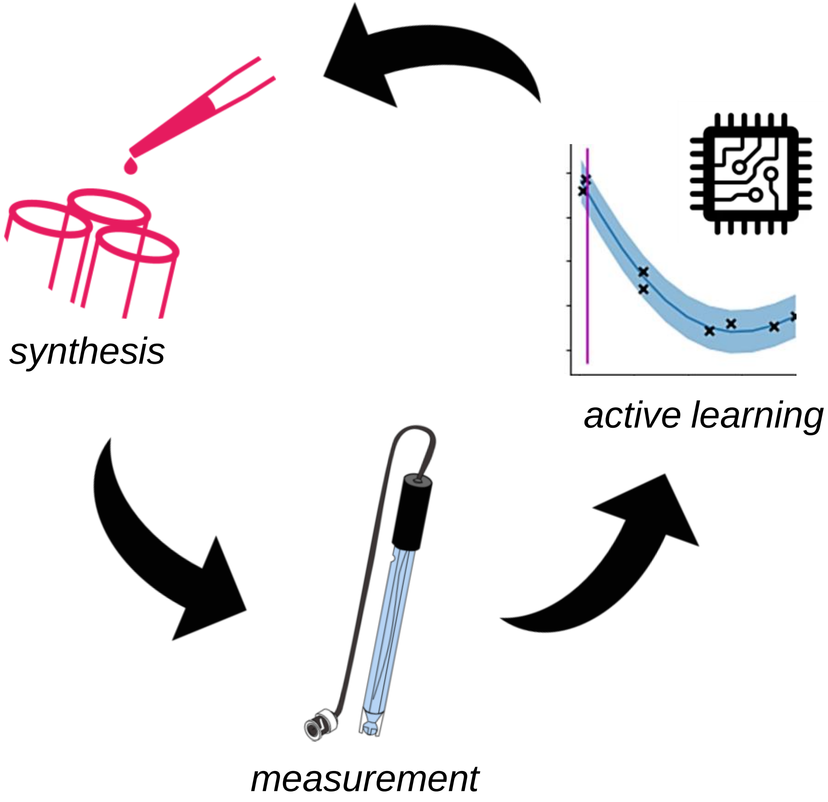
Bayesian Active Learning for Autonomous Physical Science

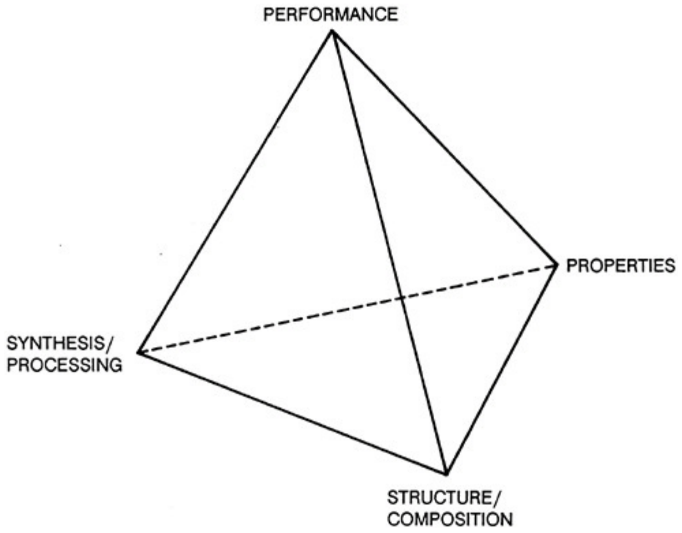


Logan Saar

Dr. Gilad Kusne (Mentor) - MML

Motivation and Goals

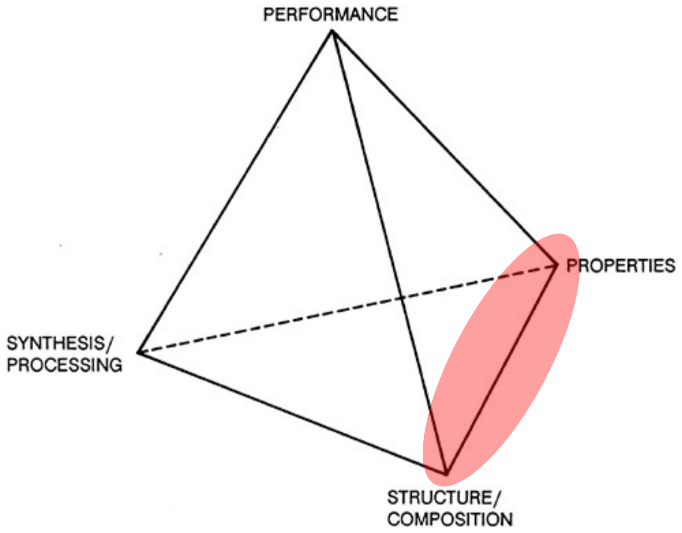
Composition - Property Relationships



Optimize Performance → Properties → Composition

Image Credit: The four elements of materials science and engineering. (n.d.). https://www.nap.edu/read/758/chapter/3.

