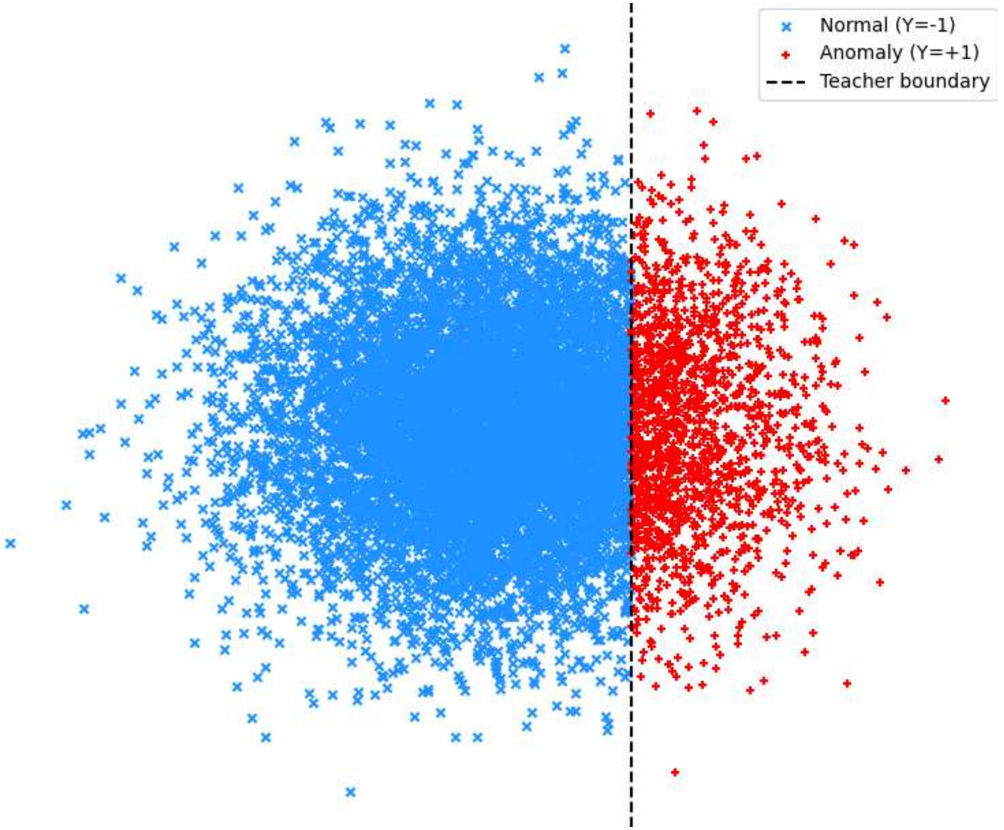


arXiv:2501.11638v2 [cs.LG] 5 Aug 2025

Class Imbalance in Anomaly Detection: Learning from an Exactly Solvable Model

F.S. Pezzicoli
TAU team, LISN
Université Paris-Saclay
CNRS, Inria
91405 Orsay, France
francesco.pezzicoli@universite-paris-saclay.fr

V. Ros
LPTMS
Université Paris Saclay
CNRS
91405 Orsay, France
valentina.ros@cnrs.fr

F.P. Landes
TAU team, LISN
Université Paris-Saclay
CNRS, Inria
91405 Orsay, France
francois.landes@inria.fr

M. Baity-Jesi
SIAM Department
Eawag (ETH)
8600 Dübendorf, Switzerland
marco.baityjesi@eawag.ch

Abstract

Class imbalance (CI) is a longstanding problem in machine learning, slowing down training and reducing performances. Although empirical remedies exist, it is often unclear which ones work best and when, due to the lack of an overarching theory. We address a common case of imbalance, that of anomaly (or outlier) detection. We provide a theoretical framework to analyze, interpret and address CI. It is based on an exact solution of the teacher-student perceptron model, through replica theory. Within this framework, one can distinguish several sources of CI: either intrinsic, train or test imbalance. Our analysis reveals that, depending on the specific problem setting, one source or another might dominate. We further find that the optimal train imbalance is generally different from 50%, with a non trivial dependence on the intrinsic imbalance, the abundance of data and on the noise in the learning. Moreover, there is a crossover between a small noise training regime where results are independent of the noise level to a high noise regime where performances quickly degrade with noise. Our results challenge some of the conventional wisdom on CI and pave the way for integrated approaches to the topic.

