Learning Imbalanced Data with Beneficial Label Noise

Guangzheng Hu, Feng Liu, Mingming Gong, Guanghui Wang, Liuhua Peng

THE UNIVERSITY OF MELBOURNE

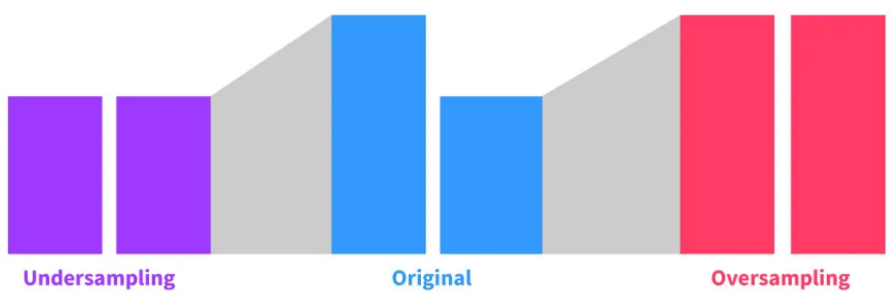
Fact: Class Imbalance biases the Decision Boundary

- Fraud transaction detection, Rare disease diagnosis

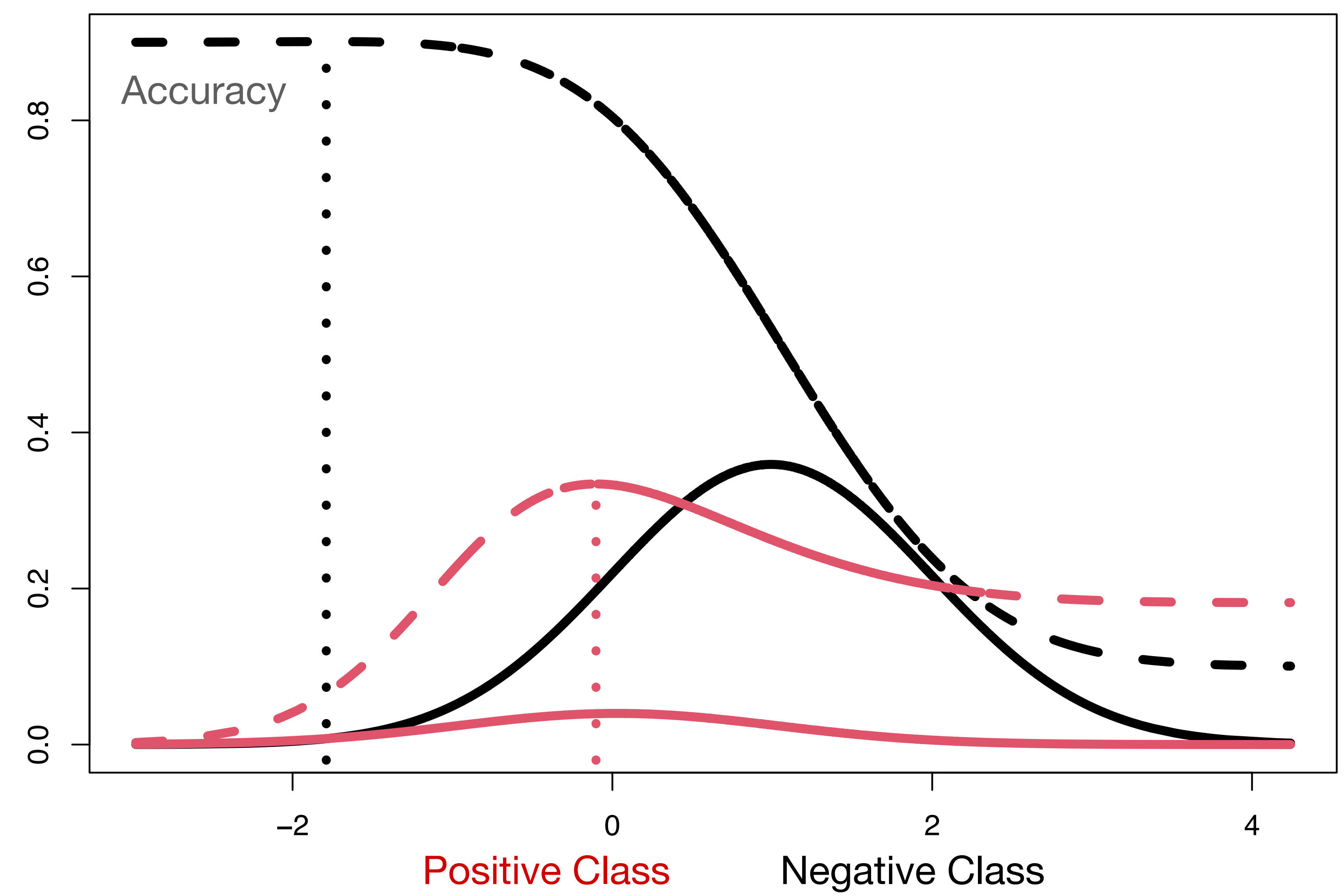
- Traditional classifier tends to classify all samples into negative class (majority class) while maximizing accuracy on imbalanced data.

