Novelty Assessment Report

Paper: Directional Convergence, Benign Overfitting of Gradient Descent in leaky ReLU two-layer Neural Networks

PDF URL: https://openreview.net/pdf?id=VOK6LNaZ3N

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

In this paper, we provide sufficient conditions of benign overfitting of fixed width leaky ReLU two-layer neural network classifiers trained on mixture data via gradient descent. Our results are derived by establishing directional convergence of the network parameters and classification error bound of the convergent direction. Our classification error bound also lead to the discovery of a newly identified phase transition. Previously, directional convergence in (leaky) ReLU neural networks was established only for gradient flow. Due to the lack of directional convergence, previous results on benign overfitting were limited to those trained on nearly orthogonal data. All of our results hold on mixture data, which is a broader data setting than the nearly orthogonal data setting in prior work. We demonstrate our findings by showing that benign overfitting occurs with high probability in a much wider range of scenarios than previously known. Our results also allow us to characterize cases when benign overfitting provably fails even if directional convergence occurs. Our work thus provides a more complete picture of benign overfitting in leaky ReLU two-layer neural networks.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Benign Overfitting in Two-Layer Neural Networks**

A total of **34 papers** were analyzed and organized into a taxonomy with **12 categories**.

Taxonomy Overview

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

- **Activation Function and Architecture Variants**
- **Data Characteristics and Noise Models**
- **Adversarial Robustness and Security**
- **Generalization Theory and Implicit Regularization**
- **Extended Architectures and Generalizations**

Complete Taxonomy Tree

- Benign Overfitting in Two-Layer Neural Networks Survey Taxonomy
- Activation Function and Architecture Variants
  - ReLU and Leaky ReLU Networks
  - Gradient Descent Training Dynamics ★ (4 papers)
    - [0] Directional Convergence, Benign Overfitting of Gradient Descent in leaky ReLU two-layer Neural Networks (Anon et al., 2026) View paper
    - [3] Benign Overfitting for Two-layer ReLU Networks (Kou, 2023) View paper
    - [5] Benign overfitting for regression with trained two-layer relu networks (Park junhyung, 2024) View paper
    - [18] Stable Minima Cannot Overfit in Univariate ReLU Networks: Generalization by Large Step Sizes (Dan Qiao, 2024) View paper
  - Hinge Loss and Margin Maximization (3 papers)
    - [8] Benign overfitting in leaky ReLU networks with moderate input dimension (Erin George, 2024) View paper
    - [12] Benign Overfitting in Linear Classifiers and Leaky ReLU Networks from KKT Conditions for Margin Maximization (Frei, 2023) View paper
    - [24] Training shallow ReLU networks on noisy data using hinge loss: when do we overfit and is it benign? (George, 2023) View paper
  - Logistic Loss and Classification (2 papers)
    - [7] Benign overfitting and grokking in relu networks for xor cluster data (Xu Zhiwei, 2023) View paper
    - [21] Benign Overfitting without Linearity: Neural Network Classifiers Trained by Gradient Descent for Noisy Linear Data (Frei, 2022) View paper
  - Convolutional Neural Networks (4 papers)
  - [1] Benign Overfitting in Two-layer Convolutional Neural Networks (Cao Yuan, 2022) View paper
  - [4] Benign Overfitting in Two-layer ReLU Convolutional Neural Networks (Y. Kou, 2023) View paper
  - [11] Initialization Matters: On the Benign Overfitting of Two-Layer ReLU CNN with Fully Trainable Layers (Meng, 2024) View paper
  - [14] Benign Overfitting for Two-layer ReLU Convolutional Neural Networks (Kou, 2023) View paper
  - Linear and Smooth Activation Networks (3 papers)
  - [10] The Interplay Between Implicit Bias and Benign Overfitting in Two-Layer Linear Networks (Chatterji, 2021) View paper
  - [15] Optimal criterion for feature learning of two-layer linear neural network in high dimensional interpolation regime (K Suzuki, 2023) View paper
  - [25] Feature learning and generalization error analysis of two-layer linear neural networks for high-dimensional inputs (Hayato Nishimori, 2024) View paper

