Hi all,

Here is your disruptive ML/AI digest for the day. Here's what you folks might find interesting today:

NOVEL ARCHITECTURES AND OBJECTIVES Renormalization Group Flow Matching for Scalable Local Generative Modeling This paper introduces renormalization group flow matching (RGFM), a generative framework that uses an exact renormalization group (RG) flow as the probability path to structure data generation across spatial scales. By exploiting the quasi-locality and scale separation properties of the RG, the authors prove that the probability flow can be approximated by local velocity fields acting over a logarithmically scaling spatial range. This enables local generative modeling with nearly linear computational cost that successfully reproduces long-range correlations far beyond its receptive field. This is a strong transfer of physics concepts to ML, offering a principled alternative to standard global flow matching that avoids its high computational costs without sacrificing global coherence.

PROBABILISTIC AND STATISTICAL METHODS The Loss Floor of Denoising Score Matching: Fisher Geometry from Schrödinger Bridges This paper isolates the irreducible excess in the denoising score matching loss and proves it is exactly the trace of the Fisher-Rao metric of the conditional endpoint family. By deriving this from a Schrödinger bridge variational principle, the authors identify the information geometry observed in diffusion latent spaces as an intrinsic component of the training loss. This provides a rigorous information-theoretic decomposition of the loss floor into an information flow determined by the data and a weight determined by the corruption schedule. It matters because it explains why raw losses from different noise ranges fail to rank models consistently, offering a new theoretical foundation for evaluating and designing diffusion objectives.

THEORY AND GENERALIZATION A Theory of Speciation in Generative Diffusion Models on Compact Riemannian Manifolds This work develops an intrinsic theory of speciation for diffusion models on compact Riemannian manifolds, moving beyond the standard symmetric pitchfork bifurcation assumptions. The authors characterize speciation via bifurcations of critical points of the evolving probability density, using spectral heat-kernel representations to explicitly incorporate manifold geometry. Poincare-Hopf and Morse theory are used to impose global topological constraints on score equilibria, revealing that generic speciation events admit an A2 fold normal form rather than a pitchfork. This matters for understanding the geometric and topological constraints on generative dynamics, providing a much deeper mechanistic explanation of how diffusion models commit to different data classes.

Generalization, memorization, and overfitting for diffusion models trained in the lazy high-dimensional regime This paper develops a generative counterpart to the theory of benign overfitting for score-based generative models trained in the lazy regime using an inner-product kernel in an RKHS. In the proportional high-dimensional regime, the authors derive exact risk trajectories under gradient flow, identifying three distinct phases: a spectral estimator that generalizes, a pure-noise score that interpolates, and an empirical Bayes estimator that memorizes. They analyze how these estimators combine along the reverse-time SDE to characterize the resulting sample distribution. This work is crucial because it formally connects implicit regularization in generative modeling to familiar supervised learning mechanisms like kernel linearization, while also uncovering a distinct phenomenology unique to diffusion models.

Until the next announcement!

P.S. In the most recent submission window: cs.LG: 138 papers stat.ML: 23 papers