Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto* C. Cousins* D. Sam S. Bach E. Upfal

AMCL - A First Look

Adversarial MultiClass Learning

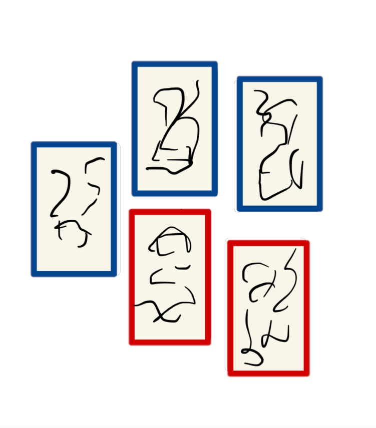
- Semi-supervised framework for **multiclass** learning from *weak labelers*
- Adversarial *feasible labeling* of unlabeled data during training
  - Feasibility: use labeled data to compute statistical constraints on weak labelers

Contribution 1st semi-supervised learner for *arbitrarily correlated* weak labelers with:
- *optimization convergence guarantees* for the adversarial multiclass learning
- *generalization bound* for the learned model

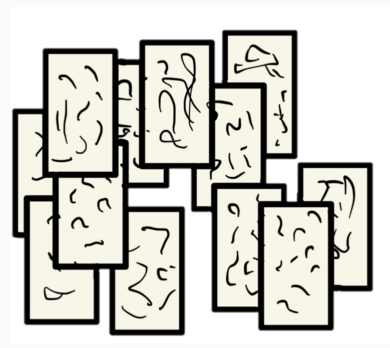
The Problem Setting

Classification task: distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$

Access to:



labeled data $X_L, Y_L \sim (\mathcal{D})^{m_L}$

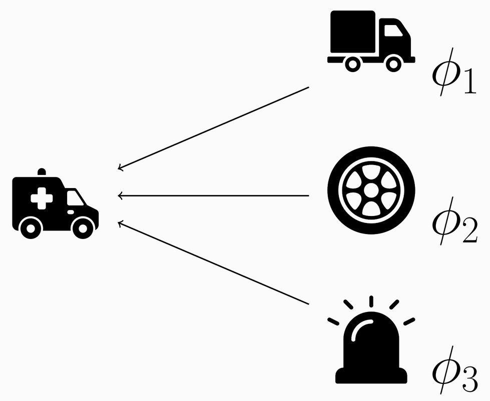


unlabeled data $X_U \sim (\mathcal{D}_\mathcal{X})^{m_U}$

$\phi_1, \ldots, \phi_n$

weak labelers

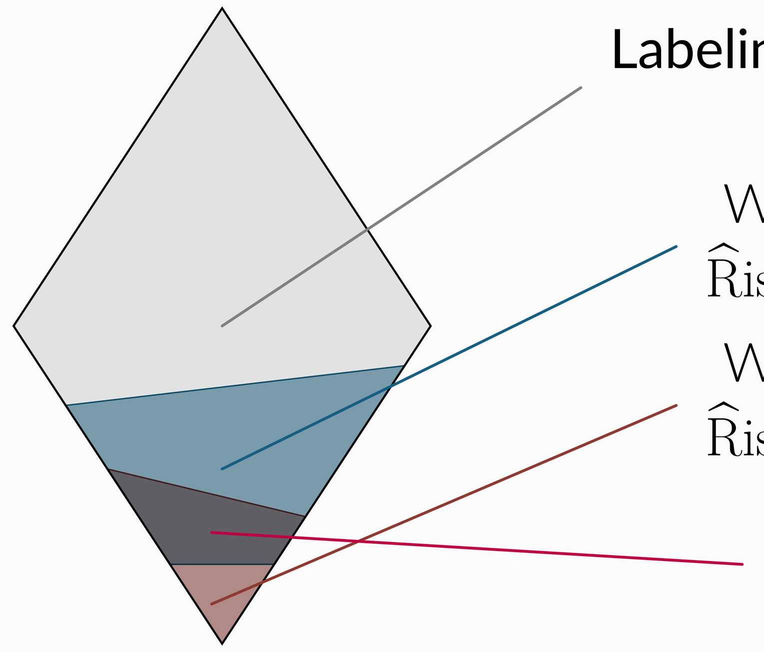
Weak labelers: classifiers for mildly related task



