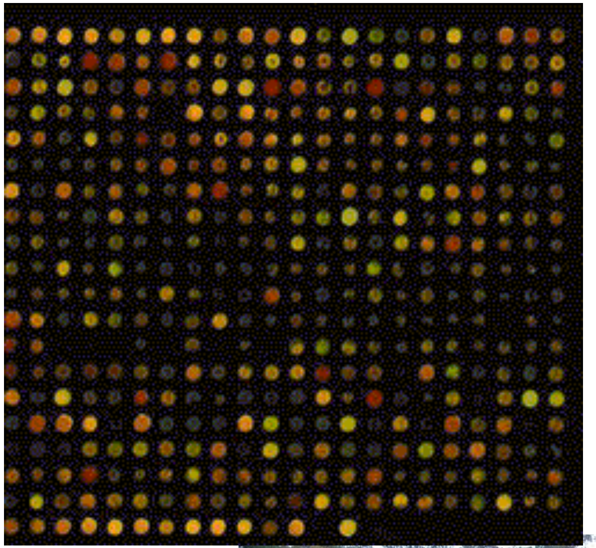
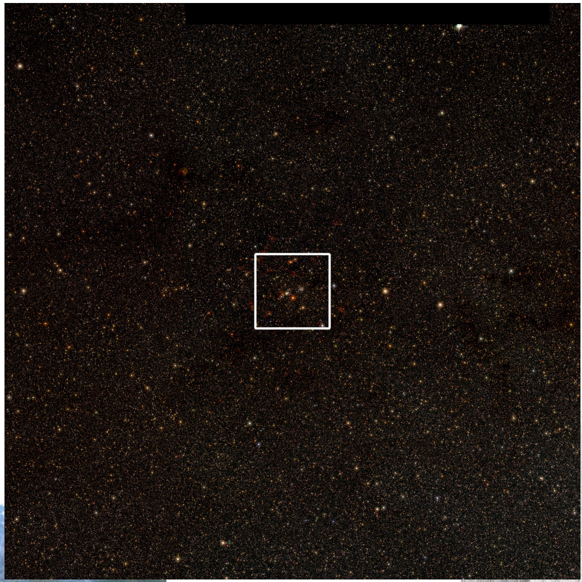
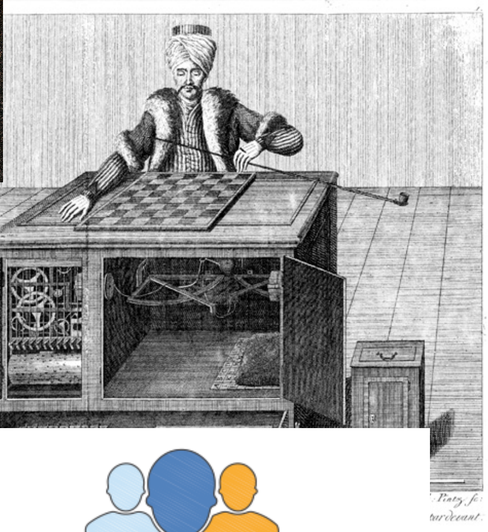
John Lipor and Laura Balzano, University of Michigan
lipor@umich.edu and girasole@umich.edu

Quantile Search for Minimum-Time Sampling

Active Learning

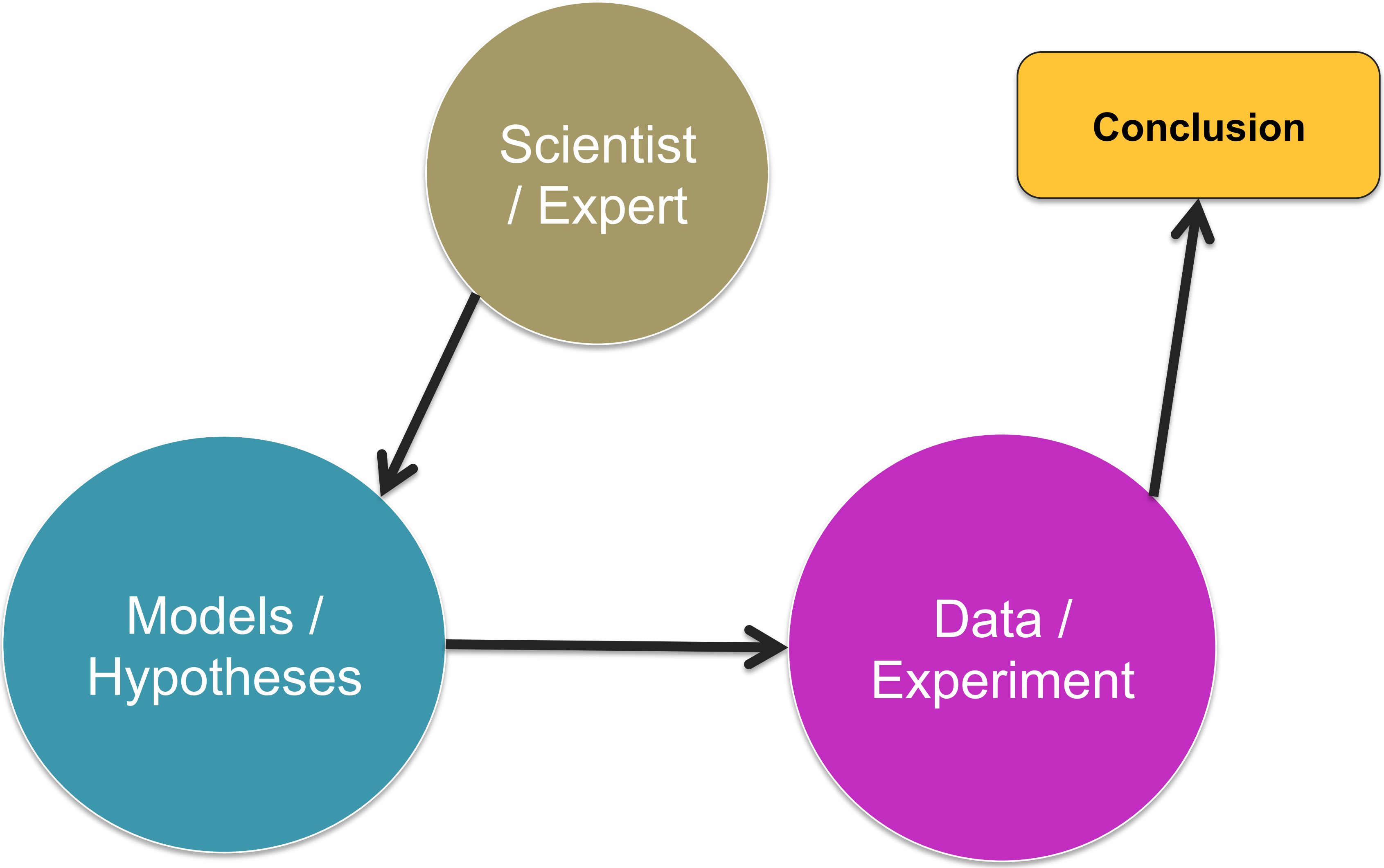






Images courtesy nih.gov, ras.org.uk, Center for Embedded Networked Sensing, wikipedia