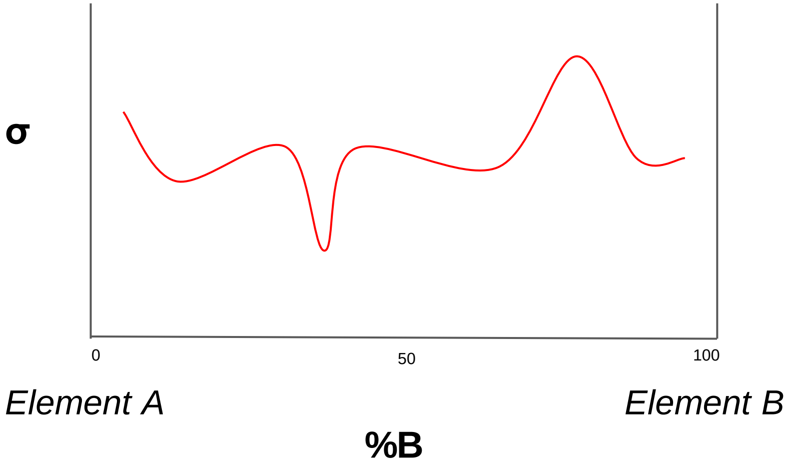
Composition - Property Relationships



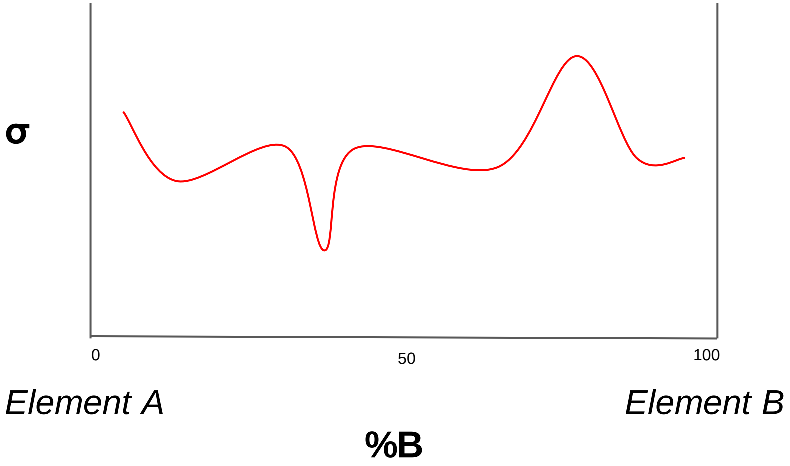
Optimize Performance → Properties → Composition

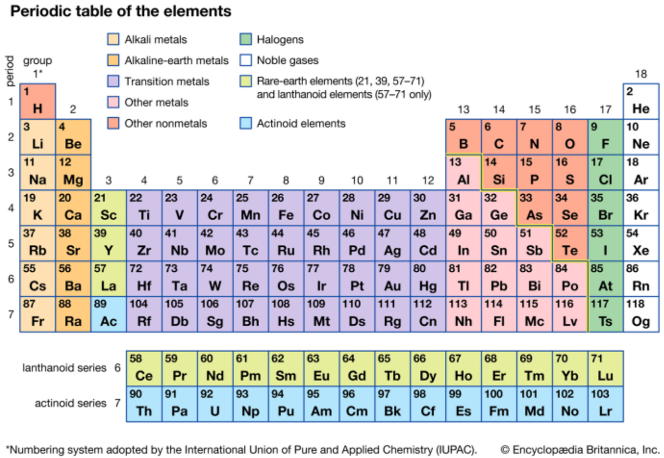
Image Credit: The four elements of materials science and engineering. (n.d.). https://www.nap.edu/read/758/chapter/3.

