Novelty Assessment Report

Paper: Variational Deep Learning via Implicit Regularization

PDF URL: https://openreview.net/pdf?id=WsN88Ns0i6

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Modern deep learning models generalize remarkably well in-distribution, despite being overparametrized and trained with little to no explicit regularization. Instead, current theory credits implicit regularization imposed by the choice of architecture, hyperparameters and optimization procedure. However, deep neural networks can be surprisingly non-robust, resulting in overconfident predictions and poor out-of-distribution generalization. Bayesian deep learning addresses this via model averaging, but typically requires significant computational resources as well as carefully elicited priors to avoid overriding the benefits of implicit regularization. Instead, in this work, we propose to regularize variational neural networks solely by relying on the implicit bias of (stochastic) gradient descent. We theoretically characterize this inductive bias in overparametrized linear models as generalized variational inference and demonstrate the importance of the choice of parametrization. Empirically, our approach demonstrates strong in- and out-of-distribution performance without additional hyperparameter tuning and with minimal computational overhead.

Core Task Landscape

This paper addresses: **Bayesian Deep Learning via Implicit Regularization of Gradient Descent**

A total of **22 papers** were analyzed and organized into a taxonomy with **15 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Theoretical Foundations of Implicit Regularization**
- **Variational Inference and Gradient-Based Sampling Methods**
- **Meta-Learning with Bayesian and Gradient-Based Approaches**
- **Applied Methods for Uncertainty Quantification and Prediction**
- **Privacy-Preserving and Specialized Neural Network Architectures**
- **Model Compression and Transfer Learning**

Complete Taxonomy Tree

- Bayesian Deep Learning via Implicit Regularization of Gradient Descent Survey Taxonomy
- Theoretical Foundations of Implicit Regularization
    - Implicit Regularization in Wide and Overparametrized Networks (3 papers)
    - [3] Loss landscapes are all you need: Neural network generalization can be explained without the implicit bias of gradient descent (P Chiang, 2022) View paper
    - [9] Wide neural networks of any depth evolve as linear models under gradient descent (Jae Hoon Lee, 2019) View paper
    - [22] Generalization in Deep Learning and Bayesian Graph Cut (Taborsky, 2022) View paper
    - Implicit Regularization as Variational Inference ★ (3 papers)
    - [0] Variational Deep Learning via Implicit Regularization (Anon et al., 2026) View paper
    - [7] Gradient regularization as approximate variational inference (Ali Ã nlÃ¼, 2021) View paper
    - [10] Implicitly bayesian prediction rules in deep learning (BK Mlodozeniec, 2024) View paper
    - General Theoretical Perspectives on Bayesian Deep Learning (3 papers)
    - [11] Inductive Bias of Neural Networks and Selected Applications (Heiss, 2024) View paper
    - [17] Deep Learning Regularization: Theory and Data Perspectives (Zhang, 2024) View paper
    - [18] Robustness and Regularization of Deep Neural Networks (Roth, 2021) View paper
- Variational Inference and Gradient-Based Sampling Methods
    - Stein Variational Gradient Descent Methods (1 papers)
    - [15] Accelerating Convergence of Stein Variational Gradient Descent via Deep Unfolding (Yuya Kawamura, 2024) View paper
    - Stochastic Gradient Descent for Gaussian Process Posteriors (1 papers)
    - [21] Sampling from Gaussian Process Posteriors using Stochastic Gradient Descent (Lin, 2023) View paper
    - Bayesian Inference for Models with Implicit Functions (1 papers)
    - [12] Gradient-bridged Posterior: Bayesian Inference for Models with Implicit Functions (Zeng Cheng, 2025) View paper
- Meta-Learning with Bayesian and Gradient-Based Approaches (2 papers)
    - [4] Scalable bayesian meta-learning through generalized implicit gradients (Yilang Zhang, 2023) View paper
    - [14] Amortized Bayesian Meta-Learning with Accelerated Gradient Descent Steps (Zhewei Zhang, 2023) View paper
- Applied Methods for Uncertainty Quantification and Prediction
    - Predictive Maintenance and Remaining Useful Life Estimation (1 papers)
    - [1] Bayesian Deep Learning for Remaining Useful Life Estimation via Stein Variational Gradient Descent (Della Libera, 2024) View paper
    - Regression with Heteroscedastic Uncertainty (2 papers)