$$S^{\mathrm{Acc}} = \left\{ x^* \in \mathcal{X} : \frac{P_1(x^*)}{P_0(x^*)} = \frac{\pi_0}{\pi_1} \right\} \qquad S^{\mathrm{F1}} = \left\{ x^* \in \mathcal{X} : \frac{P_1(x^*)}{P_0(x^*)} = \frac{\mathcal{F}1(x^*)}{2 - \mathcal{F}1(x^*)} \frac{\pi_0}{\pi_1} \right\}$$

$$S^{\mathrm{Resample}} = \left\{ x^* \in \mathcal{X} : \frac{P_1(x^*)}{P_0(x^*)} = \alpha \frac{\pi_0}{\pi_1} \right\}$$



Information Loss
Generative Error



Label-Noise-based Rebalancing Approach

Fact: Label Noise also biases the decision boundaries

$$\rho(x) = Pr(Y^* = 1 \,|\, X = x, Y = 0) \quad \propto \eta(x)$$

$$\gamma(x) = Pr(Y^* = 0 \,|\, X = x, Y = 1) \quad = 0$$

$$S^* = \left\{ x^* \in \mathscr{X} : \frac{P_1(x^*)}{P_0(x^*)} = [1 - 2\rho(x^*)]\frac{\pi_0}{\pi_1} \right\}$$

- **Intuition**: Only the majority class samples with the most similar features to the minority class are likely to be flipped.

