NEURAL INFORMATION PROCESSING SYSTEMS

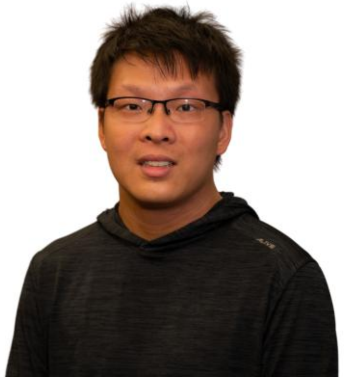
Conservative classifiers do consistently well with improving agents

Dravyansh Sharma

Alec Sun

Northwestern University

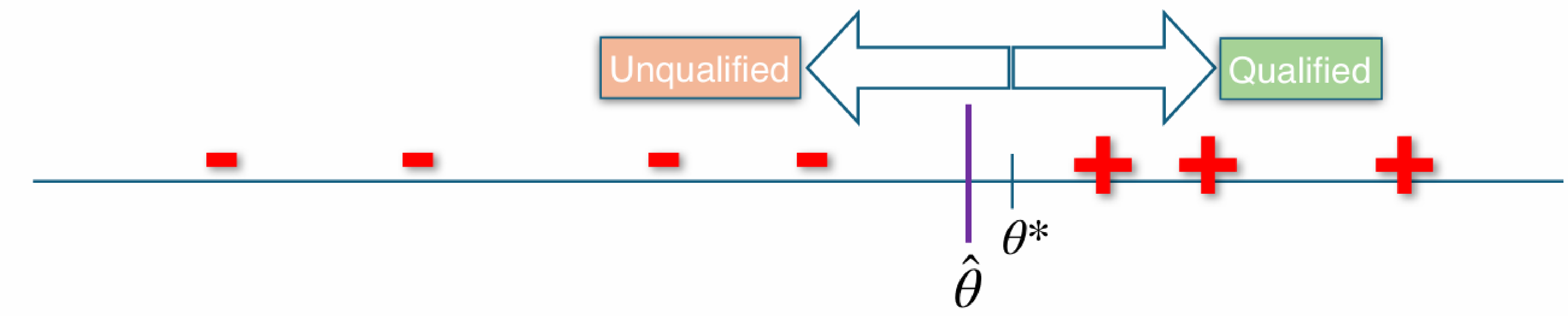
TOYOTA TECHNOLOGICAL INSTITUTE AT CHICAGO



THE UNIVERSITY OF CHICAGO

Binary classification

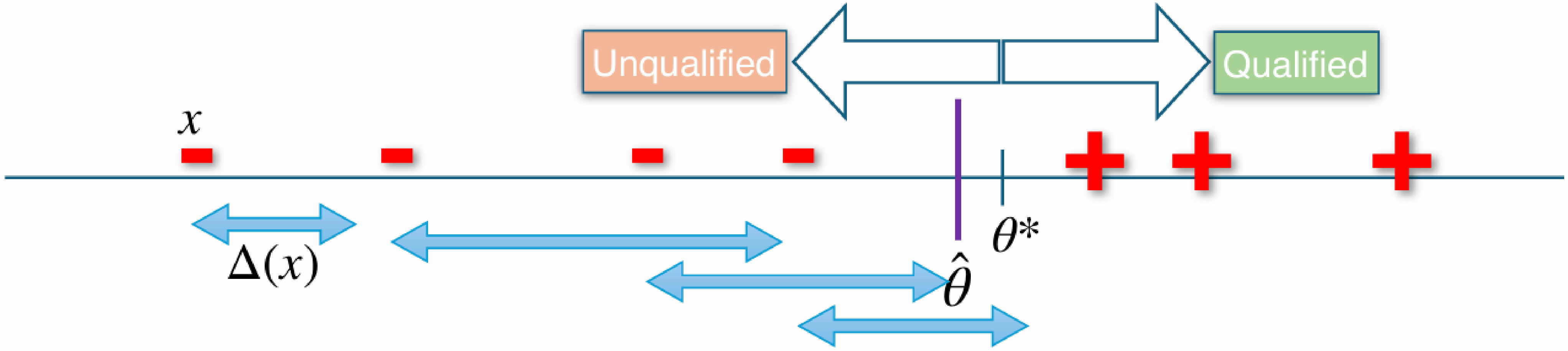
- Predict whether someone is qualified for a job



