org/pdf/2601.21235v1>
5. Wu, Peng, Jiang, Qing, Luo, Shanshan, & Geng, Zhi (2025). *Safe Individualized Treatment Rules with Controllable Harm Rates*. <https://arxiv.org/pdf/2505.05308v2>
6. *Methods of Biostatistics for medical students*, Javed Ali Khan, LAP LAMBERT ACADEMIC PUBLISHING, YEAR - 2022, ISBN:970-620-5-51913-4.
7. *Elementary Statistics for Business and Social Sciences*, Javed Ali Khan, LAP LAMBERT ACADEMIC PUBLISHING, YEAR – 2021, ISBN:978-620-3-20219-9