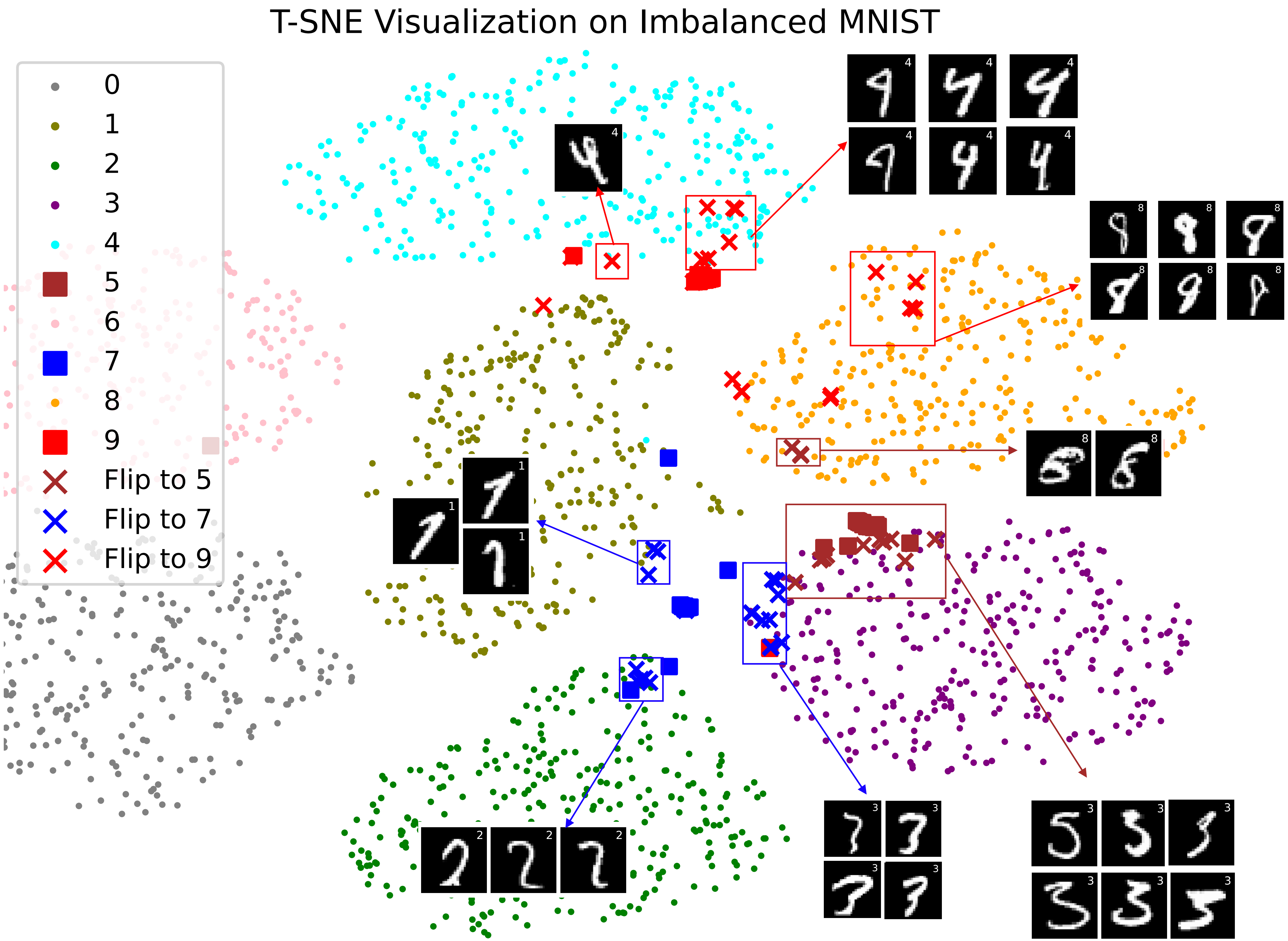
- **Advantages**: a) Minimum data-editing, b) Model/data-agnostic

- **3-Steps**: Z-score Standardlization, tanh Normalization, Label Flipping

$$\mathcal{Z}[i_{\text{majority}}] \leftarrow \frac{\hat{\eta}[i_{\text{majority}}] - \mu}{\sigma}$$

$$\rho\,[i_{\text{majority}}] \leftarrow max(\mathtt{tanh}(\mathcal{Z}[i_{\text{majority}}] - t_{flip}), 0)$$

$$\text{U} \leftarrow \text{Bernoulli}(\rho[i_{\text{majority}}])$$



- Minority classes have only 30 samples: “5”, “7”, “9”. Each majority class has 6000 samples.
- Improves the overall accuracy on balanced test set from 89.93% to 94.75%.
- Flipped samples shares similar features with the minority classes.
- Similar features are occupied by majority.
- Biased decision boundary corrected by only 65 label noises.

Experiment Results — Image Classification

	Step-wise Cifar-10		Step-wise Cifar-100	
	Acc_{MI}	$\text{Acc}_{\text{overall}}$	Acc_{MI}	$\text{Acc}_{\text{overall}}$
LDAM	66.41±0.2	77.47±0.06	19.80±0.02	45.23±0.03
LDAM+RSG	67.02±0.07	77.74±0.08	21.67±0.04	45.51±0.02
LDAM+LNR	**75.06±0.09**	**78.12±0.03**	**25.84±0.06**	**45.63±0.02**
GCL	56.78±0.08	74.80±0.04	5.48 ±0.03	43.87±0.07
GCL+LNR	**72.22±0.05**	**80.8±0.02**	**26.48±0.03**	**46.20±0.03**