- Data Characteristics and Noise Models
  - Signal-Noise Decomposition and SNR Thresholds (3 papers)
  - [6] Rethinking Benign Overfitting in Two-Layer Neural Networks (Xu Ruichen, 2025) View paper
  - [22] From Tempered to Benign Overfitting in ReLU Neural Networks (Kornowski, 2023) View paper
  - [33] Label Noise Gradient Descent Improves Generalization in the Low SNR Regime (W Huang, n.d.) View paper
  - Cluster and Class-Conditional Data (2 papers)
  - [19] The double-edged sword of implicit bias: Generalization vs. robustness in relu networks (Frei, 2023) View paper
  - [28] Towards an understanding of benign overfitting in neural networks (Li Zhu, 2021) View paper
  - Noisy Features and High-Dimensional Regimes (2 papers)
  - [26] Benign Overfitting and Noisy Features (Zhu Li, 2022) View paper
  - [27] A Classical View on Benign Overfitting: The Role of Sample Size (Park junhyung, 2025) View paper
- Adversarial Robustness and Security (3 papers)
  - [2] Benign overfitting in adversarial training of neural networks (Y Wang, 2024) View paper
  - [9] The surprising harmfulness of benign overfitting for adversarial robustness (Hao Yi-fan, 2024) View paper
  - [23] Scanning trojaned models using out-of-distribution samples (Ali Ansari, 2024) View paper
- Generalization Theory and Implicit Regularization
  - Stochastic Gradient Descent and Online Learning (3 papers)
  - [30] Dimension Independent Generalization Error with Regularized Online Optimization (Chen Xi, 2022) View paper
  - [31] Dimension independent excess risk by stochastic gradient descent (Xi Chen, 2022) View paper
  - Implicit Bias and Interpolation Mechanisms (3 papers)
  - [17] More is better: when infinite overparameterization is optimal and overfitting is obligatory (JB Simon, 2024) View paper
  - [20] Deep learning: a statistical viewpoint (Peter L. Bartlett, 2021) View paper
  - [29] Optimization Dynamics in Mildly Overparametrized Models (Zhou, 2024) View paper
- Extended Architectures and Generalizations (3 papers)
  - [13] Generalization Ability of Wide Neural Networks on (J Lai, 2023) View paper
  - [16] Unveil benign overfitting for transformer in vision: Training dynamics, convergence, and generalization (Wei Huang, 2024) View paper
  - [34] DIRECTIONAL CONVERGENCE, BENIGN OVERFITTING (DESCENT, n.d.) View paper

Narrative

Core task: benign overfitting in two-layer neural networks. This field investigates the phenomenon where overparameterized shallow networks interpolate noisy training data yet still generalize well on test data, defying classical statistical intuition. The taxonomy organizes research into several main branches: Activation Function and Architecture Variants explore how different nonlinearities (ReLU, leaky ReLU) and architectural choices influence benign overfitting; Data Characteristics and Noise Models examine the role of label noise, feature structure, and sample complexity; Adversarial Robustness and Security study whether benign overfitting persists under adversarial perturbations; Generalization Theory and Implicit Regularization analyze the implicit biases of gradient-based training that enable good generalization despite interpolation; and Extended Architectures and Generalizations broaden the scope to convolutional networks, transformers, and deeper models. Representative works such as Benign Overfitting ReLU[3] and Benign Overfitting Leaky ReLU[8] illustrate how activation choices shape the training dynamics, while studies like Benign Overfitting Adversarial[2] and Benign Overfitting Noisy Features[26] highlight the interplay between data properties and overfitting behavior.

