Online Learnability and Complexity Measures

Project Presentation - CMPUT 654, ML Theory

Alireza Masoumian

7 Dec 2023

What if instead of "batch" data, we have a stream of data?

Online Learning: Random Averages, Combinatorial Parameters, and Learnability

Alexander Rakhlin
Department of Statistics
University of Pennsylvania

Karthik Sridharan
Toyota Technological Institute at Chicago

Ambuj Tewari
Computer Science Department
University of Texas at Austin

October 29, 2010

Batch (Offline)

PAC Learning $\Longleftrightarrow$ Finite VCDim, Rademacher,...

- Existence of an efficient algorithm finding a Probably Approximate Correct hypothesis.

$$w.p. \quad 1-\delta:$$

$$err_D(\hat{f}) - err_D(f^*) \leq \epsilon$$

- Generic algorithm:
Empirical Risk Minimization (ERM)

- Oracle inequalities

Online

Sublinear Regret $\Longleftrightarrow$ Finite LDim, Seq. Rademacher,....

- Regret:

$$\frac{1}{T}\Big(\sum_{t=1}^{T} err_t(\hat{f}_t) - err_t(f^*)\Big) \in o(1)$$

- Generic algorithm:
Standard Optimal Algorithm (SOA)

- N. Littlestone, Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. 1988

Online Learning

Hypothesis Class $\mathscr{F}$ Feature Set $\mathscr{X}$

A learner interacts with an adversary over T rounds.

In each round $t = 1,2,\ldots,T$:

- Learner picks a distribution $q_t \in \mathscr{Q}$ over functions $f : \mathscr{X} \to \mathscr{Y}$
- Adversary picks a feature-label pair x_t
- Learner draws a sample $f_t \sim q_t$ and suffers loss $f_t(x_t)$

Repetitive Two-Player Game

What's its value?

Regret:

$$\mathbb{E} \inf_{f \in \mathscr{F}} \sum_{t=1}^{T} \left[f_t(x_t) - f(x_t) \right]$$

Online Learning

Hypothesis Class $\mathscr{F}$ Feature Set $\mathscr{X}$ Label Set $\mathscr{Y} \subseteq \mathbb{R}$

A learner interacts with an adversary over T rounds.

In each round $t = 1,2,\ldots,T$:

Learner picks a distribution $q_t \in \mathscr{Q}$ over functions $f : \mathscr{X} \to \mathscr{Y}$

Adversary picks a feature-label pair (x_t, y_t)

Learner draws a sample $f_t \sim q_t$ and suffers loss $err_t(f_t) = \ell(f_t(x_t) - y_t)$

Repetitive Two-Player Game

What's its value?

Regret:

$$\mathbb{E} \sum_{t=1}^{T} \left[err_t(f_t) - \inf_{f^* \in \mathscr{F}} err_t(f^*) \right]$$