Passive Learning



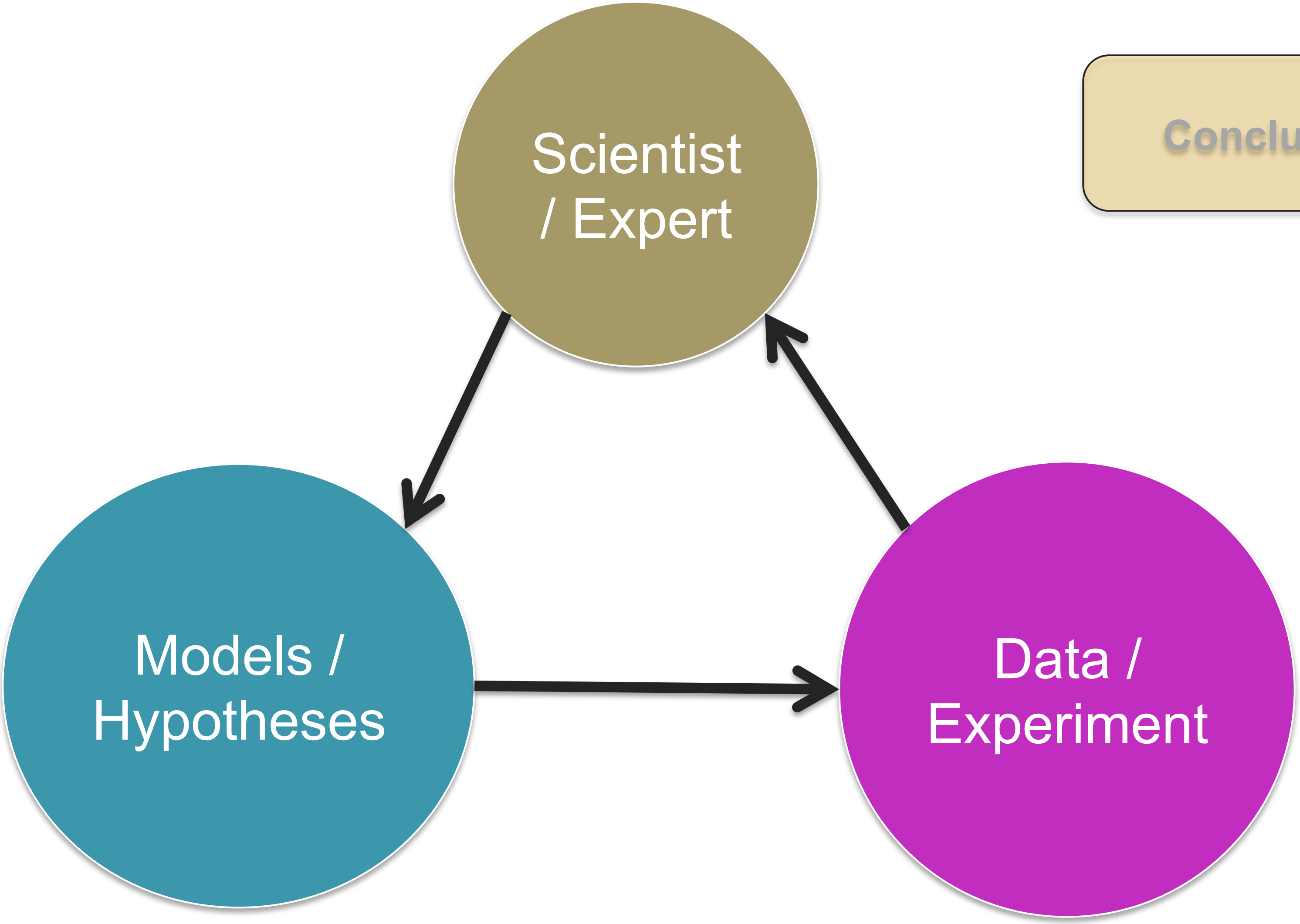
Scientist / Expert

Models / Hypotheses

Data / Experiment

Conclusion

Active Learning



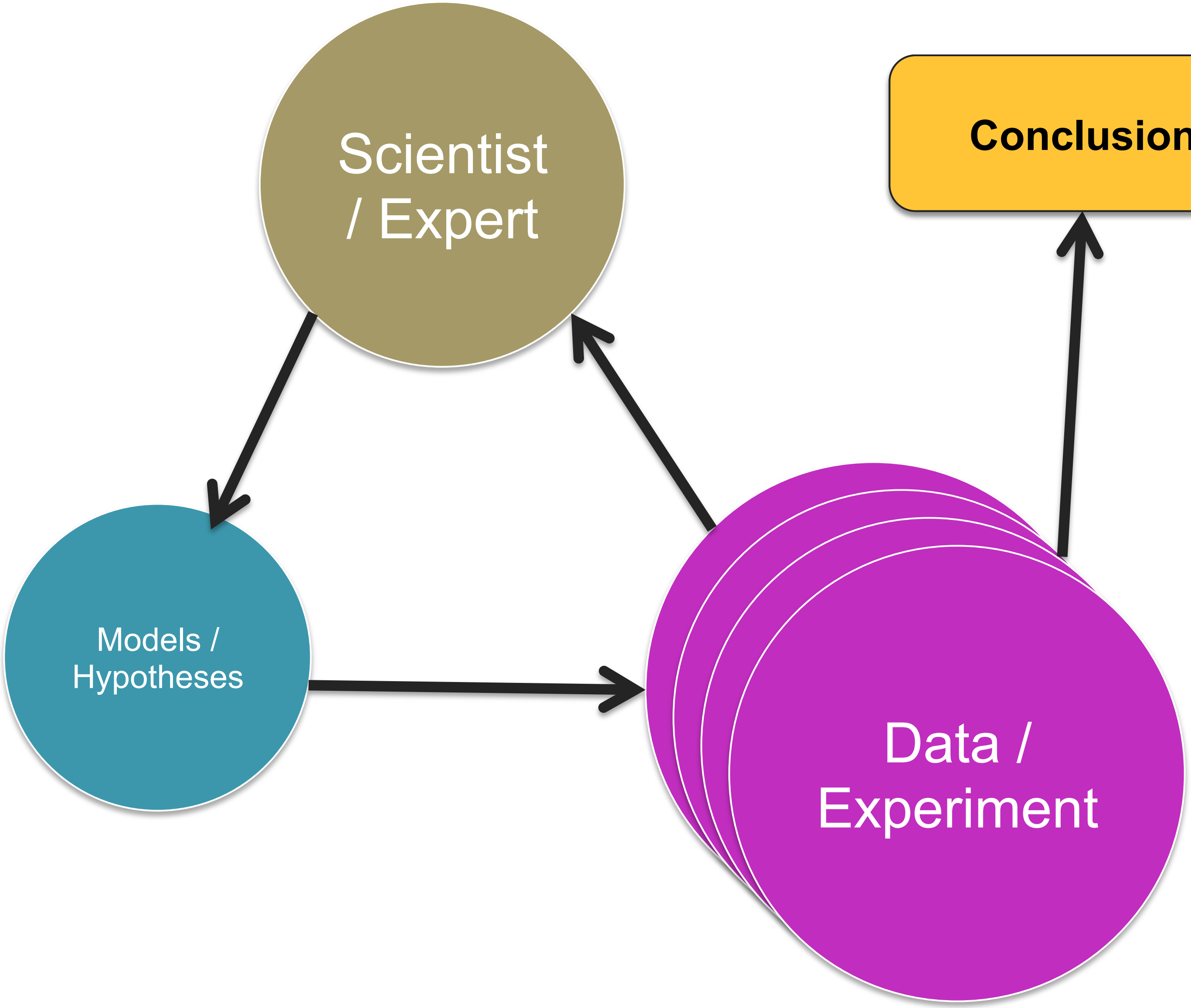
Scientist / Expert

Models / Hypotheses

Data / Experiment

Conclusion

Active Learning



Scientist / Expert

Conclusion

Models / Hypotheses

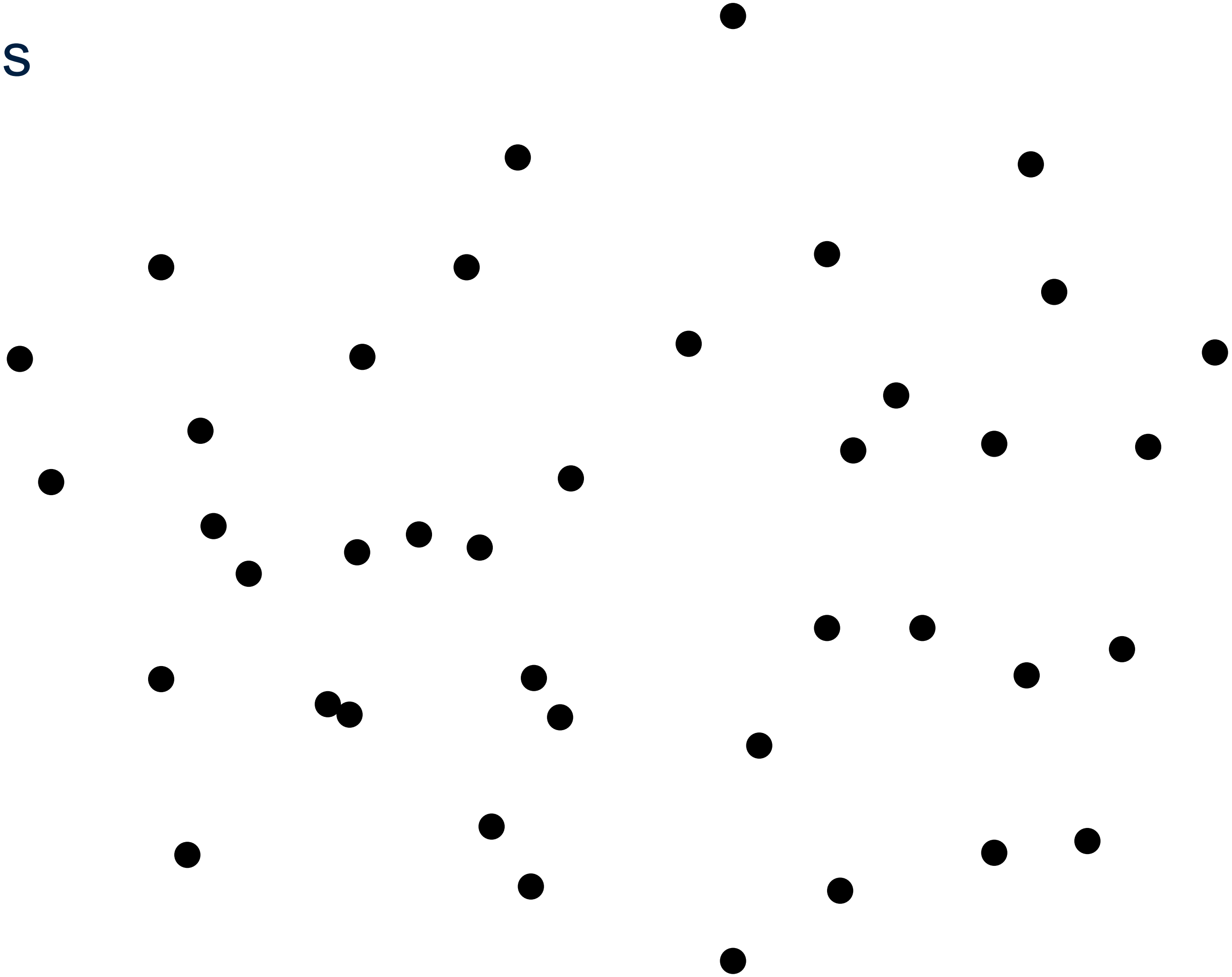
Data / Experiment

Outline

- Active Learning
- We have a big lake (the Great Lakes specifically)
  - Inter-sample spacing matters
- Problem formulation
- Deterministic Quantile Search
- Probabilistic Quantile Search

Active Classifier Learning

We seek a two class linear classifier.