A particularly active line of work focuses on gradient descent dynamics with ReLU-type activations, examining how directional convergence and implicit bias lead networks toward max-margin solutions that generalize despite perfect training fit. Directional Convergence Leaky ReLU[0] sits squarely within this branch, analyzing how leaky ReLU networks trained by gradient descent exhibit directional convergence properties that facilitate benign overfitting. This work closely relates to Benign Overfitting ReLU[3], which establishes foundational results for standard ReLU networks, and contrasts with Benign Overfitting Regression[5], which explores similar phenomena in simpler regression settings without the complexities of nonlinear activations. Meanwhile, other branches investigate whether benign overfitting extends to adversarially robust training or whether it breaks down under distribution shift, revealing trade-offs between interpolation, generalization, and robustness. Open questions remain about the precise conditions under which benign overfitting occurs, the role of initialization and architecture depth, and how these insights scale to practical deep learning scenarios.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Benign Overfitting for Two-layer ReLU Networks

Authors: Kou, Yiwen, Yiwen Kou, Chen, Zixiang, et al. (13 authors total) | **Year/Venue**: 2023 • arXiv.org | **URL**: View paper

Abstract

N/A

Relationship Analysis

Both papers belong to the same taxonomy category studying gradient descent training dynamics in ReLU networks for benign overfitting. They share the focus on establishing directional convergence and benign overfitting conditions for two-layer neural networks with ReLU-type activations trained via gradient descent. The key difference is that the original paper studies leaky ReLU networks on mixture data with general covariance structures, while the candidate paper focuses on standard ReLU convolutional neural networks with specific patch-based data structures and label-flipping noise, using different proof techniques (time-invariant coefficient ratio analysis versus directional convergence characterization).

2. Benign overfitting for regression with trained two-layer relu networks

Authors: Park junhyung, Bloebaum, Patrick, Junhyung Park, Kasiviswanathan, et al. (8 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

We study the least-square regression problem with a two-layer fully-connected neural network, with ReLU activation function, trained by gradient flow. Our first result is a generalization result, that requires no assumptions on the underlying regression function or the noise other than that they are bounded. We operate in the neural tangent kernel regime, and our generalization result is developed via a decomposition of the excess risk into estimation and approximation errors, viewing gradient f...

Relationship Analysis

Both papers belong to the same taxonomy category studying gradient descent training dynamics in ReLU networks, focusing on directional convergence and benign overfitting conditions. The original paper establishes directional convergence of gradient descent for leaky ReLU networks on mixture data with precise characterization of convergent directions, while the candidate paper studies benign overfitting for standard ReLU networks in the NTK regime for regression problems. The key difference is that the original paper proves directional convergence for gradient descent beyond the nearly orthogonal data regime without NTK assumptions, whereas the candidate paper operates strictly in the NTK/lazy training regime using gradient flow and focuses on approximation-estimation error decomposition for arbitrary regression functions.

3. Stable Minima Cannot Overfit in Univariate ReLU Networks: Generalization by Large Step Sizes

Authors: Dan Qiao, Esha Singh, Daniel Soudry, Yu-Xiang Wang, Kaiqi Zhang | **Year/Venue**: 2024 • Neural Information Processing Systems | **URL**: View paper

Abstract

We study the generalization of two-layer ReLU neural networks in a univariate nonparametric regression problem with noisy labels. This is a problem where kernels (\emph{e.g.} NTK) are provably sub-optimal and benign overfitting does not happen, thus disqualifying existing theory for interpolating (0-loss, global optimal) solutions. We present a new theory of generalization for local minima that gradient descent with a constant learning rate can \emph{stably} converge to. We show that gradient de...

Relationship Analysis

Both papers study gradient descent training dynamics in two-layer ReLU networks, sharing the same taxonomy category focused on establishing convergence and generalization conditions. While the original paper establishes directional convergence and benign overfitting conditions for leaky ReLU networks on mixture data with exponential loss, the candidate paper studies stable minima and generalization in univariate ReLU networks for nonparametric regression with squared loss, focusing on how large learning rates induce sparse linear spline fits rather than directional convergence to maximum margin solutions.

