Results on Various Models of Mistake-bounded Online Learning

Raymond Feng Andrew Lee Espen Slettnes
Mentor: Dr. Jesse Geneson

MIT PRIMES Conference

October 16, 2021

Outline

1. Introduction
2. Learning models
3. Comparing the models
4. Known results
5. Our results

Introduction

- Today, we will talk about learning algorithms.
- These are algorithms which attempt to learn information through repeated testing by an adversary (in particular, we are not the ones learning any algorithms).
    - For example, an algorithm learning to classify different images.
- Efficiency is measured by the **number of mistakes** that they make.

The Standard Model

Standard Model

The standard model is the situation of a learner attempting to classify inputs (in the set X) with labels (in the set Y) based on a number of possible functions $f \in F$ mapping X to Y.

- The learning takes place in rounds, where the adversary gives the learner a single input each round and the learner must guess the corresponding label. **The adversary then tells the learner the correct answer**.
- It is guaranteed that the information given to the learner is consistent across all rounds with some function $f \in F$. In particular, the learner will eventually never make a mistake.
- The efficiency of an algorithm can be measured by either the algorithms worst-case performance or average performance.
  - This is measured by the number of mistakes that the learner makes.

The Standard Model

- The previous model is also called strong reinforcement learning.
  - After each round, **the adversary tells the learner the correct answer**.
- In weak reinforcement learning, **the adversary only tells the learner if the given answer was correct or incorrect**.
  - This is also known as the bandit model.

An Example

- Scenario: want to learn to classify a training data set of images of dogs, foxes, and wolves.
- The set of inputs X: the data set
- The set of outputs Y: $\{\text{dog, fox, wolf}\}$
- The set of possible functions F: all $3^{|X|}$ possible functions




An Example

- Adversary gives learner an image (shown below).
- Learner says "FOX". (This is a mistake for the learner.)
- Strong reinforcement (standard model): Adversary tells the learner that the right answer is "DOG."
- Weak reinforcement (bandit model): Adversary tells the learner that they were wrong. The learner only knows that the correct answer was either "DOG" or "WOLF."

A Different Model

Delayed, Ambiguous Reinforcement Model

This model is similar to the standard weak reinforcement model, but the learner receives a fixed number r of inputs each round, and the adversary only gives the learner the next input **after they answer the previous one**. At the end of the round, the adversary says "YES" if all answers during the round were correct and "NO" otherwise.

- For $r = 1$, this model is exactly the same as the weak reinforcement learning model.
- A natural question might be whether information of future inputs is useful.

Modifying the Standard Weak Model

To analyze the importance of knowing future inputs, we introduce the following model.

Modified Standard Weak Reinforcement Model

If F is a set of functions $f : X \to Y$, we define $\text{CART}_r(F)$ to be a set of functions $f' : X^r \to Y^r$, so that each $f \in F$ has a corresponding $f' \in \text{CART}_r(F)$ such that for any $x_1, x_2, \ldots, x_r \in X$ we have

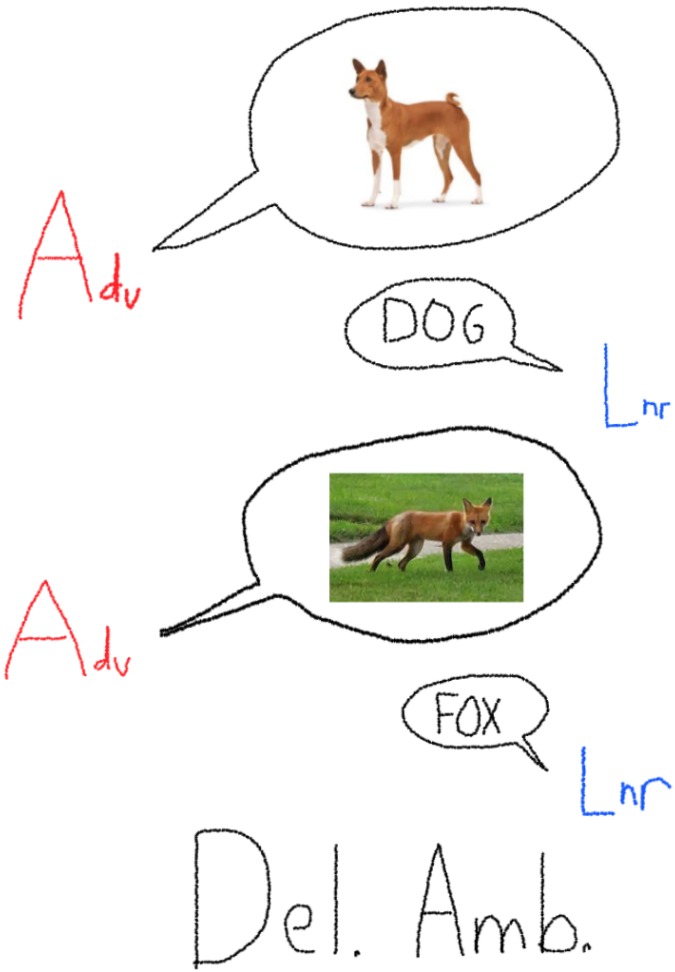
$$f'((x_1, x_2, \ldots, x_r)) = (f(x_1), f(x_2), \ldots, f(x_r)).$$

