On the Sample Complexity of Adversarial Multi-Source PAC Learning

Nikola Konstantinov, Elias Frantar, Dan Alistarh, Christoph H. Lampert

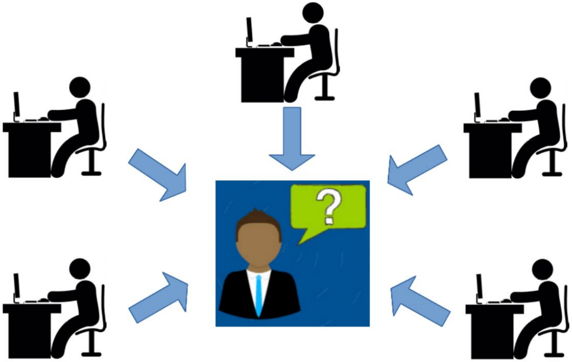
ICML, 2020

IST AUSTRIA
Institute of Science and Technology

Summary

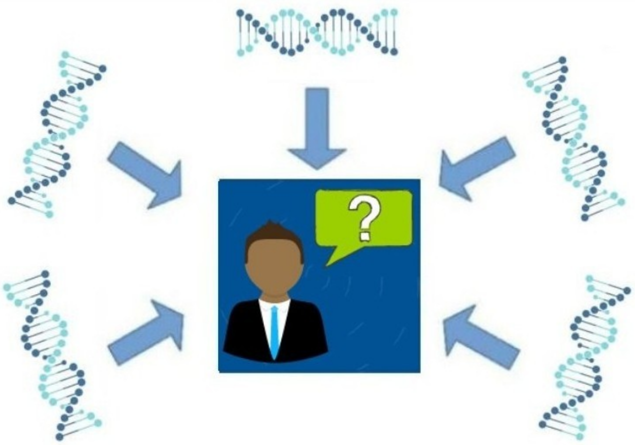
Learning from untrusted sources

Crowdsourcing



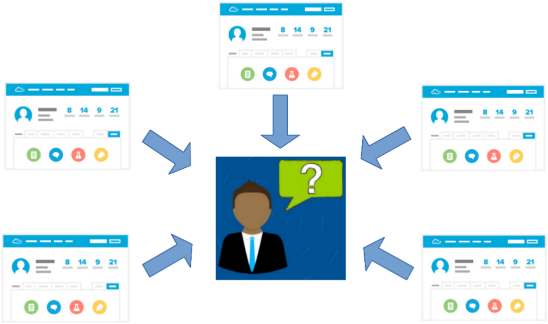
Learning from untrusted sources

Using data from multiple labs



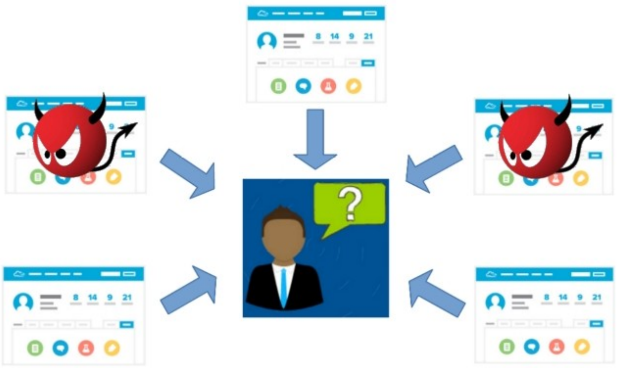
Learning from untrusted sources

Collecting data from online sources



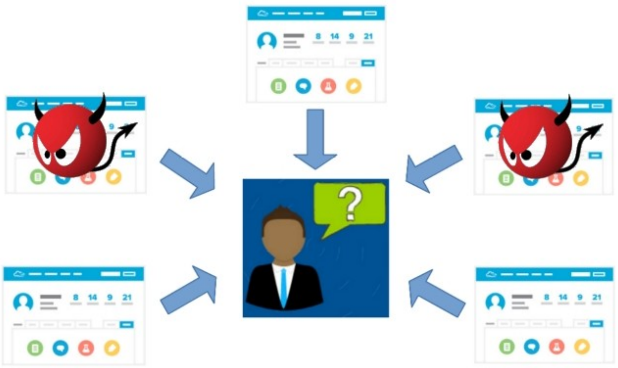
Learning from untrusted sources

Collecting data from online sources



Learning from untrusted sources

Collecting data from online sources



How much can be learnt even if some data is corrupted or manipulated?