Realizable: $\exists f^* : f^*(x_t) = y_t \quad \forall t$

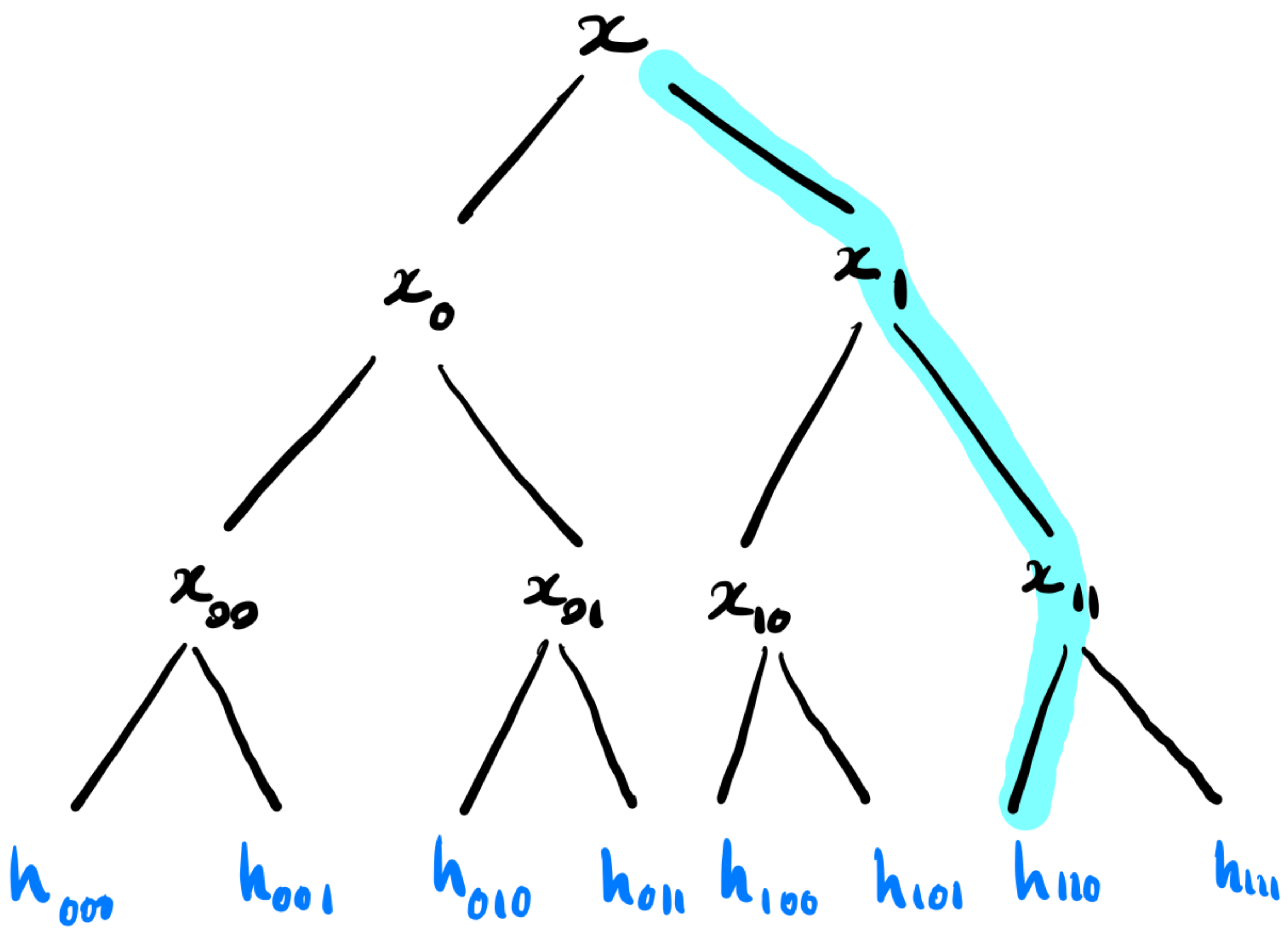
Agnostic: o.w.

Littlestone's Dimension

- Analogous to VC dimension (Binary)

Shattered Tree: ∀ path ϵ, $\exists f \in \mathscr{F}$

Ldim : = Max depth of a shattered labeled tree.



Definition 6. An $\mathcal{X}$-valued tree $\mathbf{x}$ of depth d is *shattered* by a function class $\mathcal{F} \subseteq \{\pm 1\}^{\mathcal{X}}$ if for all $\epsilon \in \{\pm 1\}^d$, there exists $f \in \mathcal{F}$ such that $f(\mathbf{x}_t(\epsilon)) = \epsilon_t$ for all $t \in [d]$. The *Littlestone dimension* $\mathrm{Ldim}(\mathcal{F}, \mathcal{X})$ is the largest d such that $\mathcal{F}$ shatters an $\mathcal{X}$-valued tree of depth d.

- Continuous version?! fat-shattering

- N. Littlestone, Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. 1988
- S. Ben David et al, Agnostic online learning, 2009

Online Learnability

- Value of the Game:

$$\mathcal{V}_T(\mathcal{F},\mathcal{X}) = \inf_{q_1\in\mathcal{Q}} \sup_{x_1\in\mathcal{X}} \mathbb{E}_{f_1\sim q_1} \cdots \inf_{q_T\in\mathcal{Q}} \sup_{x_T\in\mathcal{X}} \mathbb{E}_{f_T\sim q_T}\left[\sum_{t=1}^{T} f_t(x_t) - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f(x_t)\right]$$

Prokhorov's theorem

$$= \sup_{p_1} \mathbb{E}_{x_1\sim p_1} \ldots \sup_{p_T} \mathbb{E}_{x_T\sim p_T}\left[\sum_{t=1}^{T} \inf_{f_t\in\mathcal{F}} \mathbb{E}_{x_t\sim p_t}\left[f_t(x_t)\right] - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} f(x_t)\right]$$