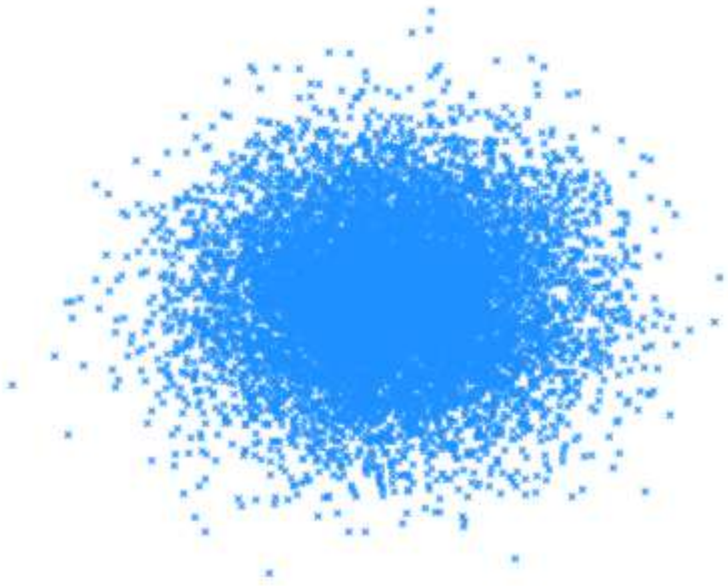
1 INTRODUCTION

Supervised learning under class imbalance (CI) is a fundamental challenge in modern machine learning, as many real-world datasets often exhibit varying degrees of imbalance [Yamanishi et al., 2000, Almeida et al., 2011, Kyathanahally et al., 2021, Schär et al., 2023]. Efforts to mitigate the detrimental effects of class imbalance have led to the development of various approaches, with the machine learning community establishing widely accepted heuristics based on empirical evaluations. These approaches can be broadly categorized into three types: those acting on the data distribution [Japkowicz and Stephen, 2002, Chawla et al., 2002, Ando and Huang, 2017, Zhang and Mani, 2003], those modifying the loss function [Xie and Manski, 1989, Kini et al., 2021, Behnia et al., 2023, Thrampoulidis et al., 2022, Menon et al., 2021], and those biasing the dynamics of the training process [Anand et al., 1993, Tang et al., 2020]. However, due to the lack of a theoretical framework for the analysis of CI, it is often unclear which of these methods work best and when or why. For this reason, recent studies tried to fill this theoretical gap, either by focusing on how imbalance influences the learning dynamics [Ye et al., 2021, Francazi et al., 2023, Kunstner et al., 2024], or how it influences the loss landscape [Mignacco et al., 2020, Loffredo et al., 2024, Mannelli et al., 2023].

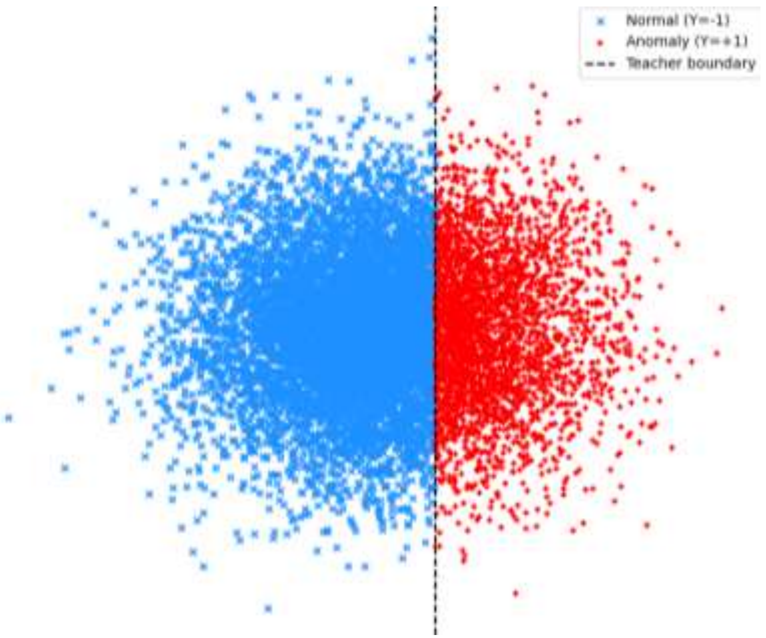
In these works, CI is treated as a single phenomenon, which can be addressed through a single formal approach. We highlight, instead, that one should distinguish between (at least) two types of imbalance. What we call *Multiple Groups* imbalance (MGI) involves samples drawn from distinctly different distributions, with the imbalance arising either from the sampling process (e.g. the toxicity of certain chemicals is tested more often than others [Schär et al., 2023]) or being intrinsic to the data itself (e.g. some species

