A Second-Order Approach to Learning with Instance-Dependent Label Noise

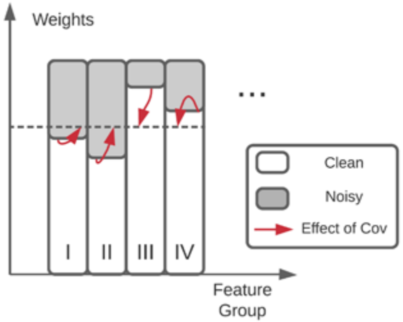
Zhaowei Zhu, Tongliang Liu, and Yang Liu

{zwzhu , yangliu}@ucsc.edu, tongliang.liu@sydney.edu.au

Code

REsponsible & Accountable Learning (REAL)
@ University of California, Santa Cruz

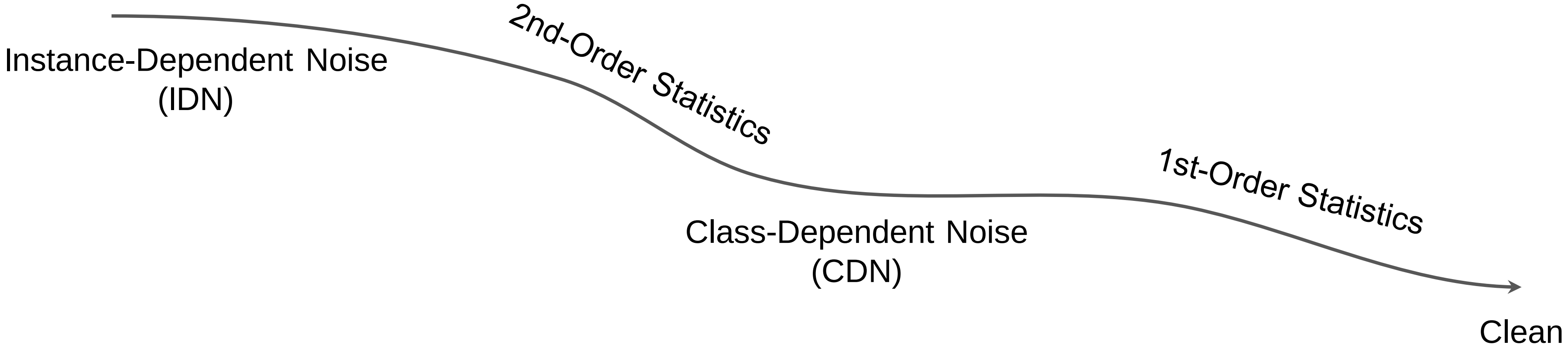
https://github.com/UCSC-REAL



UNIVERSITY OF CALIFORNIA SANTA CRUZ

THE UNIVERSITY OF SYDNEY

Covariance-Assisted Learning (CAL)

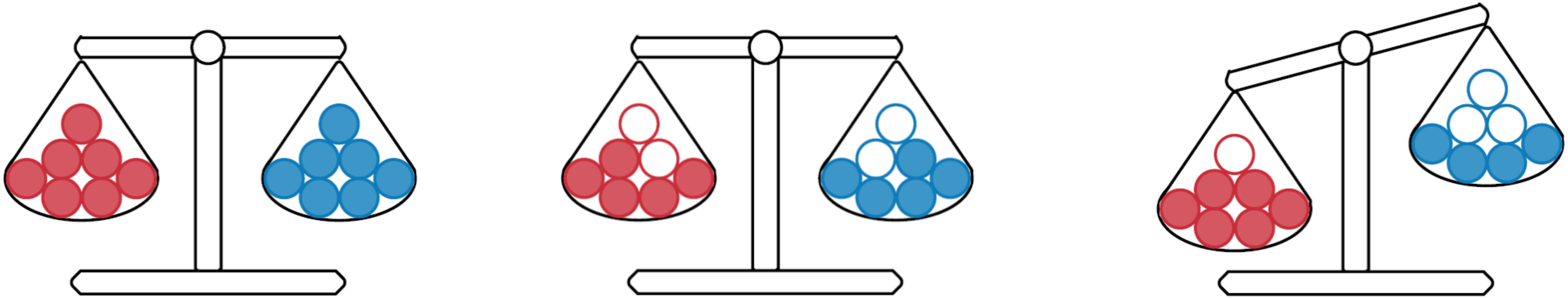


[1] N. Natarajan, et al. “Learning with noisy labels." *NeurIPS'13.*
[2] T. Liu & D. Tao. “Classification with noisy labels by importance reweighting." *TPAMI'15.*
[3] G. Patrini, et al. “Making deep neural networks robust to label noise: A loss correction approach." *CVPR'17.*

Motivation

- Two groups (not label classes) of instances with equal size

  - Empirical Risk Minimization (ERM) of instances from two groups: $\text{Loss} = \sum_{i \in \text{Group-1}} \text{Loss}_i + \sum_{j \in \text{Group-2}} \text{Loss}_j$