The Large Number Problem

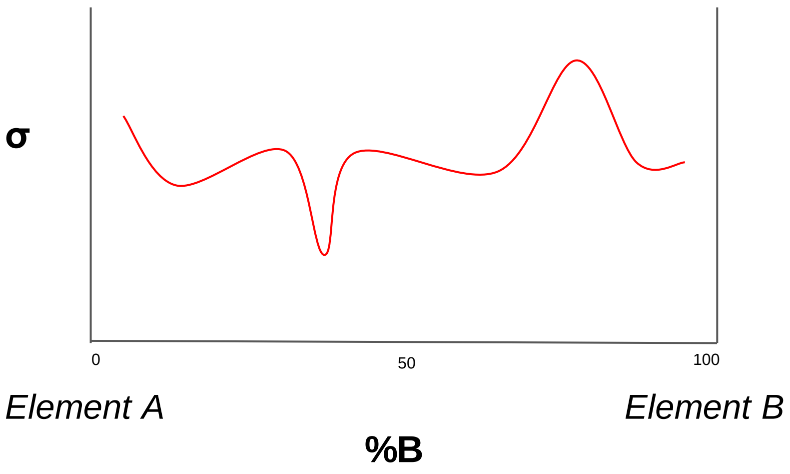


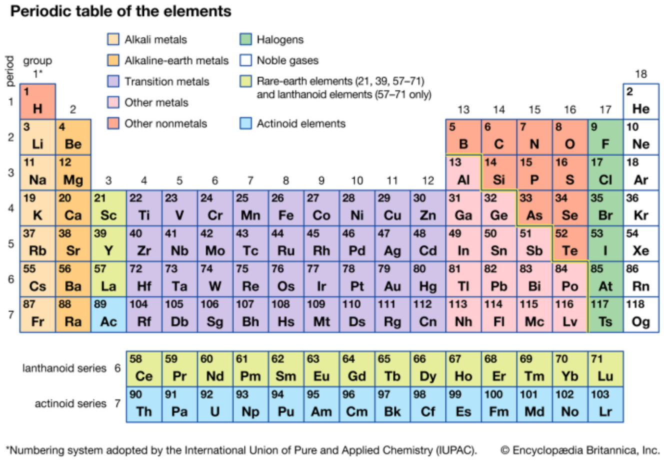
The Large Number Challenge





The Large Number Challenge



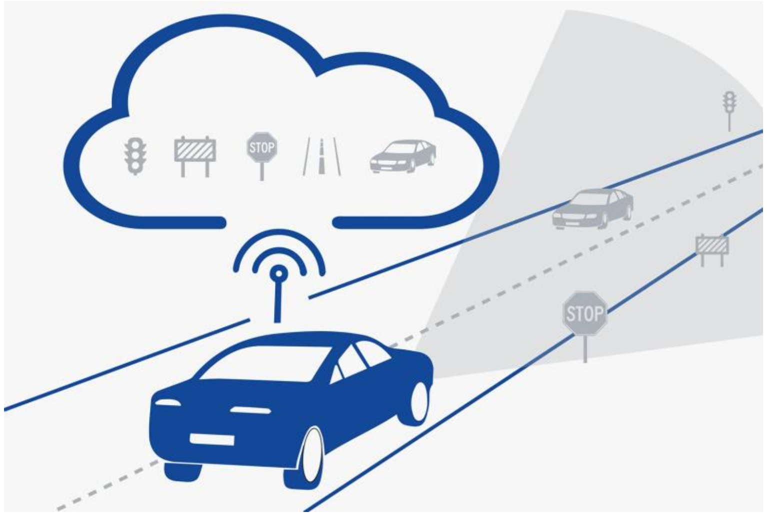


Binary: ~ *7,000*
Tertiary: ~ *250,000*
Quaternary: ~ *7.6 million*

Automated

Robot **executes** tasks

Autonomous

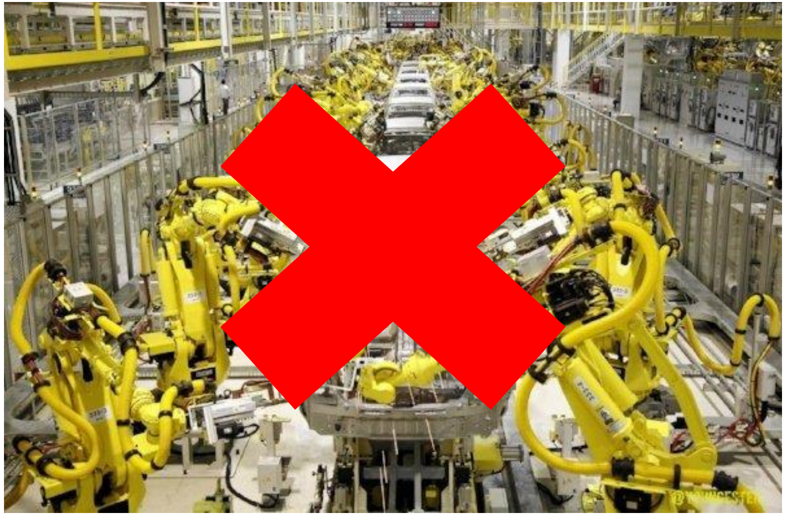


Robot **reacts** to input

Image Credit: https://blog.robotiq.com/a-brief-history-of-robots-in-manufacturing

Image Credit: Wired - Autonomous Cars

Automated



Robot **executes** tasks

Autonomous

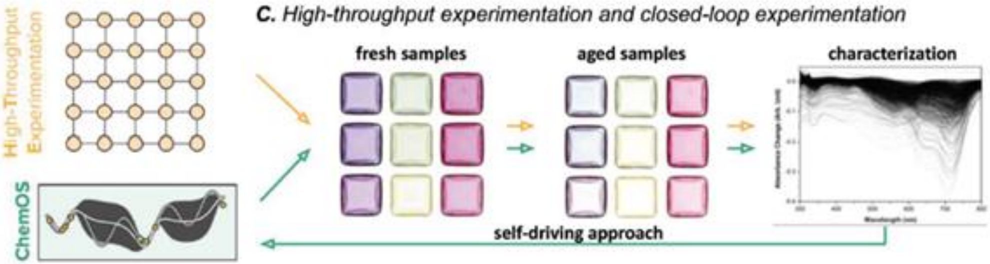
Robot **reacts** to ..
gathered data

ACTIVE LEARNING

A mobile robotic chemist

Burger et al., Nature **583**, 237 (2020)