ϕ: blue color vs. green color

Images made by Those Icons (wheel, siren), fjstudio (ambulance), Freepik (truck) from flaticon.com

Feasibility



Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$: $\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

Weak Labeler ϕ_2 **Risk Constraint**: $\widehat{\text{Risk}}(\phi_2; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \widehat{\text{Risk}}(\phi_2; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Weak Labeler ϕ_1 **Risk Constraint**: $\widehat{\text{Risk}}(\phi_1; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \widehat{\text{Risk}}(\phi_1; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Feasible Set
Intersection of all constraints

The Optimization Objective

Adversarial setting $\Longrightarrow$ Minimax Objective
- Empirical risk of a model $\boldsymbol{\theta}$ computed w.r.t. *adversarial feasible* labeling

$$\underbrace{\min_{\boldsymbol{\theta}\in\Theta}}_{\text{Subgradient Steps}} \overbrace{\max_{\text{feasible } \boldsymbol{Y}} \widehat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})}^{\text{Linear Programming}}$$

- Y is soft-labels: we use expected loss over label distributions
  $\Longrightarrow$ empirical risk is linear w.r.t. Y

Optimization Guarantees

If the *loss* is convex and L-Lipschitz Continous w.r.t. $\boldsymbol{\theta}$, and we run $T = \Omega(L^2/\varepsilon^2)$ *iterations* of the subgradient method using step size $\alpha = \varepsilon/L^2$:

$$\underbrace{\max_{\text{feasible } \boldsymbol{Y}} \widehat{\text{Risk}}(h_{\hat{\boldsymbol{\theta}}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Subgradient solution}} \le \underbrace{\min_{\boldsymbol{\theta}} \max_{\text{feasible } \boldsymbol{Y}} \widehat{\text{Risk}}(h_{\boldsymbol{\theta}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Minimax optimal solution}} + \varepsilon$$

Generalization Bound

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \max_{\text{feasible } \boldsymbol{Y}} \widehat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y}) \text{ vs } \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \text{Risk}(h_{\boldsymbol{\theta}})$$

If *loss codomain* is $[0, B]$, then w.h.p., *true labeling* is feasible, and

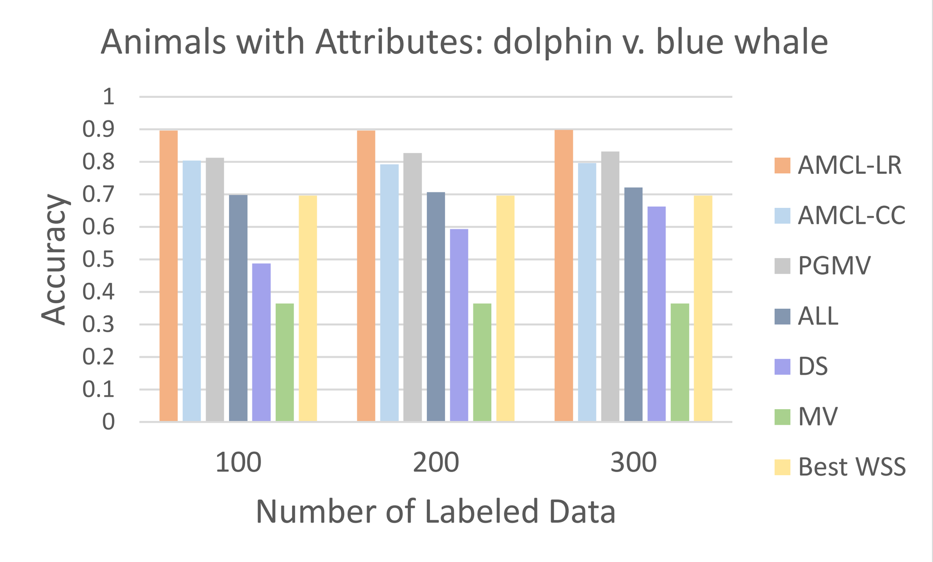
$$\underbrace{\text{Risk}(h_{\hat{\boldsymbol{\theta}}}) - \text{Risk}(h_{\boldsymbol{\theta}^*})}_{\text{Optimality Gap}} \le \underbrace{B \cdot D_f}_{\text{Adversarial Error}} + \underbrace{\sup_{\text{feasible } \boldsymbol{Y}} 4\hat{\mathfrak{R}}_{m_U}(\mathcal{L}; \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Uniform Convergence Bound}} + \underbrace{O\left(B\sqrt{\frac{\ln\frac{1}{\delta}}{m_U}}\right)}_{\text{Tail Bound}}$$