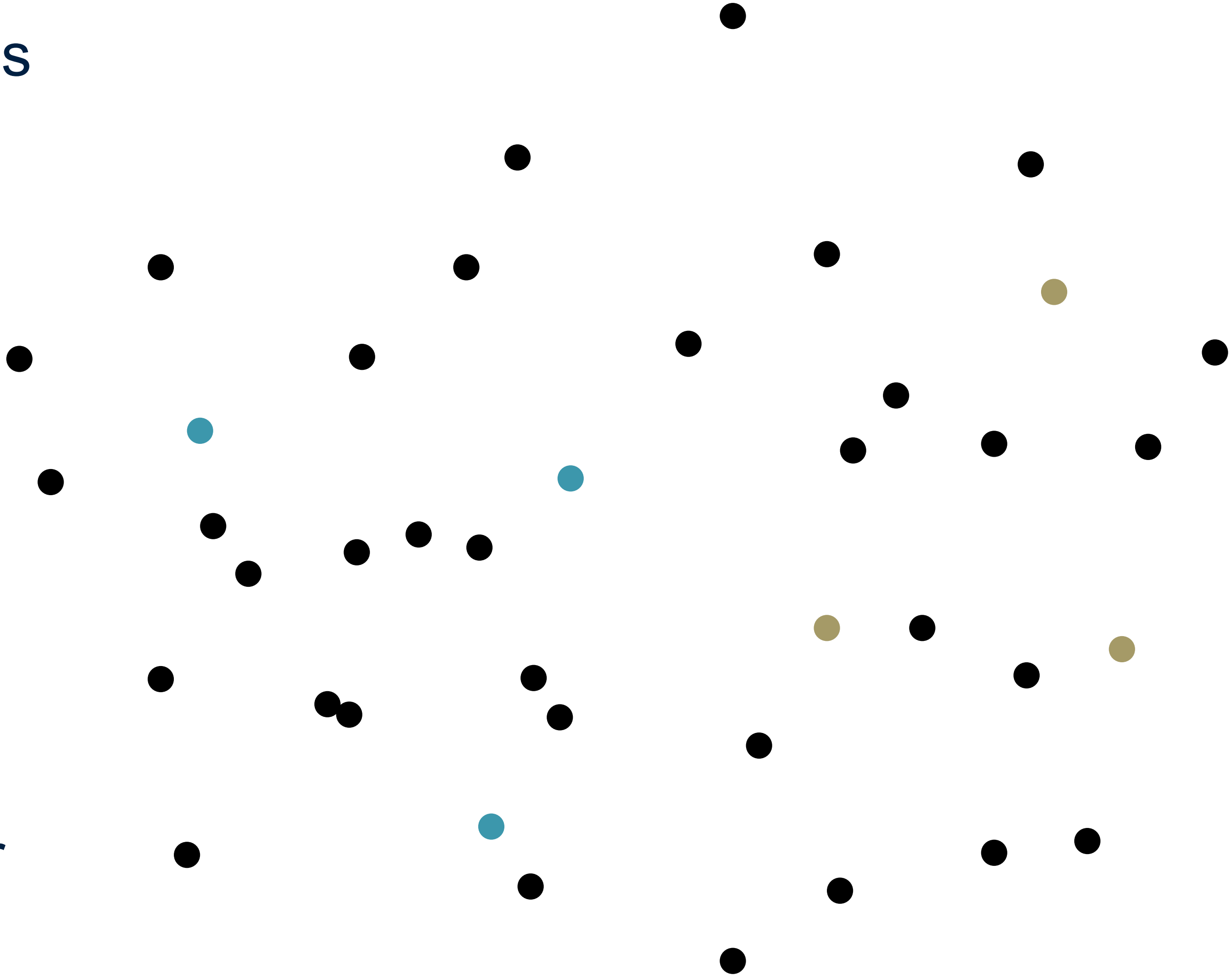


Active Classifier Learning

We seek a two class linear classifier.

Suppose we are given a small number of labels.

The idea of active learning is that we may request further labels.

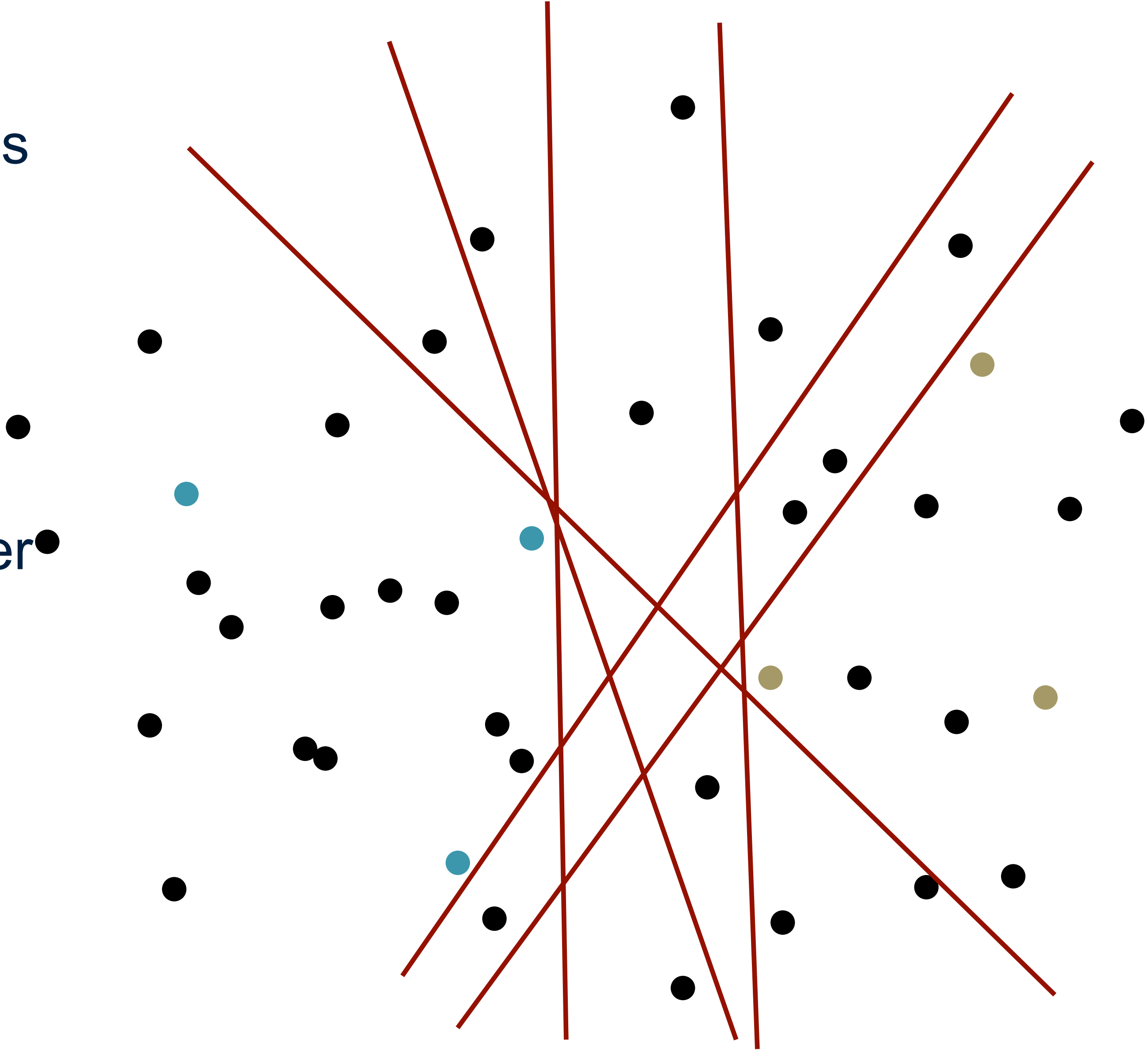


Active Classifier Learning

We seek a two class linear classifier.

Suppose we are given a small number of labels, and we can consider the remaining possible classifiers.

Where would we request a label?



Guess Who?

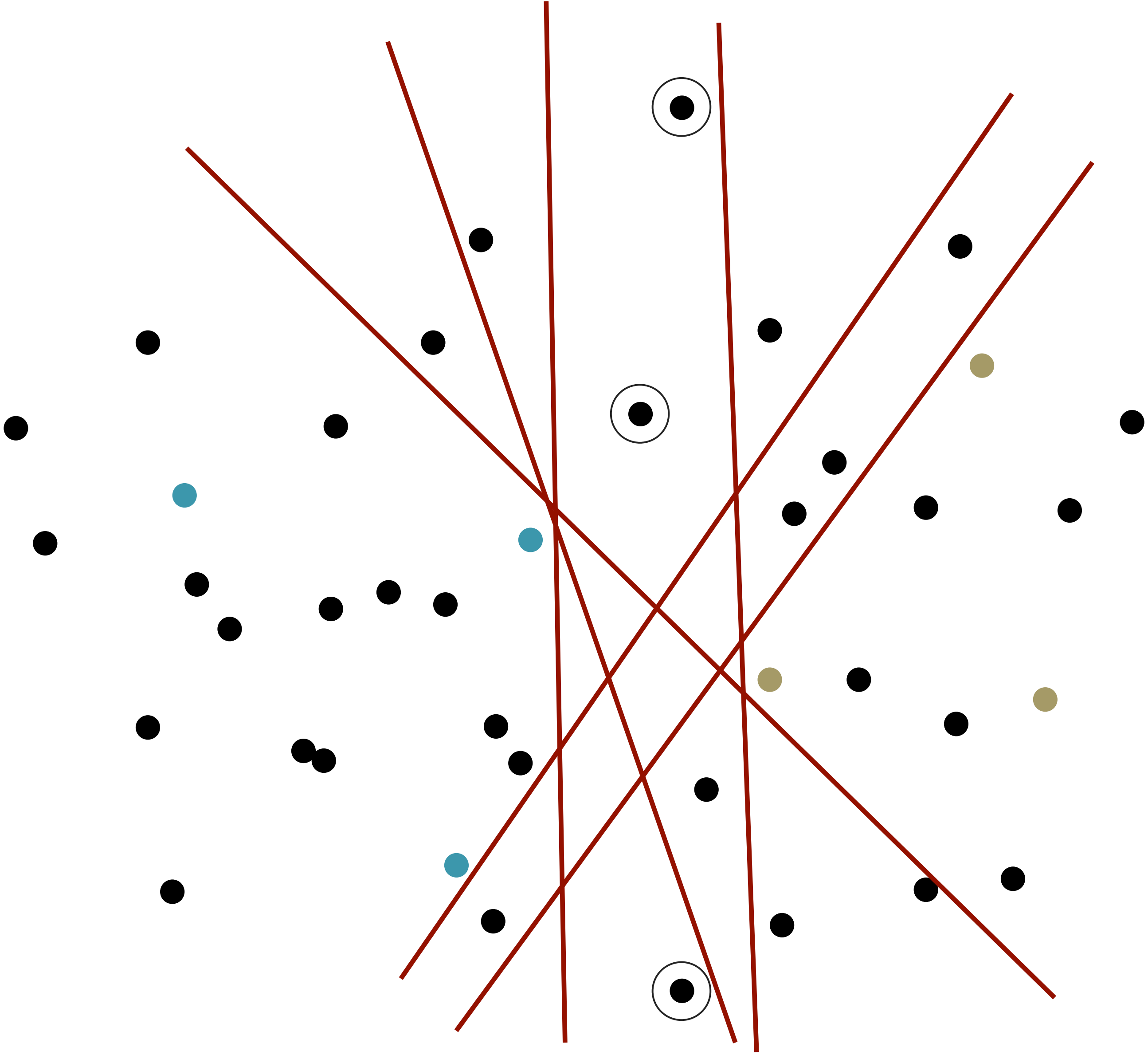
“Is the person male or female?”