- [2] Effective bayesian heteroscedastic regression with deep neural networks (A Immer, 2023) View paper
- [16] Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks paper-review & extension (J DELAVANDE, 2025) View paper
- Climate and Environmental Prediction (1 papers)
- [13] A Bayesian Deep Learning Approach to Nearâ€Term Climate Prediction (Xihaier Luo, 2022) View paper
- Inverse Problems and Image Reconstruction (1 papers)
- [19] Quantifying Model Uncertainty in Inverse Problems via Bayesian Deep Gradient Descent (Riccardo Barbano, 2021) View paper
- Unsupervised Image Restoration with Deep Image Prior (1 papers)
- [5] A retinex inspired deep image prior model for despeckling and deblurring of aerial and satellite images using proximal gradient method (Architha Shastry, 2024) View paper

- Privacy-Preserving and Specialized Neural Network Architectures
  - Differential Privacy in Deep Learning (1 papers)
  - [20] Dif-NoBa A differential privacy noise Bayesian gradient descent algorithm in Deep Learning (Panfeng zhang, 2022) View paper
  - Neural Architectures for Intuition-Based Reasoning (1 papers)
  - [6] BayesIntuit: A Neural Framework for Intuition-Based Reasoning (Mayra Bornacelly, 2025) View paper
- Model Compression and Transfer Learning (1 papers)
  - [8] Deep models, lighter footprint compressing, explaining, and transferring Bayesian neural networks (Saha, 2025) View paper

Narrative

Core task: Bayesian deep learning via implicit regularization of gradient descent. This field explores how standard gradient-based optimization in neural networks can be understood through a Bayesian lens, where the training dynamics themselves induce implicit priors and approximate posterior inference. The taxonomy reveals several complementary perspectives: theoretical foundations examine the mathematical underpinnings of implicit regularization and its connection to variational inference, while variational inference and gradient-based sampling methods develop explicit algorithms that bridge optimization and probabilistic reasoning. Meta-learning branches investigate how gradient-based updates can encode prior knowledge across tasks, and applied methods focus on practical uncertainty quantification for real-world prediction problems. Additional branches address privacy-preserving architectures and model compression, reflecting the need to deploy Bayesian principles in resource-constrained or sensitive settings. Representative works such as Gradient Regularization Inference[7] and Implicitly Bayesian Prediction[10] illustrate how gradient descent can be reinterpreted as performing approximate Bayesian updates, while studies like Loss Landscapes Generalization[3] connect optimization trajectories to generalization behavior.

A central tension in this landscape concerns whether implicit regularization alone suffices for reliable uncertainty estimates or whether explicit variational frameworks are necessary. Some lines of work, including Variational Deep Learning[0], argue that viewing gradient descent as variational inference provides a principled foundation for Bayesian deep learning, closely aligning with Gradient Regularization Inference[7] and Implicitly Bayesian Prediction[10] in emphasizing the implicit Bayesian character of standard training. In contrast, methods like Accelerating SVGD[15] and Gradient-bridged Posterior[12] develop explicit sampling or variational schemes to obtain richer posterior approximations. Variational Deep Learning[0] sits within the theoretical branch that interprets optimization dynamics as variational inference, sharing conceptual ground with its immediate neighbors but differing in how it formalizes the connection between gradient flow and posterior approximation. This positioning highlights ongoing debates about whether implicit biases of optimizers can replace or merely complement explicit Bayesian machinery for uncertainty-aware learning.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Gradient regularization as approximate variational inference

Authors: Ali Ã‚nlÃ¼, Laurence Aitchison | **Year/Venue**: 2021 | **URL**: View paper

Abstract

We developed Variational Laplace for Bayesian neural networks (BNNs), which exploits a local approximation of the curvature of the likelihood to estimate the ELBO without the need for stochastic sampling of the neural-network weights. The Variational Laplace objective is simple to evaluate, as it is the log-likelihood plus weight-decay, plus a squared-gradient regularizer. Variational Laplace gave better test performance and expected calibration errors than maximum a posteriori inference and sta...

Relationship Analysis

Both papers belong to the 'Implicit Regularization as Variational Inference' category, establishing theoretical connections between gradient-based optimization and Bayesian inference frameworks. The original paper proposes training variational neural networks by minimizing expected loss without explicit KL regularization, proving that SGD's implicit bias acts as generalized VI with Wasserstein regularization in overparametrized linear models. The candidate paper develops Variational Laplace, which approximates the ELBO using local curvature of the likelihood, resulting in a practical objective combining log-likelihood, weight-decay, and squared-gradient regularization, focusing on computational efficiency rather than characterizing implicit bias theoretically.