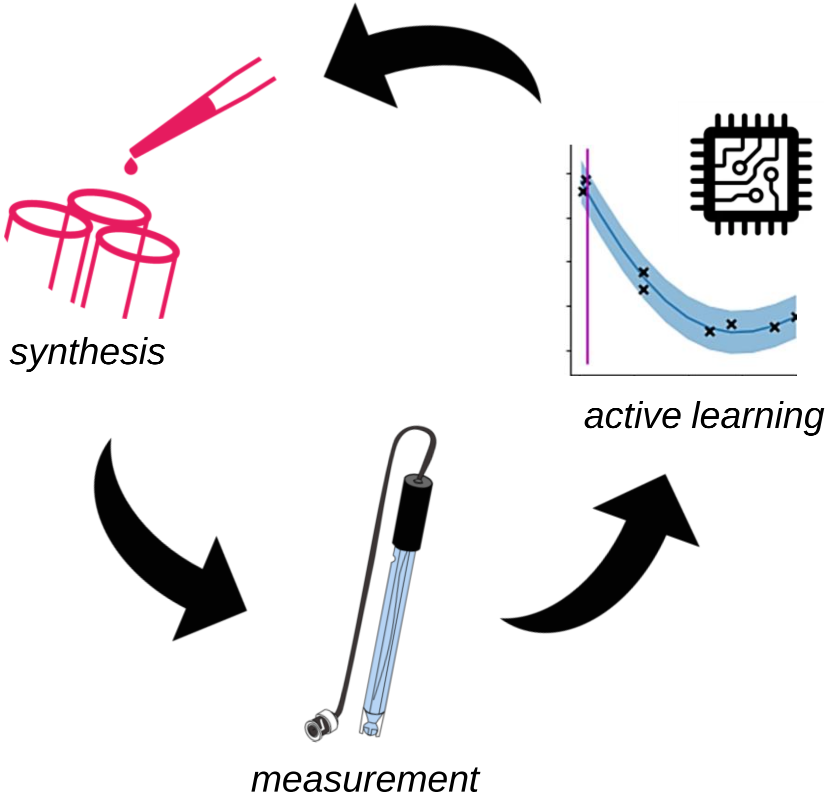
Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems

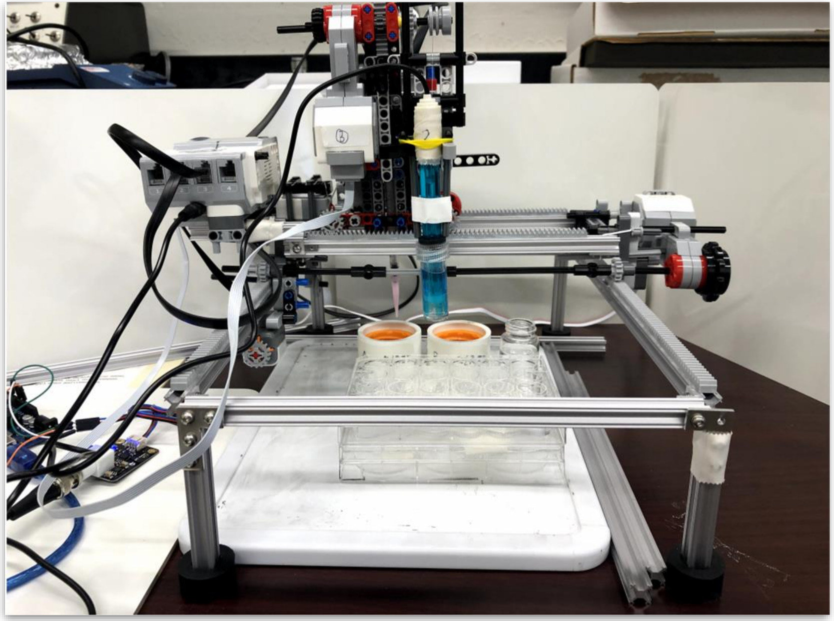


- Blending/mixing or polymers/organic molecules
- Number of experiments can be significantly reduced

Burger, B., Maffettone, P.M., Gusev, V.V. et al. A mobile robotic chemist. Nature 583, 237–241 (2020)

Low Cost Autonomous Physical Science System





Active Learning Closed Loop System

Our Mission

Composition Space
Weak Acid - *Acetic Acid* - 1 M
Conjugate Base - *Sodium Acetate Solution* - 1 M

Goal
Recover Henderson-Hasselbalch Equation.

Henderson-Hasselbalch (HH) Equation:

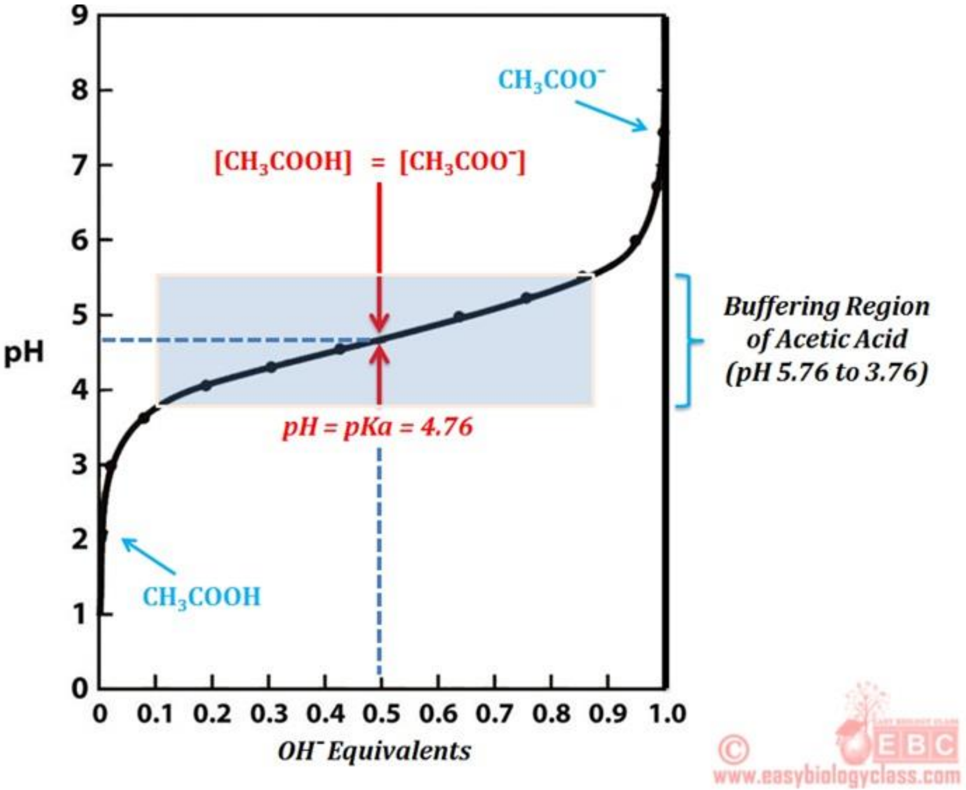
$$\mathrm{pH} = \mathrm{p}K_\mathrm{a} + \log_{10}\left(\frac{[\mathrm{Base}]}{[\mathrm{Acid}]}\right)$$

Our known parameters
(for synthesis)

Response Variable
(measured)

Dissociation Constant
(unknown to robot)

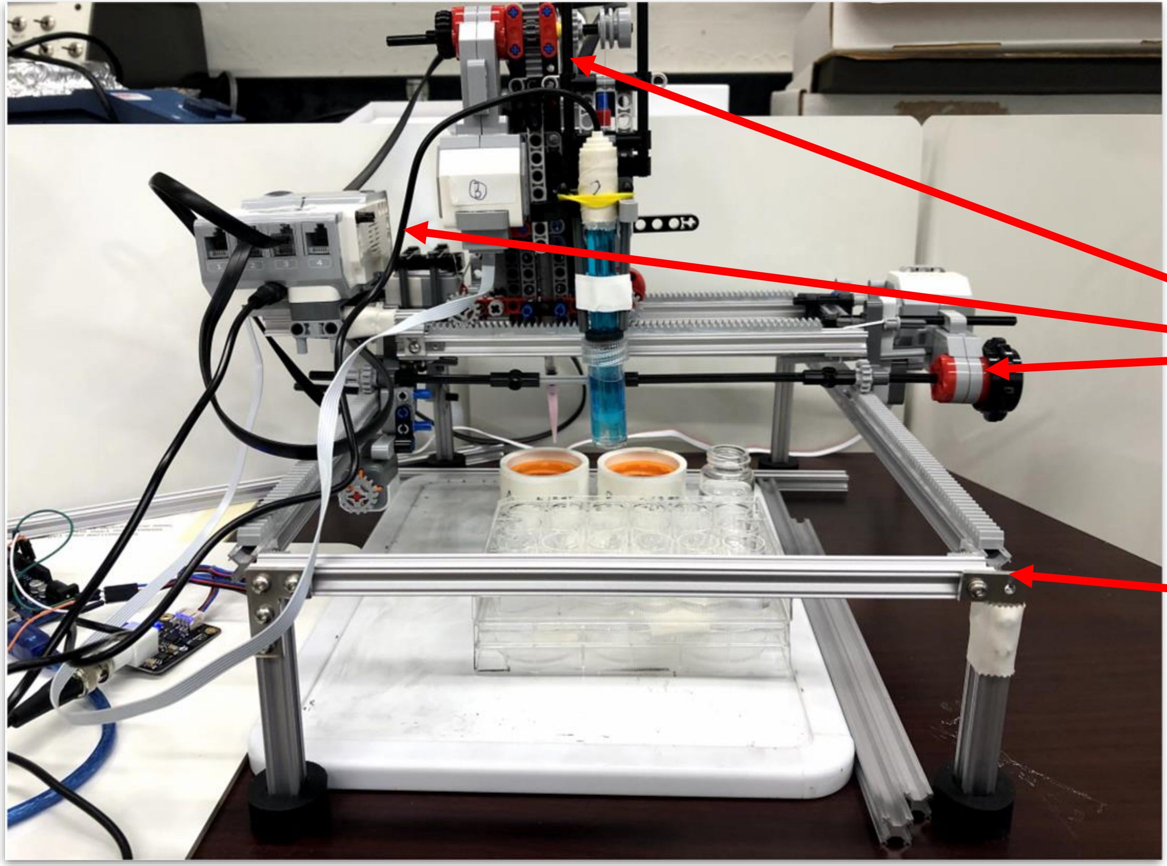
Our Mission



$$\mathrm{pH} = \mathrm{p}K_\mathrm{a} + \log_{10}\left(\frac{[\mathrm{Base}]}{[\mathrm{Acid}]}\right)$$

