CS101.2
February 5th, 2009
Paper by Sanjoy Dasgupta, Daniel Hsu, and Claire Monteleoni
Slides by David Lee

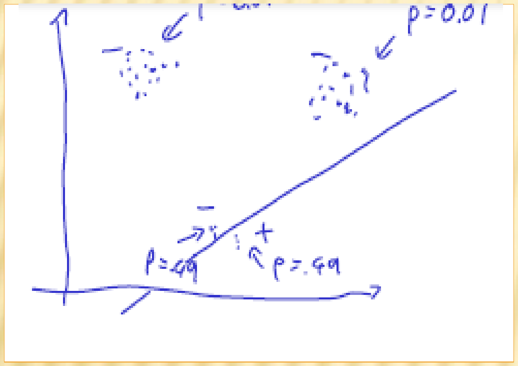
A GENERAL AGNOSTIC ACTIVE LEARNING ALGORITHM

REVIEW OF LEARNING

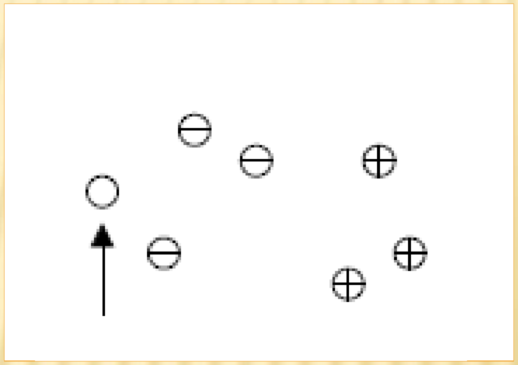
- Passive Learning
  - $n \geq \frac{1}{\epsilon}(log|H| + log\frac{1}{\delta})$
  - $n \geq \frac{1}{\epsilon^2}(log|H| + log\frac{1}{\delta})$
  - $n \gtrsim \frac{1}{\epsilon^2}(log|VC(H)| + log\frac{1}{\delta})$

