Lecture 24
How to Optimize Expensive Functions

PREDICTIVE
SCIENCE LABORATORY

Objectives

- Quantify the value of the information extracted from an experiment/simulation.
- Optimize an expensive black-box function under a limited budget.

PREDICTIVE
SCIENCE LABORATORY

The problem

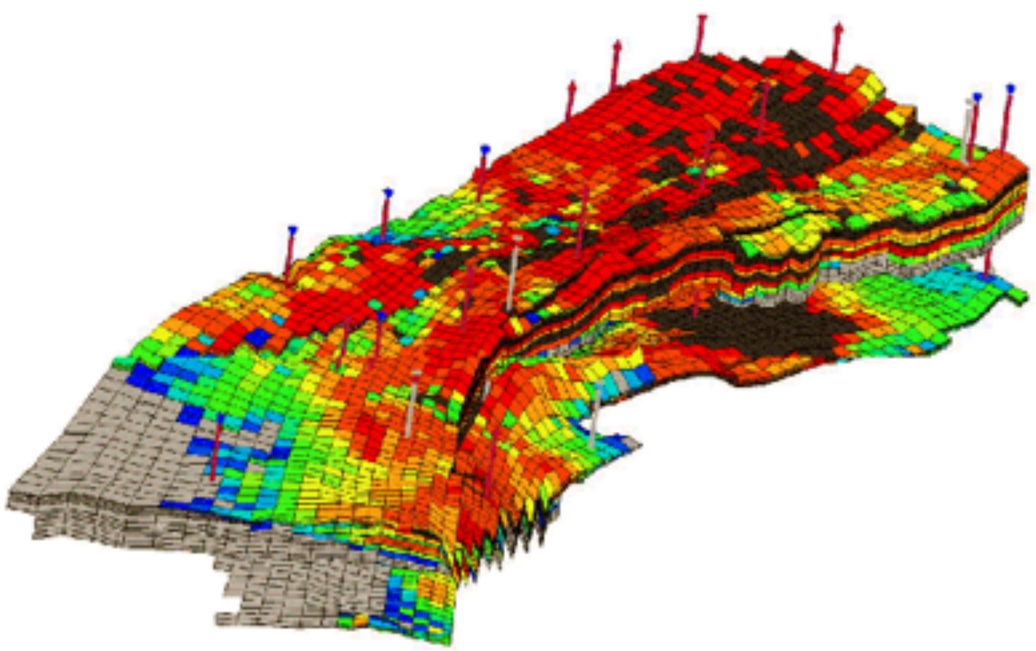
- Problem:

$$\mathbf{x}^{*} = \mathrm{argmax}_{\mathbf{x}} f(\mathbf{x})$$

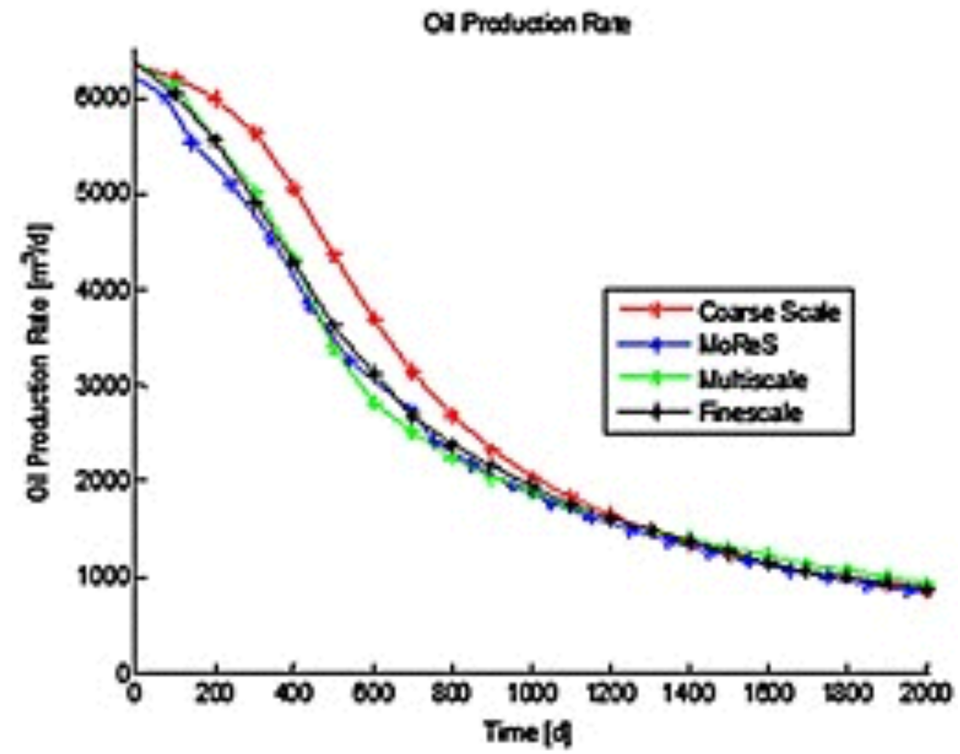
- when the objective is:
  - very expensive to evaluate
  - you don't have gradients
  - dimensionality < 30 parameters

PREDICTIVE
SCIENCE LABORATORY

Example 1: Oil-well placement problem



Simulation



x = well locations

f(**x**) = expected net present cost of investment

Pandita, Bilionis, and Panchal, 2016
http://arxiv.org/abs/1604.01147

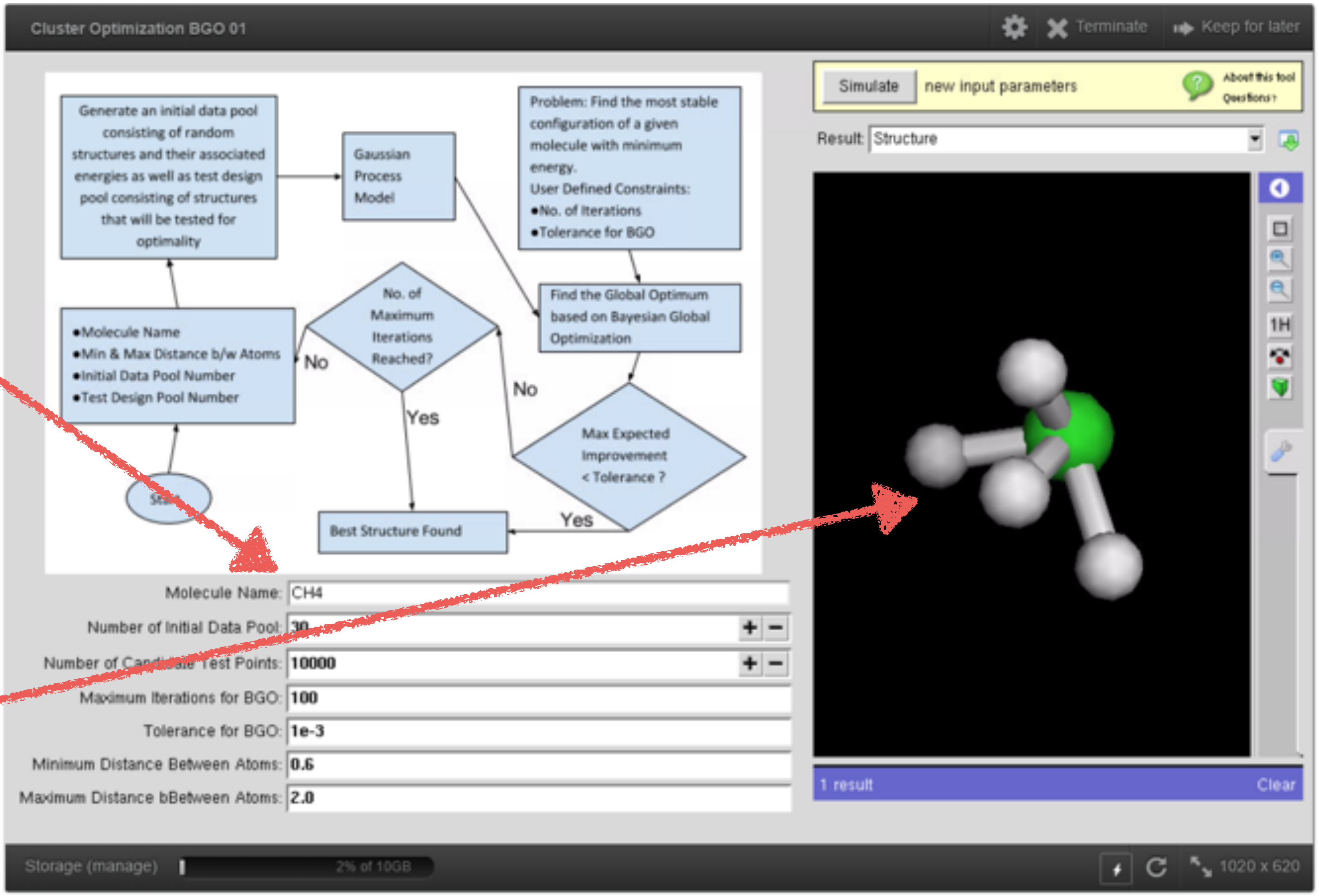
PREDICTIVE
SCIENCE LABORATORY

Example 2: Find stable structures

Surf student project

Chemical formula

Geometry (**x**) with minimum energy (y)

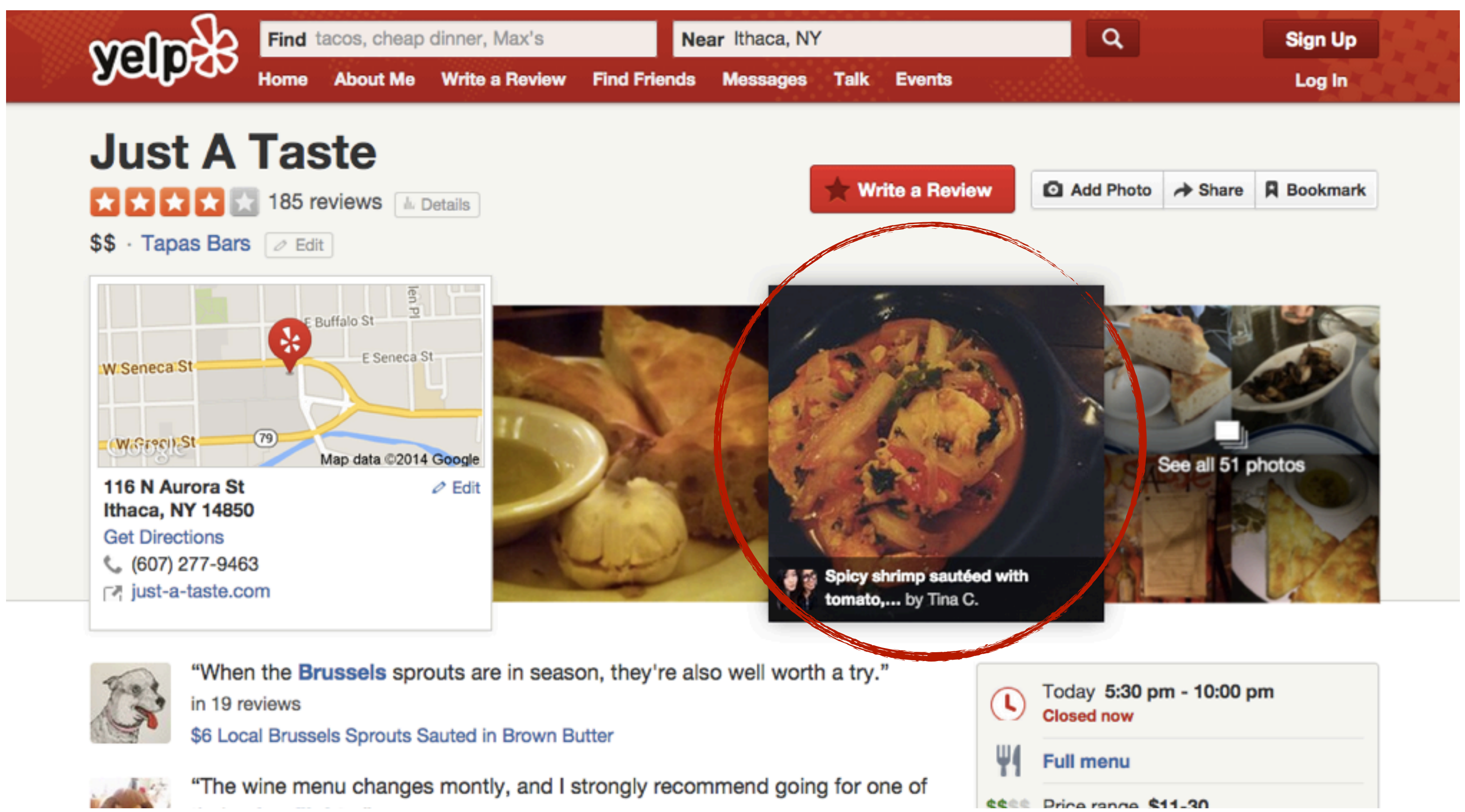


https://nanohub.org/tools/clusterbgopro

PREDICTIVE SCIENCE LABORATORY

Example 3: Web site optimization

x = web design → measurement → -y = Number of views, seconds per view, etc.



