A Potential-based Framework for Online Learning with Mistakes and Abstentions

Chicheng Zhang
joint work with Kamalika Chaudhuri

UC San Diego

NIPS Workshop on Reliable Machine Learning in the Wild
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$: 
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

Show $x_t \in \mathcal{X}$
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

Show $x_t \in \mathcal{X}$

Predict $\hat{y}_t \in \{-1, +1, \bot\}$
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

Show $x_t \in \mathcal{X}$

Predict $\hat{y}_t \in \{-1, +1, \bot\}$

Reveal $y_t \in \{-1, +1\}$
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

- Show $x_t \in \mathcal{X}$
- Predict $\hat{y}_t \in \{-1, +1, \perp\}$
- Reveal $y_t \in \{-1, +1\}$

Reliable predictions on non-abstention examples
Performance Metrics:
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

- Show $x_t \in \mathcal{X}$
- Predict $\hat{y}_t \in \{-1, +1, \perp\}$
- Reveal $y_t \in \{-1, +1\}$

Reliable predictions on non-abstention examples
Performance Metrics:

- Mistakes: $\sum_t I(\hat{y}_t = -y_t)$
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

1. Show $x_t \in \mathcal{X}$
2. Predict $\hat{y}_t \in \{-1, +1, \bot\}$
3. Reveal $y_t \in \{-1, +1\}$

Reliable predictions on non-abstention examples

Performance Metrics:

- Mistakes: $\sum_t I(\hat{y}_t = -y_t)$
- Abstentions: $\sum_t I(\hat{y}_t = \bot)$
Problem: Online Classification with Abstentions

For $t = 1, 2, \ldots$:

Show $x_t \in \mathcal{X}$

Predict $\hat{y}_t \in \{-1, +1, \perp\}$

Reveal $y_t \in \{-1, +1\}$

Reliable predictions on non-abstention examples

Performance Metrics:

- Mistakes: $\sum_t I(\hat{y}_t = -y_t)$
- Abstentions: $\sum_t I(\hat{y}_t = \perp)$
- Goal: Tradeoff mistakes and abstentions
Challenge

- [LLWS11, SZB10]: only works for finite $|\mathcal{H}|$, realizable case
Challenge

- [LLWS11, SZB10]: only works for finite $|\mathcal{H}|$, realizable case
- [ZC16]: minimax algorithm with sharp performance bounds, but intractable
Challenge

- [LLWS11, SZB10]: only works for finite $|\mathcal{H}|$, realizable case
- [ZC16]: minimax algorithm with sharp performance bounds, but intractable
- Challenge: design tractable online learning algorithms with abstentions, for general $\mathcal{H}$ and nonrealizable case
Our Contributions

- We develop such an algorithm
Our Contributions

- We develop such an algorithm
- We formalize the notion of *admissible potential function*, a "capacity measure" of hypothesis class
Our Contributions

- We develop such an algorithm
- We formalize the notion of *admissible potential function*, a “capacity measure” of hypothesis class
- We develop weighted majority-style algorithms over such potentials
Our Contributions

- We develop such an algorithm
- We formalize the notion of *admissible potential function*, a “capacity measure” of hypothesis class
- We develop weighted majority-style algorithms over such potentials
- See you at the poster :)