- Running the standard weak reinforcement model on $\text{CART}_r(F)$ would simulate the process of running rounds of r inputs, where **the adversary gives all inputs at the same time** at the beginning of each round.

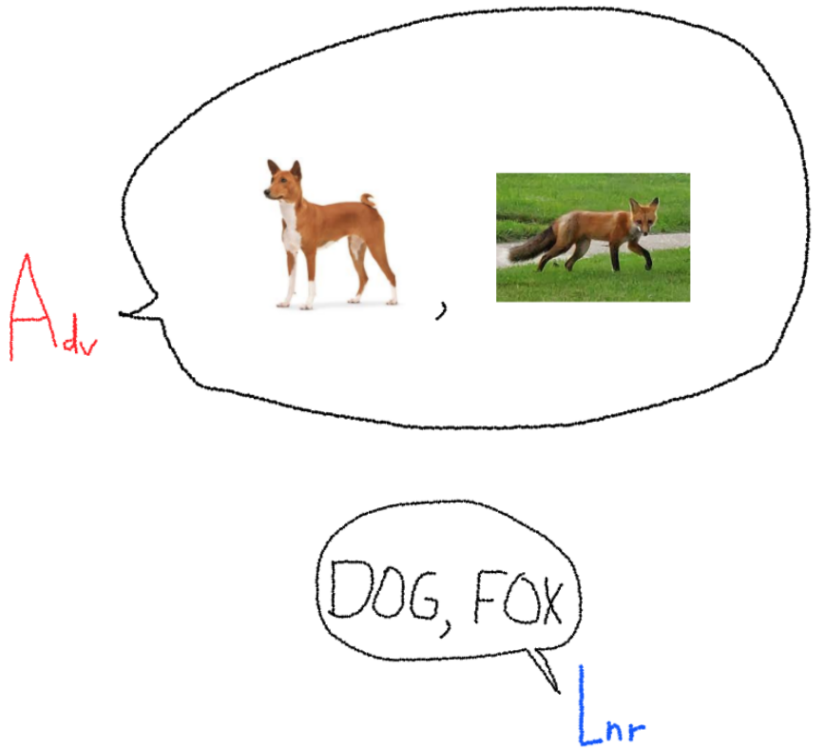
Examples

- Consider the same scenario as earlier and let $r = 2$.
- In the delayed, ambiguous model, the learner would receive an image, guess an output, receive another input, and then guess another output. The adversary would only say "YES" if both guesses were correct.
- In the modified weak reinforcement model, the adversary gives two images to the learner at the same time and then the learner guesses their outputs. The adversary says "YES" if both guesses were correct.

Examples



Del. Amb.



Mod. Weak

Measuring Efficiencies

Measuring efficiencies with opt(F)

In general, opt(F) is the maximum number of mistakes that a learner makes while learning a function from the set F. We use subscripts to indicate which learning scenario we are referring to.

- $\text{opt}_{std}(F) = \text{opt}_{strong}(F)$ refers to the standard model/strong reinforcement.
- $\text{opt}_{bandit}(F) = \text{opt}_{weak}(F)$ refers to the bandit model/weak reinforcement.
- $\text{opt}_{amb,r}(F)$ refers to delayed, ambiguous reinforcement.
- $\text{opt}_{weak}(\text{CART}_r(F))$ refers to modified weak reinforcement.