Clean — Class-dependent label noise — Instance (group)-dependent noise

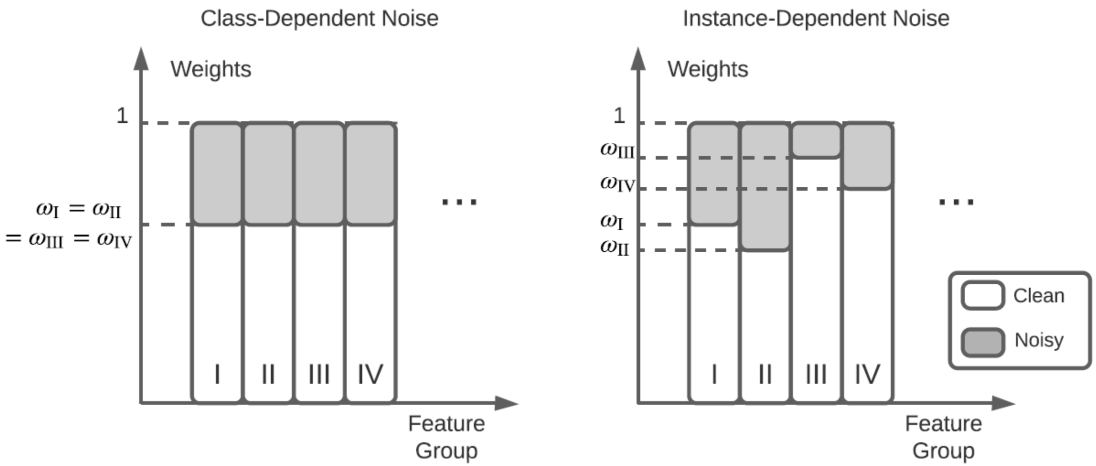
Clean: no noise
➜ equal #instances contribute to clean loss
➜ equal weights in ERM

CDN: equal noise
➜ equal #instances contribute to clean loss
➜ equal weights in ERM

IDN: Group 2: larger noise
➜ less #instances contribute to clean loss
➜ smaller weights in ERM

Insufficiency of First-Order Statistics

- **Lemma:** Peer Loss [4] is invariant to CDN: NoisyPL = $\omega \cdot$ CleanPL



Summary:

➥ IDN causes weights imbalances

➥ **CDN:**
- Only one unknown constant ω.
- Equal for all features.

➥ **IDN:**
- Multiple unknown constants ω_g.
- **Down-weight high-noise features** (Section 3.3 in our paper).

[4] Y. Liu & H. Guo. "Peer loss functions: Learning from noisy labels without knowing noise." *ICML'20.*

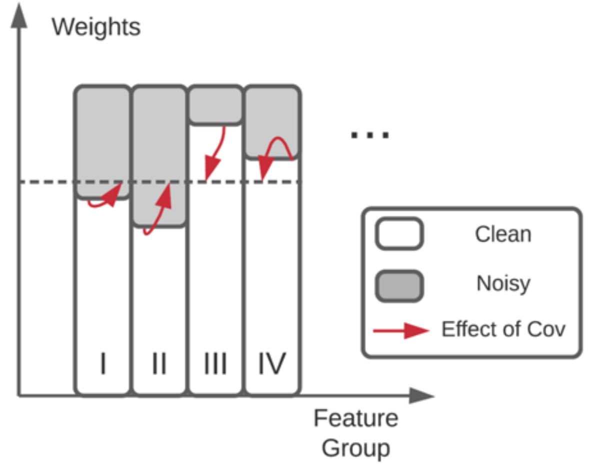
Covariance-Assisted Learning (CAL)

✨ **Challenging!**
(Details in the next slide)

- Our method:

  - Peer Loss + Covariance (requires constructing **Bayes optimal dataset** for estimating *T*):

$$\ell_{\text{CAL}}(f(x_n), \tilde{y}_n) = \ell_{\text{PL}}(f(x_n), \tilde{y}_n) - \text{Cov}(\text{Noise Trans. } T, \text{Model Pred.})$$



Summary:

➡ CAL balances weights of each feature
- High-noise (Group I, Group II): improve weights
- Low-noise (Group III, Group IV): reduce weights

➡ IDN —Covariance→ CDN —Peer Loss→ Clean

➡ Theorem 3 (in our paper):

With perfect covariance estimates, **CAL is robust to IDN**

UC SANTA CRUZ

Bayes Optimal Labels

Algorithm (Sketch)

1. Construct $\hat{D}$ (unbiased estimate of $D^* \sim \mathcal{D}^*$) with sample sieve [5]
2. Estimate (unbiased) $\hat{T}$ with $\hat{D}$ (complexity O(SampleSize))
3. [Train DNN] Implement CAL in SGD (each point $O(1)$ complexity)

- Rely on **Bayes optimal** labels
  - Unique
  - Tractable

Example:

Type	Prob. Each Class	
Clean	0.9	0.1
Noisy	0.6	0.4
Bayes opt.	1.0	0.0

Use **CORES [5]**:
A theoretically guaranteed sample sieve to find the Bayes optimal labels!

[5] H. Cheng, et al. “Learning with instance-dependent label noise: A sample sieve approach." *ICLR'21.*

Experiment

Table: Comparison of test accuracies ($\%$) using different methods.

Method	*Inst. CIFAR10*			*Inst. CIFAR100*		
	$\eta = 0.2$	$\eta = 0.4$	$\eta = 0.6$	$\eta = 0.2$	$\eta = 0.4$	$\eta = 0.6$
CE (Standard)	85.45±0.57	76.23±1.54	59.75±1.30	57.79±1.25	41.15±0.83	25.68±1.55
Forward T [2]	87.22±1.60	79.37±2.72	66.56±4.90	58.19±1.37	42.80±1.01	27.91±3.35
T-Revision [3]	90.04±0.46	84.11±2.47	72.18±2.47	58.00±0.36	43.83±8.42	36.07±9.73
Peer Loss [4]	89.12±0.76	83.26±0.42	74.53±1.22	61.16±0.64	47.23±1.23	31.71±2.06
CORES^2 [5]	91.14±0.46	83.67±1.29	77.68±2.24	66.47±0.45	58.99±1.49	38.55±3.25
CAL	**92.01±0.75**	**84.96±1.25**	**79.82±2.56**	**69.11±0.46**	**63.17±1.40**	**43.58±3.30**

Thank you !

[5] H. Cheng, et al. "Learning with instance-dependent label noise: A sample sieve approach." *ICLR'21.*

UC SANTA CRUZ