A GENERAL AGNOSTIC ACTIVE LEARNING ALGORITHM
REVIEW OF LEARNING

- Passive Learning
  + \( n \geq \frac{1}{\varepsilon} (\log |H| + \log \frac{1}{\delta}) \)
  + \( n \geq \frac{1}{\varepsilon^2} (\log |H| + \log \frac{1}{\delta}) \)
  + \( n \gtrsim \frac{1}{\varepsilon^2} (\log |VC(H)| + \log \frac{1}{\delta}) \)
REVIEW OF LEARNING

- Active Learning

Uncertainty Sampling

Pool-Based

Greedy Approach

Query by Committee
REVIEW OF LEARNING

- Leads to Agnostic Algorithm
  - Deals with noisy data
  - Good label complexity
  - No region of uncertainty
Collect unlabeled data
- Actively request labels $L$ until there is a single relevant hypothesis consistent with $L$
- Output any consistent hypothesis with those labels.

$$Pr(error_{true} \leq \epsilon) \geq 1 - \delta$$
MODIFICATIONS

actively request labels L until there is a single relevant hypothesis consistent with L

- no particular order
- request or guess label
MODIFICATIONS

Actively request labels L until there is a single relevant hypothesis consistent with L

- goal is label all points
MODIFICATIONS

Actively request labels L until there is a single relevant hypothesis *consistent with L*

- Minimize error on L
- Consistent on “guessed” points
Since the errors are similar, we request and find the is label.
(x₁, x₂, ..., xₘ) i.i.d. from Dₓ

S = ∅ and T = ∅ are sets of unlabeled and labeled data

For each x ∈ x₁..m

Let h⁺ = LEARN(Sₙ₋₁ ∪ (x, +1), Tₙ₋₁)
Let h⁻ = LEARN(Sₙ₋₁ ∪ (x, −1), Tₙ₋₁)
If |err(h⁺, Sₙ₋₁ ∪ Tₙ₋₁) − err(h⁻, Sₙ₋₁ ∪ Tₙ₋₁)| > Δₙ₋₁

y = min(err(h⁺, Sₙ₋₁ ∪ Tₙ₋₁), err(h⁻, Sₙ₋₁ ∪ Tₙ₋₁))
Sₙ = Sₙ₋₁ ∪ (x, y) and Tₙ = Tₙ₋₁

Else

y = request(x)
Sₙ = Sₙ₋₁ and Tₙ = Tₙ₋₁ ∪ (x, y)

Return h = LEARN(Sₘ, Tₘ)
ACHIEVEMENTS

- Fallback Guarantee
  - i.i.d generalization bounds for $S \cup T$
  - mathematically involved... read the paper 😊
  - $err_n(h) - err_n(h') = \hat{err}_n(h) - \hat{err}_n(h')$
  - choose $\Delta_n$
ACHIEVEMENTS

- Rate Improvement
  - Achieve exponential improvement for some situations
  - Can analyze in terms of disagreement coefficient
DISAGREEMENT COEFFICIENT

- Metric on the hypothesis space
- Measures how “different” hypothesis are
- Idea: can eliminate hypothesis that are sufficiently different from current ‘best hypothesis’
DISAGREEMENT COEFFICIENT

- Lots of symbols....!
  \[ \rho(h, h') = Pr[h(x) \neq h'(x)] \]
  \[ B(h, r) = \{ h' \in H : \rho(h, h') \leq r \} \]
  \[ \theta = \sup \left\{ \frac{Pr[\exists h \in B(h^*, r) \text{ s.t. } h(x) \neq h^*(x)]}{r} : r \geq \epsilon + \nu \right\} \]

- Maximize the percentage of points x such that there exists some hypothesis “close” to h* that doesn’t agree with h* on x (ish...)
- Turns out linear separators have \[ \theta = \sqrt{d} \]
LABEL COMPLEXITY

- Label Complexity linear in $\theta$
- Previous algorithm using this was $\theta^2$
- For linear separators, exponential improvement

$$O(\theta VC(H) \log^2(\frac{1}{\epsilon}))$$
EXPERIMENTAL RESULTS
EXPERIMENTAL RESULTS
SUMMARY

- General Agnostic Active Learning Algorithm
  - Decides whether to label each point instead of actively labeling
  - Can deal with noisy data
  - Good Label Complexity

- Disagreement Coefficient