Noise-Tolerant Interactive Learning Using Pairwise Comparisons

Yichong Xu Hongyang Zhang Kyle Miller Aarti Singh Artur Dubrawski
Carnegie Mellon University

Label vs. Comparisons

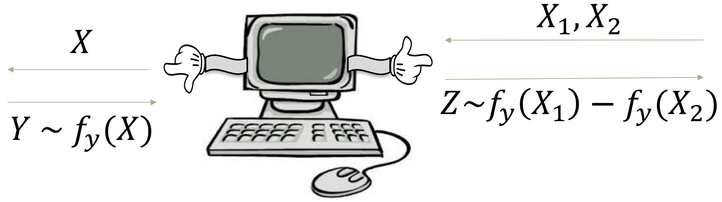
Task: Classifying old/young people portraits

Direct label query

Is the person in the image older than 30?

Comparison query

Which person looks older?



Goal: Interactive algorithm that decides which type of data to collect, when and how much

Noise Conditions

Label Distribution: $\eta(x) = P[Y = 1|X = x]$

Bayes Optimal Classifier: $h^*(x) = \text{sign}(\eta(x) - \frac{1}{2})$

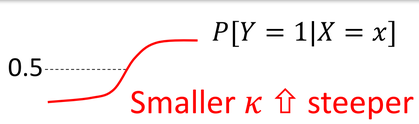
Adversarial Noise (Labels):

$$\Pr[Y \neq h^*(X)] \leq \nu.$$

Tsybakov Noise(Tsybakov 2004, Labels):

$\exists \kappa > 1, \mu > 0, \text{such that } \forall t > 0,$

$$\mathrm{P}\left(\left|\eta(X) - \frac{1}{2}\right| < t\right) < \mu t^{\frac{1}{\kappa - 1}}.$$

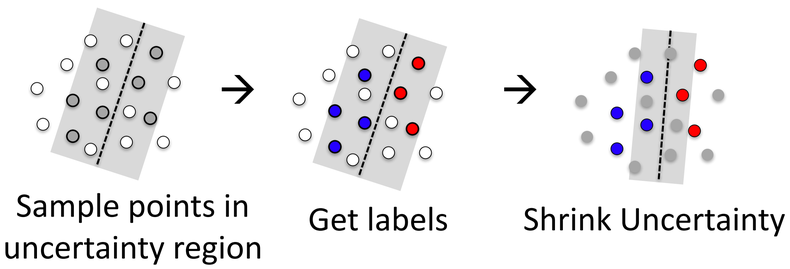


Adversarial Noise (Comparisons):

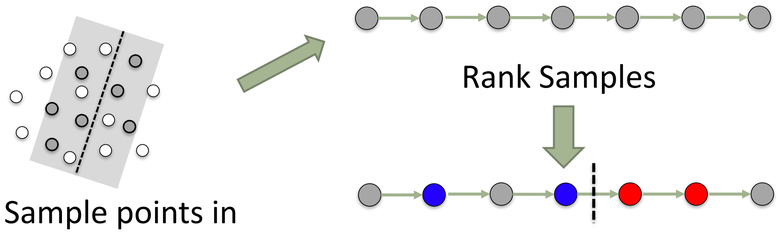
$$P[Z \neq sign(h^*(X_1) - h^*(X_2)] \leq \nu'.$$

Algorithm Description

Previous active learning algorithm: Label → Classifier



Our algorithm: Ranking → Label→ Classifier



Ranking reduces the problem to 1-dimensional problem to find threshold between pos/neg samples

⬇

Better label complexity

To adapt to noise on comparisons:
Use group-based binary search, and majority vote within groups to infer labels

Underlying active learning algorithm to combine with:

General Case: A^2 algorithm (Balcan et al., 2007)
Linear Classifiers: (Awasthi et al., 2017)

Theoretical Results

ε : classification error desired d: dimension
Tol_{comp} – Comparison noise level ν' tolerance
θ – complexity of class C

Adversarial Noise for both Label & Comparison

Work	Efficient?	#Label	#Query	Tol_{comp}
Label	No	$\mathrm{O}(d\theta\log(1/\varepsilon))$	$\mathrm{O}(d\theta\log(1/\varepsilon))$	N/A
Label	Yes	$\mathrm{O}(\mathrm{d}^2\log(d/\varepsilon))$	$\mathrm{O}(\mathrm{d}^2\log(d/\varepsilon))$	N/A
Label+ comparison	**Yes**	$\boldsymbol{O(\log(1/\varepsilon))}$	$\boldsymbol{O\left(d\log^4(d/\varepsilon)\right)}$	$\boldsymbol{\varepsilon^2}$

*Our work in bold

Tsybakov Noise for Label, Adversarial Noise for Comp

Work	Efficient?	#Label	#Query	Tol_{comp}
Label	No	$\tilde{\mathrm{O}}\left(\left(\frac{1}{\varepsilon}\right)^{2\kappa-2} d\theta\right)$	$\tilde{\mathrm{O}}\left(\left(\frac{1}{\varepsilon}\right)^{2\kappa-2} d\theta\right)$	N/A
Label+ Compa-rison	**Yes**	$\tilde{\mathrm{O}}\left(\left(\frac{1}{\varepsilon}\right)^{2\kappa-2} d\theta\right)$	$\tilde{\mathrm{O}}\left(\left(\frac{1}{\varepsilon}\right)^{2\kappa-2} \theta + d\theta\right)$	$\boldsymbol{\varepsilon^{2\kappa}}$

*No previous work exists for efficient learning under Tsybakov Noise

Proof Sketch:

1. Show that there are not too many errors in the ranking obtained from noisy comparisons.
2. Thus, Ranking -> Label -> Classifier approach achieves low error on Adversarial & Tsybakov label noise, using few label queries.
3. Combine with adversarial active learning algorithm to achieve complexity bounds.

Lower bounds

- Label complexity & Total complexity are optimal (up to log)
  Proof sketch: Reduce to the complexity of 1-dim learning
- Noise tolerance is optimal (up to log)
  Proof sketch: Assume oracle with error ν' is free, consider the best possible classifier using the oracle