CONSISTENCY OF THE k_n-NEAREST NEIGHBOR RULE UNDER ADAPTIVE SAMPLING

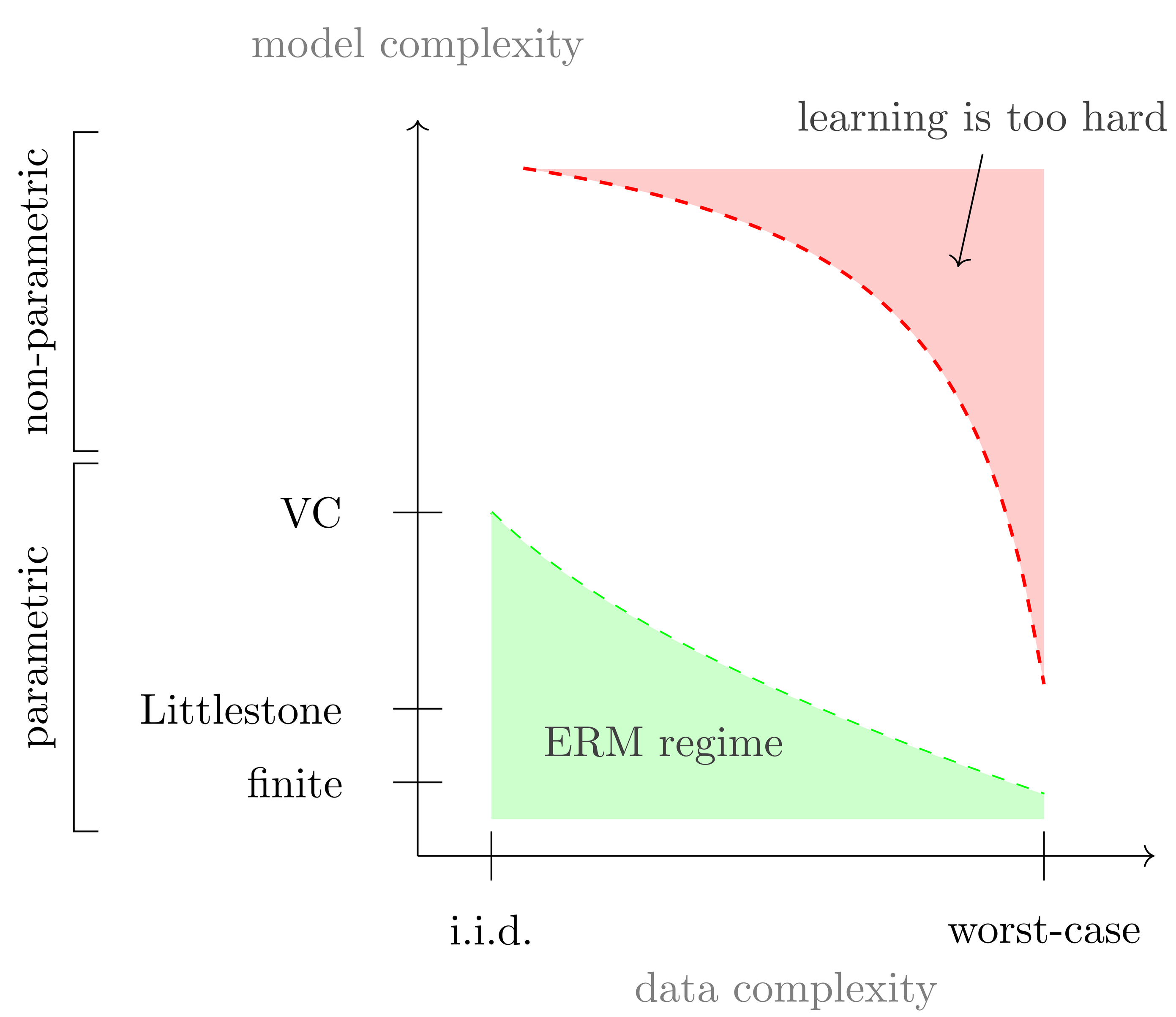
Robi Bhattacharjee[1], Sanjoy Dasgupta[2] and Geelon So[2]

[1]University of Tübingen and Tübingen AI Center, [2]University of California, San Diego