Contributions Analysis

Overall novelty summary. The paper establishes directional convergence for gradient descent (not just gradient flow) in leaky ReLU two-layer networks and derives classification error bounds revealing a phase transition in benign overfitting. It sits in the 'Gradient Descent Training Dynamics' leaf under 'ReLU and Leaky ReLU Networks', which contains four papers total. This is a moderately populated research direction within the broader taxonomy of 34 papers across the field, indicating focused but not overcrowded attention to gradient descent dynamics in ReLU-type networks.

The taxonomy shows this leaf is one of three under 'ReLU and Leaky ReLU Networks', with sibling leaves examining 'Hinge Loss and Margin Maximization' (three papers) and 'Logistic Loss and Classification' (two papers). Neighboring branches explore 'Convolutional Neural Networks' (four papers) and 'Linear and Smooth Activation Networks' (three papers). The scope_note clarifies this leaf focuses specifically on directional convergence under gradient descent/flow, excluding alternative loss functions. The paper's extension from gradient flow to gradient descent represents a technical advance within this established research direction.

Among 26 candidates examined, the contribution on directional convergence for gradient descent (10 candidates, 0 refutable) and the phase transition discovery (10 candidates, 0 refutable) appear novel within the limited search scope. However, the claim of extending results to 'broader data settings' (6 candidates examined, 2 refutable) shows more substantial prior work overlap. The statistics suggest the first two contributions face less direct competition among the examined candidates, while the data generality claim encounters existing work addressing similar mixture or non-orthogonal data scenarios.

Based on the top-26 semantic matches examined, the technical contributions on gradient descent convergence and phase transitions appear relatively novel, while the data setting extension shows clearer overlap with prior work. The analysis covers a focused subset of the literature; a broader search might reveal additional related work, particularly in the 'Data Characteristics and Noise Models' branch (nine papers) which was not the primary focus of this candidate examination.

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

Contribution 1: Directional convergence of gradient descent in leaky ReLU two-layer neural networks

Description: The authors establish directional convergence of gradient descent for leaky ReLU two-layer neural networks trained on mixture data with exponential loss, providing precise characterization of the convergent direction. This is the first such result for ReLU-type networks under gradient descent, extending beyond prior work limited to gradient flow or nearly orthogonal data.

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. Topological obstruction to the training of shallow ReLU neural networks

URL: View paper

Brief Assessment

Topological Obstruction Shallow[43] studies gradient flow trajectories constrained to invariant hyperquadrics in shallow ReLU networks, focusing on topological obstructions rather than directional convergence properties of gradient descent.

2. Learning a neuron by a shallow relu network: Dynamics and implicit bias for correlated inputs

URL: View paper

Brief Assessment

Learning Neuron Correlated[48] focuses on gradient flow (not gradient descent) for learning a single neuron with ReLU activation under correlated inputs. The candidate does not address gradient descent dynamics for leaky ReLU networks on mixture data with exponential loss.

3. Benign Overfitting in Two-layer ReLU Convolutional Neural Networks

URL: View paper

Brief Assessment

Benign Overfitting ReLU CNN[4] focuses on two-layer ReLU convolutional neural networks with label-flipping noise, not fully connected networks with leaky ReLU activation. The candidate does not establish directional convergence results for gradient descent in the setting studied by the original paper.

4. Gradient descent on two-layer nets: Margin maximization and simplicity bias

URL: View paper

Brief Assessment

Margin Maximization Simplicity[44] analyzes gradient flow (continuous-time dynamics) rather than gradient descent (discrete updates). The original paper explicitly states this is 'the first such result for ReLU-type networks under gradient descent' and distinguishes their discrete-time analysis from prior gradient flow work.