PREDICTIVE SCIENCE LABORATORY

http://yelp.github.io/MOE

Example 4: Training a robot to walk

https://www.youtube.com/watch?v=ualnbKfkc3Q

https://www.youtube.com/watch?v=GiqNQdzc5TI

Other examples

- Model calibration (if posed as an optimization problem).
- Maximize efficiency in solar cells.
- Drug development.
- …

PREDICTIVE
SCIENCE LABORATORY

Startup idea

https://sigopt.com/

Σ SIGOPT

Solutions | **Pricing** | FAQ | About | Docs | Login | Sign Up

Pricing

We have plans for teams both large and small. Start for free with no commitment or contact us and we'll help find the plan that is right for you.

SigOpt for Individuals	SigOpt for Enterprise
$999/month	**Contact for pricing**
30 day free trial, no credit card required	Free pilot
10 experiments	Customized number of experiments
Email support	Premium support
Sign Up	Contact Us

We love students here at SigOpt. So we're making our individual plan available to academics, free of charge. We also offer discounted rates on our Enterprise plans for educational usage. Contact education@sigopt.com for more information.

Free options

- https://github.com/PredictiveScienceLab/py-bgo (features stochastic and multi-objective optimization)

- https://github.com/SheffieldML/GPyOpt (features parallel optimization)

PREDICTIVE
SCIENCE LABORATORY

pip install GPyOpt

PREDICTIVE
SCIENCE LABORATORY

Idea

1. We have some observed data (designs **x** - vs - objectives **y**).

2. We fit a **statistical regression model** to the data.

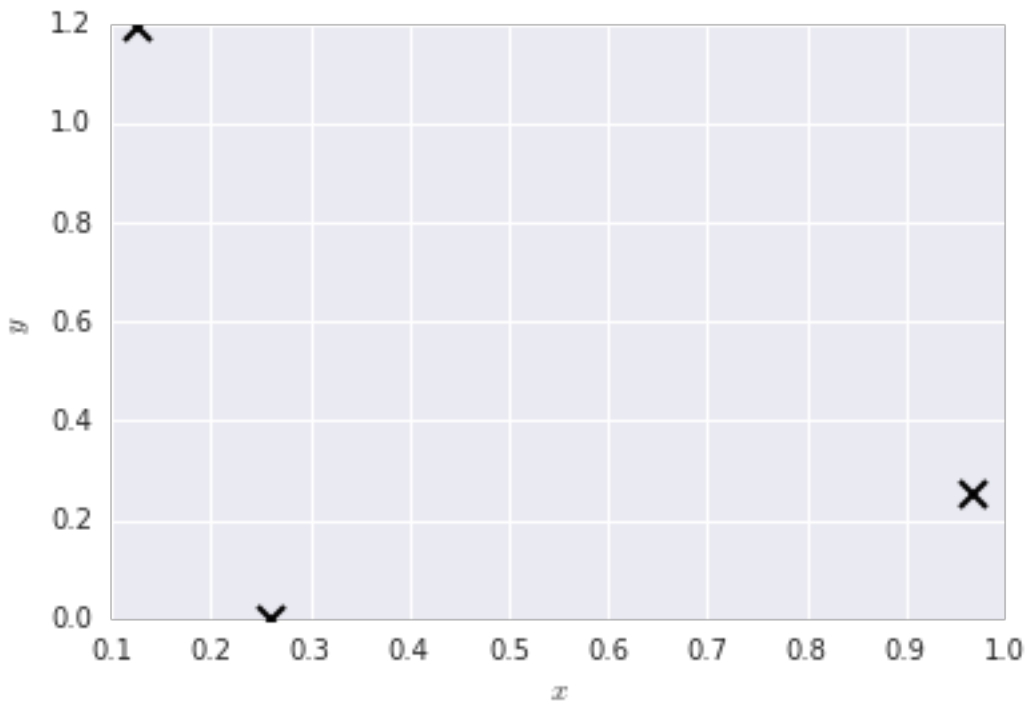
3. For each candidate design, compute the **value of information (VoI)**.

4. We find the design with the **maximum VoI**.

5. We compute the objective for this design.

Repeat until:
- budget is exhausted;
- VoI low.

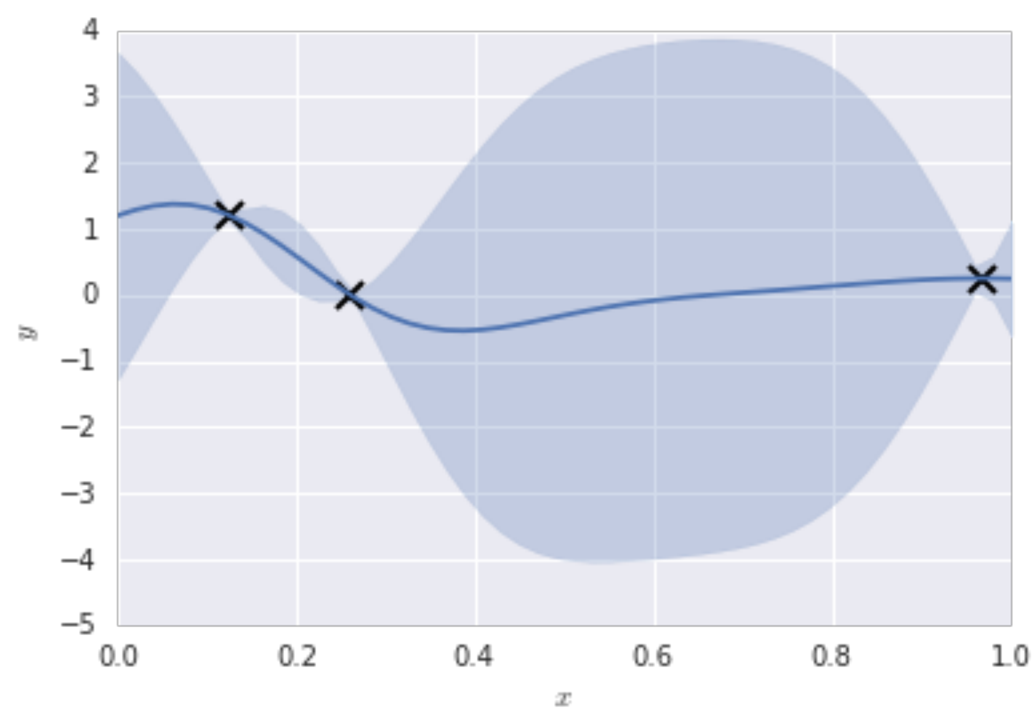
We have some data



$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$

PREDICTIVE
SCIENCE LABORATORY

We fit a statistical model



$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Gaussian process regression

- Assume that we have observed:

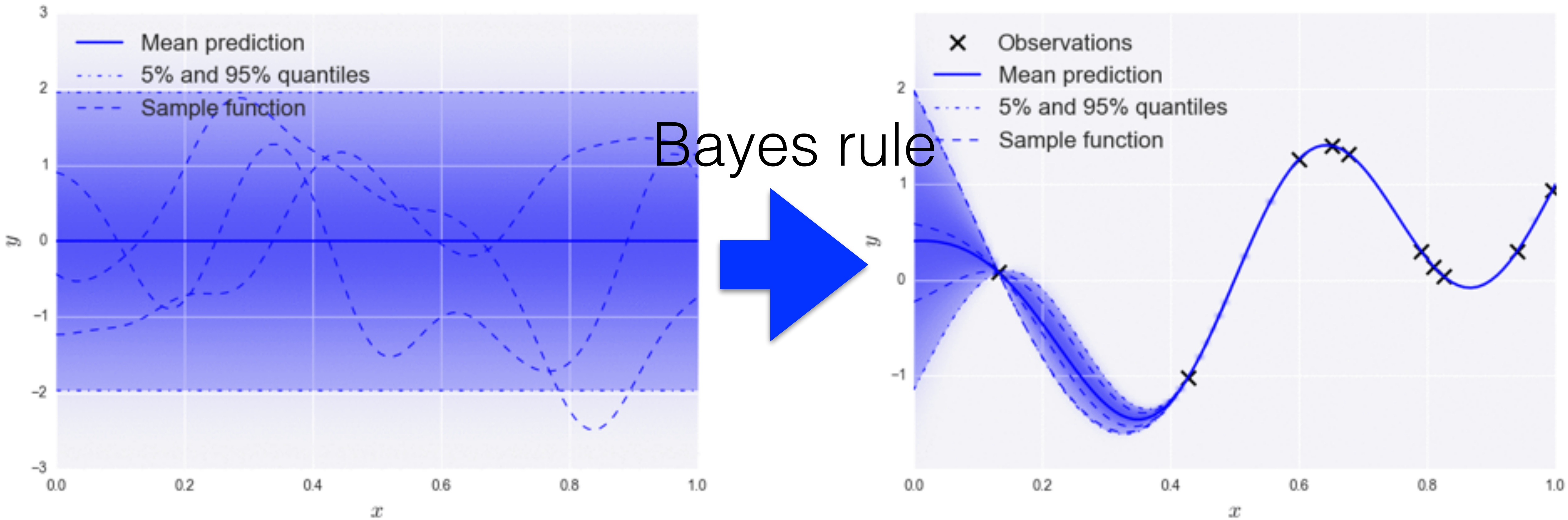
$$\mathbf{X} = \{x_1, \ldots, x_N\},$$
$$\mathbf{f} = \{f(x_1), \ldots, f(x_N)\}$$

- and that we want to make predictions at an arbitrary set of *test* inputs:

$$\mathbf{X}^* = \{x_1^*, \ldots, x_{N^*}^*\}$$
$$\mathbf{f}^* = \{f(x_1^*), \ldots, f(x_{N^*}^*)\}$$

PREDICTIVE
SCIENCE LABORATORY

Gaussian process regression



Prior GP

Posterior GP

PREDICTIVE
SCIENCE LABORATORY

The point predictive distribution

- Posterior GP:

$$f(\cdot) \mid \mathbf{X}, \mathbf{f} \sim \mathrm{GP}\left(f(\cdot) \mid \tilde{m}(\cdot), \tilde{k}(\cdot,\cdot)\right),$$