MOTIVATION

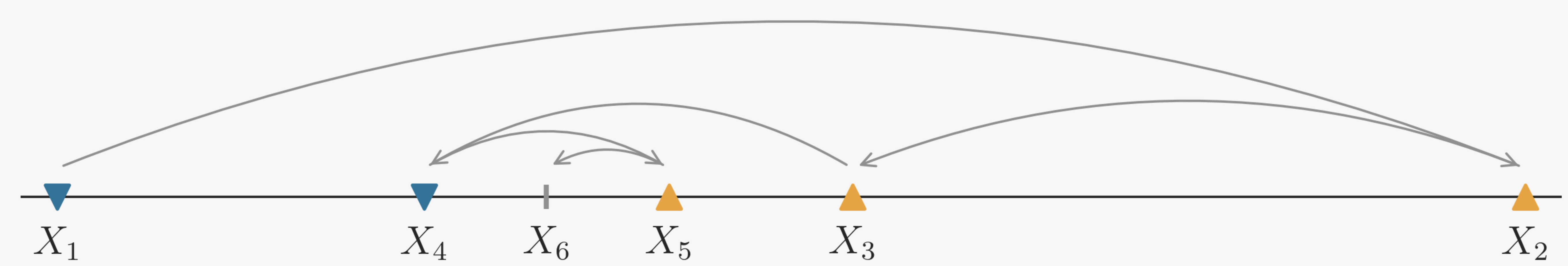
Question: can we learn from adaptive data?

- *Scientific process*—past data influences what experiment is performed next.
- *Exploratory/adaptive data analysis*—the question asked may depend on available data.
- *Online decision making*—data improves business operations, which affects further data.



TAKEAWAYS

- Adaptivity can generate **patterns out of noise**.
- Worst-case adaptive data extremely pathological.
- **Smoothed analysis** gets at more "realistic" data.
- The k_n-nearest neighbor rule is consistent under very mild statistical and structural assumptions.



Making patterns out of noise. Let $\mathscr{X} = \mathbb{R}$ and let $\eta \equiv 1/2$ everywhere (labels are generated by unbiased coin flips). If the instances X_n are sampled using the binary search procedure, this generates a data set that looks linearly separable.

LEARNING PROBLEM

Binary classification with noise

- Data space: $\mathscr{X} \times \mathscr{Y}$ where $\mathscr{Y} = \{0,1\}$
- Conditional mean label: $\eta : \mathscr{X} \to [0,1]$

$$\eta(x) \equiv \Pr(Y = 1 \,|\, X = x)$$

- Metric and measure: ρ and υ

ADAPTIVE SAMPLING

For all time $n = 1,2,\ldots$

- Adversary selects a test distribution μ_n
- Test instance $X_n \sim \mu_n$ is drawn
- Learner predicts label $\hat{Y}_n$
- Nature reveals correct label $Y_n \sim \mathrm{Ber}\big(\eta(X_n)\big)$

LEARNING GOAL: CONSISTENCY

Consistency with Bayes classifier:

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}\{\hat{Y}_n \neq Y_n^*\} = 0 \quad \mathrm{a.s.}$$

where Y_n^* is the Bayes-optimal label for X_n.

THE k_n-NEAREST NEIGHBOR RULE

Prediction rule for X_n:

- Find the k_n closest data points to X_n.
- Return the majority vote over their labels.

SMOOTHED DATA PROCESSES

Definition. A data process $\{X_n, Y_n\}_{n=1}^{\infty}$ is *uniformly dominated* υ if for all $\varepsilon > 0$, there is a $\delta > 0$ such that when a measurable set A satisfies $\upsilon(A) < \delta$, then:

$$\Pr\left(X_n \in A \,\middle|\, X_{1:n-1}, Y_{1:n-1}\right) < \varepsilon\,.$$

It is *Lipschitz dominated* or *L-smoothed* if $\varepsilon(\delta) \leq L\delta$.

MAIN RESULT

Assumptions.

- Let $(\mathscr{X}, \rho, \upsilon)$ be an upper doubling space.
- Let $(k_n)_n$ grow between $\omega(\log n)$ and $o(n/\log n)$ in a fairly smooth way.
- Let η be measurable.

Theorem. If the data process is Lipschitz dominated, then the k_n-nearest neighbor rule is consistent.