Main contributions

- Rigorous adversarial models and statistical PAC-learnability framework
- Positive results:
  - PAC-learnability is fulfilled (under minimal assumptions)
  - Explicit learning algorithm and rates
- Hardness results:
  - Sample complexity lower bound
  - The learner needs the group structure to achieve PAC-learnability

Details

Setup

Supervised learning scenario

- Input-output space $\mathcal{X} \times \mathcal{Y}$, unknown data distribution $\mathcal{D} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$
- Hypothesis space $\mathcal{H}$, loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$
- Want to find $h \in \mathcal{H}$, such that $\mathcal{R}(h) = \mathbb{E}_{\mathcal{D}}(\ell(h(x), y))$ is small

Learning from multiple sources

- Given: a set of N datasets $S = (S_1, \dots, S_N)$
- m labeled points in each: $S_i = \{(x_{i,j}, y_{i,j})\}_{j=1}^{m} \overset{\text{iid}}{\sim} \mathcal{D}$

Adversarial model

Informal description

- An adversary controls an α-fraction of the sources, $\alpha < 1/2$
- The adversary can choose the new points with full knowledge of the setup
- The learner does not know which sources are manipulated

Formal definitions

- $(\mathcal{X} \times \mathcal{Y})^{N \times m}$ - set of all unordered sequences of N sets of m points
- A fixed-set adversary is *any function* $\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^{N \times m} \to (\mathcal{X} \times \mathcal{Y})^{N \times m}$, such that:

$$(S'_1, \dots, S'_N) = \mathcal{A}(S_1, \dots, S_N) \text{ satisfies } S'_i = S_i,$$

$\forall i \in C$, where C is the set of "clean" sources and $|C| = (1 - \alpha)N$

Adversarial PAC-learnability

- A multi-source learner is a function $\mathcal{L} : (\mathcal{X} \times \mathcal{Y})^{N \times m} \to \mathcal{H}$
- Focus on fixed N and α, while $m \to \infty$
- $\mathcal{H}$ is α-adversarially PAC-learnable if $\exists m : (0,1]^2 \to \mathbb{N}$, such that for any $\epsilon, \delta \in (0,1]$, whenever $m \geq m(\epsilon, \delta)$, with probability at least $1 - \delta$:

$$\mathcal{R}(\mathcal{L}(\mathcal{A}(S)) \leq \min_{h \in \mathcal{H}} \mathcal{R}(h) + \epsilon,$$

against *any (fixed-set) adversary of power* α

Related work

Learning discrete distributions from untrusted batches

- Unsupervised version of the problem studied in (Qiao and Valiant 2018; Jain et al. 2020)

Robust PAC learning from a single dataset

- One point per source recovers the malicious noise model (Kearns et al. 1993)
- PAC-learnability is known to be impossible: minimum possible error is $\alpha/(1-\alpha)$

Byzantine-robust distributed optimization

- Practical and robust gradient optimization methods (Yin et al. 2018; Alistarh et al. 2018)
- Convergence analysis under convexity/smoothness assumptions

Collaborative learning

- Multiple parties learn *one model each*
- Adversarial PAC-learnability provably possible (Blum et al. 2017; Qiao 2018)

Adversarial PAC-learnability

Main assumption: $\mathcal{H}$ is uniformly convergent

- Given m samples $S = \{(x_1, y_1), \dots, (x_m, y_m)\} \overset{iid}{\sim} \mathcal{D}$, with probability at least $1-\delta$ over the data :
$$\sup_{h\in\mathcal{H}} |\mathcal{R}(h) - \widehat{\mathcal{R}}(h)| \leq s_{\mathcal{H},\ell}(m, \delta, S),$$
- $s_{\mathcal{H},\ell}(m, \delta, S_m) \to 0$ as $m \to \infty$, for any sequence $\{S_m\}_{m\in\mathbb{N}}$ with $S_m \in (\mathcal{X} \times \mathcal{Y})^m$