Comparing $\text{opt}_{weak}(\text{CART}_r(F))$ and $\text{opt}_{amb,r}(F)$

- Is knowing all r inputs at the same time for each round of learning helpful to the learner?
  - It is clearly not harmful, since it is extra information. Therefore, we always have
$$\text{opt}_{weak}(\text{CART}_r(F)) \leq \text{opt}_{amb,r}(F).$$
- If knowing all the inputs before the round started was not helpful, we would expect that $\text{opt}_{weak}(\text{CART}_r(F)) = \text{opt}_{amb,r}(F)$ always.
  - Indeed, for $r = 1$, when the two scenarios are the same, we always have
$$\text{opt}_{weak}(\text{CART}_1(F)) = \text{opt}_{amb,1}(F).$$

Known Results

Theorem (Auer, Long (1999))

There exist X, Y, F such that

$$\text{opt}_{weak}(\text{CART}_2(F)) < \text{opt}_{amb,2}(F).$$

Their example has $X = 1, 2, 3$, $Y = 0, 1$, and F containing the following 4 functions:

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
1	0	0	0	1
2	0	0	1	1
3	0	1	1	1

For this example, $\text{opt}_{amb,r}(F) \geq 3$ but $\text{opt}_{weak}(\text{CART}_r(F)) = 2$.
This shows that knowledge of all inputs of a round is strictly beneficial for the learner.

Our Results

We will now go over some of the results that we discovered regarding the different learning situations presented.

Modified Weak vs. Delayed Ambiguous Reinforcement

Theorem

Let X be an ordered set, $Y = \{0, 1\}$, and F be a set of *non-decreasing* functions from X to Y. Then

$$\text{opt}_{weak}(\text{CART}_r(F)) = (1 \pm o(1))r \ln(|F|)$$

and

$$\text{opt}_{amb,r}(F) = (1 - o(1))2^r \ln(|F|)$$

- Notice how the worst-case performance for the weak reinforcement model is linear in r, but it is exponential in r for the delayed, ambiguous reinforcement model.
- This is surprising! A slight difference in the flow of information had a large impact on the efficiency of learning algorithms.

Maximum Factor Gap

- Long (2019) proved that for all $M > 1$ and infinitely many k, there always exists a family of functions with a range of size k such that $\text{opt}_{std}(F) = M$ and $\text{opt}_{weak}(F)$ is about $k \ln k \, \text{opt}_{std}(F)$. He also proved this is maximum such factor gap between the models.
- We generalize this result to $\text{opt}_{weak}(CART_r(F))$. We proved that

Theorem

The maximum factor gap for general r is between $k^r \ln k$ and $rk^r \ln k$.

Relative Position Models

We can also look at models where the learner guesses a permutation $\sigma \in F$. We explored the relative position model and its variants.

Relative Position Model

In the relative position model, the adversary chooses a special input x and set of r inputs S. The learner guesses the relative position of x in S, i.e. the number of elements y of S such that $\sigma(x) > \sigma(y)$.

Delayed Relative Position Model

Instead of giving all of S at once, the adversary gives the elements **one by one**, and for each input the learner guesses if it is higher or lower. The learner is correct if the number of "lower"s equals x's relative position.

We denote the maximum number of mistakes in the nondelayed version by $\text{opt}(\text{RPOS}_r(F))$ and in the delayed version by $\text{opt}^{\text{RPOS}}_{amb,r}(F)$.

Relative Position Models

If $F = S_n$, an optimal strategy for the adversary is to force the learner to do an insertion sort. For the learner, the insertion step is then to guess $\sigma(n)$, the value of the last input.

If we represent a lower value as a 0 and a greater value as a 1, we see that this is equivalent to guessing a non-decreasing function, giving:

Theorem

For the nondelayed version of the relative position model,

$$\mathrm{opt}(\mathrm{RPOS}_r(S_n)) = (1 + o(1))r \ln |S_n|,$$

whereas for the delayed version,

$$\mathrm{opt}^{\mathrm{RPOS}}_{amb,r}(S_n) = (1 + o(1))2^r \ln |S_n|.$$

Acknowledgements

We thank

- Our mentor Dr. Jesse Geneson of San Jose State University for valuable guidance and feedback.
- Professor Pavel Etingof, Dr. Slava Gerovitch, Dr. Tanya Khovanova, the MIT Math Department, and the MIT PRIMES program, for providing us with the opportunity to work on this project.
- You for listening.