5. SGD Learns Over-parameterized Networks that Provably Generalize on Linearly Separable Data

URL: View paper

Brief Assessment

Over-parameterized Provably Generalize[47] focuses on generalization guarantees for SGD on linearly separable data, not on establishing directional convergence with precise characterization of the convergent direction for gradient descent.

6. Training two-layer RELU networks with gradient descent is inconsistent

URL: View paper

Brief Assessment

Two-layer Inconsistent[50] studies inconsistency and local minima in gradient descent training, not directional convergence properties. The paper focuses on failure modes rather than convergence guarantees.

7. The Implicit Bias of Minima Stability in Multivariate Shallow ReLU Networks

URL: View paper

Brief Assessment

Minima Stability Implicit[51] focuses on stability analysis of shallow ReLU networks with quadratic loss, characterizing stable solutions through smoothness properties. It does not address directional convergence of gradient descent for leaky ReLU networks on mixture data with exponential loss, which is the core novelty claim of the original paper.

8. Feature selection and low test error in shallow low-rotation relu networks

URL: View paper

Brief Assessment

Feature Selection Low-rotation[45] focuses on gradient flow and SGD in ReLU networks with low-rotation constraints and margin maximization, not on establishing directional convergence of gradient descent for leaky ReLU networks on mixture data with exponential loss.

9. Non-Singularity of the Gradient Descent map for Neural Networks with Piecewise Analytic Activations

URL: View paper

Brief Assessment

Non-Singularity Gradient Map[49] focuses on proving non-singularity of the gradient descent map for neural networks with piecewise analytic activations, not on establishing directional convergence with precise characterization of convergent direction for leaky ReLU networks on mixture data.

10. Towards understanding learning in neural networks with linear teachers

URL: View paper

Brief Assessment

Linear Teachers Understanding[46] focuses on proving global optimization of SGD for cross-entropy loss and characterizing convergence to linear decision boundaries through weight clustering. It does not establish directional convergence of gradient descent with precise characterization of the convergent direction for mixture data with exponential loss, which is the specific contribution claimed by the original paper.

Contribution 2: Classification error bounds revealing phase transition in benign overfitting

Description: The authors derive classification error bounds for the convergent direction that reveal a phase transition between weak signal and strong signal regimes. They provide both upper and lower bounds for Gaussian mixtures, showing when benign overfitting occurs or provably fails even with directional convergence.

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. Universal scaling laws of absorbing phase transitions in artificial deep neural networks

URL: View paper

Brief Assessment

Universal Scaling Laws[36] studies phase transitions in signal propagation dynamics at the 'edge of chaos' in deep neural networks using statistical mechanics frameworks. This is fundamentally different from the original paper's analysis of classification error bounds and benign overfitting in gradient descent training of two-layer networks.

2. Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models

URL: View paper

Brief Assessment

Memorizing without Overfitting[35] studies phase transitions in bias-variance trade-off for over-parameterized models but does not address benign overfitting classification error bounds with directional convergence in neural networks trained via gradient descent on mixture data.

3. Rethinking Benign Overfitting in Two-Layer Neural Networks

URL: View paper

Brief Assessment

Rethinking Benign Overfitting[6] studies two-layer CNNs with heterogeneous class-dependent noise and focuses on long-tailed data distributions, while the original paper examines leaky ReLU fully-connected networks with mixture data. The phase transitions identified differ in their underlying mechanisms and data settings.

4. Benign, tempered, or catastrophic: Toward a refined taxonomy of overfitting

URL: View paper

Brief Assessment

Refined Taxonomy Overfitting[40] studies phase transitions between benign, tempered, and catastrophic overfitting regimes in kernel regression and neural networks, but does not provide classification error bounds with directional convergence for leaky ReLU networks on mixture data as in the original paper.

5. Benign overfitting in adversarially robust linear classification

URL: View paper

Brief Assessment

Adversarially Robust Linear[38] focuses on adversarial training settings with adversarial perturbations, not the phase transition phenomenon in standard benign overfitting for neural networks studied in the original paper.