REVIEW OF LEARNING

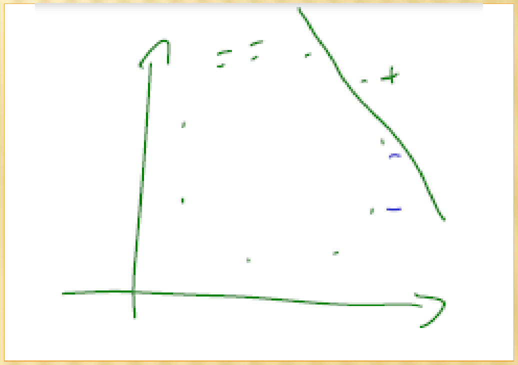
- Active Learning



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

ORIGINAL POOL-BASED

- 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

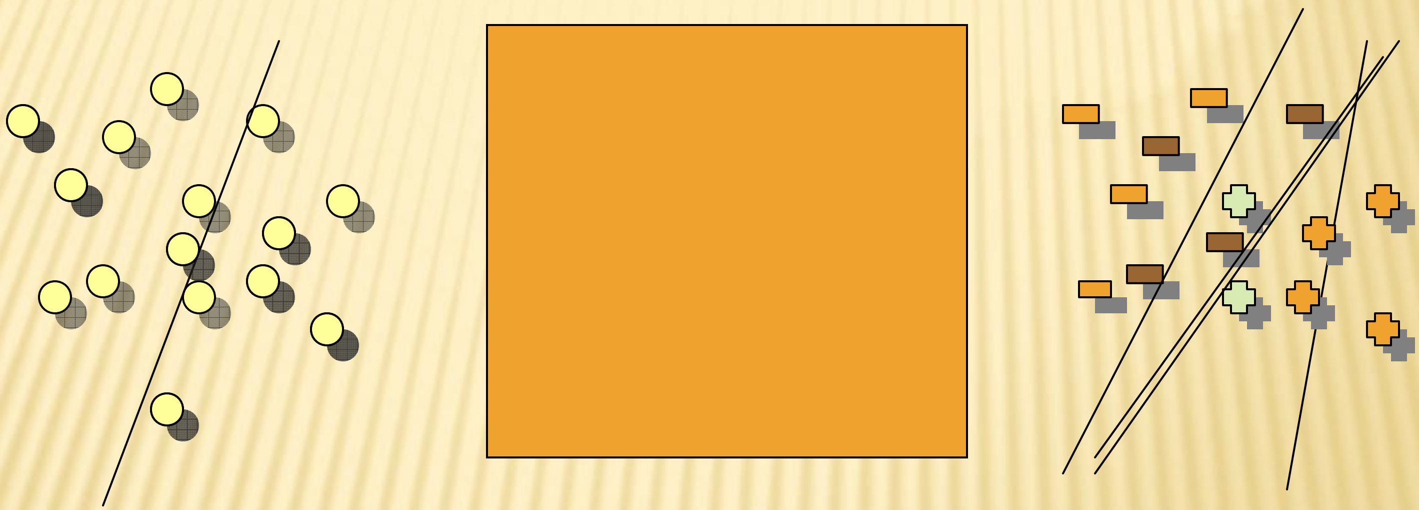
- goal is label all points

MODIFICATIONS

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

- Minimize error on L
- Consistent on "guessed" points

VISUAL ALGORITHM



Since the number of + is larger, we guess that it is +

Since the neighbors are all “unfamiliar”, we request and find label

FORMAL ALGORITHM

$(x_1,\ x_2,\ \ldots,\ x_m)$ i.i.d. from D_x

$S = \emptyset$ and $T = \emptyset$ are sets of unlabeled and labeled data

For each $x \in x_{1..m}$

Let $h_+ = LEARN(S_{n-1} \cup (x, +1), T_{n-1})$

Let $h_- = LEARN(S_{n-1} \cup (x, -1), T_{n-1})$

If $|err(h_+, S_{n-1} \cup T_{n-1}) - err(h_-, S_{n-1} \cup T_{n-1})| > \triangle_{n-1}$

$y = min(err(h_+, S_{n-1} \cup T_{n-1}), err(h_-, S_{n-1} \cup T_{n-1}))$

$S_n = S_{n-1} \cup (x, y)$ and $T_n = T_{n-1}$

Else

$y = request(x)$

$S_n = S_{n-1}$ and $T_n = T_{n-1} \cup (x, y)$

Return $h = LEARN(S_m, T_m)$

ACHIEVEMENTS

- Fallback Guarantee
  - i.i.d generalization bounds for $S \cup T$
  - mathematically involved… read the paper ☺
  - $err_n(h) - err_n(h') = \widehat{err}_n(h) - \widehat{err}_n(h')$
  - choose $\triangle_n$

