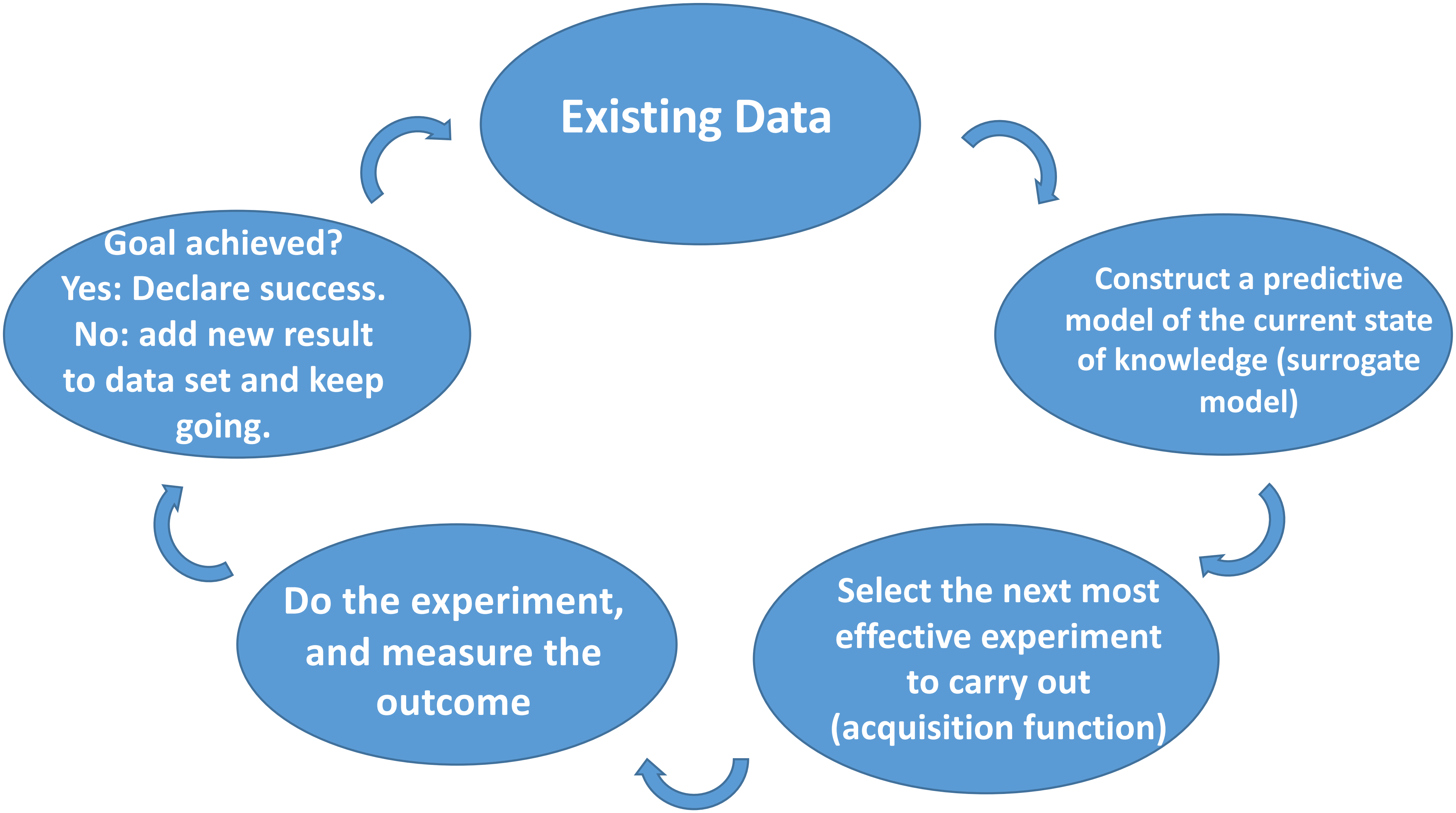
Closed Loop Search



Surrogate Model

- This is just building a model that describes the current data that you have.
- Many ways to do this, we will use Gaussian Process Regression
  - A non-parametric, Bayesian approach to regression where we infer a probability distribution for an unknown function using the samples we have

$$\underbrace{p(f|D)}_{\text{posterior}} = \frac{\overbrace{p(D|f)}^{\text{Likelihood}}\overbrace{p(f)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Marginal likelihood}}}$$