6. Unveil benign overfitting for transformer in vision: Training dynamics, convergence, and generalization

URL: View paper

Brief Assessment

Transformer Vision Overfitting[16] studies benign overfitting in vision transformers with softmax attention, while the original work focuses on leaky ReLU two-layer neural networks. The phase transition conditions differ fundamentally: the candidate uses $n \cdot \mathrm{SNR}^2 = \omega(1)$ for transformers, whereas the original characterizes phase transitions between weak signal ($n\|\mu\|^2 \lesssim r$) and strong signal ($n\|\mu\|^2 \gtrsim r$) regimes for neural networks with different architectural components and training dynamics.

7. Benign overfitting in leaky ReLU networks with moderate input dimension

URL: View paper

Brief Assessment

Benign Overfitting Leaky ReLU[8] studies benign overfitting in leaky ReLU networks with hinge loss, focusing on signal-to-noise ratio conditions. The original paper uses exponential loss and derives phase transitions between weak/strong signal regimes based on $n\|\mu\|^2$ versus r, which is a different analytical framework.

8. Benign Overfitting without Linearity: Neural Network Classifiers Trained by Gradient Descent for Noisy Linear Data

URL: View paper

Brief Assessment

Benign Overfitting Noisy Linear[21] studies two-layer neural networks with log-concave distributions and adversarial label noise, while the original paper focuses on Gaussian mixture models with specific phase transitions between weak/strong signal regimes. The distributional assumptions and technical approaches differ substantially.

9. Benign overfitting of non-smooth neural networks beyond lazy training

URL: View paper

Brief Assessment

Non-smooth Beyond Lazy[37] focuses on binary classification with non-smooth activations and provides generalization bounds, but does not establish phase transition phenomena between weak and strong signal regimes with both upper and lower bounds as claimed in the original work.

10. Understanding generalization in transformers: Error bounds and training dynamics under benign and harmful overfitting

URL: View paper

Brief Assessment

Transformers Error Bounds[39] focuses on transformers with label-flipping noise and attention mechanisms, not leaky ReLU two-layer neural networks on mixture data. The phase transitions studied differ fundamentally in architecture and data settings.

Contribution 3: Extension of benign overfitting results to broader data settings

Description: The authors extend benign overfitting results beyond the nearly orthogonal data regime studied in prior work to general mixture data settings, including polynomially tailed distributions. Their deterministic conditions allow proving benign overfitting with high probability under weaker distributional assumptions than previous sub-Gaussian requirements.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. DIRECTIONAL CONVERGENCE, BENIGN OVERFITTING

URL: View paper

Brief Assessment

Directional Convergence Benign[34] is the same paper as the original submission (identical abstract, methodology, and results). This is a self-comparison rather than a distinct prior work that could refute novelty claims.

2. Benign overfitting and grokking in relu networks for xor cluster data

URL: View paper

Brief Assessment

Benign Overfitting Grokking[7] focuses on XOR cluster data with label noise in neural networks, demonstrating benign overfitting in a non-linearly separable setting. The original paper extends benign overfitting to mixture data beyond nearly orthogonal settings with polynomially tailed distributions, which is a different distributional generalization.

3. Binary Classification of Gaussian Mixtures: Abundance of Support Vectors, Benign Overfitting, and Regularization

URL: View paper

Brief Assessment

Gaussian Mixtures Classification[42] studies binary linear classification under Gaussian mixture models, while the original paper focuses on two-layer leaky ReLU neural networks trained via gradient descent. The candidate's focus on linear classifiers (SVM and min-norm interpolating) represents a fundamentally different model class than the neural network architecture studied in the original work.

4. Benign overfitting in leaky ReLU networks with moderate input dimension

URL: View paper

Prior Art Analysis

Benign Overfitting Leaky ReLU[8] explicitly claims to extend benign overfitting results beyond nearly orthogonal data settings, requiring only $d = \Omega(n)$ instead of $d = \Omega(n^2 \log n)$. This directly challenges the original paper's claim of being first to extend beyond nearly orthogonal regimes. Both papers study mixture data settings and relax orthogonality assumptions, with the candidate achieving this with moderate input dimension requirements.