References I

Peter Auer and Philip M. Long. "Structural Results About On-line Learning Models With and Without Queries". In: *Machine Learning* 36 (3 Sept. 1999), pp. 147–181. DOI: 10.1023/A:1007614417594.

Peter Auer et al. "On the complexity of function learning". In: *Machine Learning* 18 (2 Feb. 1995), pp. 187–230. ISSN: 1573-0565. DOI: 10.1007/BF00993410.

Avrim Blum. "On-Line Algorithms in Machine Learning". In: *Online Algorithms: The State of the Art*. Ed. by Amos Fiat and Gerhard J. Woeginger. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 306–325. ISBN: 978-3-540-68311-7. DOI: 10.1007/BFb0029575. URL: https://doi.org/10.1007/BFb0029575.

J. Lawrence Carter and Mark N. Wegman. "Universal Classes of Hash Functions (Extended Abstract)". In: *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*. STOC '77. Boulder, Colorado, USA: Association for Computing Machinery, 1977, pp. 106–112. ISBN: 9781450374095. DOI: 10.1145/800105.803400. URL: https://doi.org/10.1145/800105.803400.

K. Crammer and C. Gentile. "Multiclass classification with bandit feedback using adaptive regularization". In: *Machine Learning* 90 (2012), pp. 347–383.

Varsha Dani, Thomas Hayes, and Sham Kakade. "Stochastic Linear Optimization under Bandit Feedback.". In: Jan. 2008, pp. 355–366.

Amit Daniely and Tom Helbertal. "The price of bandit information in multiclass online classification". In: *CoRR* abs/1302.1043 (2013). arXiv: 1302.1043. URL: http://arxiv.org/abs/1302.1043.

References II

Jacob Fox. "Stanley-Wilf limits are typically exponential". In: (2013). arXiv: 1310.8378 [math.CO].

Jesse Geneson. "A note on the price of bandit feedback for mistake-bounded online learning". In: *CoRR* abs/2101.06891 (2021). arXiv: 2101.06891. URL: https://arxiv.org/abs/2101.06891.

Elad Hazan and Satyen Kale. "NEWTRON: An efficient bandit algorithm for online multiclass prediction". English (US). In: *Advances in Neural Information Processing Systems* (Dec. 2011). 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011 ; Conference date: 12-12-2011 Through 14-12-2011, pp. 891–899.

Nick Littlestone. "Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm". In: *Machine Learning* 2 (4 Apr. 1988), pp. 285–318. ISSN: 1573-0565. DOI: 10.1023/A:1022869011914.

Philip M. Long. "New bounds on the price of bandit feedback for mistake-bounded online multiclass learning". In: *Theoretical Computer Science* 808 (2020). Special Issue on Algorithmic Learning Theory, pp. 159–163. ISSN: 0304-3975. DOI: https://doi.org/10.1016/j.tcs.2019.11.017. URL: https://www.sciencedirect.com/science/article/pii/S0304397519307297.

Michael Luby and Avi Wigderson. "Pairwise Independence and Derandomization". In: *Found. Trends Theor. Comput. Sci.* 1.4 (Aug. 2006), pp. 237–301. ISSN: 1551-305X. DOI: 10.1561/0400000009. URL: https://doi.org/10.1561/0400000009.

References III

C. R. Rao. "Hypercubes of strength "d" leading to confounded designs in factorial experiments". In: *Bulletin of the Calcutta Mathematical Society* (38 1946), pp. 67–68. ISSN: 0008-0659.

C. Radhakrishna Rao. "Factorial Experiments Derivable from Combinatorial Arrangements of Arrays". In: *Supplement to the Journal of the Royal Statistical Society* 9.1 (1947), pp. 128–139. DOI: 10.2307/2983576.

Amitai Regev. "Asymptotic values for degrees associated with strips of young diagrams". In: *Advances in Mathematics* 41.2 (1981), pp. 115–136. ISSN: 0001-8708. DOI: https://doi.org/10.1016/0001-8708(81)90012-8. URL: https://www.sciencedirect.com/science/article/pii/0001870881900128.

H. Shvaytser. "Linear manifolds are learnable from positive examples". Unpublished manuscript. 1988.