Image Credit: https://www.easybiologyclass.com/titration-curve-of-a-weak-acid-and-its-pka-biochemistry-notes/

Active Learning Closed Loop System - Our Robot



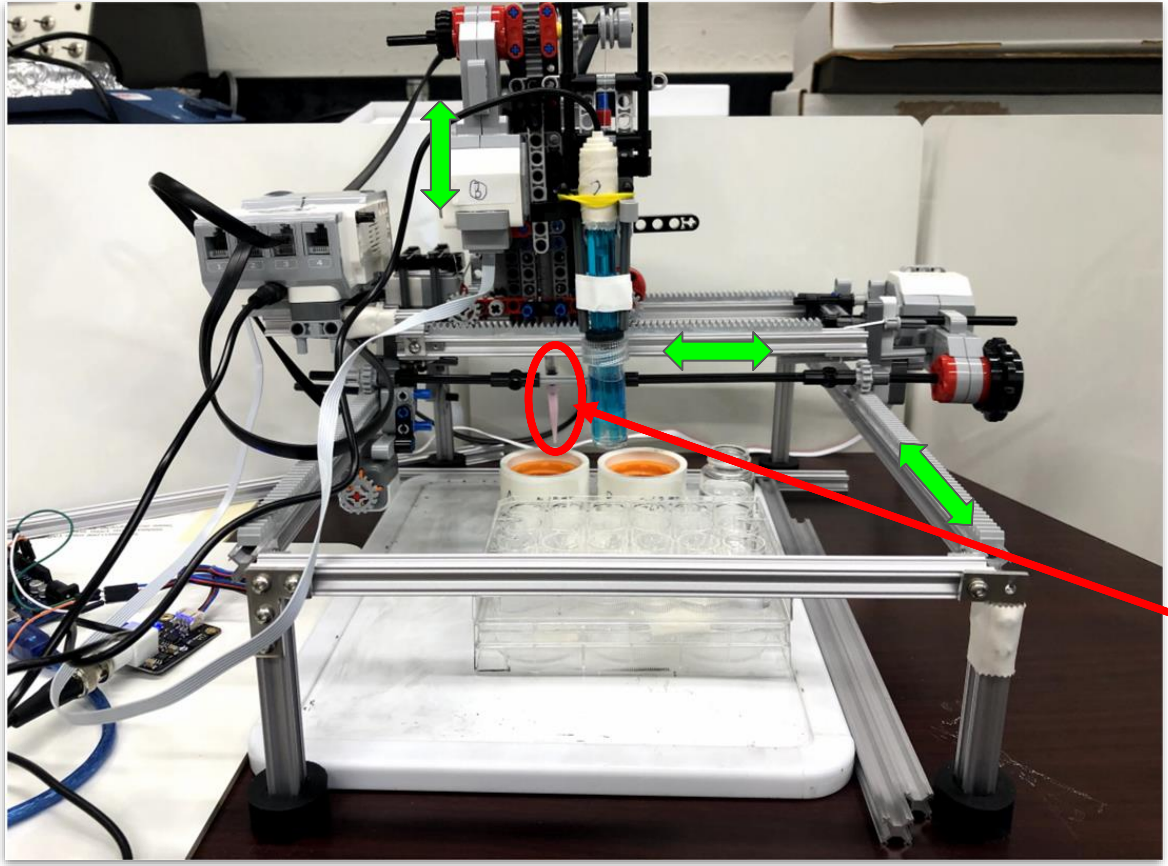
Construction

Lego Mindstorm Components

Aluminum Frame

Gerber, L. C. (2017). Liquid-handling Lego robots and experiments for STEM education and research. PLOS Biology, 15(3)