2. Implicitly bayesian prediction rules in deep learning

Authors: BK Mlodozeniec, RE Turner | **Year/Venue**: 2024 | **URL**: View paper

Abstract

â€¦ Bayesian principles into deep learning. In this paper, we propose how to measure how close a general prediction rule is to being implicitly Bayesian, â€¦ fit with gradient descent following â€¦

Relationship Analysis

Both papers belong to the category of establishing theoretical connections between gradient-based implicit regularization and variational Bayesian inference frameworks. They share the common goal of understanding how optimization procedures in neural networks relate to Bayesian principles without explicit priors. The original paper focuses on characterizing the implicit bias of SGD as generalized variational inference with a 2-Wasserstein regularizer in overparameterized linear models, while the candidate paper takes a predictive approach by defining and measuring 'implicitly Bayesian' prediction rules through exchangeability properties, without requiring explicit posterior computation.

Contributions Analysis

Overall novelty summary. The paper proposes Implicit Bias Variational Inference (IBVI), which regularizes variational neural networks by relying solely on the implicit bias of stochastic gradient descent rather than explicit priors or hyperparameter tuning. It resides in the 'Implicit Regularization as Variational Inference' leaf, which contains only three papers total (including this one). This leaf sits within the broader 'Theoretical Foundations of Implicit Regularization' branch, indicating the work occupies a relatively sparse research direction focused on establishing formal connections between optimization dynamics and Bayesian inference frameworks.

The taxonomy reveals neighboring leaves addressing related but distinct perspectives: 'Implicit Regularization in Wide and Overparametrized Networks' examines learning dynamics without explicit variational framing, while 'General Theoretical Perspectives on Bayesian Deep Learning' provides broader reviews. The sibling papers in the same leaf share the goal of interpreting gradient descent as variational inference, but the taxonomy's scope notes clarify that this leaf excludes purely empirical applications (which belong in 'Applied Methods') and meta-learning extensions. The paper thus connects to theoretical characterizations of implicit bias while diverging from explicit sampling methods found in the 'Variational Inference and Gradient-Based Sampling Methods' branch.

Among 25 candidates examined across three contributions, the IBVI method shows one refutable candidate out of 10 examined, suggesting some prior work addresses similar algorithmic ideas. The theoretical characterization of implicit bias as generalized variational inference examined 5 candidates with none refutable, indicating this formalization may offer fresh perspective. The extension of maximal update parametrization to probabilistic networks examined 10 candidates with none refutable, suggesting this parametrization choice is relatively unexplored in the Bayesian setting. The limited search scope (25 candidates, not exhaustive) means these assessments reflect top semantic matches rather than comprehensive field coverage.

Based on the top-25 semantic matches examined, the work appears to contribute novel theoretical framing and parametrization insights within a sparse research direction, though the IBVI method itself encounters some prior overlap. The taxonomy structure confirms this area remains less crowded than applied uncertainty quantification branches, which contain more papers. The analysis does not cover broader optimization literature or recent preprints outside the search scope, so the novelty assessment remains provisional pending deeper investigation of gradient-based Bayesian methods.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Implicit Bias Variational Inference (IBVI) method

Description: The authors introduce a method for training variational neural networks by maximizing the expected log-likelihood without explicit KL regularization to the prior. Instead, the method exploits the implicit regularization of SGD to prevent uncertainty collapse and achieve robust generalization.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Subspace inference for Bayesian deep learning

URL: View paper

Brief Assessment

Subspace Inference[29] focuses on performing Bayesian inference in low-dimensional subspaces of neural network parameter space (e.g., PCA subspaces of SGD trajectories), not on exploiting implicit regularization from SGD to train variational networks without explicit KL regularization.

---

2. Stochastic gradient descent as approximate bayesian inference

URL: View paper

Brief Assessment

SGD Bayesian Inference[28] focuses on using constant-rate SGD as approximate Bayesian inference by tuning learning rates to match stationary distributions to posteriors. The original paper's IBVI exploits implicit regularization from SGD initialization and parametrization in overparametrized models, characterizing it as generalized VI with Wasserstein regularization—a distinct theoretical framework and application context.

---

3. LLM Unlearning via Loss Adjustment with Only Forget Data

URL: View paper

Brief Assessment

LLM Unlearning[31] focuses on machine unlearning for large language models using loss adjustment techniques, not on variational inference with implicit regularization from SGD for Bayesian deep learning.