“Is she wearing a hat?”

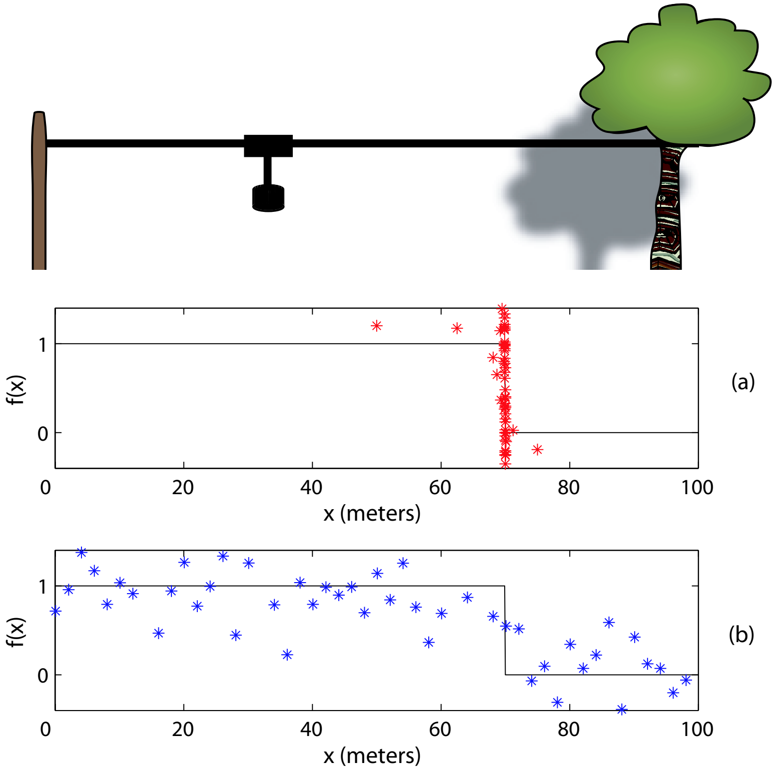
We try as best as we can to cut the remaining possibilities *in half*.

Active Classifier Learning

The same principle applies here: We may choose the query that *throws away half* of the remaining possible linear classifiers.

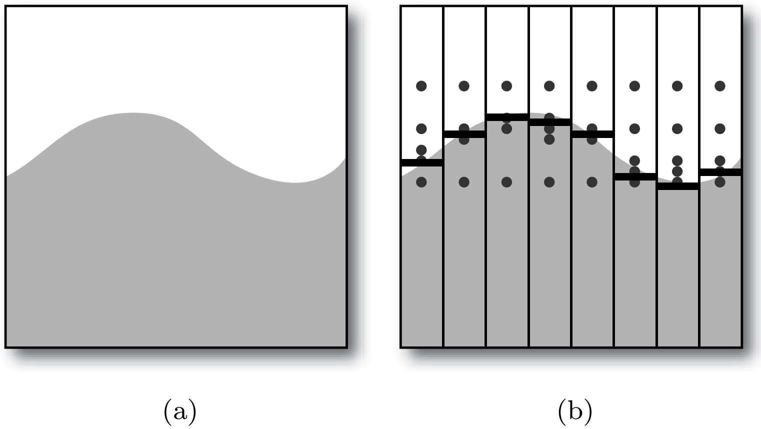


Active Threshold Estimation



Learn a 1D transition.

Figure from “Active Learning for Adaptive Mobile Sensing Networks” by Singh, Nowak, and Ramanathan. IPSN 2006



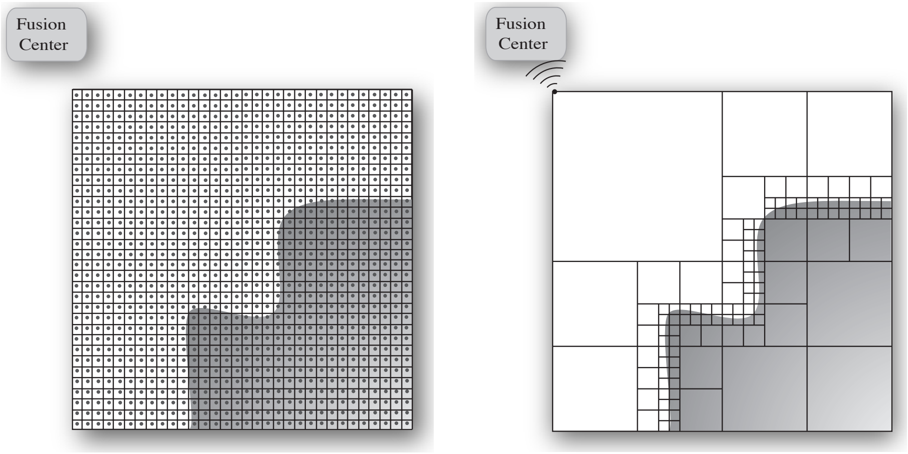
Learn a 2D boundary fragment.

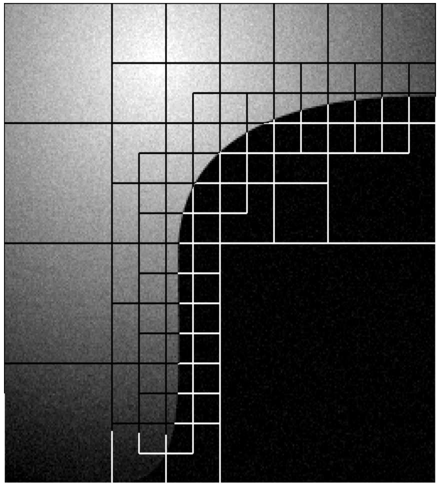
Figure from “Active Learning and Sampling” by Castro and Nowak in *Foundations and Applications of Sensor Management*, 2008.

Active Function Estimation

Estimate a function.

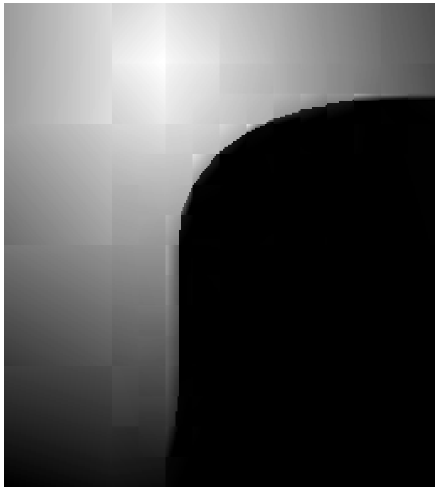
Figure from “Backcasting: Adaptive Sampling for Sensor Networks” by Willett, Martin, and Nowak. IPSN 2004





Preview subset of 256x256 sensors

All sensors used



Only 20% sensors used

Outline

- Active Learning
- We have a big lake (the Great Lakes specifically)
  - Inter-sample spacing matters
- Problem formulation
- Deterministic Quantile Search
- Probabilistic Quantile Search

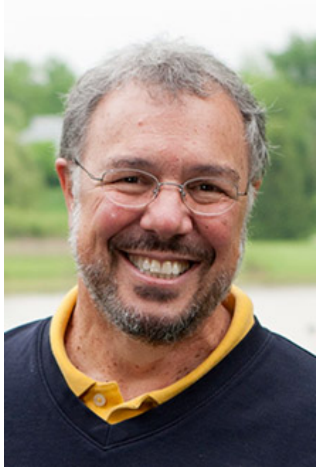
Sampling the Great Lakes

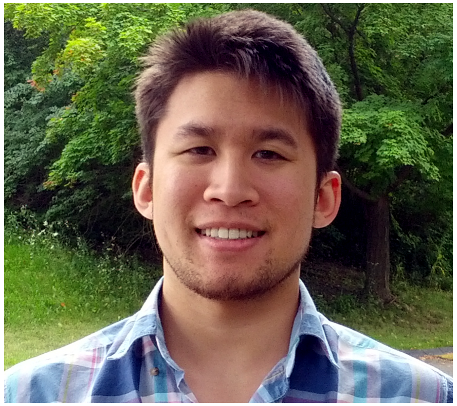
MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

Source: NYTimes 2012

Sampling the Great Lakes

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

Hypoxia in Lake Erie

The hypoxic region is at the lake bottom, where oxygen is scarce. We seek to estimate the spatial extent.