- Looking at just one point, we get the *point predictive distribution:*

$$y \mid \mathbf{x}, \mathbf{X}, \mathbf{f} \sim \mathcal{N}\left(y \mid \tilde{m}(\mathbf{x}), \tilde{\sigma}^2(\mathbf{x})\right),$$

$$\tilde{\sigma}^2(\mathbf{x}) = \tilde{k}(\mathbf{x}, \mathbf{x}).$$

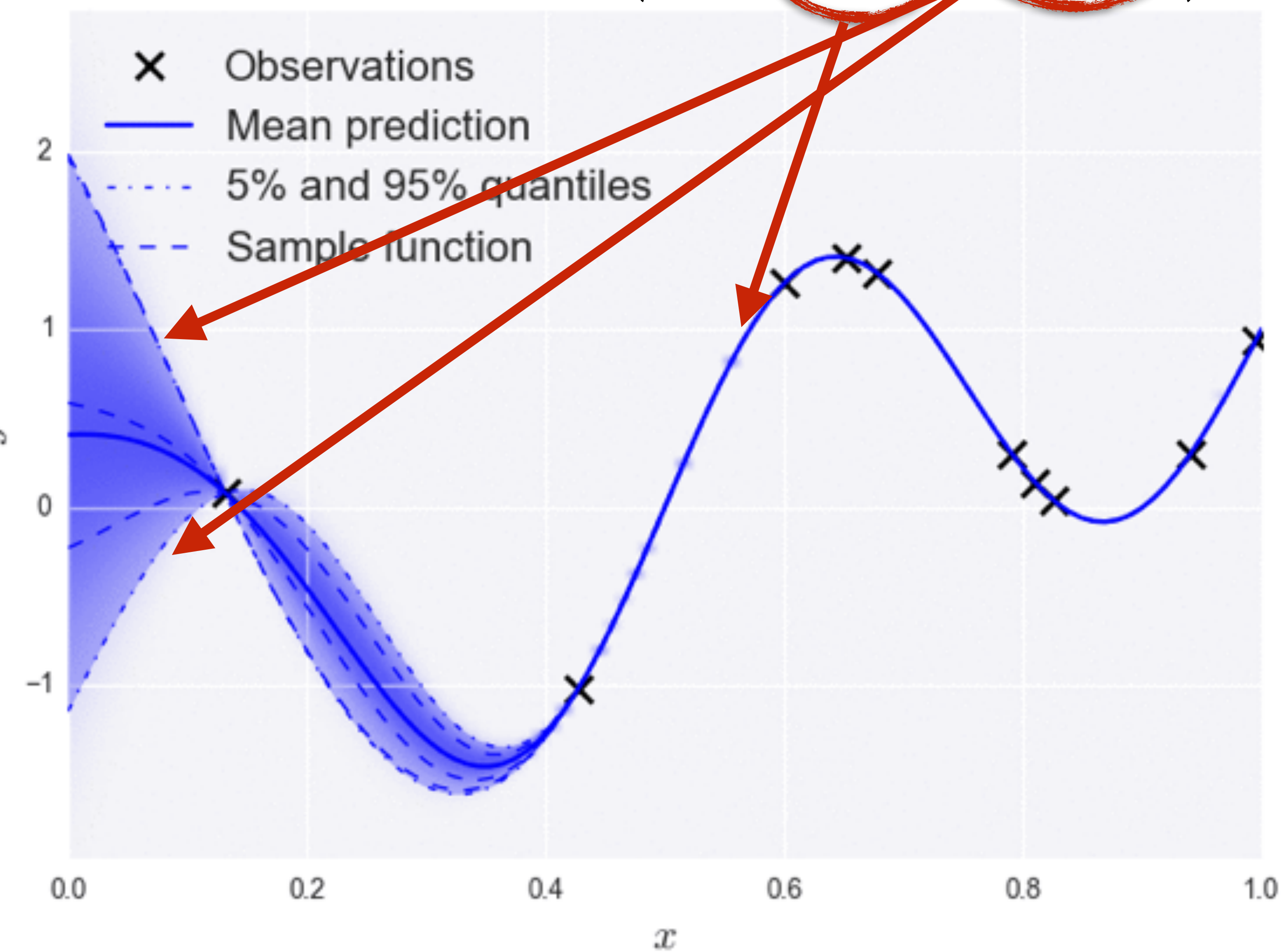
- You may use the mean as a surrogate.

PREDICTIVE SCIENCE LABORATORY

Gaussian process regression

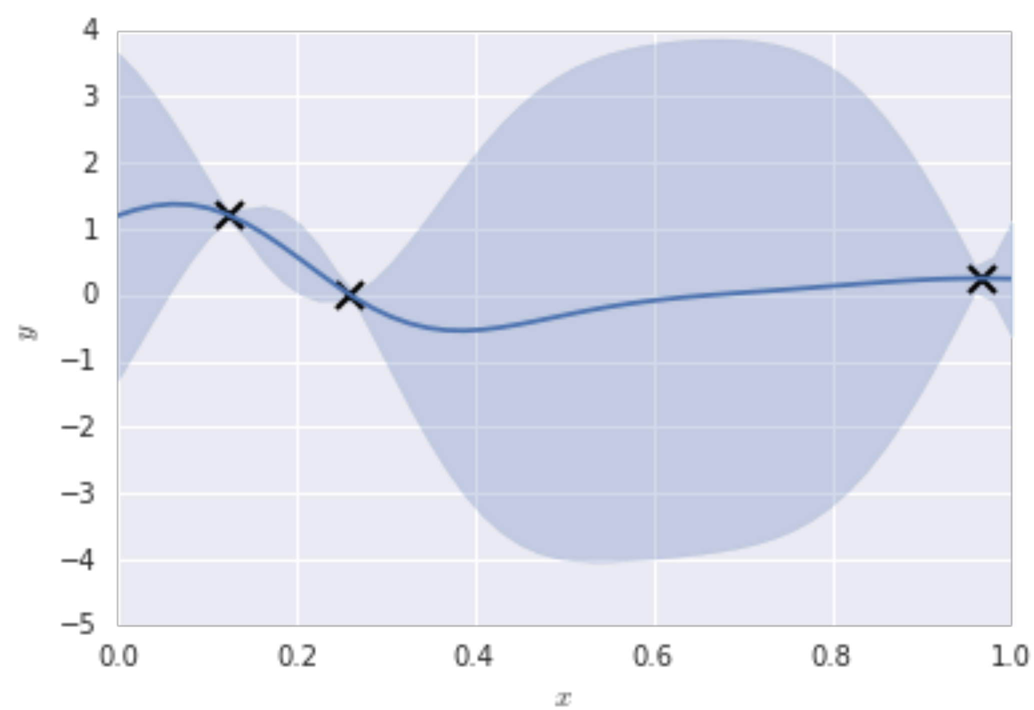
$$y \mid \mathbf{x}, \mathbf{X}, \mathbf{f} \sim \mathcal{N}\left(y \mid \tilde{m}(\mathbf{x}), \tilde{\sigma}^2(\mathbf{x})\right),$$

$$f(\mathbf{x}) = \tilde{m}(\mathbf{x}) \pm 2\tilde{\sigma}(\mathbf{x})$$



PREDICTIVE
SCIENCE LABORATORY

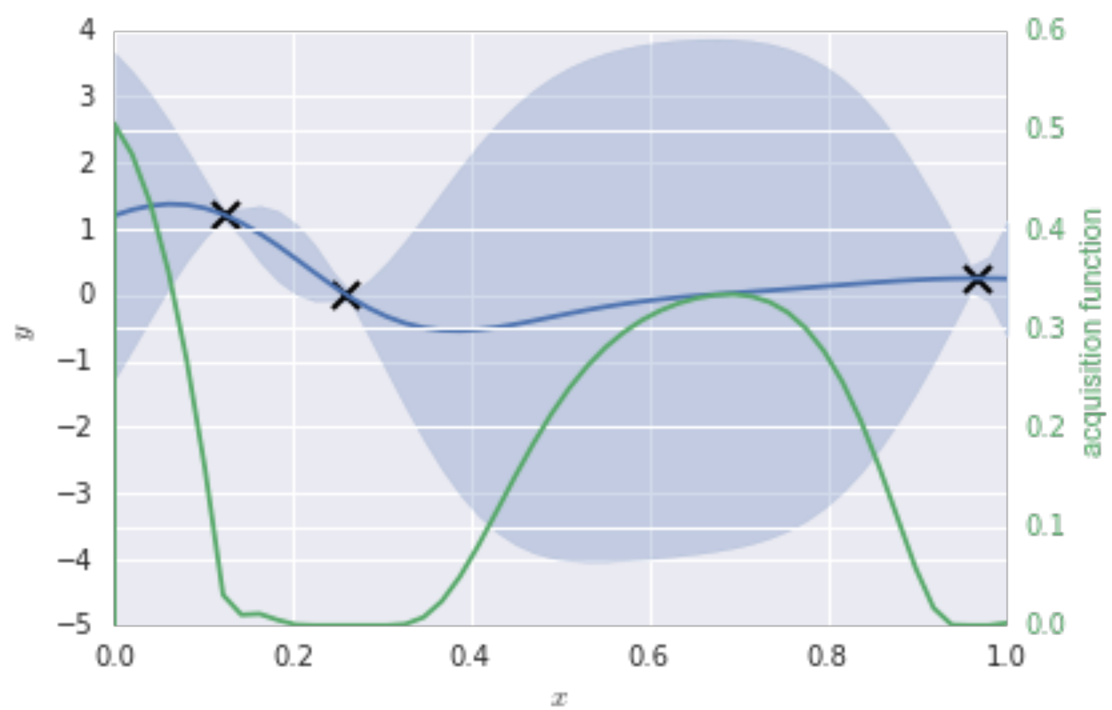
We fit a statistical model



$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Quantify the value of information via an acquisition function

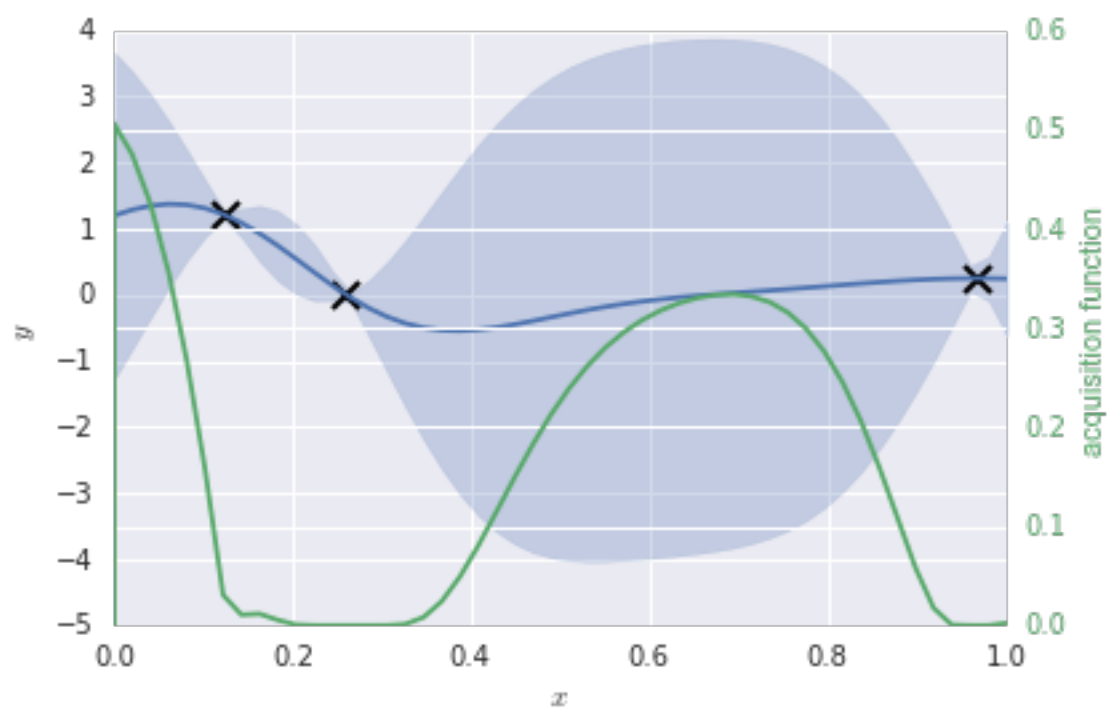


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Quantify the value of information via an acquisition function