---

4. Semi-Implicit Variational Inference via Kernelized Path Gradient Descent

URL: View paper

Brief Assessment

Kernelized Path Gradient[36] focuses on semi-implicit variational inference with kernelized KL divergence estimation, not on exploiting implicit regularization from SGD to train variational networks without explicit KL regularization to the prior.

---

5. Implicit bias of SGD in -regularized linear DNNs: One-way jumps from high to low rank

URL: View paper

Brief Assessment

Implicit Bias SGD[33] focuses on implicit bias in L2-regularized linear deep networks for matrix completion tasks, studying rank transitions between local minima. This is fundamentally different from the original paper's variational inference framework for uncertainty quantification in neural networks.

---

6. Semi-Implicit Variational Inference

URL: View paper

Brief Assessment

Semi-Implicit Variational[32] focuses on expanding variational families through mixing distributions with flexible densities, not on exploiting implicit regularization from SGD optimization dynamics.

---

7. Implicitly bayesian prediction rules in deep learning

URL: View paper

Brief Assessment

Implicitly Bayesian Prediction[10] focuses on characterizing prediction rules as implicitly Bayesian through exchangeability properties, not on variational inference methods that exploit implicit regularization from SGD for training neural networks.

---

8. Neural Operator Variational Inference Based on Regularized Stein Discrepancy for Deep Gaussian Processes

URL: View paper

Brief Assessment

Neural Operator Inference[35] focuses on deep Gaussian processes using Stein discrepancy and neural generators for posterior approximation, not on exploiting implicit regularization from SGD for variational neural networks in general settings.

---

9. Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks

URL: View paper

Prior Art Analysis

SGD Variational Inference[34] demonstrates that stochastic gradient descent inherently performs variational inference by minimizing an average potential over a posterior distribution with entropic regularization, without requiring explicit KL regularization. This work predates the ORIGINAL paper and establishes that SGD's implicit regularization naturally prevents uncertainty collapse through its inherent dynamics, which directly challenges the novelty claim that the authors were first to exploit implicit regularization of SGD for variational neural networks.

Evidence

Evidence 1 - **Rationale**: The candidate proves that SGD performs variational inference with an energetic-entropic split without explicit KL regularization, establishing the theoretical foundation for the approach claimed as novel in the original paper. - **Original**: we propose to learn a variational distribution over the weights of a deep neural network by maximizing the expected log-likelihood in analogy to training via maximum likelihood in the standard case. however, in contrast to variational bayes, there is no explicit regularization via a kullback-leibler... - **Candidate**: theorem 5 (sgd performs variational inference). the functional f(ρ) =β-1 kl (ρ ||ρss) decreases monotonically along the trajectories of the fokker-planck equation (fp) and converges to its minimum, which is zero, at steady-state. moreover, we also have an energetic-entropic split f(ρ) =ex∈ρ [φ(x) ...

Evidence 2 - **Rationale**: The candidate demonstrates that SGD's implicit bias toward entropy maximization prevents uncertainty collapse, which is the core mechanism claimed as novel in the original paper. - **Original**: surprisingly, we show theoretically and empirically that training this way does not cause uncertainty to collapse away from the training data, if initialized and parametrized correctly. - **Candidate**: the second shows that sgd has an implicit bias towards solutions that maximize the entropy of the distribution ρ. note that the energetic term in (11) has potential φ(x), instead of f (x). this is an important fact and the crux of this paper.

---

10. A simple baseline for bayesian uncertainty in deep learning

URL: View paper

Brief Assessment

Simple Baseline Uncertainty[30] focuses on SWA-Gaussian for uncertainty estimation using SGD iterates with explicit covariance computation, not on implicit regularization from SGD without KL divergence to the prior.

Contribution 2: Theoretical characterization of implicit bias as generalized variational inference

Description: The authors prove that for overparametrized linear models, the implicit bias of SGD when training via the expected loss is equivalent to generalized variational inference with a 2-Wasserstein regularizer penalizing deviations from the prior, extending prior results for non-probabilistic models.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. 3.2 Coresets and Sketches for Regression Problems on Data Streams and Distributed Data

URL: View paper

Brief Assessment

Coresets Sketches Regression[25] focuses on coresets and sketches for regression problems on data streams and distributed data, not on implicit bias in overparameterized models or variational inference frameworks.