Pezzicoli, F.S., Ros, V., Landes, F.P., Baity-Jesi, M.: Class imbalance in anomaly detection: Learning from an exactly solvable model (2025), https://arxiv.org/abs/2501.11638

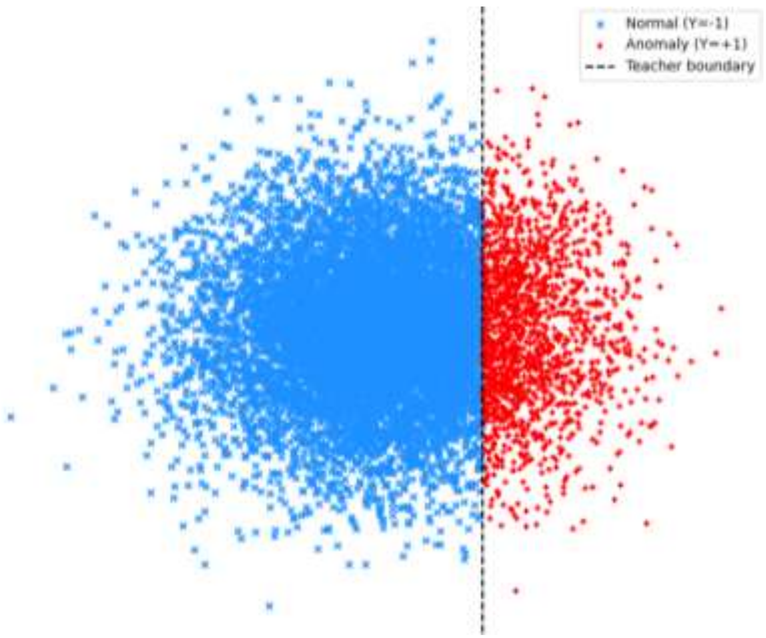
REVISITING OPTIMAL CLASS RATIOS IN IMBALANCED LEARNING



Gaussian data (unlabeled)



Gaussian data, labeled with teacher (bias = -0.5)



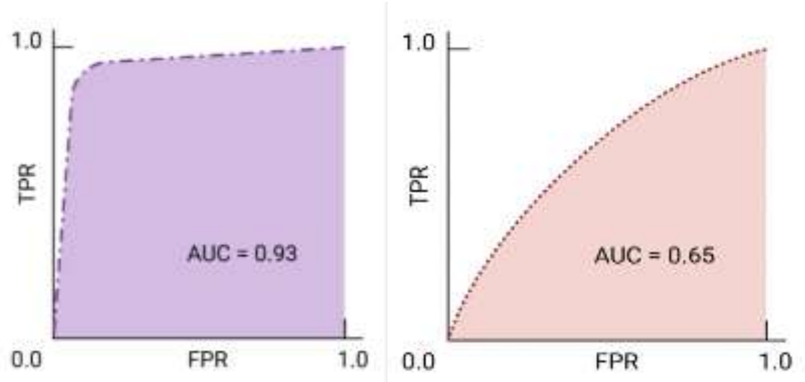
Gaussian data, labeled with teacher (bias = -1.0)