- Don't know θ^* but have past data
- Publish a test cutoff $\hat{\theta}$

Learning with improvements

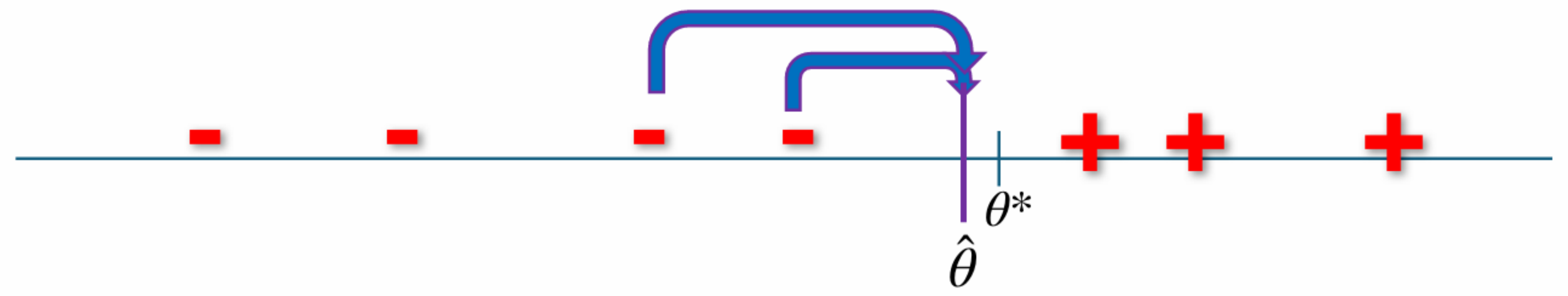
- **Assumption:** People put in effort to improve their qualification



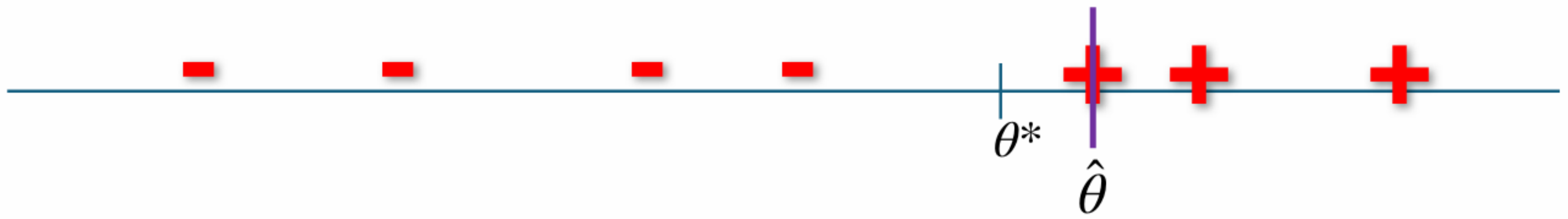
- Agent x can improve in region $\Delta(x)$
- How to set $\hat{\theta}$ under improvements?

Where to set cutoff

- **Cutoff too low:** Agents improve to the cutoff but are not qualified



- **Cutoff too high:** It’s fine! No false positives (everyone hired is qualified) nor false negatives (positives improve to the cutoff)



Formal model

- Ground-truth classifier f^* from hypothesis class $\mathcal{H}$
- Agent x has improvement region $\Delta(x)$

1. Design a classifier h and publish it
2. If $h(x) = 0$ but there is some $x' \in \Delta(x)$ for which $h(x') = 1$, x moves to such a x' (breaking ties arbitrarily)

Comparison with strategic classification

- Strategic classification
  - Agents manipulate their features and deceive the classifier
  - Movements are not genuine

- Learning with improvements
  - Agents genuinely improve to meet the classifier’s threshold
  - Movements in the feature space are real