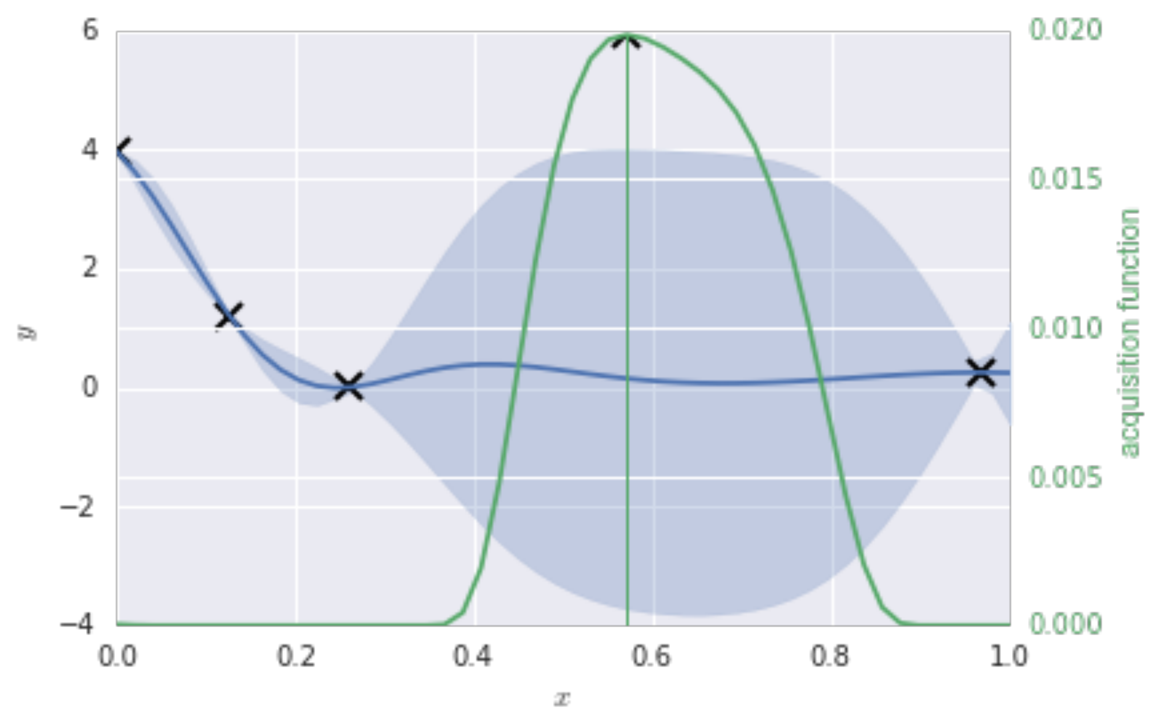


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \dots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 2)

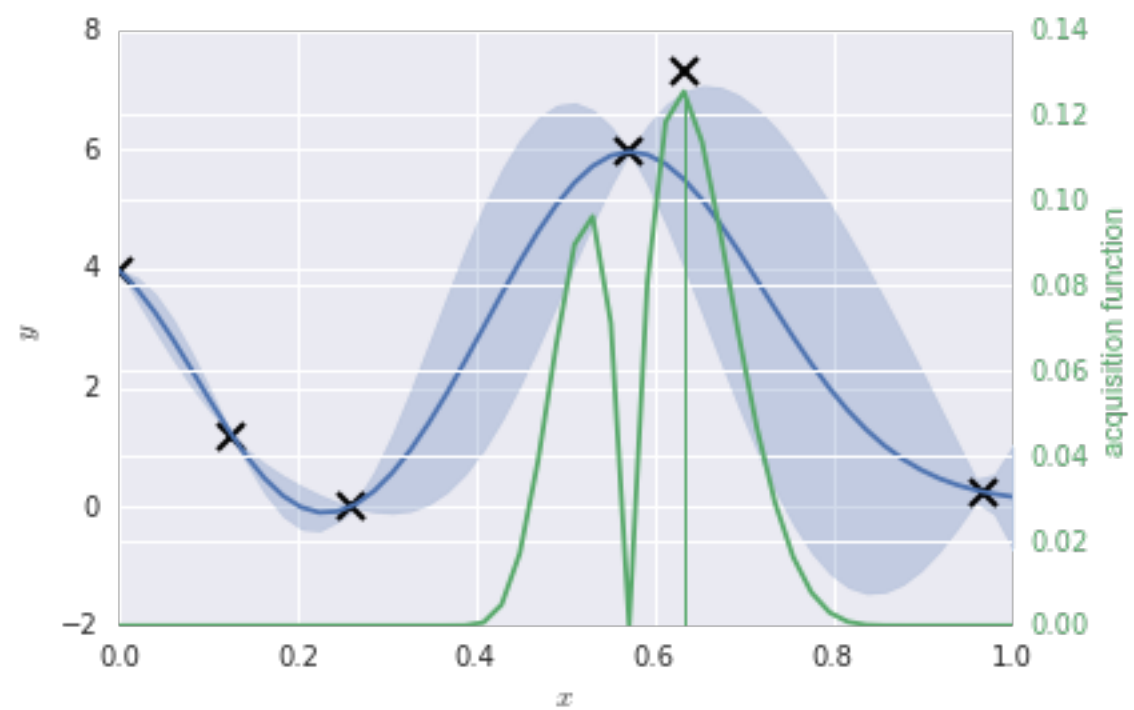


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 3)

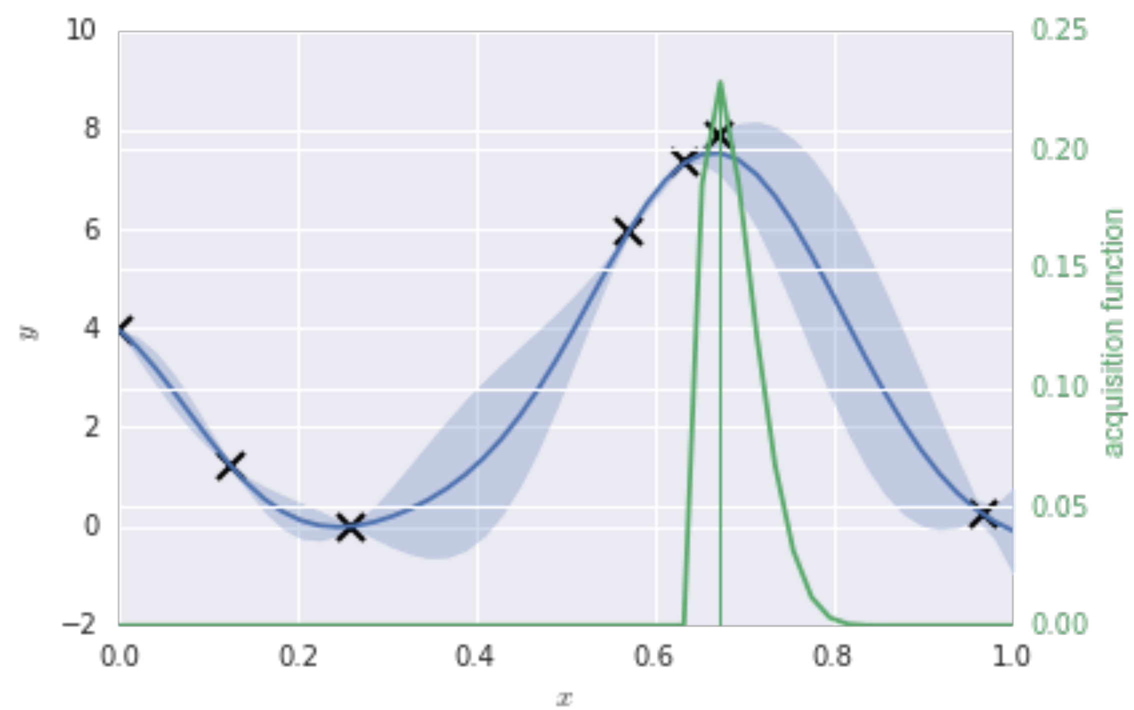


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 3)

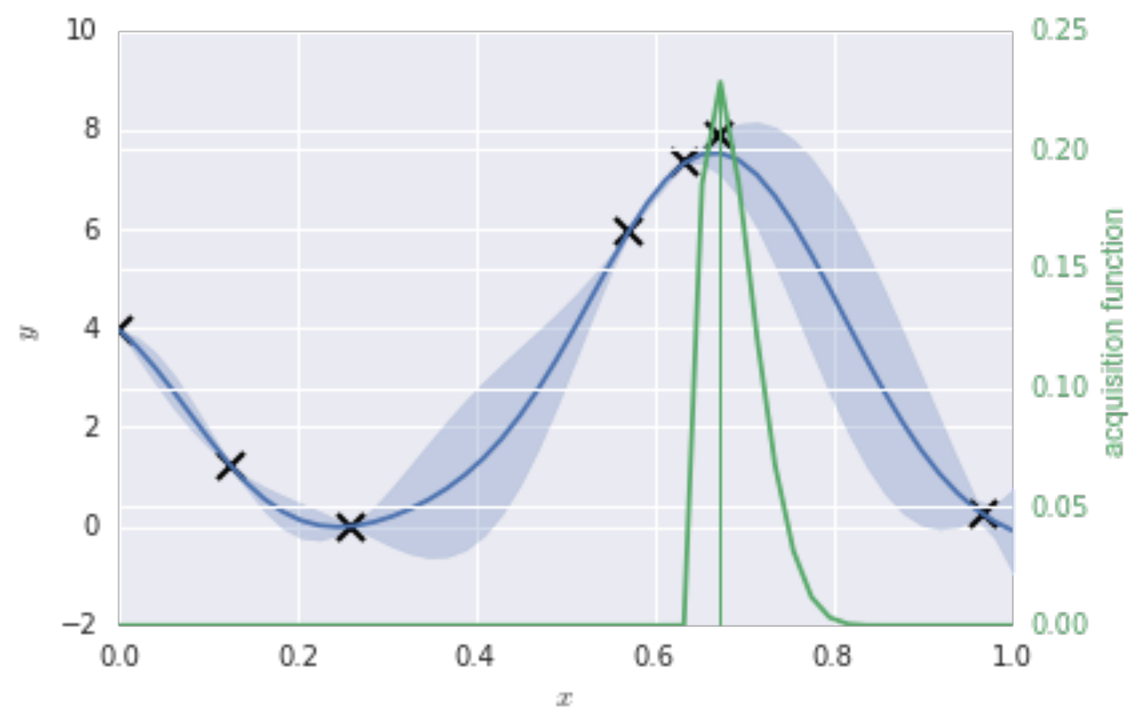


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 4)

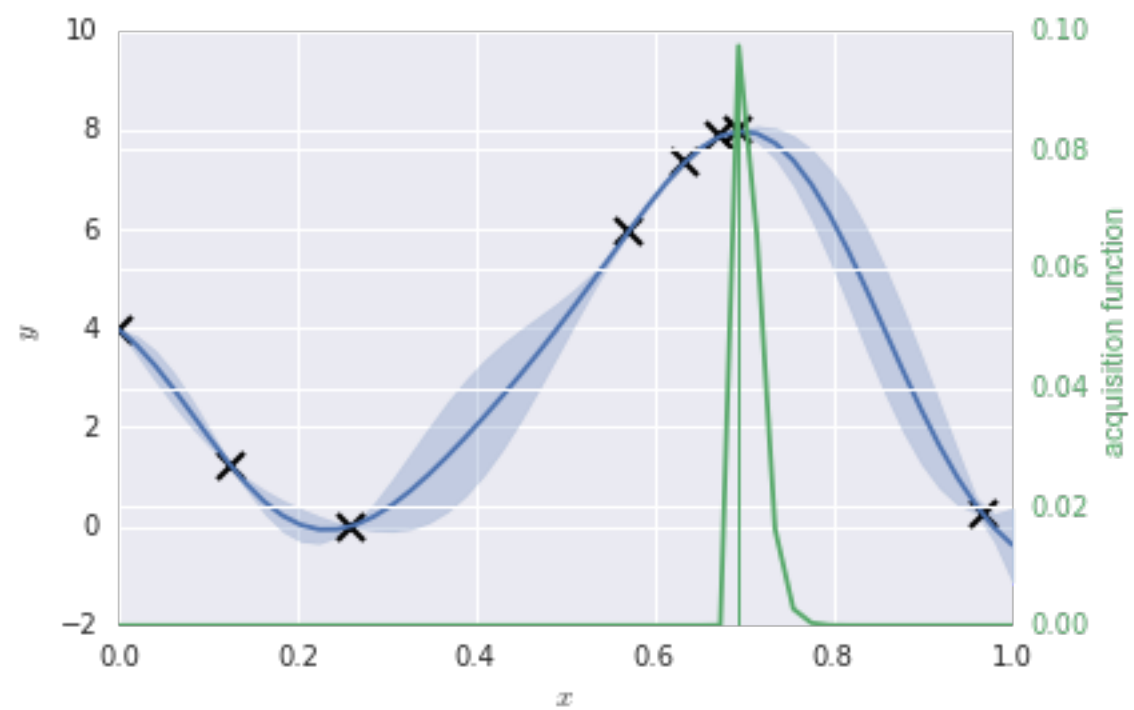


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 5)



$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$

$$y_{1:n} = \{y_1, \ldots, y_n\}$$

$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

The problem

- Problem:

$$\mathbf{x}^* = \text{argmax}_{\mathbf{x}} f(\mathbf{x})$$

- when the objective is:
  - very expensive to evaluate
  - you don't have gradients
  - dimensionality < 30 parameters

PREDICTIVE
SCIENCE LABORATORY

Dynamic vs Myopic Information Acquisition

- Optimal information acquisition policies…
- => Dynamic programming/control theory.
- Too hard mathematical/computational problems.
- What if, we just pick one piece of information at a time?
- Myopic (one-step-look-ahead) policies.