In black we see the largest observed hypoxic zone of Erie for a given year.

Estimates based on ~4 sampling cruises per year, each giving ~360 samples. Estimated ~2500 pixel values.

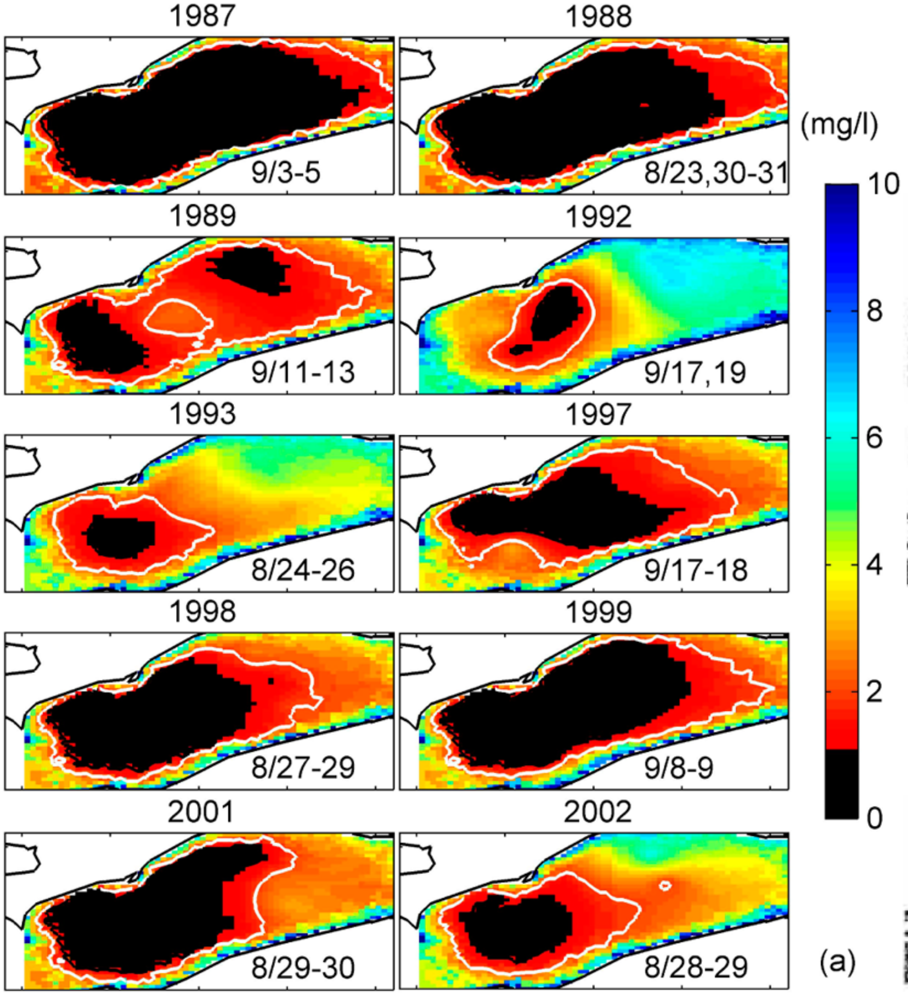
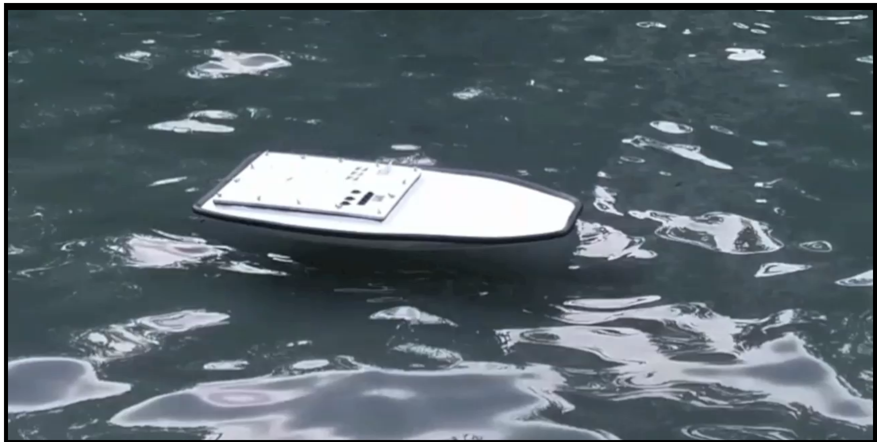


Figure from “Spatial and Temporal Trends in Lake Erie Hypoxia, 1987−2007”. Zhou, Obenour, Scavia, Johengen, Michalak. ACS Journal of Environmental Science and Technology, 2012.

The sampling system



Lurie Fountain

The Grove

Images courtesy Branko Kerkez, senseplatypus.com, sontek.com, arduino.cc, forbes.com, amazon.com, arc-ts.umich.edu

Outline

- Active Learning
- We have a big lake (the Great Lakes specifically)
  - Inter-sample spacing matters
- Problem formulation
- Deterministic Quantile Search
- Probabilistic Quantile Search

Problem Formulation

Our function comes from the class of step functions:

$$\mathcal{F} = \{ f : [0,1] \to \mathbb{R} | f(x) = \mathbf{1}_{[0,\theta)}(x) =: f_\theta(x) \} \,.$$

We wish to recover θ from n noisy measurements Y_i, $i = 1, \ldots, n$ where

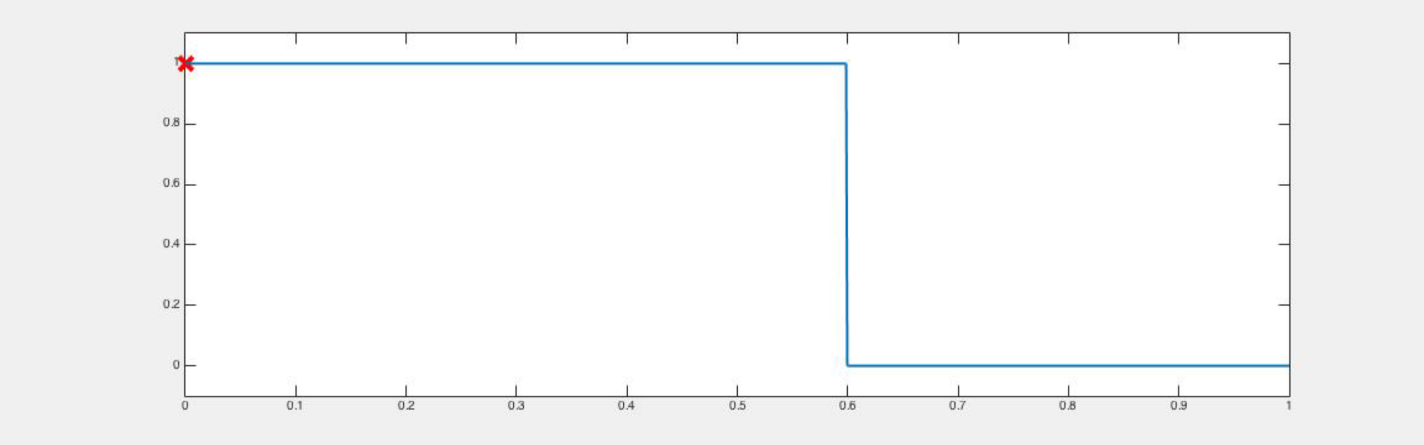
$$Y_i = \begin{cases} f_\theta(X_i) & \text{with probability } 1-p \\ 1 - f_\theta(X_i) & \text{with probability } p \end{cases}$$

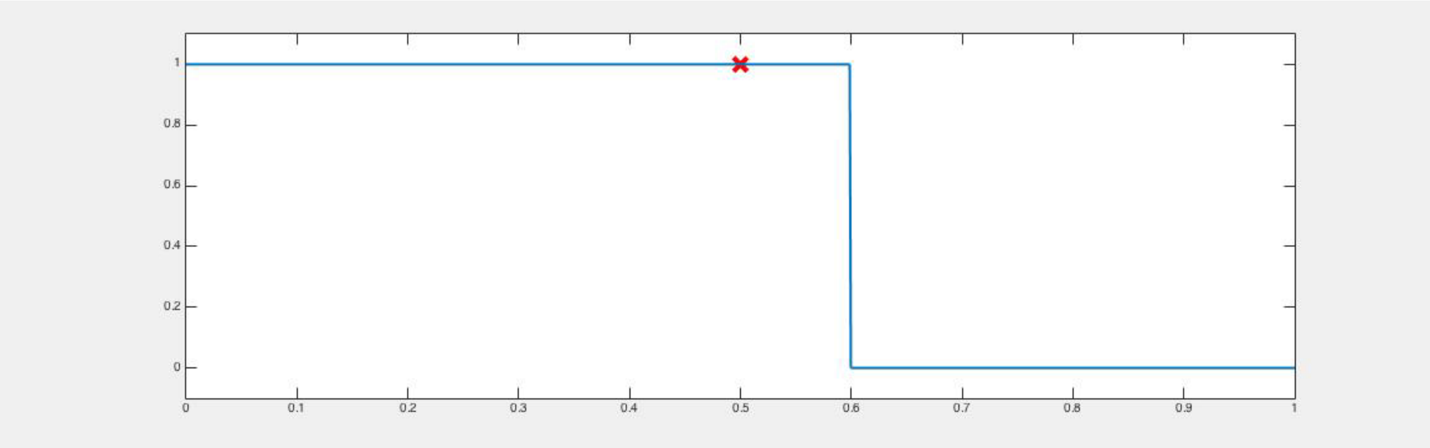
and the X_i can be chosen based on previous pairs (X_j, Y_j), $j = 1, \ldots, i-1$.

