**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

<table>
<thead>
<tr>
<th>Is the person in the image older than 30?</th>
</tr>
</thead>
</table>

Comparison query

<table>
<thead>
<tr>
<th>Which person looks older?</th>
</tr>
</thead>
</table>

**Algorithm Description**

Previous active learning algorithm: Label → Classifier

Sample points in uncertainty region

Get labels

Shrink Uncertainty

Our algorithm: Ranking → Label → Classifier

Rank Samples

Sample points in uncertainty region

Binary Search to find the boundary

Infer Other Labels

---

**Theoretical Results**

\( \varepsilon : \) classification error desired \( d : \) dimension  
\( \text{To}_{\text{comp}} : \) Comparison noise level \( \nu' : \) tolerance  
\( \theta : \) complexity of class \( C \)

**Adversarial Noise for both Label & Comparison**

<table>
<thead>
<tr>
<th>Work</th>
<th>Efficient?</th>
<th>#Label</th>
<th>#Query</th>
<th>( \text{To}_{\text{comp}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>No</td>
<td>( O(d\log(1/\varepsilon)) )</td>
<td>( O(d\log(1/\varepsilon)) )</td>
<td>N/A</td>
</tr>
<tr>
<td>Label</td>
<td>Yes</td>
<td>( O(d^3\log(d/\varepsilon)) )</td>
<td>( O(d^3\log(d/\varepsilon)) )</td>
<td>N/A</td>
</tr>
<tr>
<td>Label+ Comparison</td>
<td>Yes</td>
<td>( O(\log(1/\varepsilon)) )</td>
<td>( O(d\log^4(1/\varepsilon)) )</td>
<td>( \varepsilon^2 )</td>
</tr>
</tbody>
</table>

*Our work in bold

**Tsybakov Noise for Label, Adversarial Noise for Comp**

<table>
<thead>
<tr>
<th>Work</th>
<th>Efficient?</th>
<th>#Label</th>
<th>#Query</th>
<th>( \text{To}_{\text{comp}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>No</td>
<td>( \Theta \left( \frac{1}{\varepsilon} \right) 2^x - 2 ) ( d\theta )</td>
<td>( \Theta \left( \frac{1}{\varepsilon} \right) 2^x - 2 ) ( d\theta )</td>
<td>N/A</td>
</tr>
<tr>
<td>Label+ Comparison</td>
<td>Yes</td>
<td>( \Theta \left( \frac{1}{\varepsilon} \right) 2^x - 2 ) ( d\theta )</td>
<td>( \Theta \left( \frac{1}{\varepsilon} \right) 2^x - 2 ) ( \theta + d\theta )</td>
<td>( \varepsilon^{2x} )</td>
</tr>
</tbody>
</table>

*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)
- Noise tolerance is optimal (up to log)
- Proof sketch: Assume oracle with error \( \nu' \) is free, consider the best possible classifier using the oracle