The value of information

- The value of information (VoI) depends on what you want to do.
- Can be quantified objectively if:
  - you have assigned probabilities over all possibilities.
  - you can quantify your profit/loss if any of the possibilities happen.

PREDICTIVE
SCIENCE LABORATORY

The value of information

VoI of **x** = how much expected gain if I measure at **x**

= expected profit if I measure at **x**
- current best alternative

= expected income if I measure at **x**
- cost of measuring **x**
- current best alternative

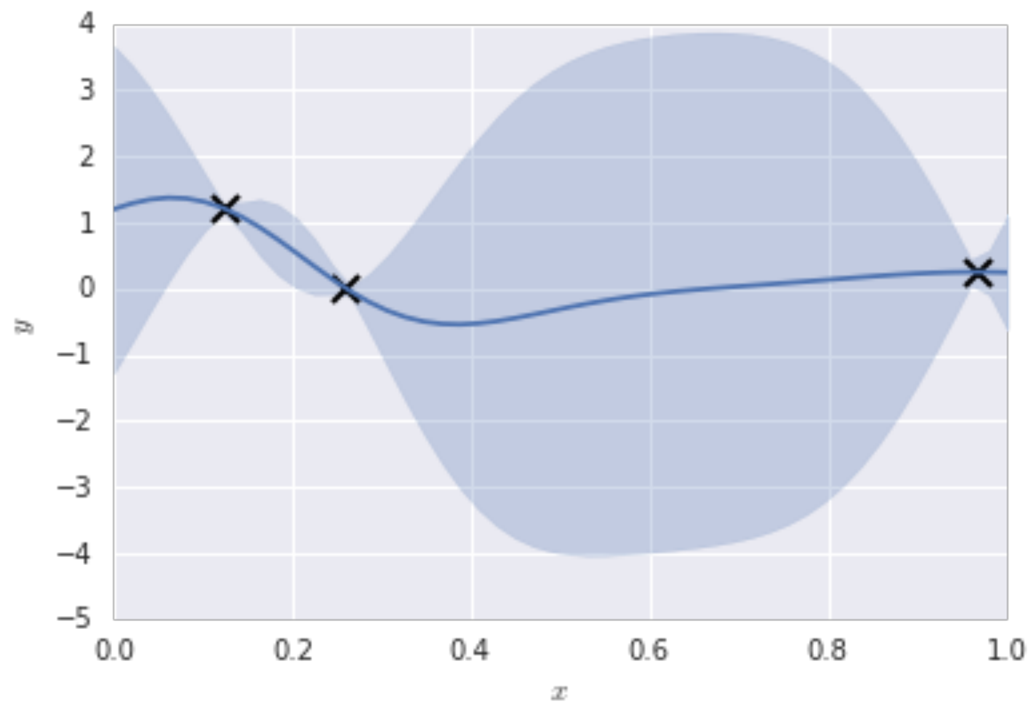
PREDICTIVE
SCIENCE LABORATORY

Acquisition function

- Most of the times, we don’t have the details to find the VoI.
- We use heuristic approximations to VoI such as:
    - the probability of improvement
    - the expected improvement
    - the knowledge gradient
    - the expected information gain
    - …

PREDICTIVE
SCIENCE LABORATORY

Maximum Mean



Add the point with the maximum expected mean:

$$a(\mathbf{x}) = m(\mathbf{x})$$

Tries to exploit what we know.

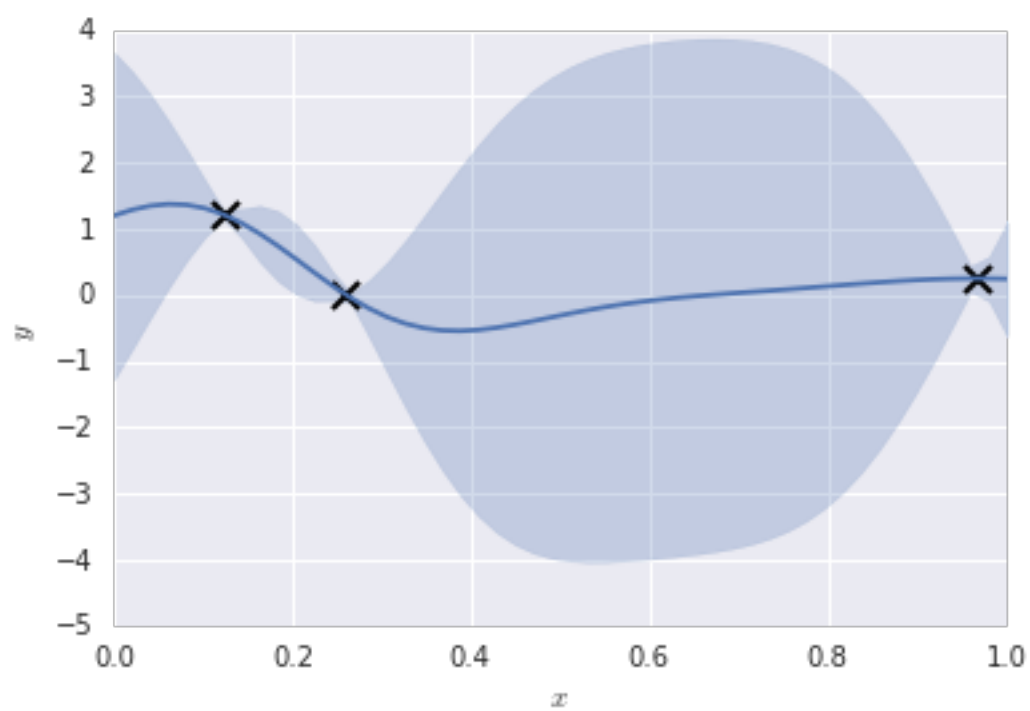
It does not converge.

How can we add some elements of exploration?

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \dots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Maximum Upper Interval



Use the variance to explore:

$$a(\mathbf{x}) = m(\mathbf{x}) + \psi\sigma(\mathbf{x})$$

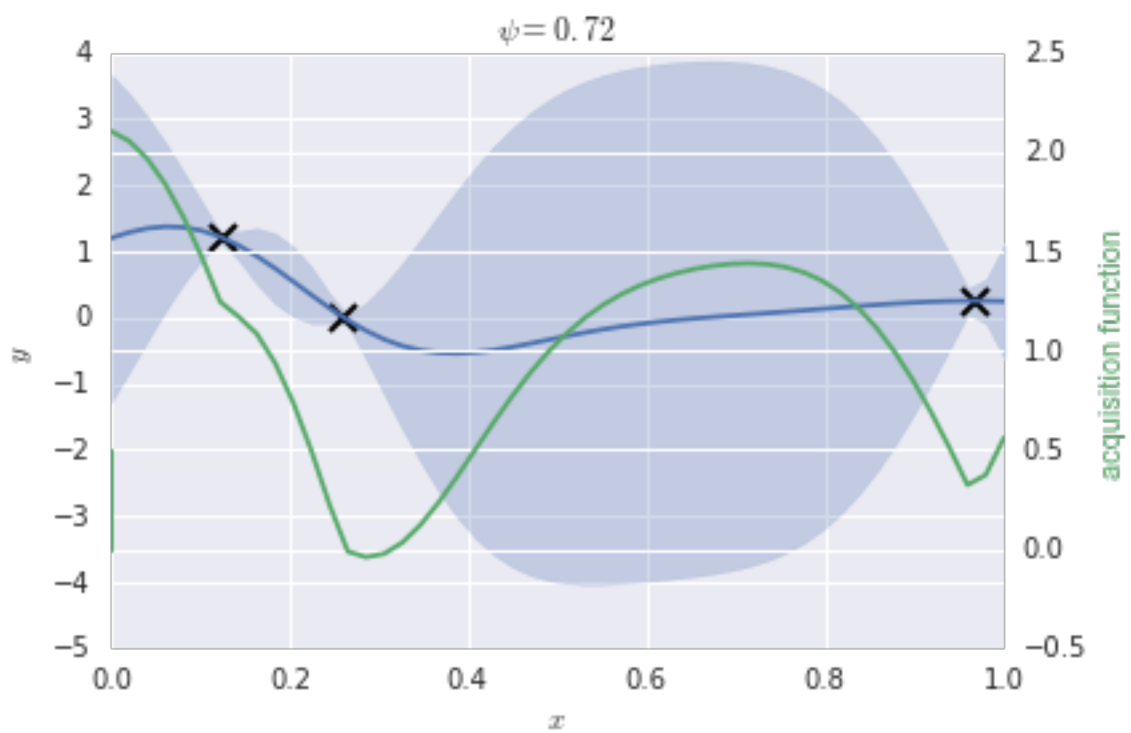
Adds upper quantile.

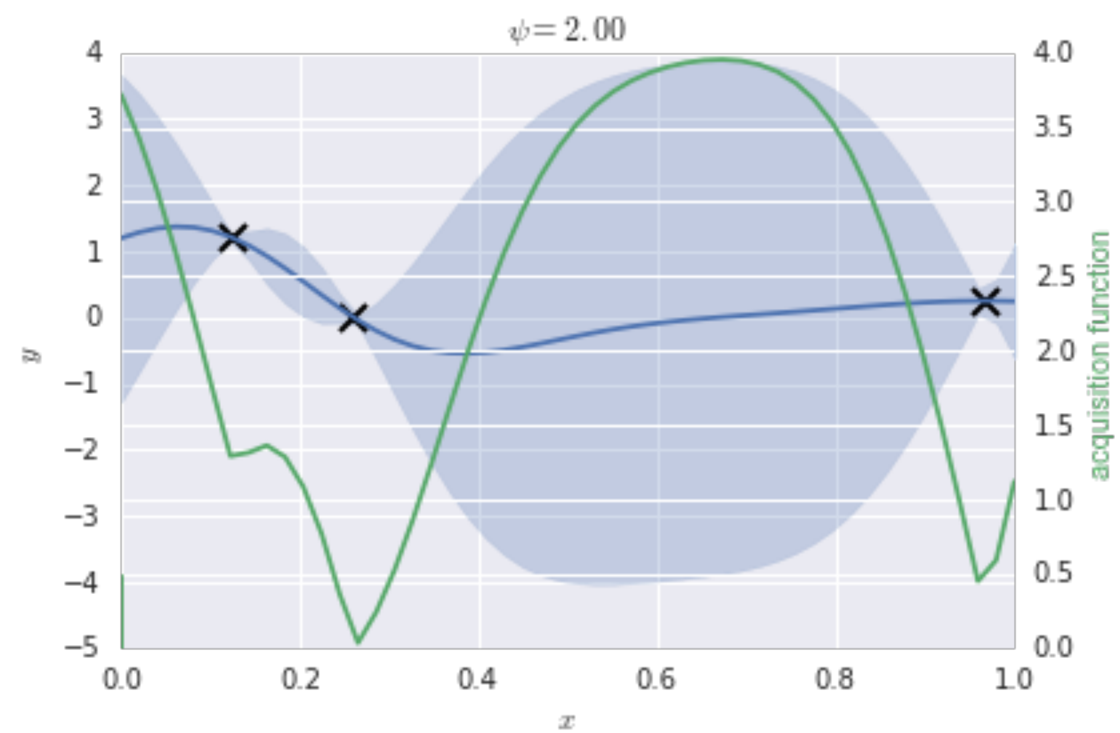
Provable convergence to the global maximum!

