Lower Bounds for Passive and Active Learning

Maxim Raginsky
Duke / UIUC

Alexander Rakhlin
UPenn

Presenter: Maxim Raginsky

Two Learning Paradigms

$(X_1, Y_1), \dots, (X_n, Y_n)$

Passive Learning

X_1

Y_1

$\dots$

X_n

Y_n

Active Learning

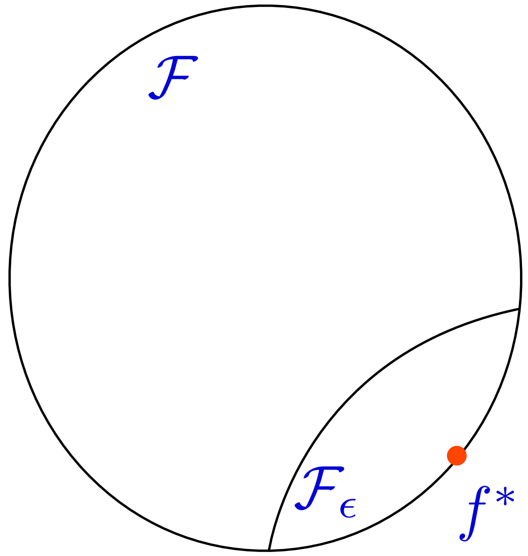
What governs the learning rate?

VC dimension   Disagreement Coefficient

Wanted:
A Unified Lower
Bound Analysis

- Vapnik-Chervonenkis Class $\text{VC-dim}(\mathcal{F}) = d$
- Hard Margin Parameter $\left| \mathbb{E}[Y|X = x] - \frac{1}{2} \right| > \frac{h}{2}$

Not all VC classes are created equal:



Alexander's Capacity Function $\tau(\epsilon)$

measure of X's on which functions in $\mathcal{F}_\epsilon$ disagree.

Supremum of this function is the disagreement coefficient

Passive Learning

$h \neq 1 \qquad n = \Omega\left(\frac{(1-\delta) d \log \tau(\varepsilon)}{\varepsilon h^2} + \frac{\log \frac{1}{\delta}}{\varepsilon h^2} \right)$

$h = 1 \qquad n = \Omega\left(\frac{(1-\delta) d}{\varepsilon} \right)$

Active Learning

$$n = \Omega\left(\frac{(1-\delta) d \log \tau(\varepsilon)}{h^2} + \frac{\tau(\varepsilon) \log \frac{1}{\delta}}{h^2} \right)$$

Tools from Information Theory

Can phrase the problem in terms of information gain on every round

Data Processing Inequality for ϕ-Divergences

$$D_\phi(\mathbb{P}_Z \| \mathbb{Q}_Z) \leq D_\phi(\mathbb{P} \| \mathbb{Q})$$

Classical Fano inequality is a consequence, but not enough for our purposes.

Freedom to choose ϕ is key

A new packing lemma allows to consider *any* active learning method