Evidence

Evidence 1 - **Rationale**: Both papers claim to extend benign overfitting beyond nearly orthogonal data settings. The candidate explicitly states they do not require nearly orthogonal training data and achieve this with significantly weaker dimensional requirements ($d = \Omega(n)$ vs $d = \Omega(n^2 \log n)$), demonstrating that similar extensions were achieved in prior work. - **Original**: all of our results hold on mixture data, which is a broader data setting than the nearly orthogonal data setting in prior work. we demonstrate our findings by showing that benign overfitting occurs with high probability in a much wider range of scenarios than previously known. - **Candidate**: in contrast to prior work we do not require the training data to be nearly orthogonal. notably, for input dimension D and training sample size N, while results in prior work require $d = \omega(n^2 \log n)$, here we require only $d = \omega\left(n\right)$.

Evidence 2 - **Rationale**: Both papers study mixture data models where inputs decompose into signal and noise components. The candidate's work on such mixture models without requiring nearly orthogonal data demonstrates that extensions to broader data settings were already established in prior work. - **Original**: our work establishes benign overfitting beyond the nearly orthogonal data regime and can be applied to a wider class of mixtures than the prior work such as polynomially tailed mixture. - **Candidate**: we consider input data that can be decomposed into the sum of a common signal and a random noise component, that lie on subspaces orthogonal to one another.

5. Benign overfitting in multiclass classification: All roads lead to interpolation

URL: View paper

Brief Assessment

Multiclass Interpolation[41] focuses on multiclass linear classification with Gaussian mixture models and multinomial logistic models, while the original paper studies two-layer leaky ReLU neural networks on mixture data. These are fundamentally different model classes and settings.

6. Benign overfitting of non-smooth neural networks beyond lazy training

URL: View paper

Prior Art Analysis

Non-smooth Beyond Lazy[37] demonstrates that benign overfitting results can be extended beyond nearly orthogonal data regimes to general mixture data settings under weaker distributional assumptions. The candidate paper explicitly states it removes restrictions to nearly orthogonal data and extends to mixture models with only logarithmic Sobolev constant assumptions rather than sub-Gaussian requirements, which directly overlaps with the original paper's claimed contribution of extending beyond nearly orthogonal settings to polynomially tailed distributions.

Evidence

Evidence 1 - **Rationale**: Both papers claim to extend benign overfitting results to mixture data beyond nearly orthogonal settings with weaker distributional assumptions than sub-Gaussian, showing the candidate established this extension prior to the original paper. - **Original**: all of our results hold on mixture data, which is a broader data setting than the nearly orthogonal data setting in prior work. - **Candidate**: remark 1 (log-sobolev assumption). the assumption on the logarithmic sobolev constant (ledoux, 2001) ofpz is a purely technical one; it is used to derive lipschitz concentration properties (cf. ledoux (2001); please refer to the supplement material of this paper for details) for a simple proof of th...