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1,\dots,\mathbf{x}_n\}$$
$$y_{1:n} = \{y_1,\dots,y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Maximum Upper Interval





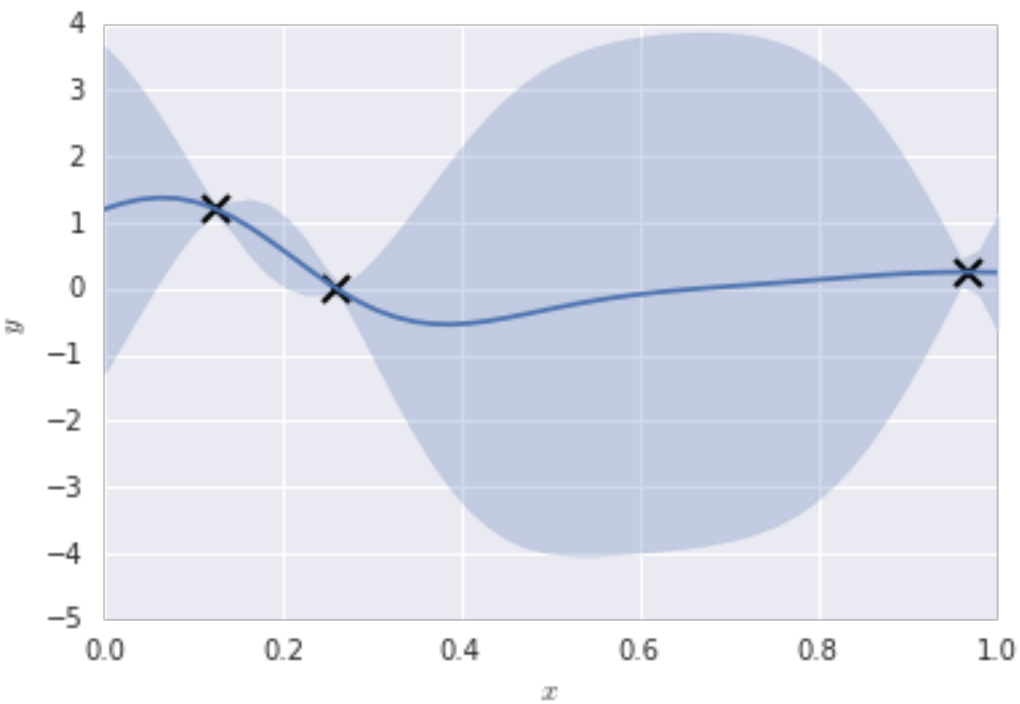
Exploits more

Explores more

Too much exploration…

PREDICTIVE
SCIENCE LABORATORY

Probability of Improvement



Current best:

$$\tilde{y}_n = \max_{1\le i\le n} y_i$$

Hypothetical simulation at **x** yields y

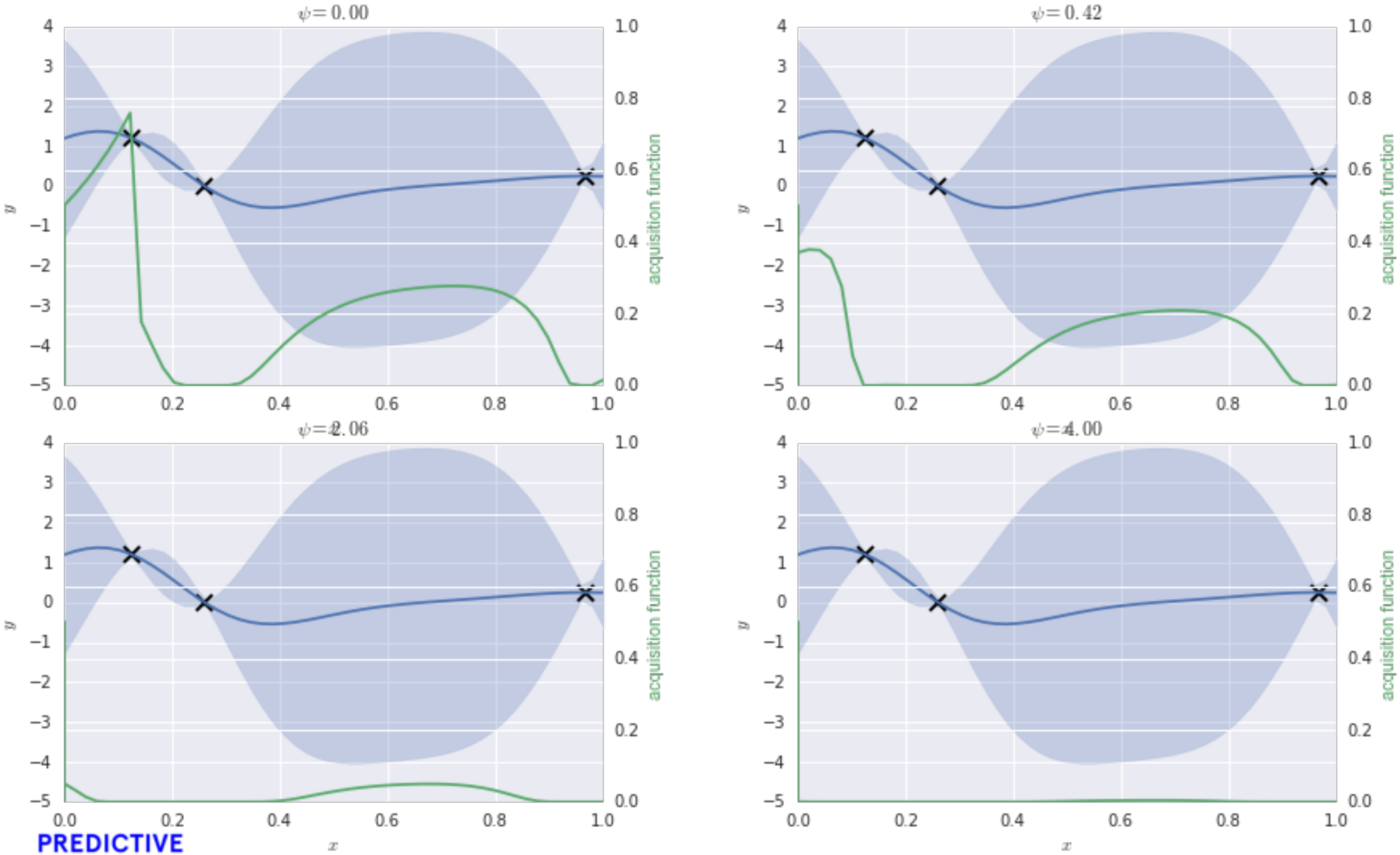
The probability of improvement is:

$$\begin{aligned} a(\mathbf{x}) &= P[y > \tilde{y}_n + \psi \mid \mathbf{x}] \\ &= \int_{\tilde{y}_n+\psi}^{\infty} p(y \mid \mathbf{x}) dy \\ &= 1 - \Phi\left(\frac{\tilde{y}_n + \psi - m(\mathbf{x})}{\sigma(\mathbf{x})}\right) \end{aligned}$$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \dots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Probability of Improvement

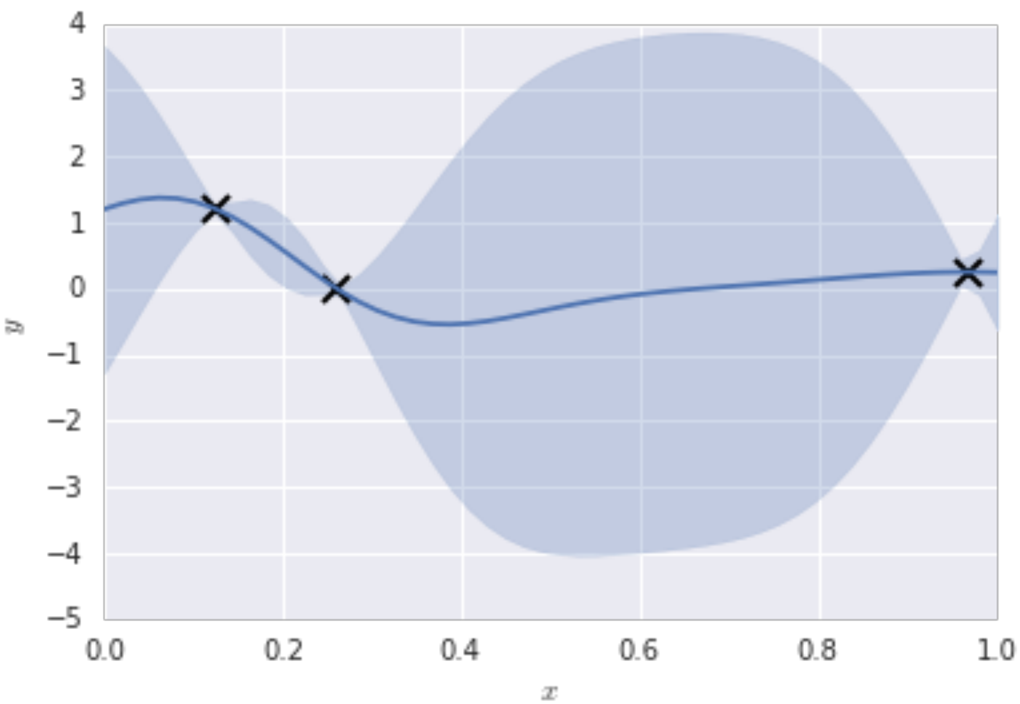


PREDICTIVE
SCIENCE LABORATORY

Probability of Improvement - Why not use it?

- Large value of psi -> exploration.
- Small value of psi -> exploitation.
- But how to you pick it?

Expected Improvement



Current best:

$$\tilde{y}_n = \max_{1\le i\le n} y_i$$

Hypothetical simulation at **x** yields y

The improvement is:

$$a(\mathbf{x}, y) = \max\{0, y - \tilde{y}_n\}$$

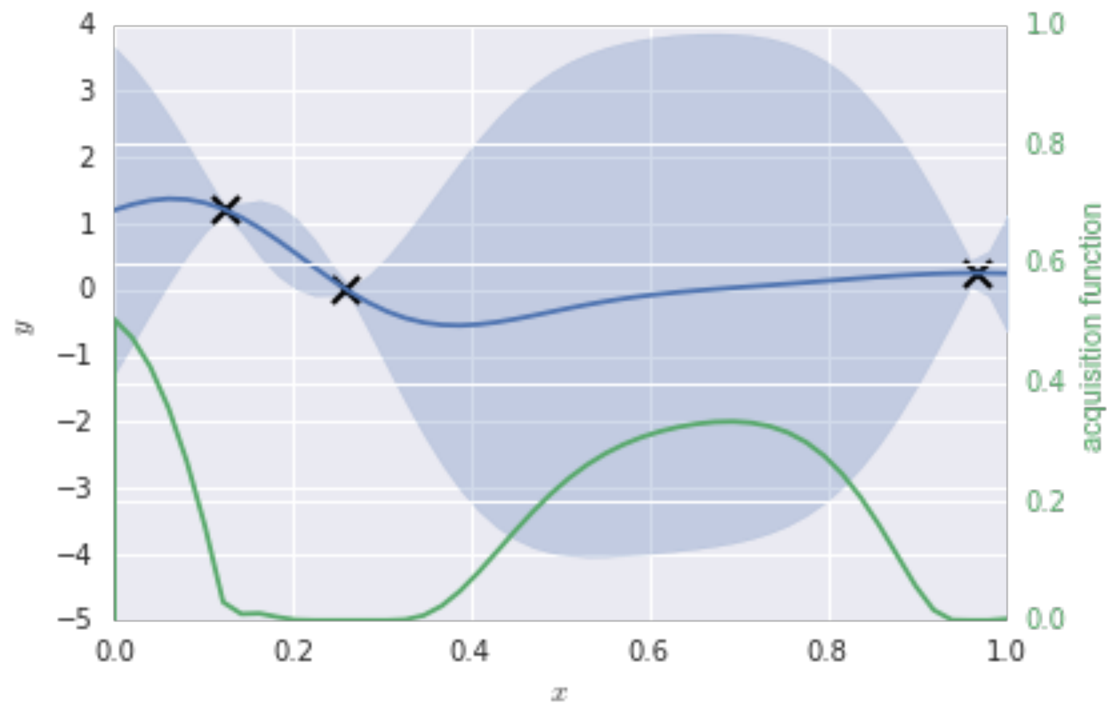
Integrate over y:

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1,\dots,\mathbf{x}_n\}$$
$$y_{1:n} = \{y_1,\dots,y_n\}$$
$$p(y\mid\mathbf{x}) \approx \mathcal{N}\left(y\mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

$$a(\mathbf{x}) = \int_{-\infty}^{\infty} \max\{0, y - \tilde{y}_n\} p(y\mid\mathbf{x})dy$$
$$= (m(\mathbf{x}) - \tilde{y}_n)\Phi\left(\frac{m(\mathbf{x}) - \tilde{y}_n}{\sigma(\mathbf{x})}\right) + \sigma(\mathbf{x})\phi\left(\frac{m(\mathbf{x}) - \tilde{y}_n}{\sigma(\mathbf{x})}\right)$$

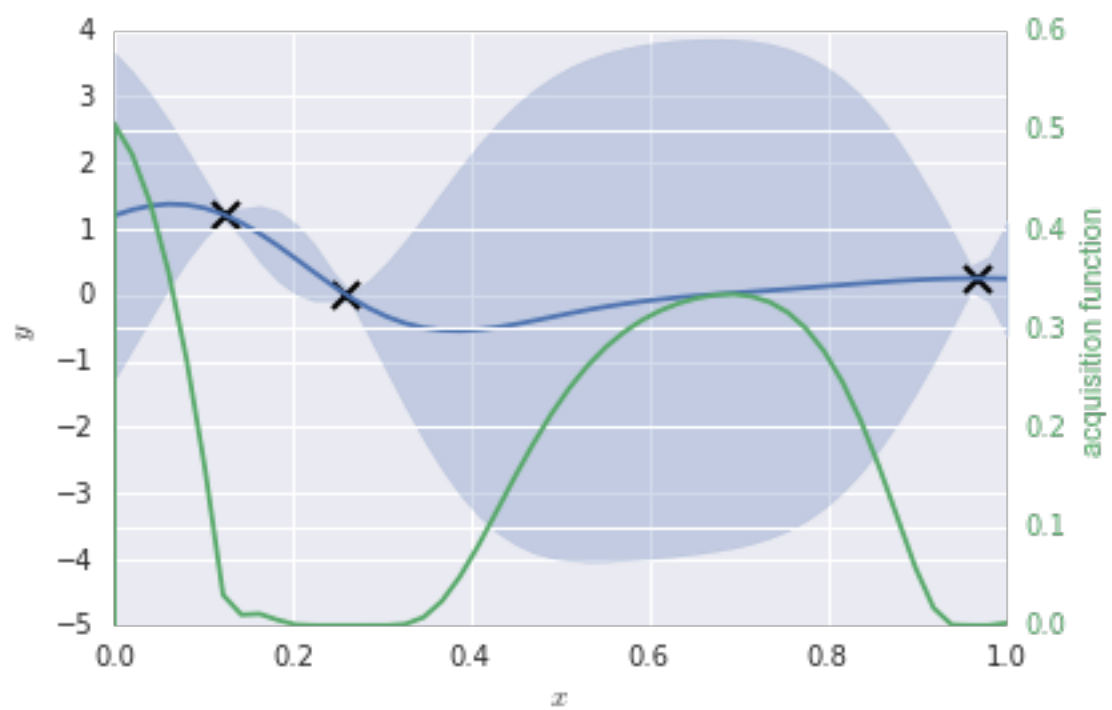
PREDICTIVE
SCIENCE LABORATORY

Expected Improvement



Automatic exploration vs exploitation…

Quantify the value of information via an acquisition function

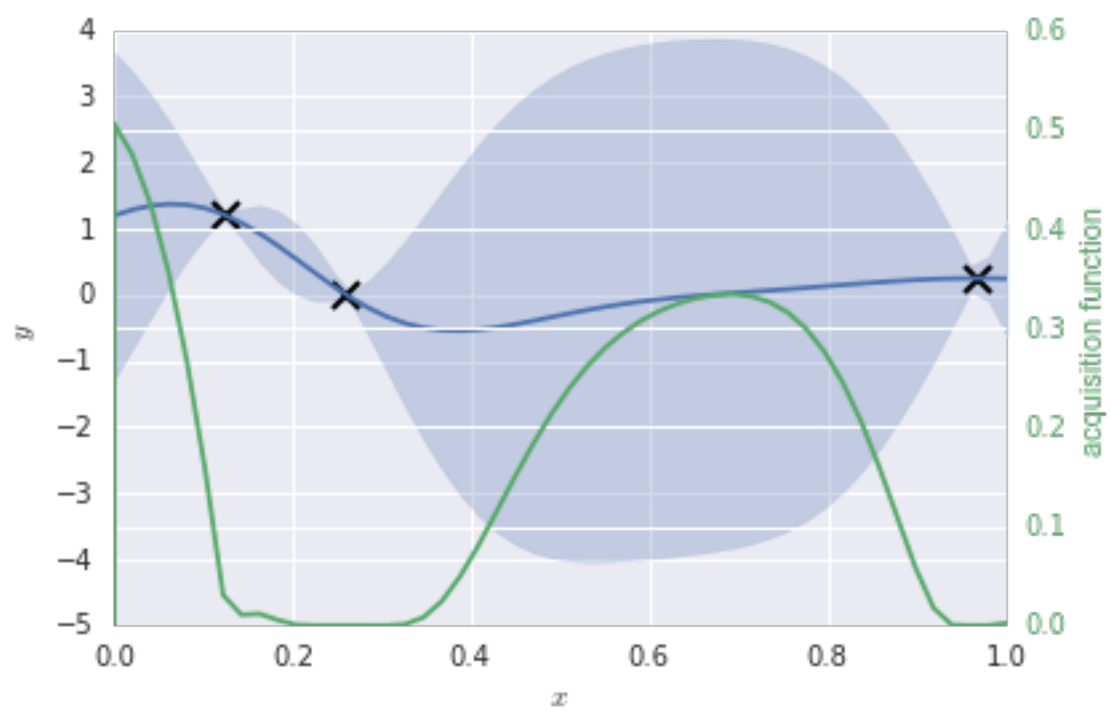


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$

PREDICTIVE
SCIENCE LABORATORY

Quantify the value of information via an acquisition function

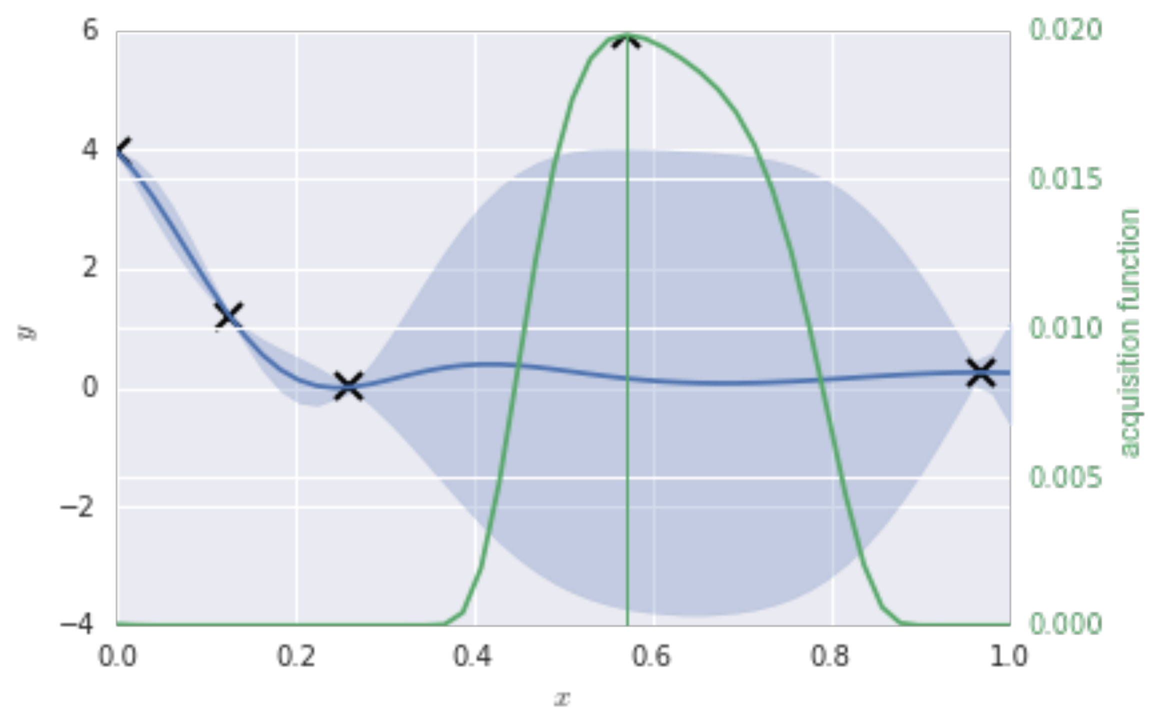


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \dots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 2)

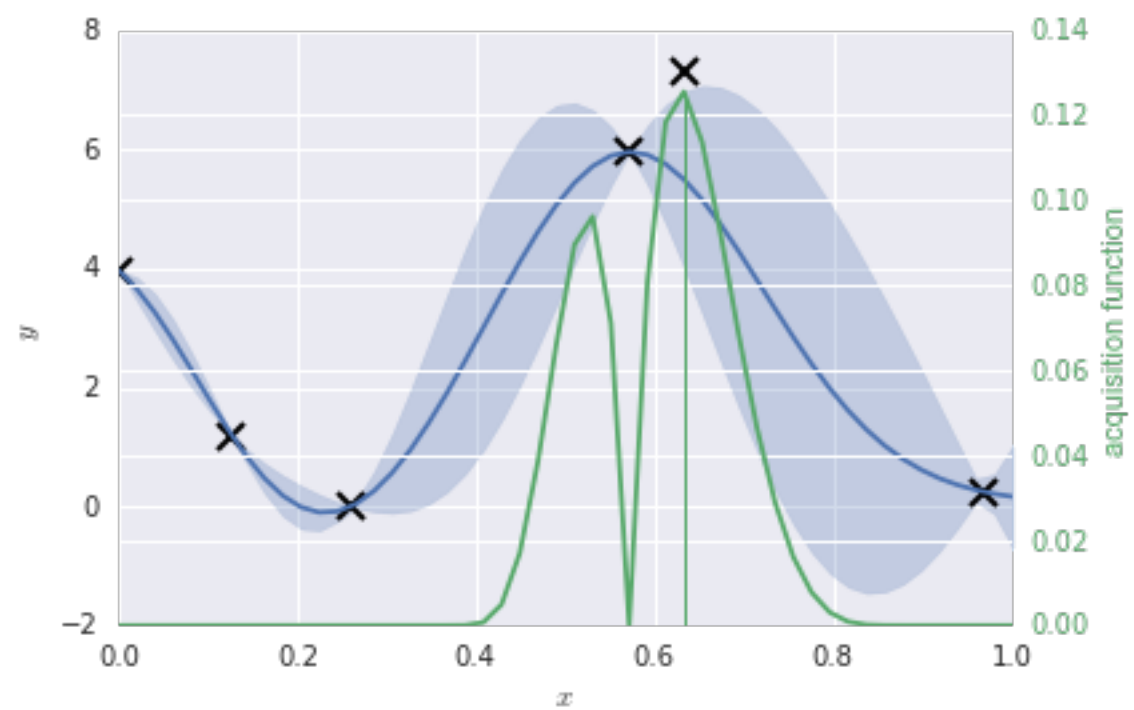


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 3)

