Revisiting Optimal Class Ratios in Imbalanced Learning

Baptiste Pras

Université Paris-Saclay

Abstract. Class imbalance is a common challenge in classification tasks, particularly in anomaly detection, where one class may be extremely underrepresented. While it is often assumed that a 50-50 class ratio in training is optimal, recent theoretical work challenges this assumption, showing that the optimal class ratio can differ from 0.5 in simplified settings.
In this project, we examine whether this conclusion holds under more practical conditions. We replace the non-standard loss used in the theoretical model with standard ones (e.g., Hinge, Perceptron), and train models using Gradient Descent and Langevin dynamics on synthetic Gaussian data.
Our findings confirm that the optimal training class ratio deviates from 0.5 even with standard loss functions, reinforcing the robustness of earlier theoretical predictions.

Keywords: Machine Learning · Class Imbalance · Anomaly Detection

1 Motivation

This work explores the impact of class imbalance in the training set on classification performance. It builds on the recent paper *Class Imbalance in Anomaly Detection: Learning from an Exactly Solvable Model* [2], which shows that the optimal training class ratio ρ^*_{train} is not necessarily 0.5, contrary to standard intuition.

The original model has two limitations: it assumes perfectly Gaussian input distributions and uses a non-standard loss function that cannot be optimized via gradient descent. This project addresses the second limitation by replacing the paper's non-trainable loss (an Error-Counting loss) with standard alternatives such as the *Hinge* and *Perceptron* losses. We then analyze the learning dynamics under both Gradient Descent (GD) and Langevin dynamics [1].

The central question is: does the theoretical result $\rho^*_{\text{train}} \neq 0.5$ persist when standard, differentiable loss functions are used?

We show that while the choice of loss does affect the optimization landscape, the core result remains valid: the optimal class ratio is typically not 0.5, even with widely-used loss functions.

2 Methods: Gaussian Data Setup

We study the effect of loss functions on class-imbalanced learning using a synthetic dataset. Inputs $X \in \mathbb{R}^D$ are sampled from a standard Gaussian distribution and labeled by a fixed linear teacher with a negative bias to induce class imbalance. We generate $N = \alpha D$ samples with $\alpha = 8$, ensuring a favorable sample-to-dimension ratio. Let $X \sim \mathcal{N}(0, I_D)$.

We denote by ρ_{train} the proportion of positive samples in the training set, and by ρ^*_{train} the optimal value maximizing balanced accuracy on the original test set. The intrinsic class ratio ρ_0 corresponds to the natural balance before any resampling. We generate 42 datasets with $\rho_{\text{train}} \in (0, 1)$ via under- and over-sampling while keeping N fixed.

We compare three loss functions:

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

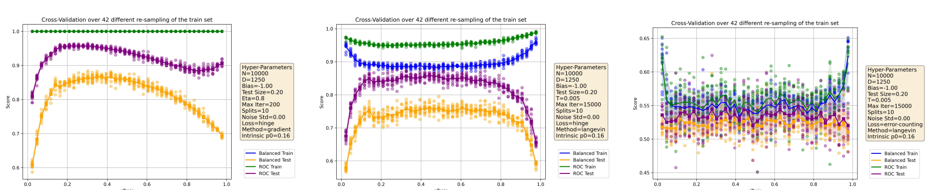
$$f_{\text{Perceptron}}(X_n, Y_n) = \max(0, -Y_n(w \cdot X_n + b)) \tag{2}$$

$$f_{\text{EC}}(X_n, Y_n) = \frac{1}{2}\left(\text{sign}(w \cdot X_n + b) - Y_n\right)^2 \tag{3}$$

We minimize the empirical loss $E(w, b) = \sum_{n=1}^{N} f(X_n, Y_n)$ using either GD or Langevin dynamics. The Error-Counting loss is only used with Langevin, as it is non-differentiable. All hyperparameters were tuned for optimal performance. For GD, we use a learning rate of $\eta = 0.8$ and run for 150 iterations. For Langevin dynamics, we apply a temperature schedule with $T = 0.05$ for the first 1500 iterations, then $T = 0.005$ for the remaining 13 500 iterations (15 000 steps in total).

3 Results

We compare the Error-Counting loss (Eq. 3) and the Hinge loss (Eq. 1), trained using GD or Langevin dynamics. Figure 1 presents the main results.



(a) Gradient Descent using the Hinge loss

(b) Langevin dynamics using the Hinge loss

(c) Langevin dynamics using the Error-Counting loss

Fig. 1: Comparison of optimization methods and loss functions on synthetic Gaussian data, with 10-fold cross-validation (mean ± std) with N = 10000 and D = 1250, using the balanced and ROC accuracy

Our main findings are as follows:

- With the Hinge loss, GD produces a clear optimum at $\rho^*_{\text{train}} < 0.5$, supporting the original theoretical result. The deviation from 0.5 is robust and consistent.
- Langevin dynamics with the Hinge loss yields a flatter curve, though a slight asymmetry persists, indicating reduced but non-negligible sensitivity to imbalance.
- The Error-Counting loss leads to irregular and noisy results under Langevin, with no clear optimum. Efforts to reproduce the smooth behavior reported in the original paper were unsuccessful.
- Overall, the Hinge loss outperforms the Error-Counting loss in terms of balanced and ROC accuracy. Similar results were observed with the Perceptron loss, confirming the robustness of margin-based approaches.

4 Conclusion

We have extended prior theoretical work on class-imbalanced learning by evaluating standard loss functions under practical optimization regimes. Our results show that both Hinge and Perceptron losses recover an optimal training ratio $\rho^*_{\text{train}} \neq 0.5$, especially when trained with GD. While Langevin dynamics flattens the loss landscape, it does not eliminate the effect. Margin-based losses also outperform the non-differentiable Error-Counting loss.

Acknowledgments

This work was conducted as part of a research internship during my third year of undergraduate studies, under the supervision of François Landes, whom I thank for his guidance and feedback throughout the project. I also thank Martin Leiva and Vladimir Herrera Nativi, with whom I collaborated closely. Their complementary work focused on extending this study to real-world datasets beyond the Gaussian setting explored here.

The source code of all our experiments is available at: https://gitlab.inria.fr/flandes/cilossdada

References

1. Borysenko, O., Byshkin, M.: Coolmomentum: A method for stochastic optimization by langevin dynamics with simulated annealing. Scientific Reports **11**(1), 10705 (2021)
2. 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