Appendix: Text Similarity Detection

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

References

- [0] Directional Convergence, Benign Overfitting of Gradient Descent in leaky ReLU two-layer Neural Networks View paper
- [1] Benign Overfitting in Two-layer Convolutional Neural Networks View paper
- [2] Benign overfitting in adversarial training of neural networks View paper
- [3] Benign Overfitting for Two-layer ReLU Networks View paper
- [4] Benign Overfitting in Two-layer ReLU Convolutional Neural Networks View paper
- [5] Benign overfitting for regression with trained two-layer relu networks View paper
- [6] Rethinking Benign Overfitting in Two-Layer Neural Networks View paper
- [7] Benign overfitting and grokking in relu networks for xor cluster data View paper
- [8] Benign overfitting in leaky ReLU networks with moderate input dimension View paper
- [9] The surprising harmfulness of benign overfitting for adversarial robustness View paper
- [10] The Interplay Between Implicit Bias and Benign Overfitting in Two-Layer Linear Networks View paper
- [11] Initialization Matters: On the Benign Overfitting of Two-Layer ReLU CNN with Fully Trainable Layers View paper
- [12] Benign Overfitting in Linear Classifiers and Leaky ReLU Networks from KKT Conditions for Margin Maximization View paper
- [13] Generalization Ability of Wide Neural Networks on View paper
- [14] Benign Overfitting for Two-layer ReLU Convolutional Neural Networks View paper
- [15] Optimal criterion for feature learning of two-layer linear neural network in high dimensional interpolation regime View paper
- [16] Unveil benign overfitting for transformer in vision: Training dynamics, convergence, and generalization View paper
- [17] More is better: when infinite overparameterization is optimal and overfitting is obligatory View paper
- [18] Stable Minima Cannot Overfit in Univariate ReLU Networks: Generalization by Large Step Sizes View paper
- [19] The double-edged sword of implicit bias: Generalization vs. robustness in relu networks View paper
- [20] Deep learning: a statistical viewpoint View paper
- [21] Benign Overfitting without Linearity: Neural Network Classifiers Trained by Gradient Descent for Noisy Linear Data View paper
- [22] From Tempered to Benign Overfitting in ReLU Neural Networks View paper
- [23] Scanning trojaned models using out-of-distribution samples View paper
- [24] Training shallow ReLU networks on noisy data using hinge loss: when do we overfit and is it benign? View paper
- [25] Feature learning and generalization error analysis of two-layer linear neural networks for high-dimensional inputs View paper
- [26] Benign Overfitting and Noisy Features View paper
- [27] A Classical View on Benign Overfitting: The Role of Sample Size View paper

- [28] Towards an understanding of benign overfitting in neural networks View paper
- [29] Optimization Dynamics in Mildly Overparametrized Models View paper
- [30] Dimension Independent Generalization Error with Regularized Online Optimization View paper
- [31] Dimension independent excess risk by stochastic gradient descent View paper
- [32] Dimension Independent Generalization Error by Stochastic Gradient Descent View paper
- [33] Label Noise Gradient Descent Improves Generalization in the Low SNR Regime View paper
- [34] DIRECTIONAL CONVERGENCE, BENIGN OVERFITTING View paper
- [35] Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models View paper
- [36] Universal scaling laws of absorbing phase transitions in artificial deep neural networks View paper
- [37] Benign overfitting of non-smooth neural networks beyond lazy training View paper
- [38] Benign overfitting in adversarially robust linear classification View paper
- [39] Understanding generalization in transformers: Error bounds and training dynamics under benign and harmful overfitting View paper
- [40] Benign, tempered, or catastrophic: Toward a refined taxonomy of overfitting View paper
- [41] Benign overfitting in multiclass classification: All roads lead to interpolation View paper
- [42] Binary Classification of Gaussian Mixtures: Abundance of Support Vectors, Benign Overfitting, and Regularization View paper
- [43] Topological obstruction to the training of shallow ReLU neural networks View paper
- [44] Gradient descent on two-layer nets: Margin maximization and simplicity bias View paper
- [45] Feature selection and low test error in shallow low-rotation relu networks View paper
- [46] Towards understanding learning in neural networks with linear teachers View paper
- [47] SGD Learns Over-parameterized Networks that Provably Generalize on Linearly Separable Data View paper
- [48] Learning a neuron by a shallow relu network: Dynamics and implicit bias for correlated inputs View paper
- [49] Non-Singularity of the Gradient Descent map for Neural Networks with Piecewise Analytic Activations View paper
- [50] Training two-layer RELU networks with gradient descent is inconsistent View paper
- [51] The Implicit Bias of Minima Stability in Multivariate Shallow ReLU Networks View paper