---

2. Why do Overparameterized Neural Networks Generalize?

URL: View paper

Brief Assessment

Overparameterized Networks Generalize[24] is a thesis overview discussing general generalization phenomena in overparameterized networks, not a research paper proving specific theoretical results about implicit bias in linear models as generalized Bayesian inference with Wasserstein regularizers.

---

3. Optimal Implicit Bias in Linear Regression

URL: View paper

Brief Assessment

Optimal Implicit Bias[26] focuses on linear regression models and characterizes implicit bias through convex optimization potentials, not through generalized variational inference with Wasserstein regularizers for overparameterized probabilistic models as in the original paper.

---

4. Computationally Efficient Posterior Inference with Langevin Monte Carlo and Early Stopping

URL: View paper

Brief Assessment

Langevin Early Stopping[27] focuses on early-stopped Langevin Monte Carlo for posterior sampling, not on characterizing implicit bias of SGD in overparameterized models as generalized variational inference with Wasserstein regularization.

---

5. On the Optimal Weighted Regularization in Overparameterized Linear Regression

URL: View paper

Brief Assessment

Optimal Weighted Regularization[23] focuses on weighted ridge regression in overparameterized linear models, not on characterizing implicit bias of SGD as generalized variational inference with Wasserstein regularizers for probabilistic models.

Contribution 3: Extension of maximal update parametrization to probabilistic networks

Description: The authors extend the maximal update parametrization (μP) to variational neural networks, enabling hyperparameter transfer from small to large models and ensuring feature learning even as network width increases, which is demonstrated empirically on CIFAR-10.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. LO: Compute-Efficient Meta-Generalization of Learned Optimizers

URL: View paper

Brief Assessment

Meta-Generalization Learned Optimizers[46] focuses on extending maximal update parametrization (μP) to learned optimizer architectures for meta-generalization, not to variational/probabilistic neural networks as in the original paper.

2. u-P: The Unit-Scaled Maximal Update Parametrization

URL: View paper

Brief Assessment

Unit-Scaled Maximal Update[39] focuses on combining maximal update parametrization with unit scaling for deterministic neural networks and low-precision training, not on extending μP to variational/probabilistic networks with uncertainty quantification as in the original paper.

3. Tuning large neural networks via zero-shot hyperparameter transfer

URL: View paper

Brief Assessment

Zero-Shot Hyperparameter Transfer[38] focuses on hyperparameter transfer in deterministic neural networks using maximal update parametrization (μP), not on extending μP to variational/probabilistic networks for Bayesian deep learning.

4. Sparse maximal update parameterization: A holistic approach to sparse training dynamics

URL: View paper

Brief Assessment

Sparse Maximal Update[37] focuses on sparse neural network training dynamics and hyperparameter transfer for sparsity levels, not on extending μP to variational/probabilistic networks as in the original paper.

5. Practical efficiency of muon for pretraining

URL: View paper

Brief Assessment

Practical Efficiency Muon[41] focuses on applying maximal update parametrization (μP) to the muon optimizer for deterministic neural networks in pretraining contexts, not on extending μP to variational/probabilistic neural networks with uncertainty quantification as in the original paper.

6. Scaling Diffusion Transformers Efficiently via Î¼P

URL: View paper

Brief Assessment

Scaling Diffusion Transformers[45] focuses on extending μP to diffusion transformers for vision generative models, not to variational neural networks for Bayesian deep learning as in the original paper.

7. Maximal Update Parametrization and Zero-Shot Hyperparameter Transfer for Fourier Neural Operators

URL: View paper

Brief Assessment

Maximal Update FNO[43] focuses on extending μP to Fourier Neural Operators for PDE solving, not to variational/probabilistic neural networks. The candidate addresses hyperparameter transfer in a completely different domain (operator learning) with different network architectures and does not challenge the novelty of applying μP to variational Bayesian networks.

8. Feature learning in infinite-depth neural networks

URL: View paper

Brief Assessment

Feature Learning Infinite-Depth[42] focuses on depth scaling in deterministic residual networks, not on extending maximal update parametrization to variational/probabilistic neural networks as claimed in the original paper.

9. Deep Linear Network Training Dynamics from Random Initialization: Data, Width, Depth, and Hyperparameter Transfer

URL: View paper

Brief Assessment

Deep Linear Training[44] focuses on theoretical characterization of gradient descent dynamics in deep linear networks with random initialization and data, not on extending maximal update parametrization to variational/probabilistic neural networks as in the original paper.