DATA

$$f_{\mathrm{Perceptron}}(X_n, Y_n) = \max(0, -Y_n(w \cdot X_n + b))$$

$$f_{\mathrm{Hinge}}(X_n, Y_n) = \max(0, 1 - Y_n(w \cdot X_n + b))$$

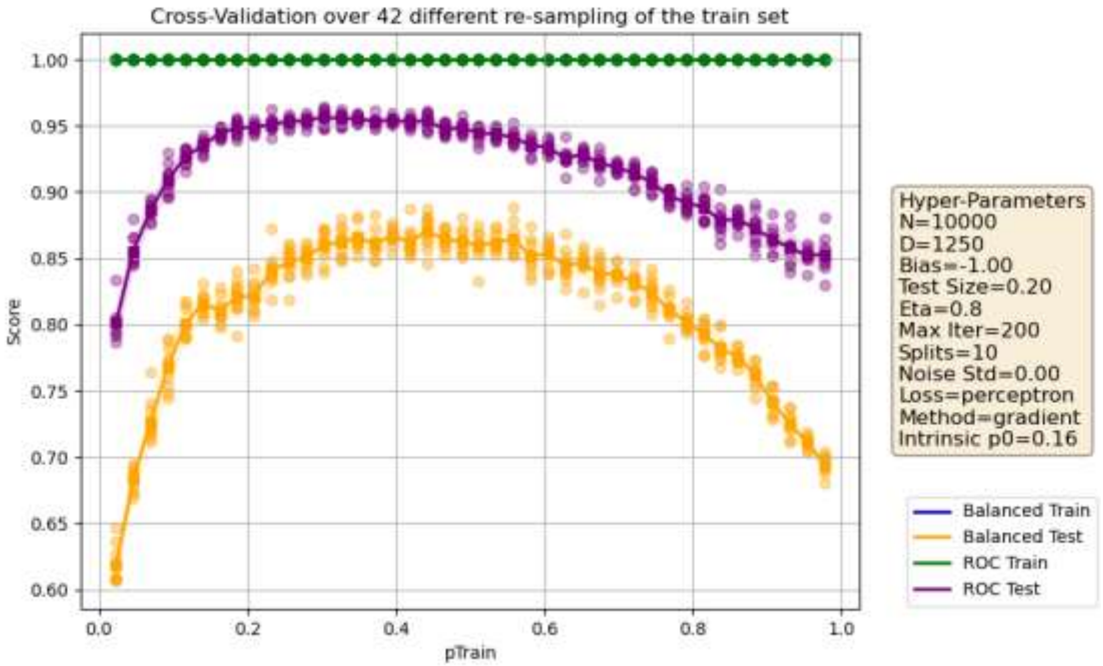
$$f_{\mathrm{EC}}(X_n, Y_n) = \frac{1}{2}\left(\mathrm{sign}(w \cdot X_n + b) - Y_n\right)^2$$



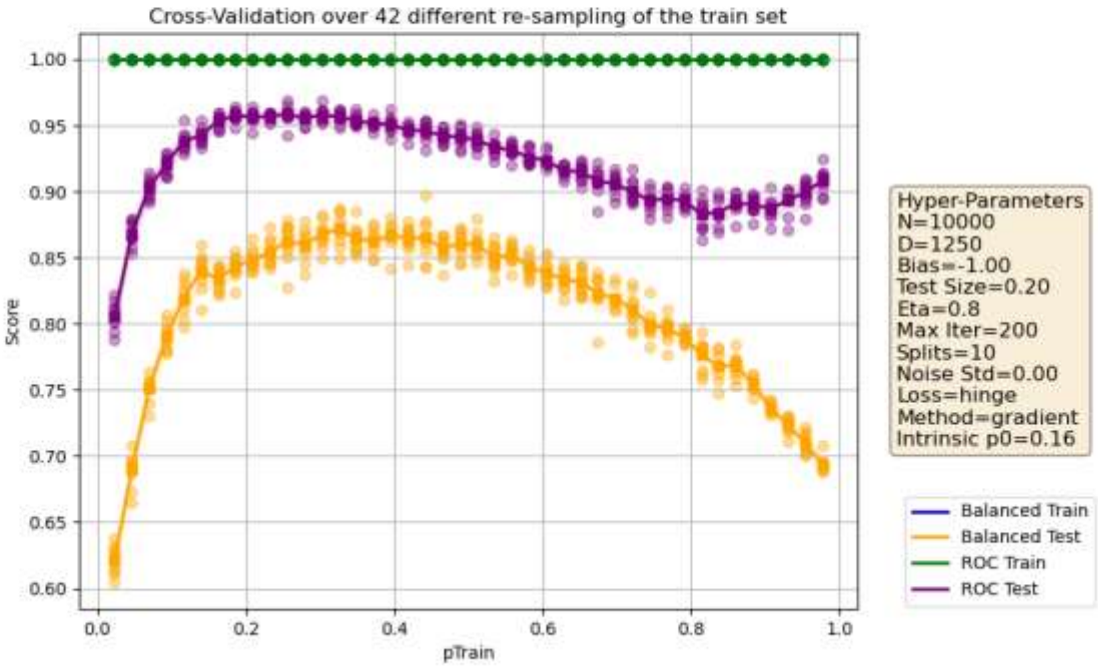
ρ_{train} ρ^*_{train} ρ_0

Borysenko, O., Byshkin, M.: Coolmomentum: A method for stochastic optimization by langevin dynamics with simulated annealing. Scientific Reports 11(1), 10705 (2021)