Theorem

$\mathcal{H}$ *- uniformly convergent* $\implies$ $\mathcal{H}$ *- adversarially PAC-learnable.*

Adversarial PAC-learnability

Main assumption: $\mathcal{H}$ is uniformly convergent

- Given m samples $S = \{(x_1, y_1), \dots, (x_m, y_m)\} \overset{iid}{\sim} \mathcal{D}$, with probability at least $1-\delta$ over the data :
$$\sup_{h\in\mathcal{H}} |\mathcal{R}(h) - \widehat{\mathcal{R}}(h)| \leq s_{\mathcal{H},\ell}(m, \delta, S),$$
- $s_{\mathcal{H},\ell}(m, \delta, S_m) \to 0$ as $m \to \infty$, for any sequence $\{S_m\}_{m\in\mathbb{N}}$ with $S_m \in (\mathcal{X} \times \mathcal{Y})^m$

Theorem

$\mathcal{H}$ - uniformly convergent $\implies$ $\mathcal{H}$ - adversarially PAC-learnable.

Holds even against a stronger adversary that can choose which sources to corrupt

Sample complexity upper bound

- In many situations $s_{\mathcal{H},\ell}(m, \delta, S) = \mathcal{O}(1/\sqrt{m})$
- There exists a learning algorithm, such that with probability at least $1-\delta$:

$$\mathcal{R}(\mathcal{L}(\mathcal{A}(S))) - \min_{h\in\mathcal{H}} \mathcal{R}(h) \leq \widetilde{\mathcal{O}}\Big(\frac{1}{\sqrt{(1-\alpha)Nm}} + \alpha\frac{1}{\sqrt{m}}\Big),$$

against any fixed-set adversary[1]

[1] $\widetilde{\mathcal{O}}$ hides constants and logarithmic factors

Hardness results (for formal statements see paper)

Sample complexity lower bound

- No learning algorithm can achieve against any adversary error less than:

$$\mathcal{O}\left(\frac{1}{\sqrt{(1-\alpha)Nm}} + \alpha\frac{1}{m}\right)$$

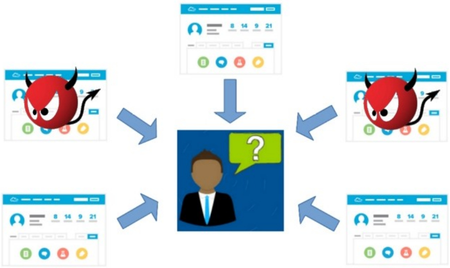
- If m is constant and $\alpha > 0$, $N \to \infty$ does not guarantee PAC-learnability

The learner has to use the group structure

- No learning algorithm that ignores the group structure can guarantee error less than $\mathcal{O}(\alpha/(1-\alpha))$

Summary

- Learning from multiple unreliable sources now commonplace
- Setup modeled as a PAC-learning problem with an adversary
- Group structure enables PAC-learnability, even against a strong adversary
- Describe fundamental limitations on the learner



Summary

- Learning from multiple unreliable sources now commonplace
- Setup modeled as a PAC-learning problem with an adversary
- Group structure enables PAC-learnability, even against a strong adversary
- Describe fundamental limitations on the learner

Thank you for your attention!

Meet us at the poster session for more details.

References I

Alistarh, Dan, Zeyuan Allen-Zhu, and Jerry Li (2018). "Byzantine stochastic gradient descent". In: *NeurIPS*.

Blum, Avrim, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao (2017). "Collaborative PAC Learning". In: *NIPS*.

Jain, Ayush and Alon Orlitsky (2020). "Optimal Robust Learning of Discrete Distributions from Batches". In: *ICML*.

Kearns, Michael and Ming Li (1993). "Learning in the presence of malicious errors". In: *SIAM Journal on Computing*.

Qiao, Mingda (2018). "Do Outliers Ruin Collaboration?" In: *ICML*.

Qiao, Mingda and Gregory Valiant (2018). "Learning Discrete Distributions from Untrusted Batches". In: *ITCS*.

Yin, Dong, Yudong Chen, Kannan Ramchandran, and Peter Bartlett (2018). "Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates". In: *ICML*.