10. Scaling exponents across parameterizations and optimizers

URL: View paper

Brief Assessment

Scaling Exponents Parameterizations[40] focuses on scaling exponents across different parameterizations (standard, NTK, μP, mean-field) for deterministic neural networks with optimizers like SGD and Adam, not on extending μP specifically to variational/probabilistic networks as claimed in the original contribution.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Variational Deep Learning via Implicit Regularization View paper
- [1] Bayesian Deep Learning for Remaining Useful Life Estimation via Stein Variational Gradient Descent View paper
- [2] Effective bayesian heteroscedastic regression with deep neural networks View paper
- [3] Loss landscapes are all you need: Neural network generalization can be explained without the implicit bias of gradient descent View paper
- [4] Scalable bayesian meta-learning through generalized implicit gradients View paper
- [5] A retinex inspired deep image prior model for despeckling and deblurring of aerial and satellite images using proximal gradient method View paper
- [6] BayesIntuit: A Neural Framework for Intuition-Based Reasoning View paper

- [7] Gradient regularization as approximate variational inference View paper
- [8] Deep models, lighter footprint compressing, explaining, and transferring Bayesian neural networks View paper
- [9] Wide neural networks of any depth evolve as linear models under gradient descent View paper
- [10] Implicitly bayesian prediction rules in deep learning View paper
- [11] Inductive Bias of Neural Networks and Selected Applications View paper
- [12] Gradient-bridged Posterior: Bayesian Inference for Models with Implicit Functions View paper
- [13] A Bayesian Deep Learning Approach to Nearâ□□Term Climate Prediction View paper
- [14] Amortized Bayesian Meta-Learning with Accelerated Gradient Descent Steps View paper
- [15] Accelerating Convergence of Stein Variational Gradient Descent via Deep Unfolding View paper
- [16] Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks paper-review & extension View paper
- [17] Deep Learning Regularization: Theory and Data Perspectives View paper
- [18] Robustness and Regularization of Deep Neural Networks View paper
- [19] Quantifying Model Uncertainty in Inverse Problems via Bayesian Deep Gradient Descent View paper
- [20] Dif-NoBa A differential privacy noise Bayesian gradient descent algorithm in Deep Learning View paper
- [21] Sampling from Gaussian Process Posteriors using Stochastic Gradient Descent View paper
- [22] Generalization in Deep Learning and Bayesian Graph Cut View paper
- [23] On the Optimal Weighted Regularization in Overparameterized Linear Regression View paper
- [24] Why do Overparameterized Neural Networks Generalize? View paper
- [25] 3.2 Coresets and Sketches for Regression Problems on Data Streams and Distributed Data View paper
- [26] Optimal Implicit Bias in Linear Regression View paper
- [27] Computationally Efficient Posterior Inference with Langevin Monte Carlo and Early Stopping View paper
- [28] Stochastic gradient descent as approximate bayesian inference View paper
- [29] Subspace inference for Bayesian deep learning View paper
- [30] A simple baseline for bayesian uncertainty in deep learning View paper
- [31] LLM Unlearning via Loss Adjustment with Only Forget Data View paper
- [32] Semi-Implicit Variational Inference View paper
- [33] Implicit bias of SGD in -regularized linear DNNs: One-way jumps from high to low rank View paper
- [34] Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks View paper
- [35] Neural Operator Variational Inference Based on Regularized Stein Discrepancy for Deep Gaussian Processes View paper
- [36] Semi-Implicit Variational Inference via Kernelized Path Gradient Descent View paper
- [37] Sparse maximal update parameterization: A holistic approach to sparse training dynamics View paper
- [38] Tuning large neural networks via zero-shot hyperparameter transfer View paper
- [39] u-P: The Unit-Scaled Maximal Update Parametrization View paper
- [40] Scaling exponents across parameterizations and optimizers View paper
- [41] Practical efficiency of muon for pretraining View paper
- [42] Feature learning in infinite-depth neural networks View paper
- [43] Maximal Update Parametrization and Zero-Shot Hyperparameter Transfer for Fourier Neural Operators View paper
- [44] Deep Linear Network Training Dynamics from Random Initialization: Data, Width, Depth, and Hyperparameter Transfer View paper
- [45] Scaling Diffusion Transformers Efficiently via Î¼P View paper
- [46] LO: Compute-Efficient Meta-Generalization of Learned Optimizers View paper