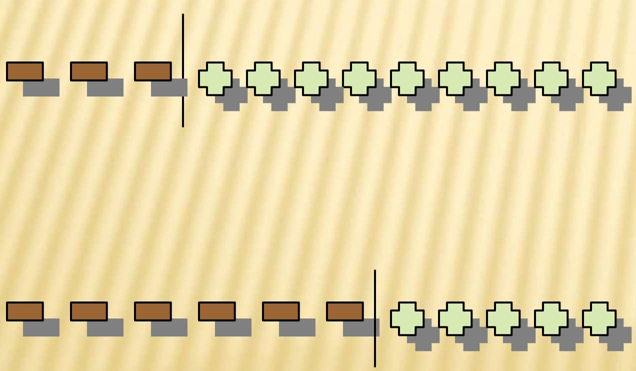
ACHIEVEMENTS

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

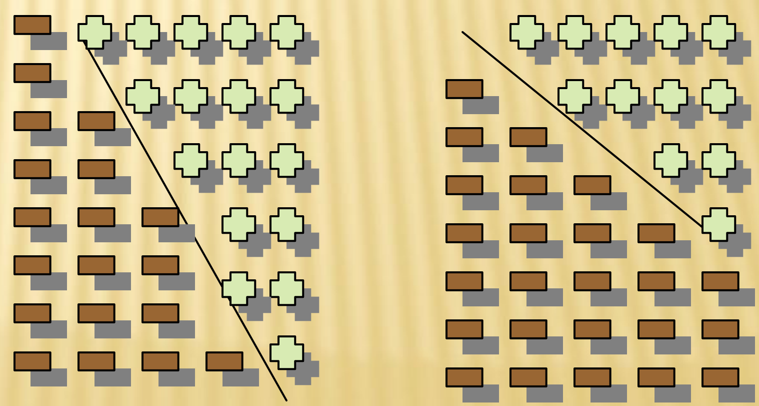
DISAGREEMENT COEFFICIENT

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

$\mathbb{R}$



$\mathbb{R}^2$



DISAGREEMENT COEFFICIENT

- Lots of symbols….!

$$\rho(h,h') = Pr[h(x) \neq h'(x)]$$

$$B(h,r) = \{h' \in H : \rho(h,h') \leq r\}$$

$$\theta = sup\{\frac{Pr[\exists h \in B(h^*,r) \text{ s.t. } h(x) \neq h^*(x)]}{r} : r \geq \epsilon + \nu\}$$

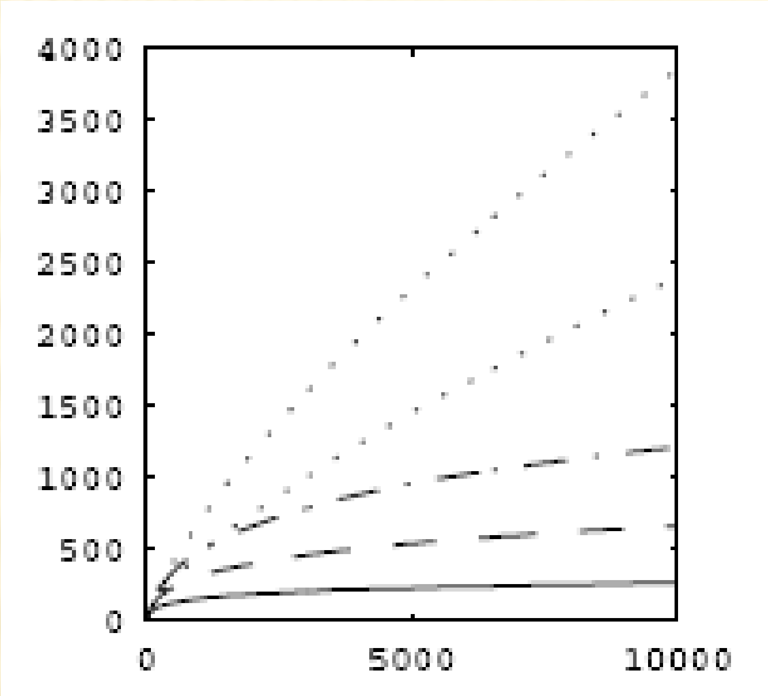
- Maximize the percentage of points x such that there exists some hypothesis "close" to h* that doesn't agree with h* on x (ish…)
- Turns out linear separators have $\theta = \sqrt{d}$

