Overfitting Mitigation in Neural Networks: Could ergodic regularization be the future of AI?

Axentiva

Yovani Torres Favier[1], Antonio Monleon[2], Jesús Cuervo[3], Carlos Crespo[1,2]

[1] Axentiva Solutions SL, Barcelona, Spain; [2] University of Barcelona, GM Statistics Dept, Barcelona, Spain; [3] Axentiva Solutions SL, Oviedo, Spain

#MSR161

Background & Objective

Overfitting refers to the tendency of estimation procedures or models to learn patterns, including noise or dataset specific quirks, that do not generalize to new, unseen data. In other words, overfitting describes a model's tendency to adapt too closely to the training data, mistaking specific details or noise for generalizable structures, thereby failing to perform well in broader contexts. This can lead to poor performance, high variance, and low robustness.

Overfitting in artificial neural networks (ANN) poses significant challenges in biomedical applications with limited sample sizes. Traditional regularization methods like weight decay and dropout often prove insufficient for complex medical datasets, where robust generalization is critical for accelerating patient access.

This study aimed to develop a novel theoretical framework based on ergodic principles that prevents overfitting.

Methods

We established conditions under which ANN become ergodic transformations, ensuring that generalization gaps vanish. We derived an "ergodic regularizer" that penalizes deviations from volume preservation by controlling the Jacobian determinant of network layers.

$$L_{\text{ergodic}} = \sum_{i} (\det J_f(x_i) - 1)^2$$

We evaluated this approach across multilayer perceptron, convolutional networks, recurrent networks, and transformers architectures on Wisconsin Breast Cancer and diabetes data.

We compared two regularization schemes:

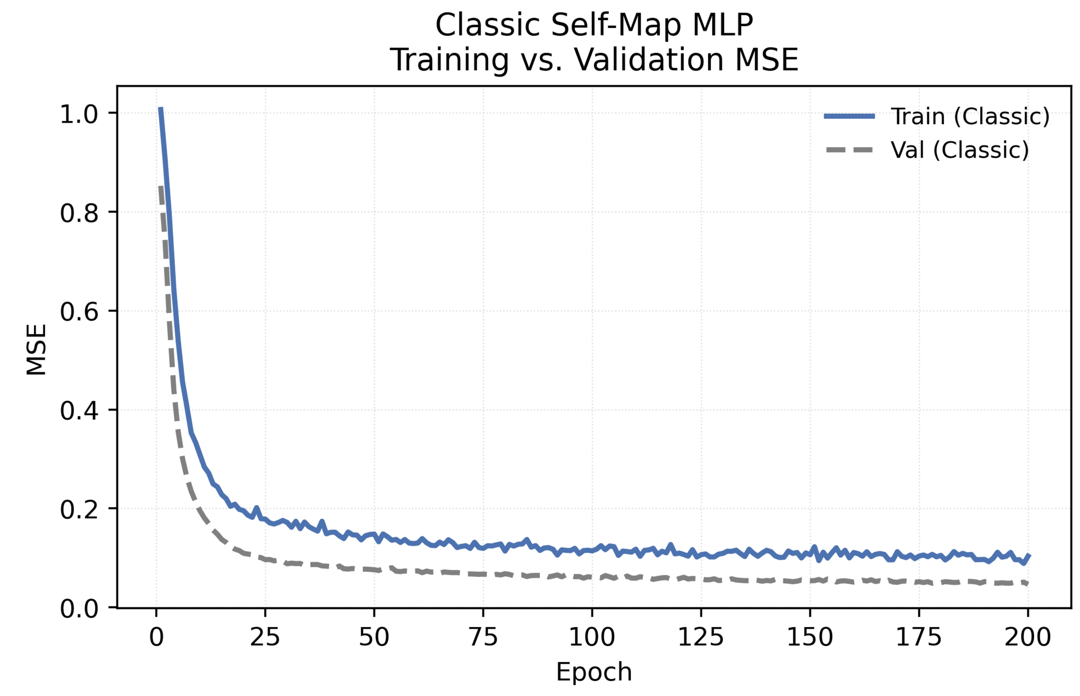
- Classic (L2 only): Weight decay (ℓ_2 penalty) in Adam, no Jacobian constraint.
- Ergodic (Jacobian penalty): Very small ℓ_2 weight decay plus a conditional penalty on the Frobenius-norm of the Jacobian $\partial f/\partial \mathbf{x}$ whenever its batch-mean exceeds a threshold.

For each ANN and configuration, we measured training and inference time, along with the evolution of training loss, test loss, and final test accuracy. This direct comparison on SMNNs allowed us to isolate the impact of ergodic regularization in a controlled feed-forward setting.

Analyses were deployed in Python.

Results

Across all tested architectures, ergodic regularization consistently outperformed traditional methods. In breast cancer feature reconstruction, the ergodic approach achieved 38% lower test error compared to standard regularization.