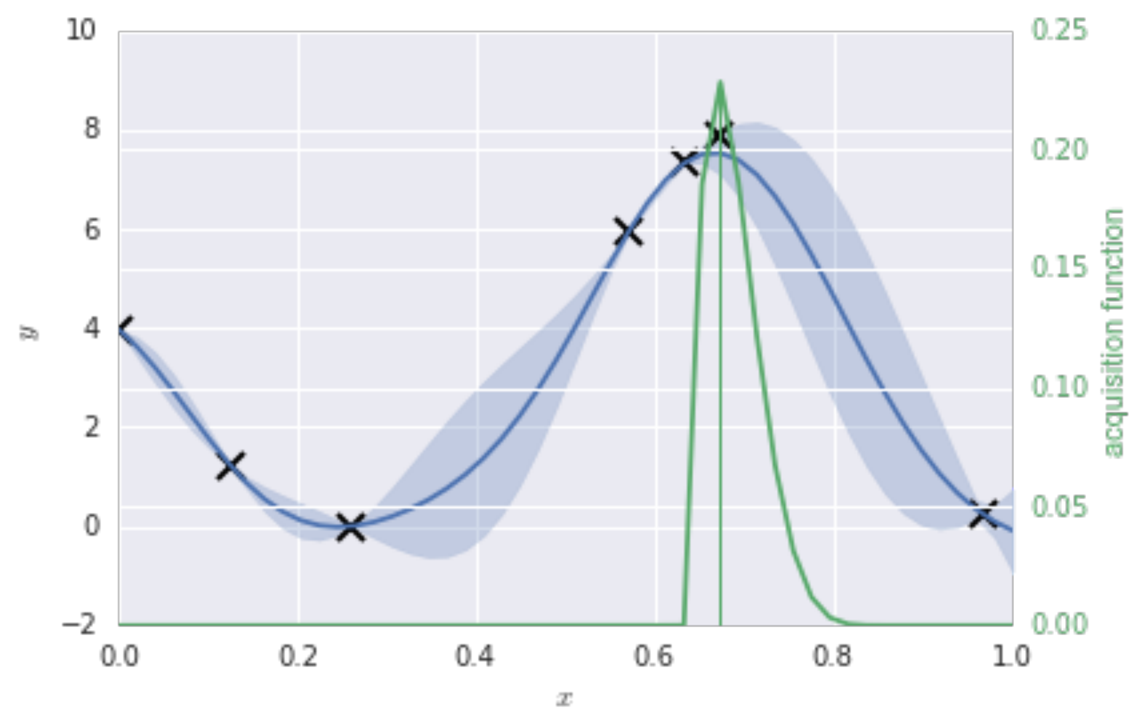


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 3)

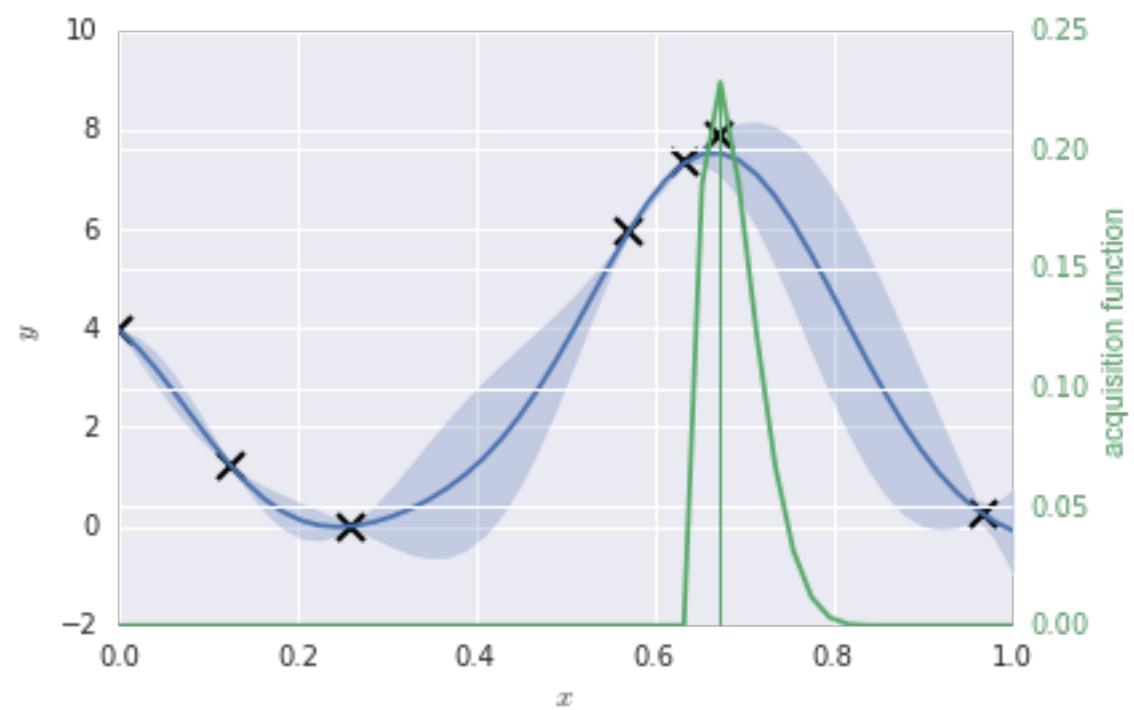


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 4)

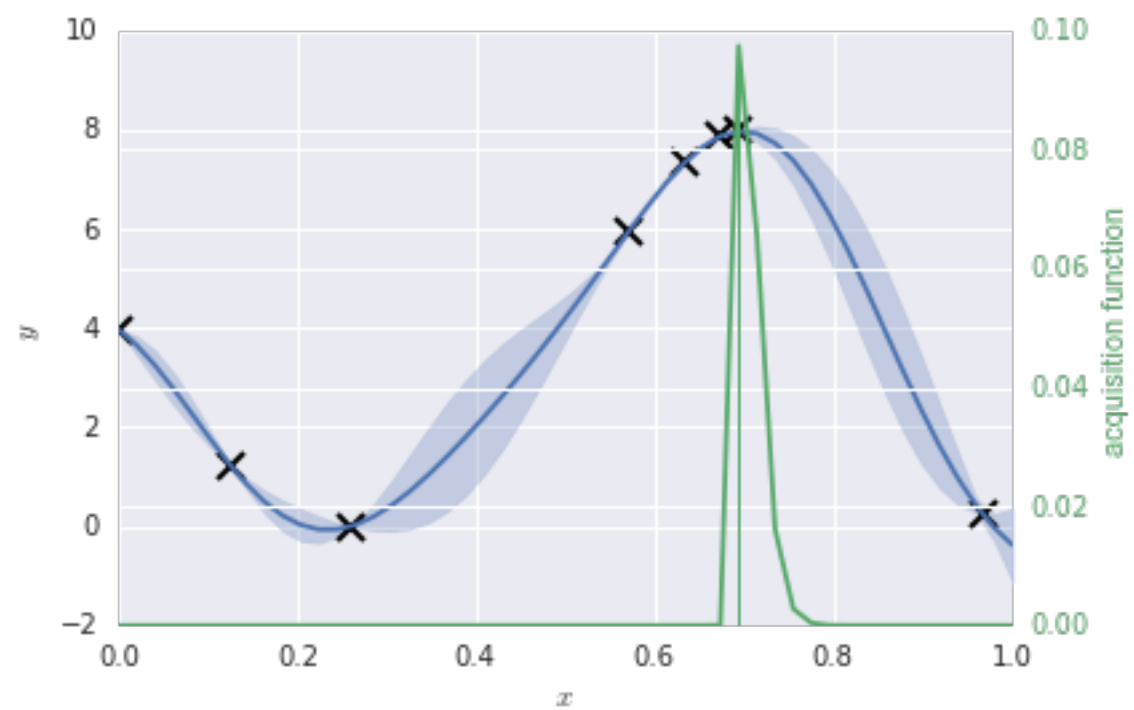


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

PREDICTIVE
SCIENCE LABORATORY

Repeat (Iteration 5)

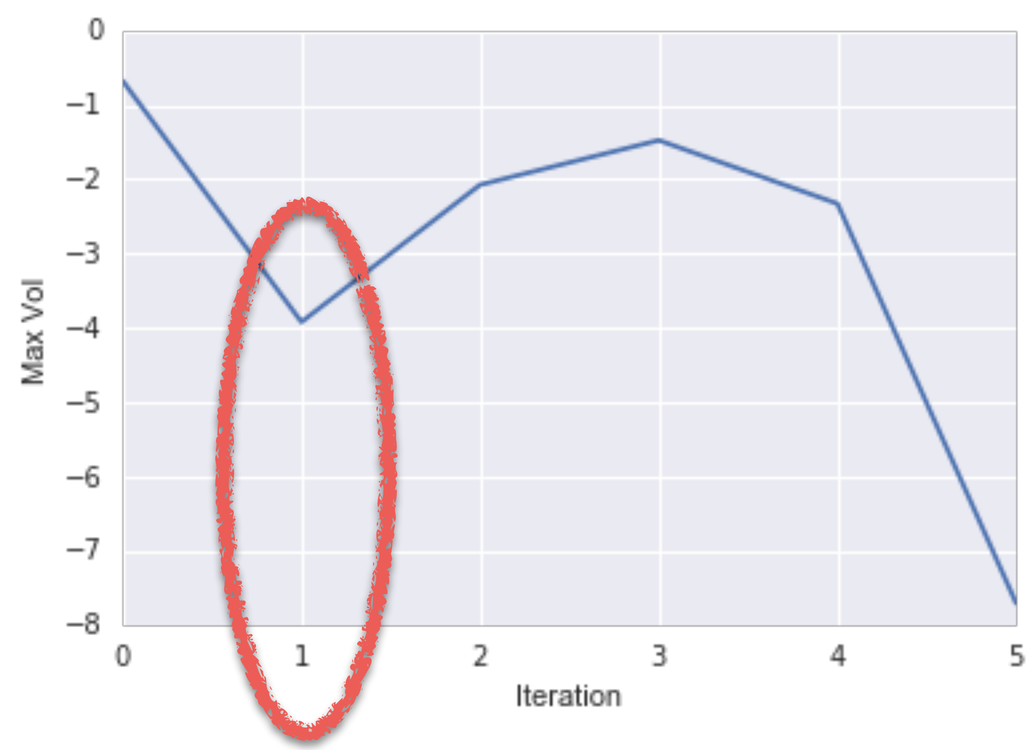


$a(\mathbf{x})$

$$\mathbf{x}_{1:n} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$y_{1:n} = \{y_1, \ldots, y_n\}$$
$$p(y \mid \mathbf{x}) \approx \mathcal{N}\left(y \mid m(\mathbf{x}), \sigma^2(\mathbf{x})\right)$$
$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$$

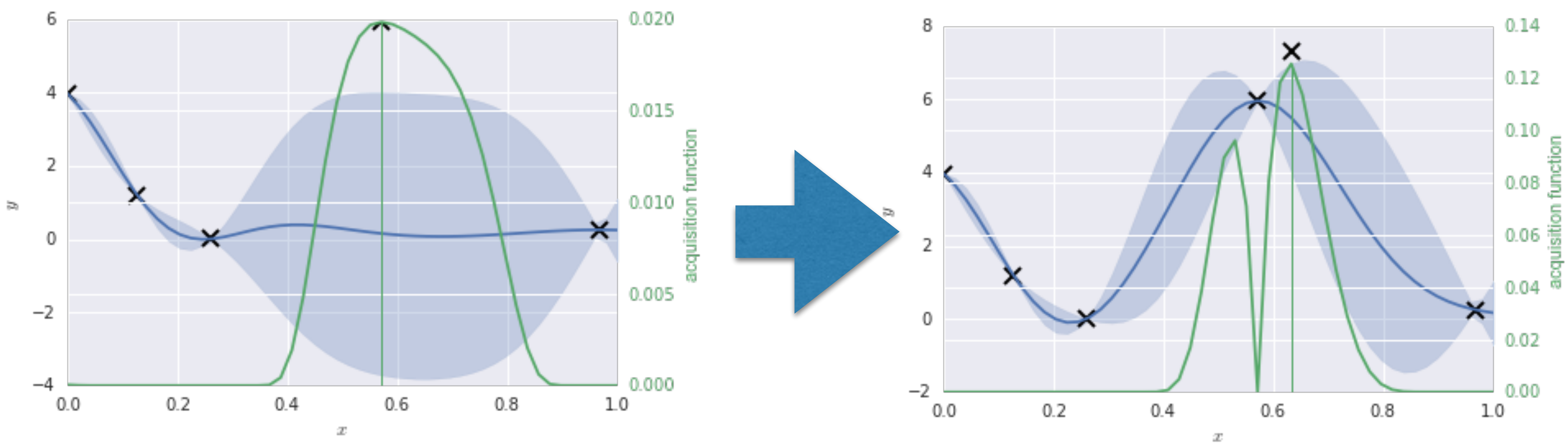
PREDICTIVE
SCIENCE LABORATORY

The S-curve Phenomenon of Information



Unexpected information

The S-curve Phenomenon of Information



PREDICTIVE
SCIENCE LABORATORY

Advanced Topics for Next Time

- Enforcing optimization constraints.
- Information acquisition for multi-objective optimization.
- Information acquisition under uncertainty.
- Information acquisition from various heterogeneous sources.

Start the notebook of lecture 24