Denote our estimate of θ after n samples as $\widehat{\theta}_n$. We consider either the worst-case or expected error,

$$\sup_{\theta \in [0,1]} \left| \widehat{\theta}_n - \theta \right| \quad \text{or} \quad \mathbb{E}\left[\left| \widehat{\theta}_n - \theta \right| \right] \,.$$

Sample Complexity p=0





Optimal Sample Complexity

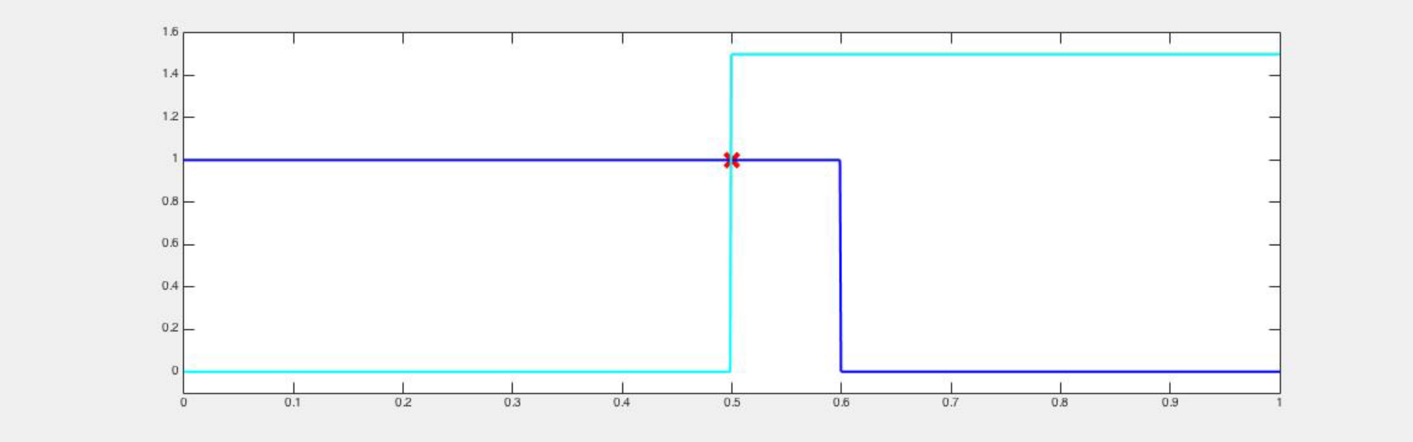
	worst case	expected (with uniform prior on θ)
adaptive (binary bisection optimal)	$\sup_{\theta\in[0,1]} \left\lvert \widehat{\theta}_n - \theta \right\rvert \le \frac{1}{2^{n+1}}$	$\mathbb{E}\left[\left\lvert \widehat{\theta}_n - \theta \right\rvert\right] = \frac{1}{2^{n+2}}$
non-adaptive (uniform grid optimal)	$\sup_{\theta\in[0,1]} \left\lvert \widehat{\theta}_n - \theta \right\rvert \le \frac{1}{2}\frac{1}{(n+1)}$	$\mathbb{E}\left[\left\lvert \widehat{\theta}_n - \theta \right\rvert\right] = \frac{1}{4}\frac{1}{(n+1)}$

… but worst case distance

	worst case $\sup_{\theta\in[0,1]} \left\lvert\widehat{\theta}_n - \theta\right\rvert$	expected (with uniform prior on θ) $\mathbb{E}\left[\left\lvert\widehat{\theta}_n - \theta\right\rvert\right]$
adaptive sample complexity (binary bisection optimal)	$\leq \frac{1}{2^{n+1}}$	$= \frac{1}{2^{n+2}}$
adaptive distance traveled (fix desired error ϵ, start at $X_i = 0$)	$= 1 - \epsilon$	$= 1 - 2\epsilon$

Quantile Search

Binary search is nothing other than taking a sample at the 2-quantile of the posterior distribution for θ.

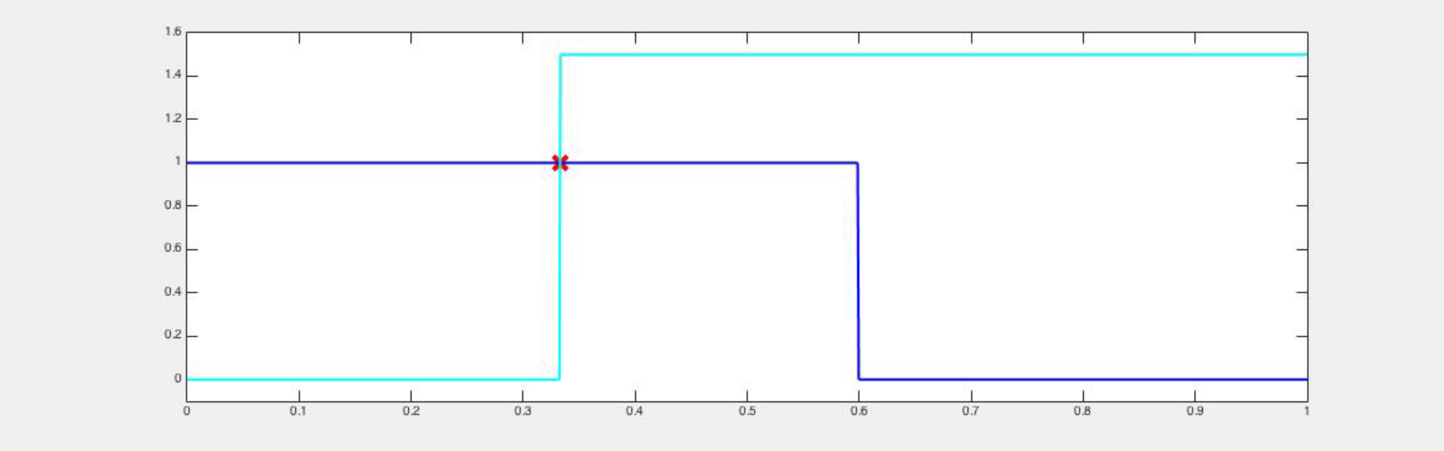


Quantile search generalizes this by instead taking a sample at the first m-quantile of the posterior distribution for θ.

Quantile Search

Quantile search generalizes this by instead taking a sample at the first m-quantile of the posterior distribution for θ.

E.g. m=3:



Choosing m allows for a tradeoff between number of samples and distance traveled.

Quantile Search Cost

	worst case $\sup_{\theta\in[0,1]} \left\|\widehat{\theta}_n - \theta\right\|$	expected (with uniform prior on θ) $\mathbb{E}\left[\left\|\widehat{\theta}_n - \theta\right\|\right]$
bisection sample complexity	$\leq \frac{1}{2^{n+1}}$	$= \frac{1}{2^{n+2}}$
m-quantile search sample complexity $\rho = \frac{m-1}{m}$	$\leq \frac{1}{2}\rho^n$	$= \frac{1}{4}\left(\rho^2 + (1-\rho)^2\right)^n$
bisection distance traveled (as desired error $\epsilon \to 0$)	$= 1$	$= 1$
m-quantile search distance traveled (as desired error $\epsilon \to 0$)	$= 1$	$= \frac{m}{2m-2}$

Sample-Distance Tradeoff

Goal: Minimize the total sampling time subject to a given reconstruction error.

$$\begin{aligned} \min \quad & T = \gamma N + \eta D \\ \text{subject to} \quad & \left| \theta - \hat{\theta}_n \right| \leq \epsilon \end{aligned}$$

$$\begin{aligned} \gamma &= \text{time/sample} \\ N &= \text{number of samples} \\ \eta &= \text{time/distance} \\ D &= \text{distance traveled} \end{aligned}$$

Sample-Distance Tradeoff

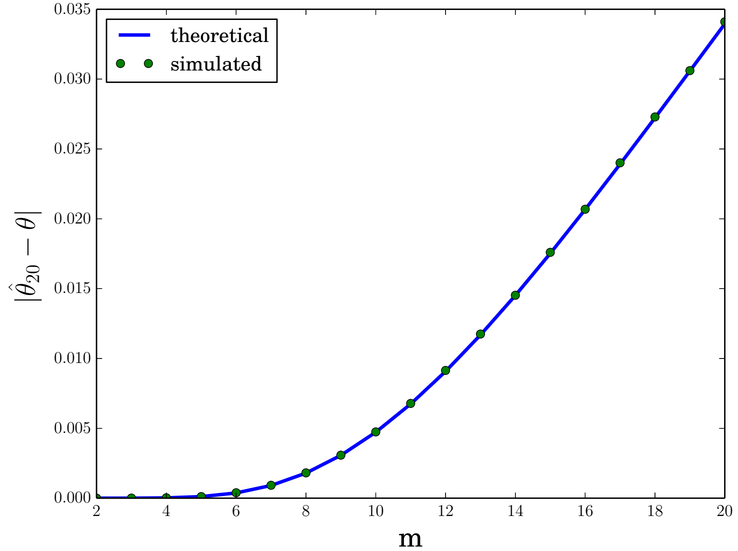
Let N be a random variable denoting the samples required to achieve an error ε, and D the distance. Rearranging the expected sample complexity, we have

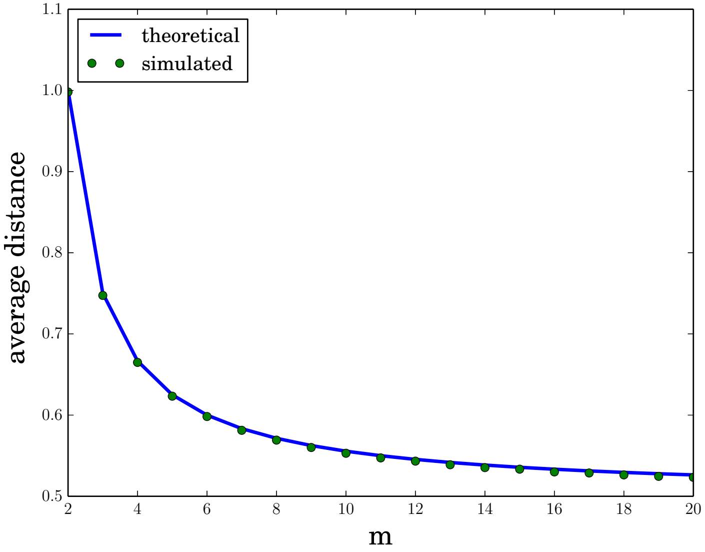
$$\mathbb{E}[N] = \frac{\log(4\varepsilon)}{\log\left(\left(\frac{m-1}{m}\right)^2 + \frac{1}{m^2}\right)} \equiv n'$$