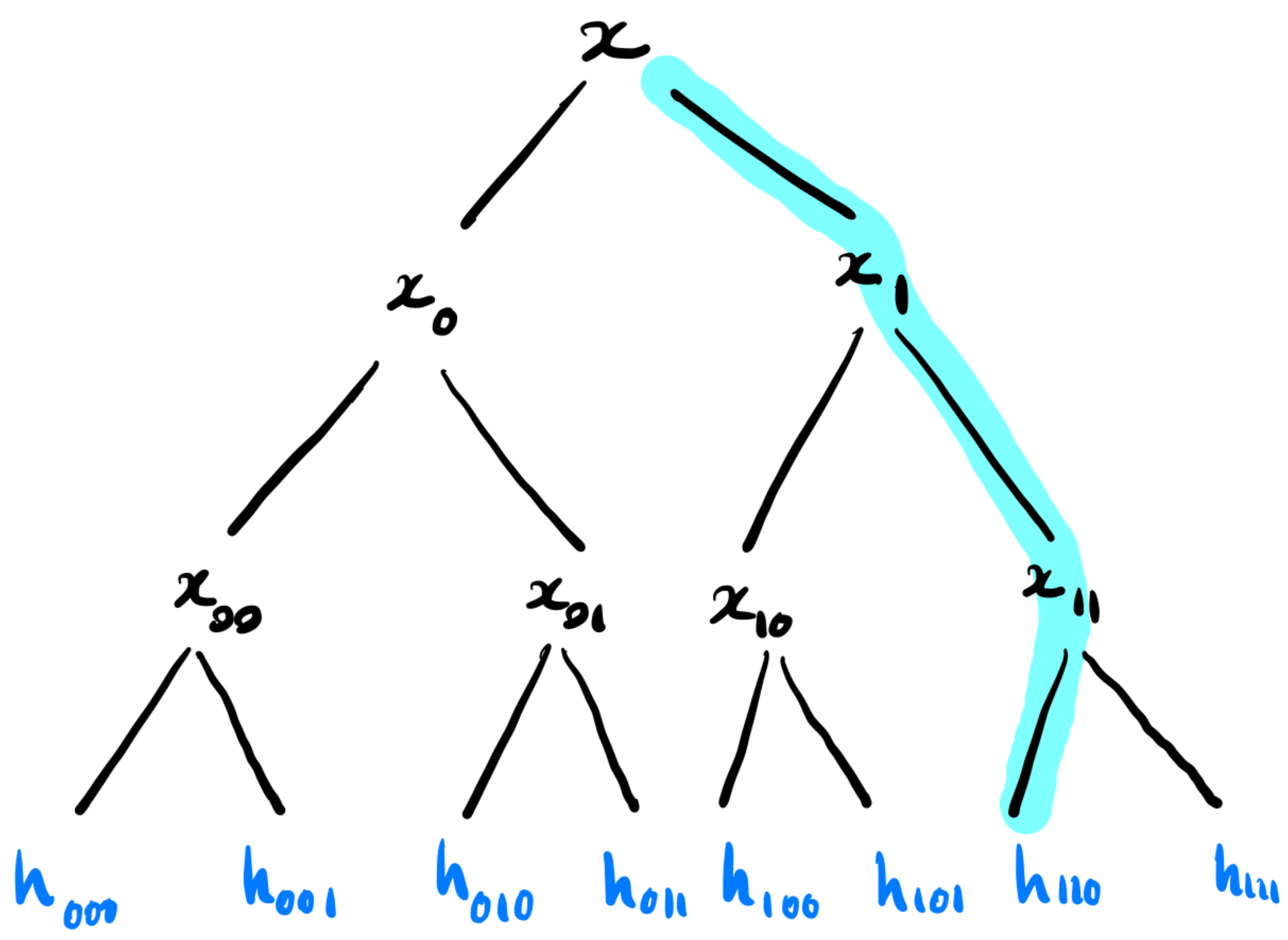
- $\mathscr{F}$ is "online learnable" w.r.t $\mathscr{X}$ if:

$$\limsup_{T\to\infty} \frac{\mathcal{V}_T(\mathcal{F},\mathcal{X})}{T} = 0 \ .$$

Sequential Rademacher Complexity

- Like offline, but ϵ and samples are no longer independent:

$$\mathfrak{R}_T(\mathcal{F}) = \sup_{\mathbf{x}} \mathbb{E}_\epsilon \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\mathbf{x}_t(\epsilon)) \right]$$



- Minimax value is bounded with Sequential Rademacher complexity:

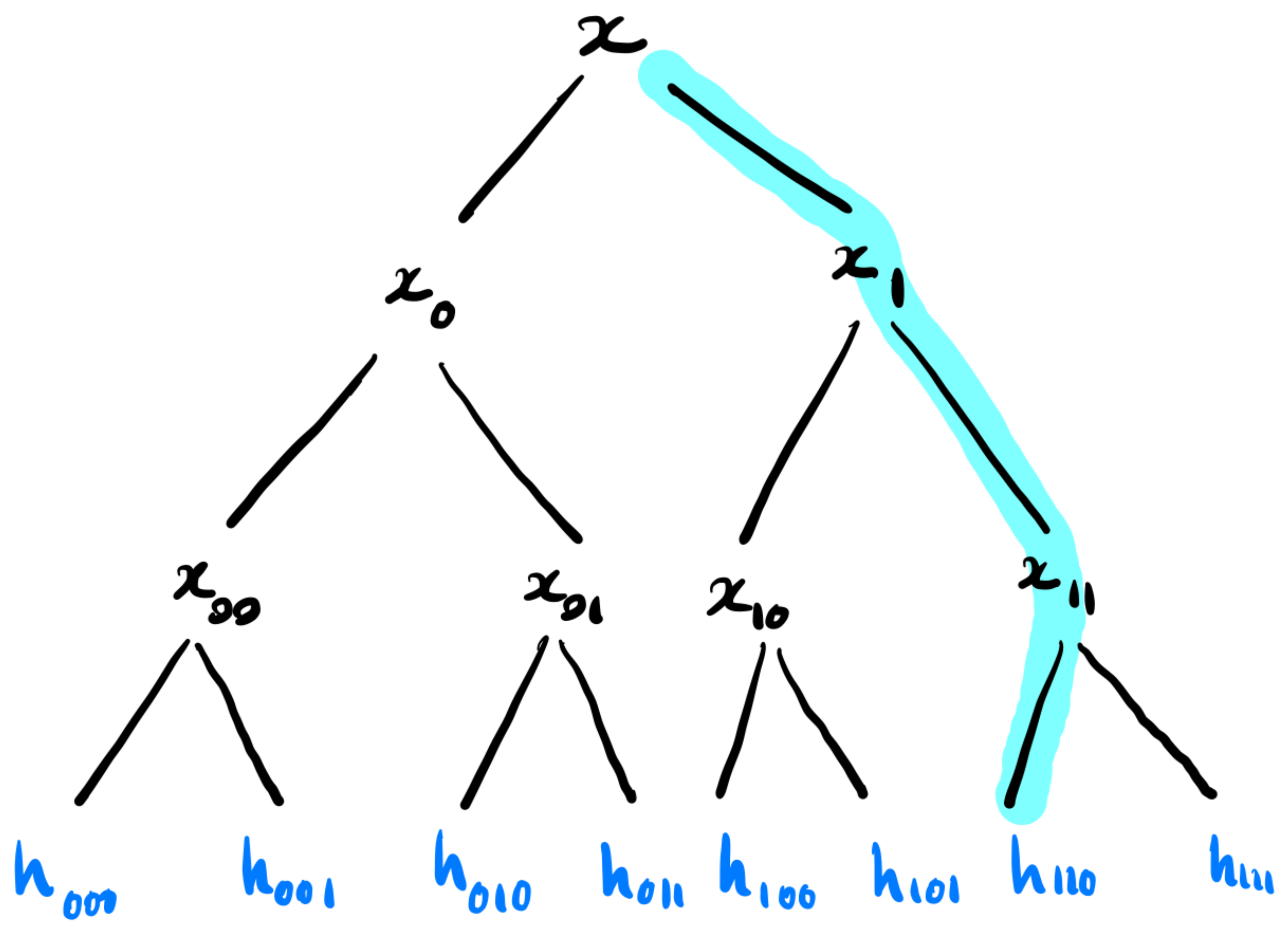
$$\mathcal{V}_T(\mathcal{F}) \leq 2\mathfrak{R}_T(\mathcal{F})$$

- A bit more tricky symmetrization...

Standard Optimal Algorithm (SOA)

- ERM of online learning

- Compute the little stone dimension of a sub-hypothesis class



- Realizable setting:

$$Reg \leq Ldim$$

- Agnostic setting:

$$Reg \leq \tilde{\mathcal{O}}(\sqrt{Ldim \,.\, T})$$

- Making T^{Ldim} many experts s.t. ∃ optimal experts → Expert Advise problem

- S. Ben David et al, Agnostic online learning, 2009

What if I use ERM in Online Learning?

- Not too sublinear!
- Still need exponentially number of queries to ERM

Online Learning and Solving Infinite Games with an ERM Oracle

Angelos Assos * Idan Attias † Yuval Dagan ‡ Constantinos Daskalakis §
Maxwell Fishelson ¶

July 11, 2023

Unapproximability of Littlestone's Dimension

- Hard to approximate!

Improved Inapproximability of VC Dimension and Littlestone's Dimension via (Unbalanced) Biclique

Pasin Manurangsi
Google Research, Thailand
pasin@google.com

November 4, 2022

Thank you