Lower Bounds for Passive and Active Learning

Maxim Raginsky
Duke / UIUC

Alexander Rakhlin
UPenn

Presenter: Maxim Raginsky
Two Learning Paradigms

Passive Learning

(\(X_1, Y_1\), \ldots, \(X_n, Y_n\))

Active Learning

\(X_1\)
\(Y_1\)
\ldots
\(X_n\)
\(Y_n\)

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**

\[ h \neq 1 \quad n = \Omega \left( \frac{(1-\delta)d \log \tau(\epsilon)}{\epsilon h^2} + \frac{\log \frac{1}{\delta}}{\epsilon h^2} \right) \]

\[ h = 1 \quad n = \Omega \left( \frac{(1-\delta)d}{\epsilon} \right) \]

**Active Learning**

\[ n = \Omega \left( \frac{(1-\delta)d \log \tau(\epsilon)}{h^2} + \frac{\tau(\epsilon) \log \frac{1}{\delta}}{h^2} \right) \]
Tools from Information Theory

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

Data Processing Inequality for $\phi$-Divergences

$$D_\phi(P_Z \parallel Q_Z) \leq D_\phi(P \parallel Q)$$

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

Freedom to choose $\phi$ is key

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