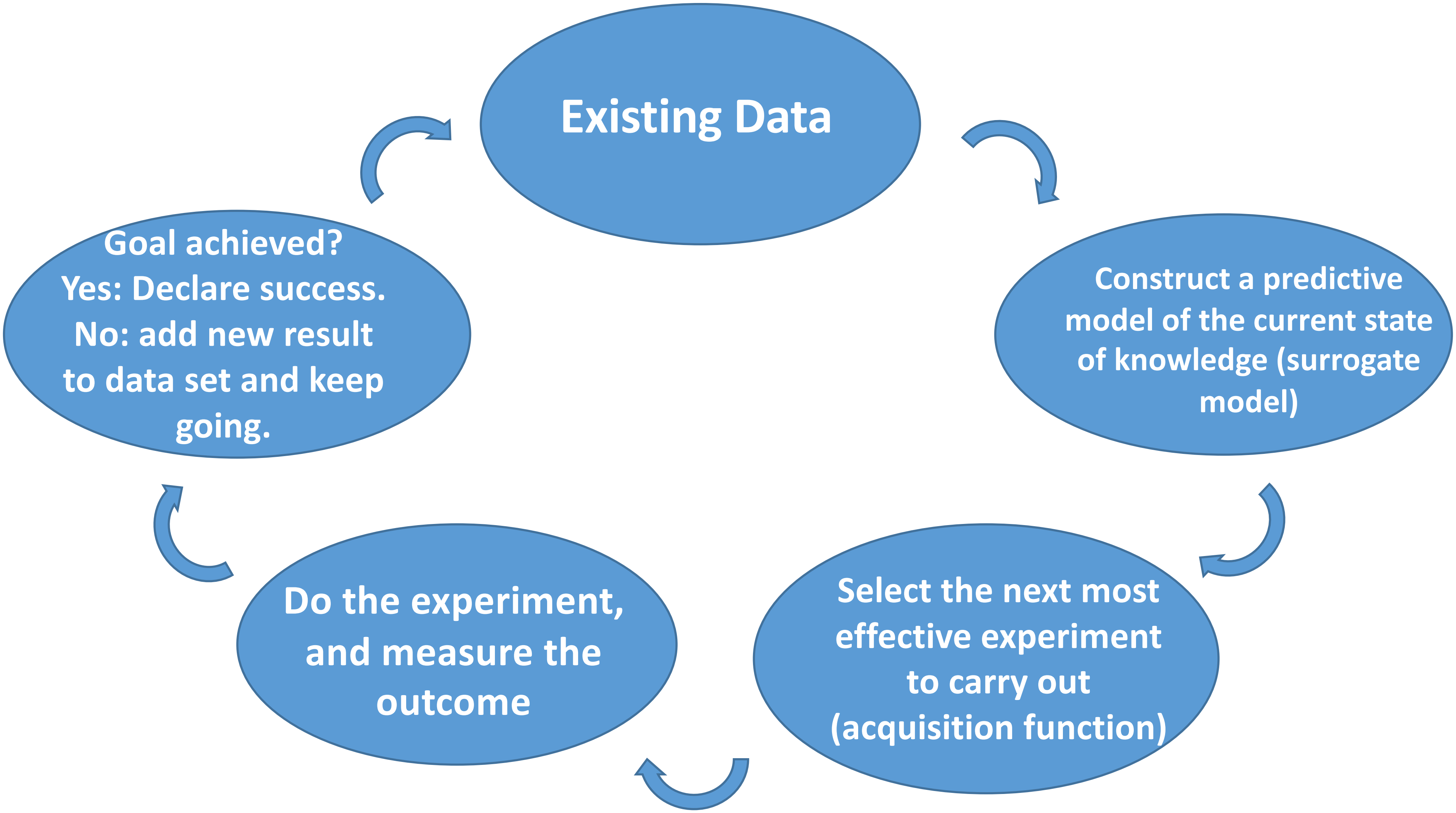
- For now, treat it as a Black Box. We will go into details on Thursday.

Surrogate Model – Gaussian Process Regression

$f^*(GP(x_i^*))$ allows you to predict the value of the function f* for any input x_i*

$\sigma(GP(x_i^*))$ and gives us confidence intervals σ for our estimate of f*

Closed Loop Search



Acquisition Function

Lots of possible ways to choose what is the next best experiment to try:

- Maximum expected improvement (exploitation)

$$x^* = argmax\,(f^*(GP(x_i^*)))$$

- Maximum uncertainty (exploration)

$$x^* = argmax\,(\sigma(GP(x_i^*)))$$

- Maximum probability of improvement

$$x^* = argmax\,\left(\frac{f^*\big(GP(x_i^*)\big) - f^*(GP(x_{best}))}{\sigma(GP(x_i^*))}\right)$$

Acquisition Function

Exercise

- We will explore Bayesian search methods to find the maximum value of a function
- The function is called the objective function:

$$f(x) = x\sin(x) \quad \text{for} \quad x \in [0,10]$$

- We will see how many experiments (evaluations of the real function) we need to find the maximum
- We will explore how noisy measurements of the objective function affect our ability to learn it.
- (For later: to really see the power of the approach try making the function more complex, higher-dimensional, etc.)

Link & Questions

https://colab.research.google.com/drive/1evDv8_yUl8vduslZguYJYbV9f6D_zqK6?usp=sharing

Question 1. In Part 1 – Imagine that you sample the objective function, but your measurements are noisy. What happens to your ability to learn the underlying function as the noise increases? Should you just pick your highest sampled point as the maximum?

Question 2. In Part 2, make sure you understand what all the lines on the plots are. How well does the Gaussian Process Regressor work as a surrogate function? How does your answer change if you vary the noise from 0.01,0.1,1,10?

Question 3. In Part 2, does choosing input values x to measure the function randomly make it easier or harder to estimate the function? When do you think randomly selecting experiments would be useful?

Question 4. In Part 3, what does the sequence of plots in the bottom row show? How does the next data point get selected, from one plot to the next?

Question 5. In Part 3, modify the code to vary the acquisition function, between exploration, exploitation, and maximum probability of improvement. How does it affect the performance of BO?