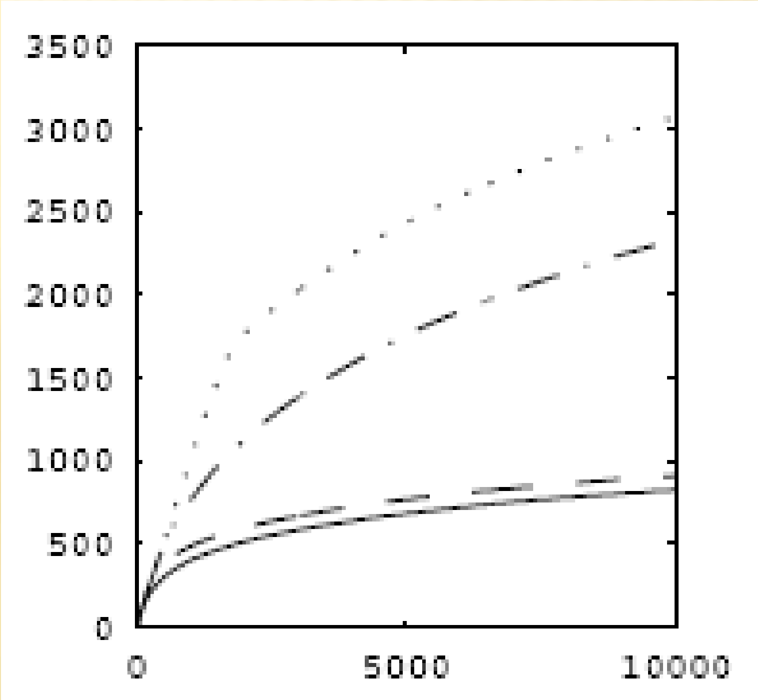
LABEL COMPLEXITY

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

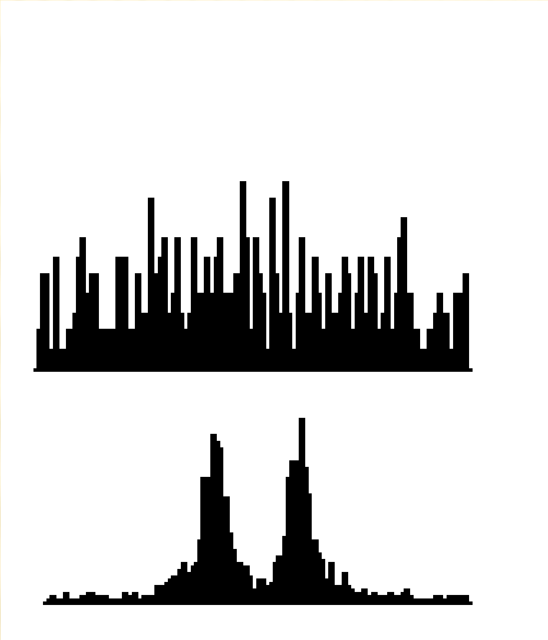
$$O(\theta VC(H) log^2(\frac{1}{\epsilon}))$$

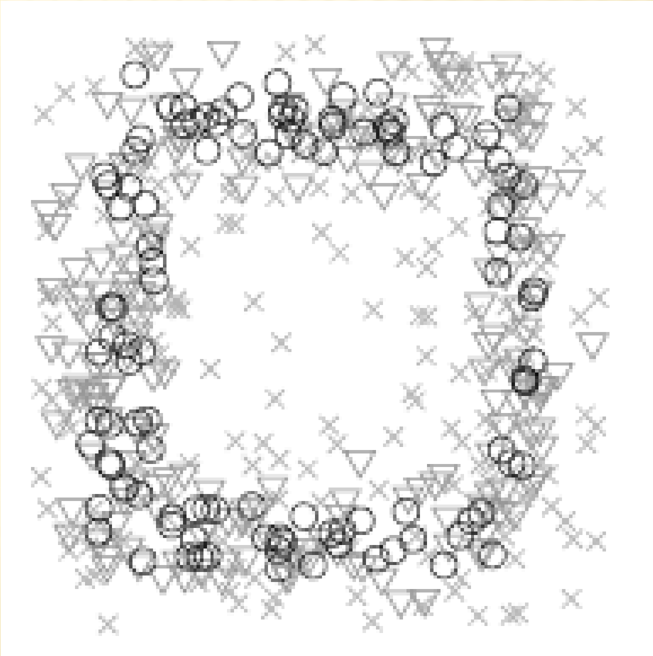
EXPERIMENTAL RESULTS





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