Active Learning Closed Loop System - Our Robot

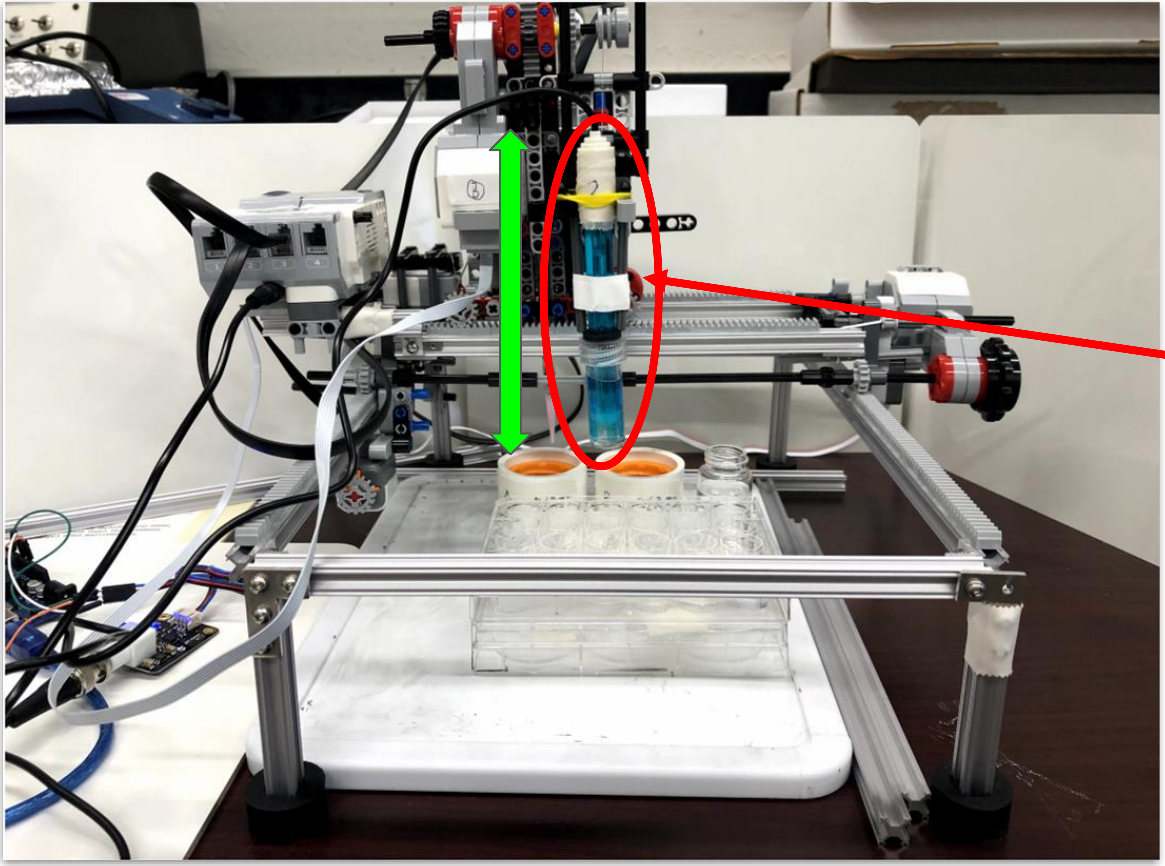


Sample Synthesis

X - Y - Z mobility

Syringe for sample collection / deposition (volume controlled)

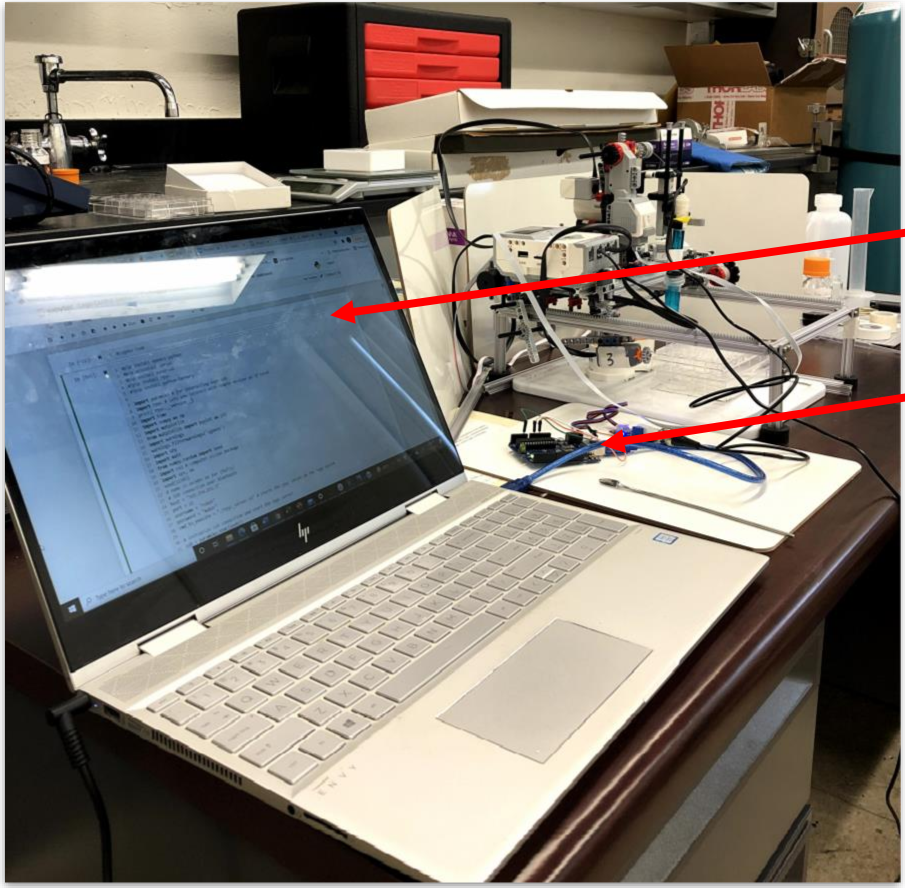
Active Learning Closed Loop System - Our Robot



Measurement

Arduino electrochemical
pH probe
(voltage readings)