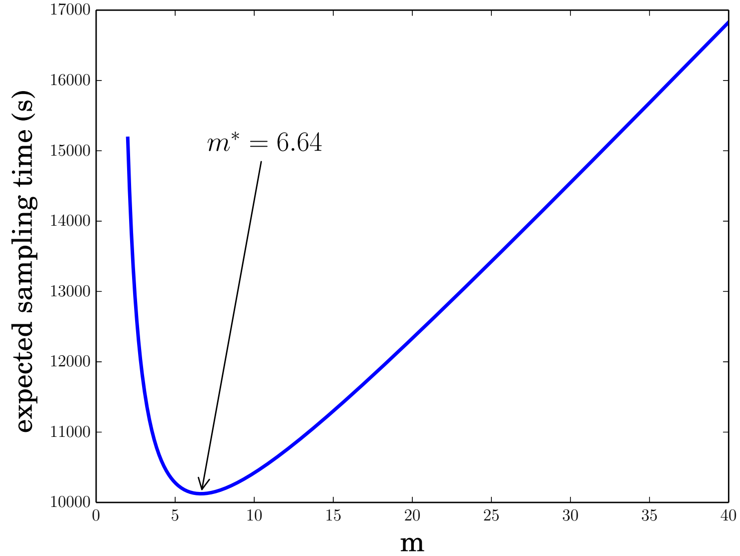
Denote the sampling time T. Then

$$\begin{aligned}
\mathbb{E}[T] &= \gamma\mathbb{E}[N] + \eta\mathbb{E}[D] \\
&= \gamma n' + \eta\left(\frac{m}{2m-2} - \frac{1}{(2m-2)(2m-1)^{n'}}\right) \\
&\approx \gamma n' + \eta\frac{m}{2m-2}
\end{aligned}$$

Sample-Distance Tradeoff





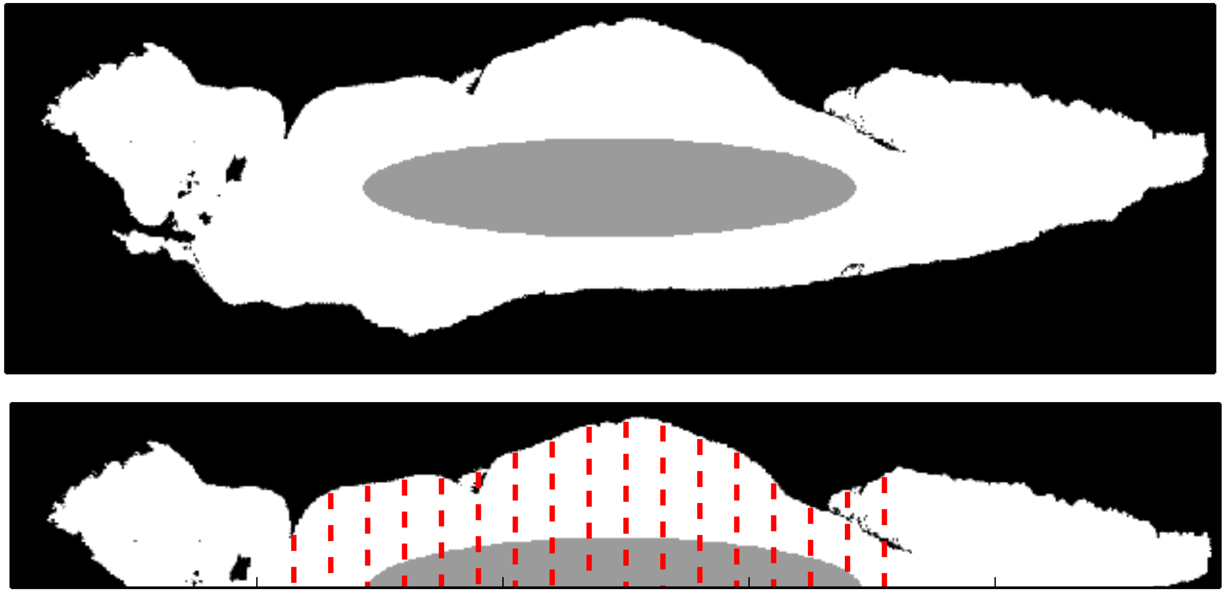


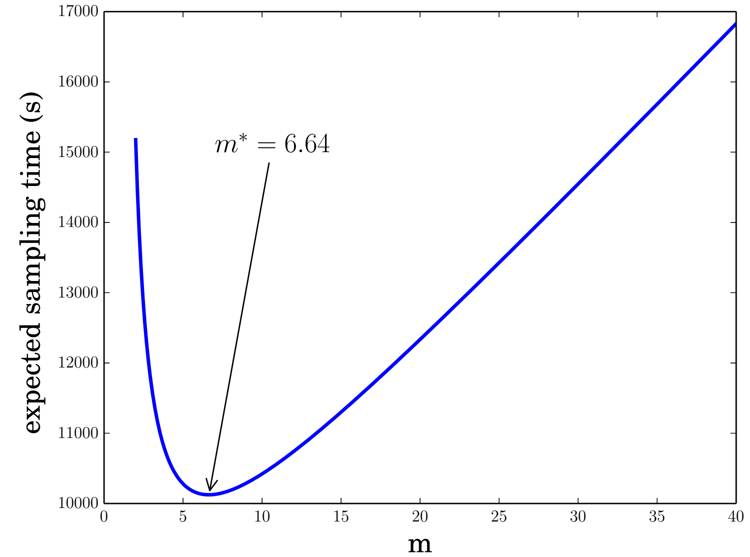
(Top left) Simulated and theoretical expected error after 20 samples as we vary m,

(Top right) Simulated and theoretical distance traveled as we vary m, and

(Bottom left) optimal m for $\gamma = 60\frac{s}{samp}$ and $\eta = \frac{1}{4}\frac{s}{m}$.

Sample-Distance Tradeoff



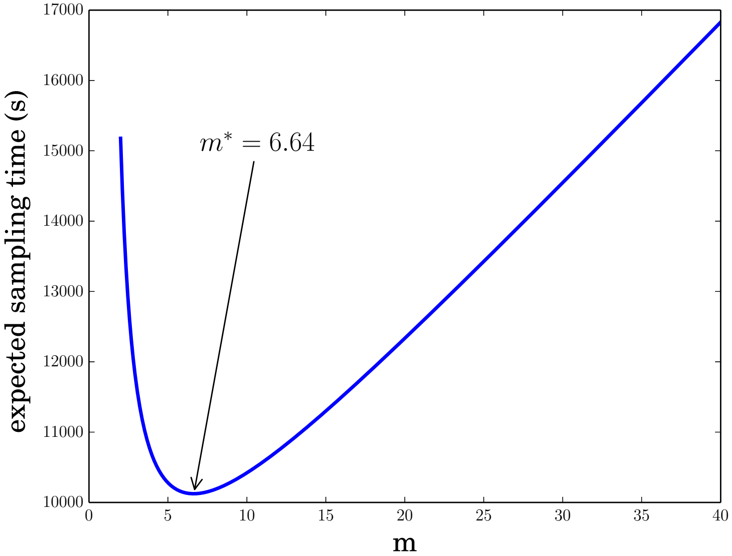


To simulate sampling all of Lake Erie, we split it first in half and then into 16 strips and perform DQS. In most cases we can sample the entire boundary in 2-3 days; fast enough to assume a stationary hypoxic zone.

(Bottom left) optimal m for $\gamma = 60\frac{s}{samp}$ and $\eta = \frac{1}{4}\frac{s}{m}$.

Sample-Distance Tradeoff

Sampling Time (s)	Speed (m/s)	m	Total Time (hrs)
60	4	2	62
60	4	6.64	43
60	2	2	123
60	2	8.92	81
10	4	2	61
10	4	14.63	35
10	2	2	122
10	2	20.26	64



To simulate sampling all of Lake Erie, we split it first in half and then into 16 strips and perform DQS. In most cases we can sample the entire boundary in 2-3 days; fast enough to assume a stationary hypoxic zone.

(Bottom left) optimal m for $\gamma = 60\frac{s}{samp}$ and $\eta = \frac{1}{4}\frac{s}{m}$.