Previous work

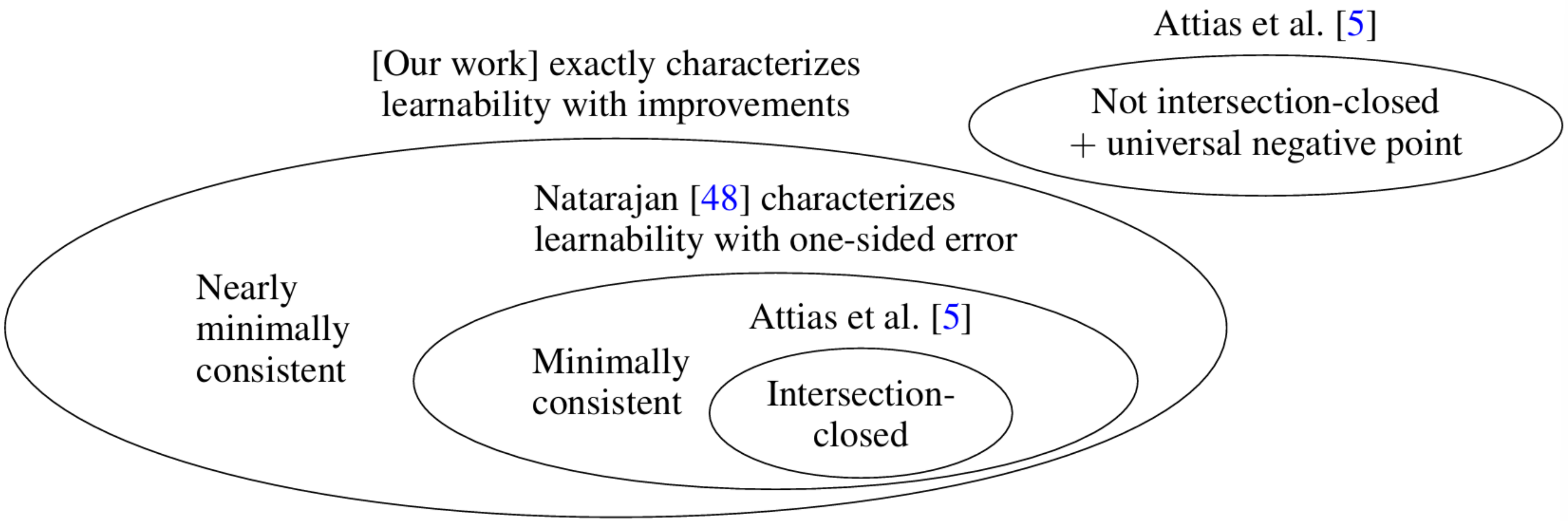
PAC Learning with Improvements

Idan Attias[1,2] Avrim Blum[2] Keziah Naggita[2]
Donya Saless[2] Dravyansh Sharma[2,3] Matthew Walter[2*]

[1]University of Illinois at Chicago
[2]Toyota Technological Institute at Chicago
[3]Northwestern University
{idan, avrim, knaggita, donya, dravy, mwalter}@ttic.edu

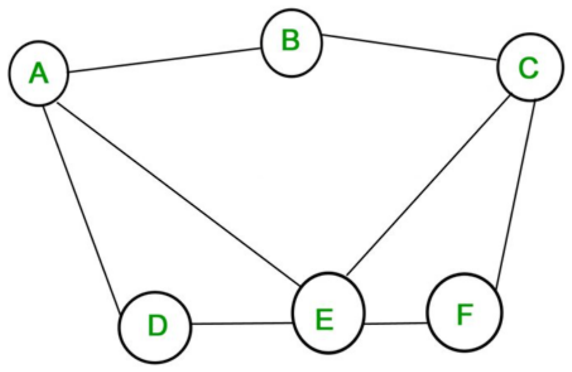
Results: proper, realizable learning

- Complete characterization of proper (published classifier must be in the hypothesis class), realizable (hypothesis class must contain the ground truth) learnability for any improvement function



Results: beyond proper, realizable learning

- Improper learning (published classifier may lie outside hypothesis class)
- Learning with label noise
- Online learning (agents are vertices on a graph and can improve to neighbors)



Results: online learning

- For both realizable and agnostic (hypothesis class may not contain the ground truth) settings:
  - New algorithm based on risk-averse majority vote
  - Nearly tight mistake bounds

	Realizable setting	Agnostic setting
Mistake upper bound	$(\Delta_G + 1)\log\|\mathcal{H}\|$	$O\left(\Delta_G \cdot (\mathsf{OPT} + \log\|\mathcal{H}\|)\right)$
Mistake lower bound	$\Delta_G - 1$	$\Delta_G \cdot \mathsf{OPT}$

$\Delta_G =$ Maximum degree of vertex in G

Conclusion

- Characterize statistical and online learning under improvements in many natural but challenging settings
  - Proper learning
  - Improper learning
  - Learning with noise
  - Online learning
- **Moral of the story:** “conservative” classifiers perform well