METHODS

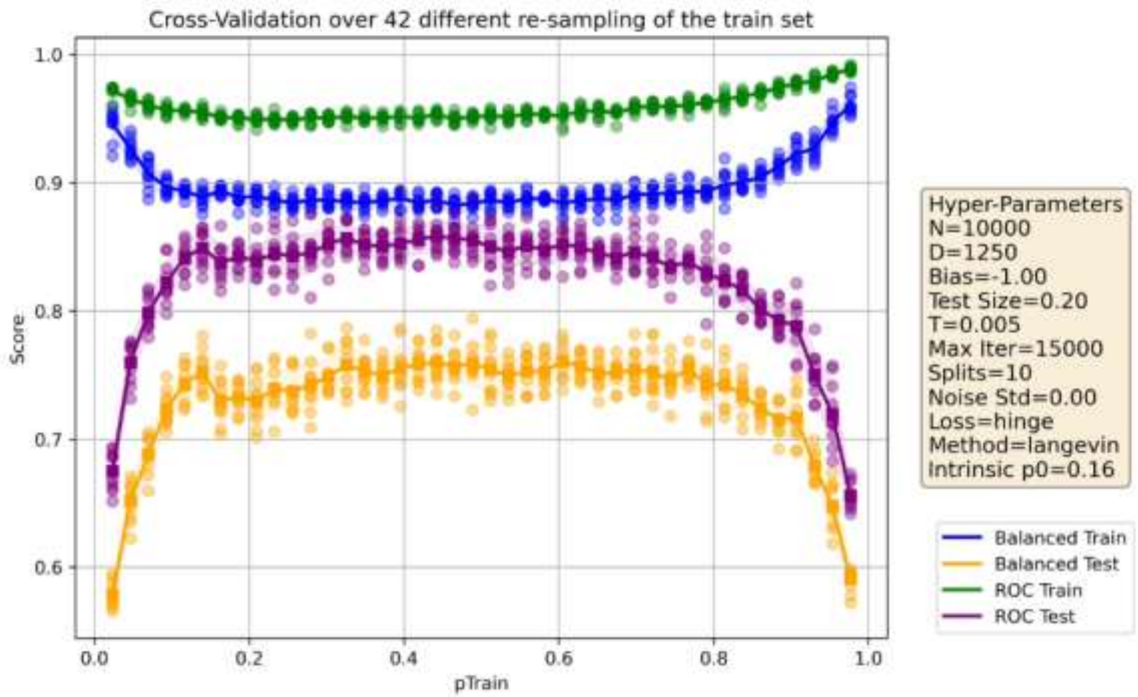


Gradient Descent using the Perceptron loss

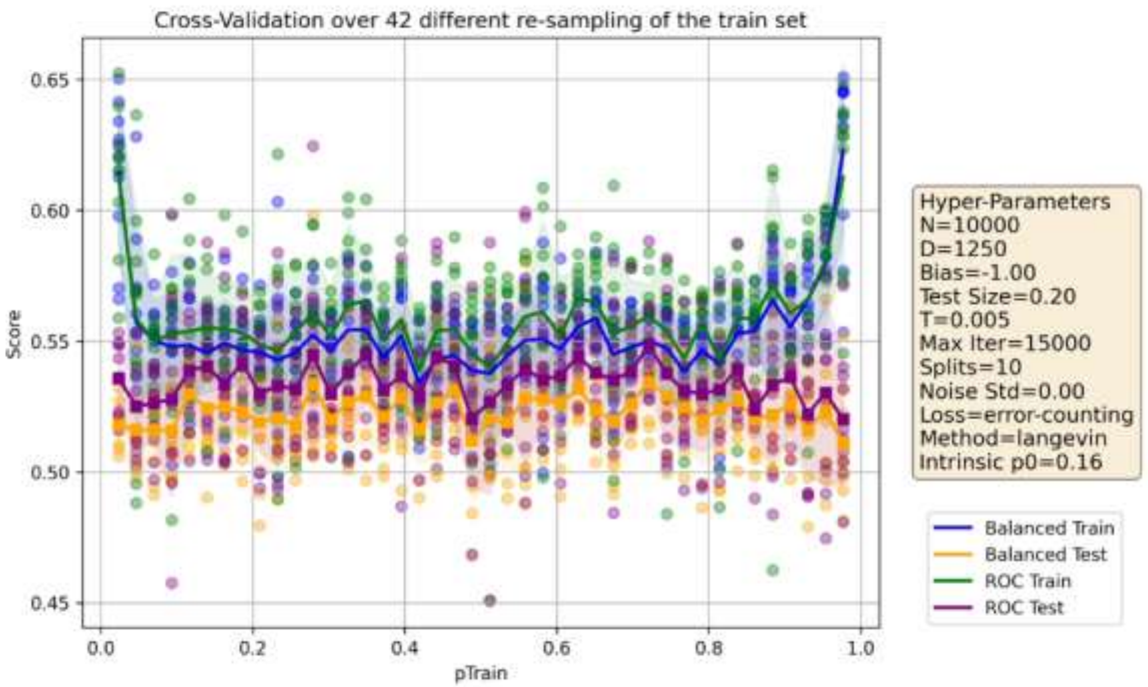


Gradient Descent using the Hinge loss

RESULTS



Langevin Dynamics using the Hinge loss



Langevin Dynamics using the Error-Counting loss

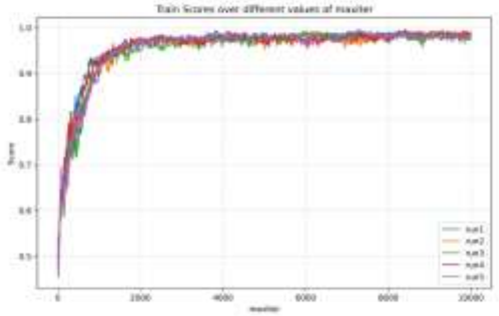
RESULTS

OPTIMAL IMBALANCE RATIO ρ^*_{train} IN THE TRAIN SET:

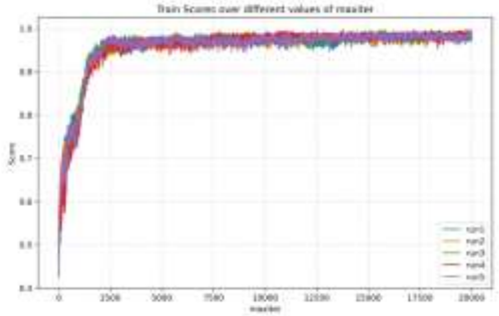
~~0.5~~

TUNABLE HYPERPARAMETER

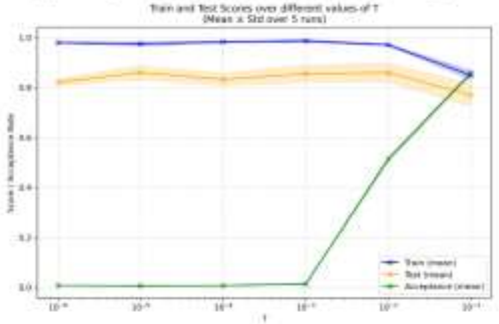
CONCLUSION



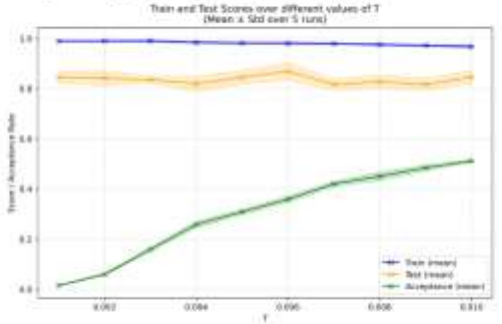
(a) Evolution of the train score across 5 runs during Langevin dynamics using the Error-Counting loss



(b) Evolution of the train score across 5 runs during Langevin dynamics using the Hinge loss



(c) Train and test performance (mean ± std over 5 runs) for Langevin Dynamics for 6 different temperature values $T \in [10^{-6}, 10^{-1}]$, using the Hinge loss, showing also the acceptance rate across runs



(d) Train and test performance (mean ± std over 5 runs) for Langevin Dynamics for 10 different temperature values $T \in [10^{-3}, 10^{-2}]$, using the Hinge loss, showing also the acceptance rate across runs

Hyperparameters	GD	Langevin
maxiter	200	15,000
Learning rate	0.8	/
T	/	0.05 -> 0.005*

*0.05 for the first 1,500 iterations, 0.005 then

$$\exp(-\frac{\Delta E}{T})$$

ADDITIONAL INFOS