Outline

- ✧Active Learning
- ✧We have a big lake (the Great Lakes specifically)
  - ✧ Inter-sample spacing matters
- ✧Problem formulation
- ✧Deterministic Quantile Search
- ✧Probabilistic Quantile Search

Problem Formulation

Our function comes from the class of step functions:

$$\mathcal{F} = \{f : [0,1] \to \mathbb{R} | f(x) = \mathbf{1}_{[0,\theta)}(x) =: f_\theta(x)\} .$$

We wish to recover θ from n noisy measurements Y_i, $i = 1, \ldots, n$ where

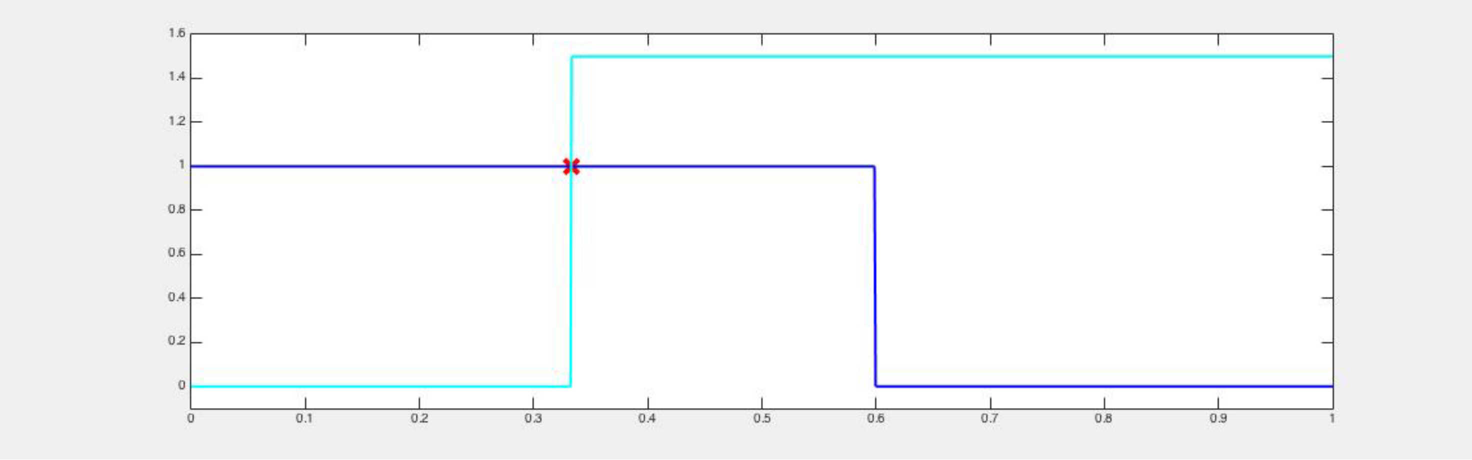
$$Y_i = \begin{cases} f_\theta(X_i) & \text{with probability } 1-p \\ 1 - f_\theta(X_i) & \text{with probability } p \end{cases}$$

and the X_i can be chosen based on previous pairs (X_j, Y_j), $j = 1, \ldots, i-1$.

You may think of this noise model as modeling a detector where the probability of false alarm or false detection are the same, *e.g.* detection after additive Gaussian noise.

Noisy case p>0

If we were to update our posterior according to the measurements without taking the noise probability into consideration, we'd get lost:



Probabilistic Quantile Search

Rather than sampling at 1/m into the feasible interval, we sample 1/m into the posterior distribution on θ (See Burnashev and Zigangirov, "An interval estimation problem for controlled observations," Problems in Information Transmission, 1974 for special case bisection algorithm).

Algorithm 1 Probabilistic Quantile Search (PQS)

```
initialize prior density $\pi_0(\theta) = 1$ for $\theta \in [0, 1]$
while not converged do
    choose $X_n$ such that $\int_0^{X_n} \pi_n(x)dx = 1/m$
    $Y_n \leftarrow f(X_n)$
    perform Bayesian update to obtain $\pi_{n+1}(x)$
end while
return $\hat{\theta}_n$ such that $\int_0^{\hat{\theta}_n} \pi_{n+1}(x)dx = 1/2$
```

Probabilistic Quantile Search

MICHIGAN ENGINEERING

UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

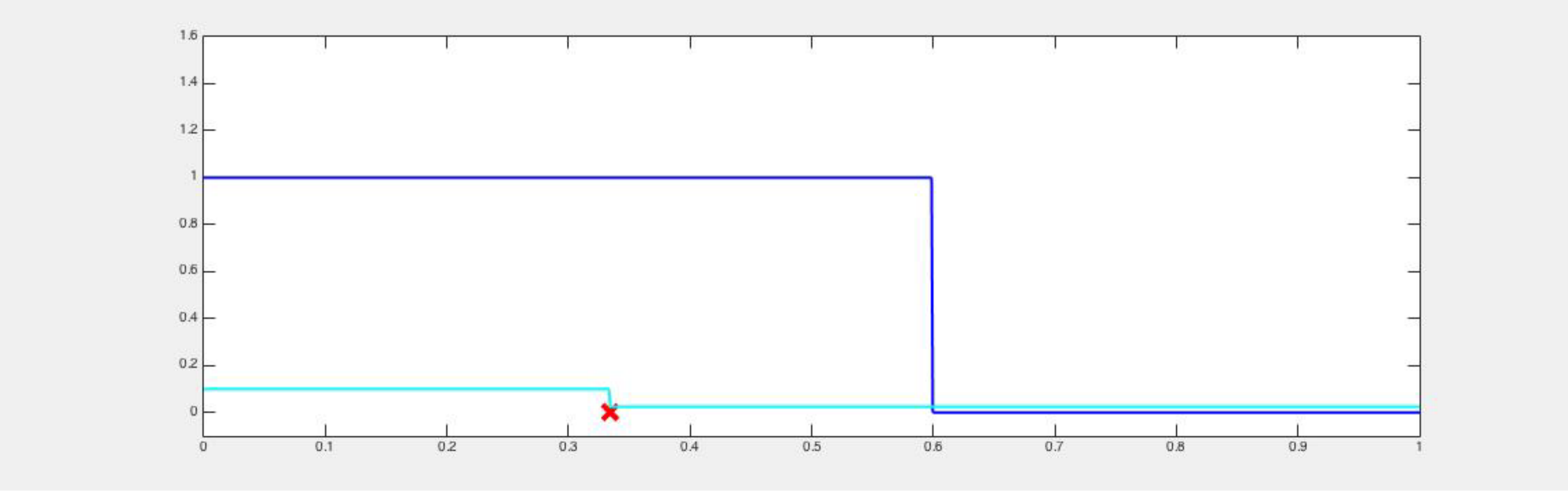
Algorithm 1 Probabilistic Quantile Search (PQS)

```
initialize prior density $\pi_0(\theta) = 1$ for $\theta \in [0,1]$
while not converged do
    choose $X_n$ such that $\int_0^{X_n} \pi_n(x)dx = 1/m$
    $Y_n \leftarrow f(X_n)$
    Perform Bayesian update to obtain $\pi_{n+1}(x)$:
    if $Y_n = 0$ then
        $$\pi_{n+1}(x) = \begin{cases} (1-p)\left(\frac{m}{1+(m-2)p}\right)\pi_n(x) & x \leq 1/m \\ p\left(\frac{m}{1+(m-2)p}\right)\pi_n(x) & x > 1/m \end{cases}$$
    else
        $$\pi_{n+1}(x) = \begin{cases} p\left(\frac{m}{1+(m-2)p}\right)\pi_n(x) & x \leq 1/m \\ (1-p)\left(\frac{m}{1+(m-2)p}\right)\pi_n(x) & x > 1/m \end{cases}$$
    end if
end while
return $\hat{\theta}_n$ such that $\int_0^{\hat{\theta}_n} \pi_{n+1}(x)dx = 1/2$
```

Probabilistic Quantile Search

Rather than sampling at 1/m into the feasible interval, we sample 1/m into the posterior distribution on θ.

For m=3 , p=0.2:



Sample complexity

The discretized PQS algorithm for $m \geq 2$ satisfies

$$\sup_{\theta \in [0,1]} \mathbb{E}\left[|\widehat{\theta}_n - \theta| \right] \leq 2 \left(\frac{m-1}{m} + \frac{2\sqrt{p(1-p)}}{m} \right)^{n/2}.$$

This bound matches Castro and Nowak for m=2. In this case, comparing to O(1/n) error bound given by grid sampling, we have a $(½)^{n/2}$ bound.

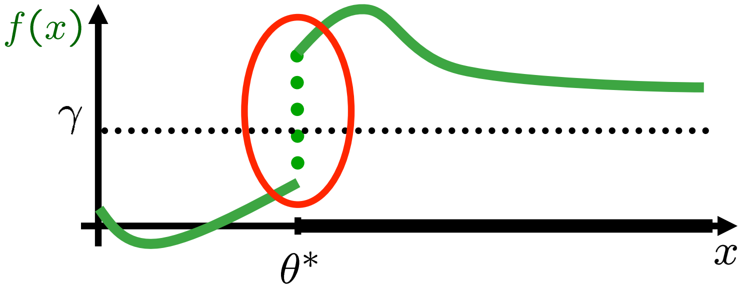
It's loose in practice, but finding the expected number of samples and any distance bounds or expectations is significantly more technical, because of the less obvious interaction of sample locations with the posterior in PQS.

Future: Hölder smooth

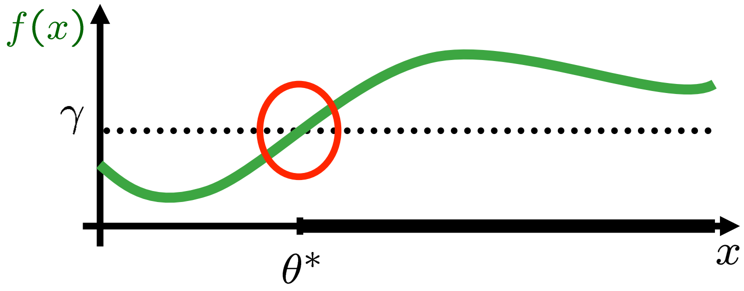
To generalize we could consider a more general class of functions, Hölder smooth f el set at γ. A function is α-Hölder smooth for c $\delta > 0$ so that $\forall x : |f(x) - \gamma| \leq \delta$ we have:

$$|f(x) - \gamma| \geq c|x - \theta^*|^\alpha \ .$$

This is wrong: the idea is that we have something that is unsmooth at the transition.



$\alpha = 0$



$\alpha = 1$

Figures courtesy Rui Castro

Results on sample complexity only are known in the case of binary bisection and for d-dimensional signals that have the additional assumption of smooth level sets.

Thank you!

Questions?