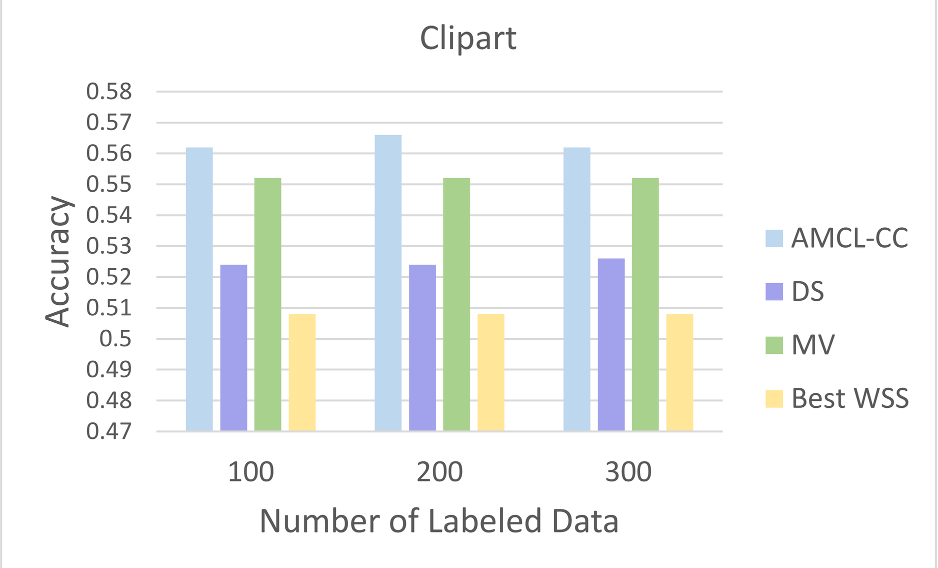
Average diameter of the feasible set
Geometrically quantifies weak labeler information

$$D_f = \frac{1}{m_U} \sup_{\text{feasible } \boldsymbol{Y}', \boldsymbol{Y}''} \sum_{j=1}^{m_U} \|\boldsymbol{y}'_j - \boldsymbol{y}''_j\|_1$$

Experiments

- We run experiments over two *image* datasets
  - Animals with Attributes (binary)
  - DomainNet (multiclass)
- We use two practical instantiations of our general framework:
  1. AMCL-CC: convex combination of the output of the weak labelers
  2. AMCL-LR: logistic regression (softMax) over images' features





Our method often outperforms the **baselines**, and state-of-the-art algorithms:
- **MV** (Majority Vote), **DS** (Dawid-Skene), **Best WSS** (Best weak labeler)
- ALL (Adversarial Label Learning), PGMV (Performance-Guaranteed Majority Vote)

alessio_mazzetto@brown.edu cyrus_cousins@brown.edu