	Long-tailed Cifar-10			
	Many-shot	**Medium-shot**	**Few-shot**	**Overall**
LDAM	**82.62±0.06**	76.12±0.1	75.01±0.1	78.39±0.03
LDAM+RSG	81.56±0.15	**77.03±0.1**	77.30±0.1	78.93±0.02
LDAM+LNR	81.17±0.08	76.42±0.01	**79.83±0.1**	**79.34±0.01**
GCL	**88.60±0.04**	**79.57±0.01**	70.08±0.2	80.55±0.03
GCL+LNR	88.20±0.04	79.50±0.07	**77.60±0.2**	**82.41±0.03**
MiSLAS	**91.00±0.14**	80.17±0.22	75.72±0.19	82.10±0.12
MiSLAS+ReMix	90.04±0.20	79.82±0.16	79.78±0.20	82.92±0.10
MiSLAS+SelMix(10k)	86.81±0.22	80.50±0.17	83.52±0.21	83.29±0.07
MiSLAS+SelMix(1k)	81.61±0.14	79.89±0.13	**87.60±0.20**	82.72±0.22
MiSLAS+SelMix(imb)	82.21±0.13	81.44±0.11	81.9±0.21	81.8 ±0.09
MiSLAS+LNR	84.62±0.22	**80.90±0.22**	86.07±0.19	**83.56**±0.08

	Long-tailed Cifar-100			
	Many-shot	**Medium-shot**	**Few-shot**	**Overall**
LDAM	**62.21±0.05**	43.28±0.08	20.83±0.03	42.98±0.03
LDAM+RSG	60.46±0.05	**43.88±0.1**	22.57±0.09	43.08±0.07
LDAM+LNR	61.04±0.03	43.36±0.02	**24.11±0.04**	**43.58±0.02**
GCL	**67.16±0.03**	46.63±0.06	13.57±0.06	43.90±0.03
GCL+LNR	57.11±0.07	**51.38±0.07**	**25.02±0.09**	**45.48±0.05**
MiSLAS	**62.05±0.09**	48.42±0.11	26.07±0.12	46.85±0.09
MiSLAS+ReMix	59.06±0.21	49.22±0.09	27.93±0.10	46.59±0.15
MiSLAS+SelMix(10k)	60.93±0.12	**52.06±0.17**	25.10±0.13	47.43±0.10
MiSLAS+SelMix(1k)	61.27±0.08	50.82±0.18	21.34±0.12	46.04 ±0.11
MiSLAS+SelMix(imb)	56.66±0.12	51.17±0.06	25.31±0.21	45.65 ±0.23
MiSLAS+LNR	56.26±0.24	51.46±0.22	**35.34±0.21**	**48.52** ±0.12

- 99% of samples from minority classes are removed with the imbalance ratio = 100.
- 94 label noises delivers effective trade-off between the classification performance of the head class (yellow) and the tail class (green).

	0	1	2	3	4	5	6	7	8	9
0	**-26**	+2	-4	0	+1	-2	0	-1	**+13**	**+16**
1	**1**	**-4**	0	-1	+1	0	0	0	**0**	**+4**
2	5	+1	**-4**	-2	0	-7	+1	+3	-1	+3
3	-2	-1	+3	**+18**	-4	-13	-5	+2	0	+5
4	0	0	-2	-3	**+3**	-5	0	+5	+1	+1
5	2	0	-1	-3	0	**-1**	+1	+2	0	+2
6	1	+2	-2	+6	-1	-5	**-4**	+2	-1	+1
7	-5	0	+3	-20	-1	-8	+4	**+22**	+1	+5
8	**-91**	-6	+2	-2	-1	-1	-2	0	**+80**	+20
9	**-42**	**-48**	-3	-5	+1	-2	-3	-2	+3	**+102**

Table 3: Confusion matrix comparison of GCL and LNR. The signed values denote the changes made by LNR.

Experiment Results — KEEL Binary Classification

- 32 KEEL imbalanced datasets (tabular data) with range of imbalance ratios from 1.82 to 49.6.
- Compared with matured resampling methods with relative ranking on F1, G-mean, and AUC.
- LNR does not involve adding or removing samples, significantly enhancing F1 score and G-mean scores without compromising AUC performance.