METHOD	MEAN SQUARE ERROR	
	TRAIN	TEST
Classic	0.10340	0.05223
Ergodic	0.05800	0.03232

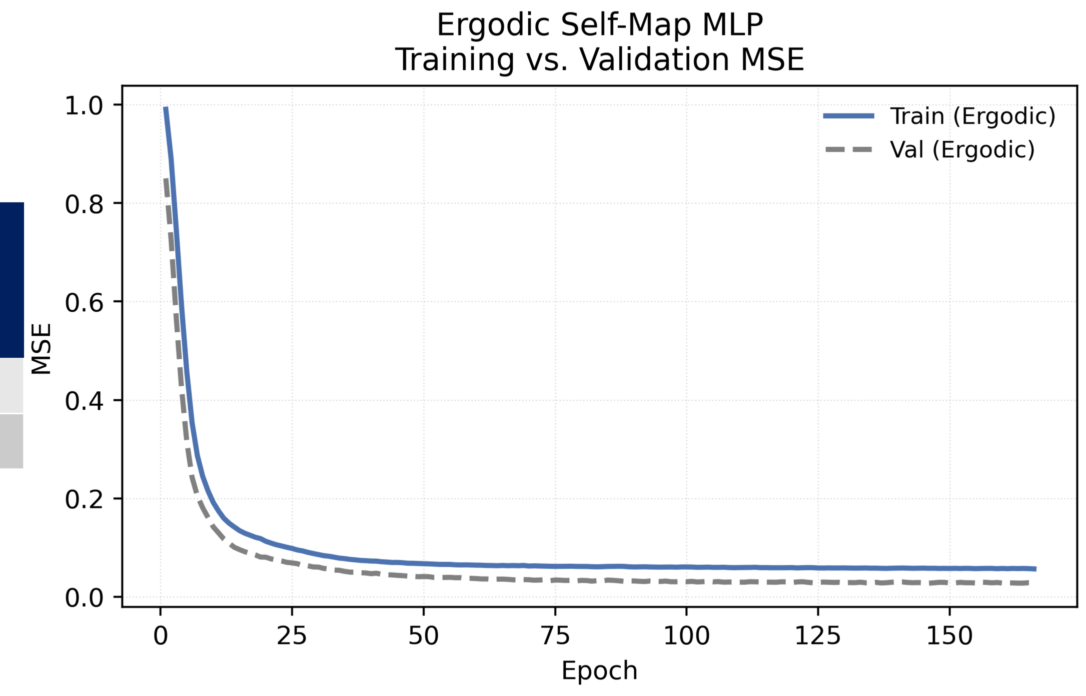
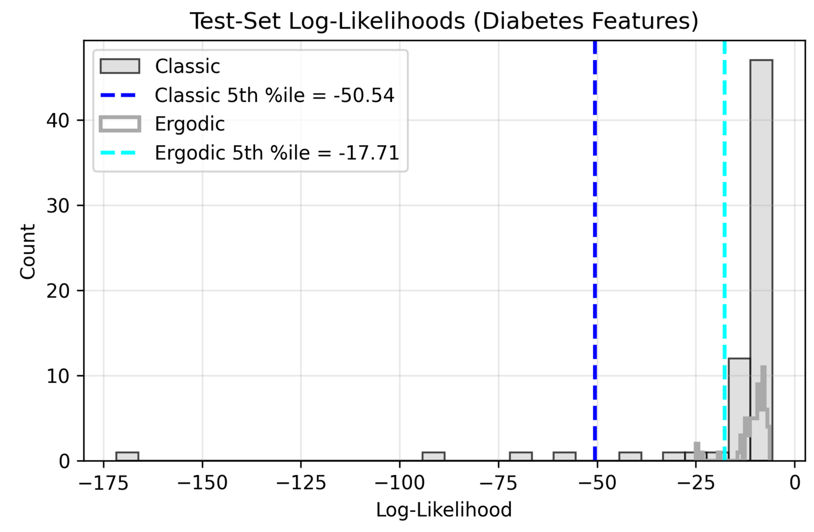


Figure 1. Training and validation MSE curves for Classic and Ergodic self-map MLPs. The Jacobian penalty yields a smoother validation trajectory and lower final test MSE.

For diabetes density, ergodic regularization produced significantly higher log-likelihood scores, indicating better capture of underlying data distributions without overfitting.



METHOD	LOG-LIKELIHOOD
Classic	-15.5959
Ergodic	-10.2403

Figure 2. Test-Set Log-Likelihoods (Diabetes Features). Shaded bars denote counts for Classic (light gray) and Ergodic (dark gray) RealNVP. Blue dashed line marks the Classic model's 5th percentile (−50.54); teal dashed line marks the Ergodic model's 5th percentile (−17.71).

The method showed strength in scenarios with limited training data.

Conclusions

Ergodic regularization has been demonstrated to be an **effective** approach **to mitigating overfitting** in a variety of machine learning problems and a reduction in bias which is critical for models in clinical decision support or pharmacoeconomic modeling. By enforcing volume preservation properties, this method ensures ANN learn genuine data patterns rather than dataset-specific artifacts, showing particular promise for real-world science applications where robust generalization is essential for clinical reliability and impacting the drug discovery pipeline.

ISPOR Europe 2025- Glasgow
Tuesday, 11 November, 16:00 - 19:00

ccrespo@axentiva.com