Active Learning Closed Loop System - Our Robot



Setup

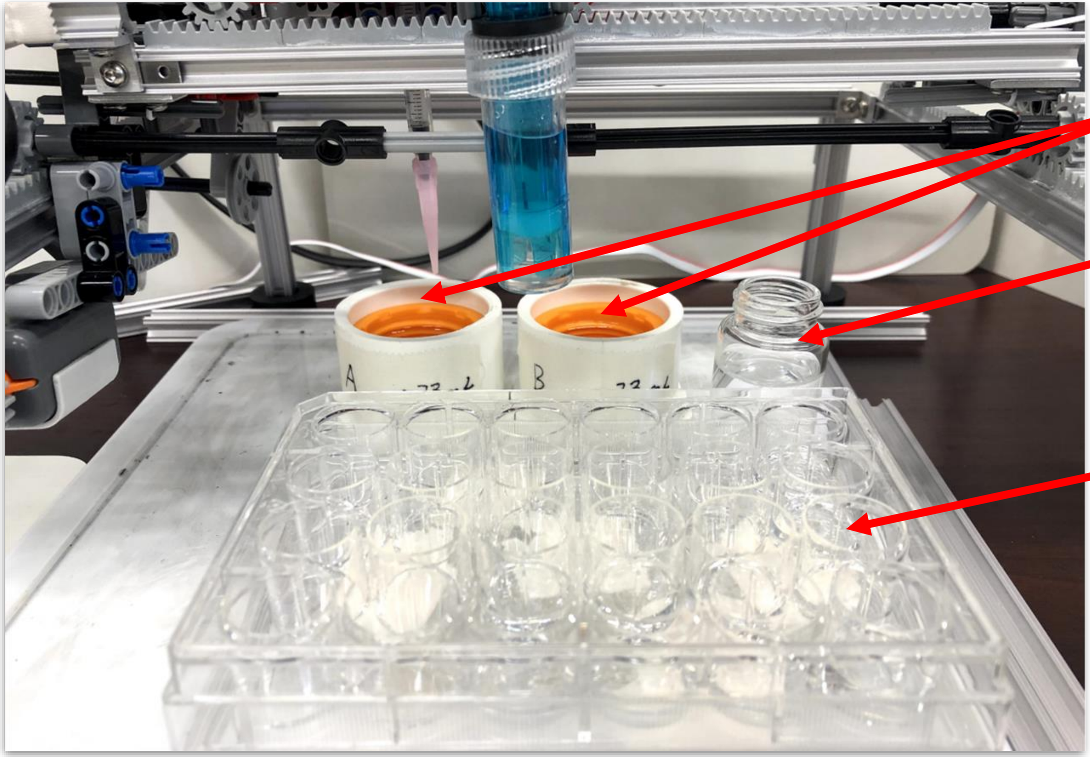
Python Script

Arduino pH meter/USB

PC/Robot BT connection

Active Learning Closed Loop System - Our Robot

Process

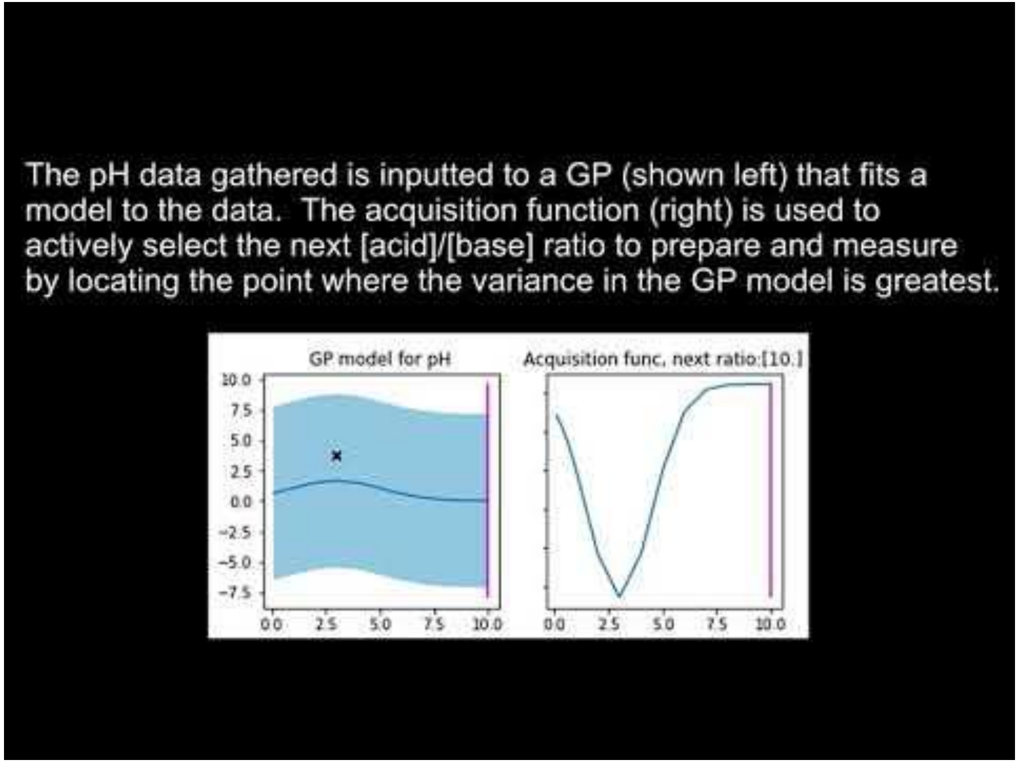


Acid and base reservoirs

DI water (probe cleaning)

Plastic sample wells

Active Learning Closed Loop System - Our Robot

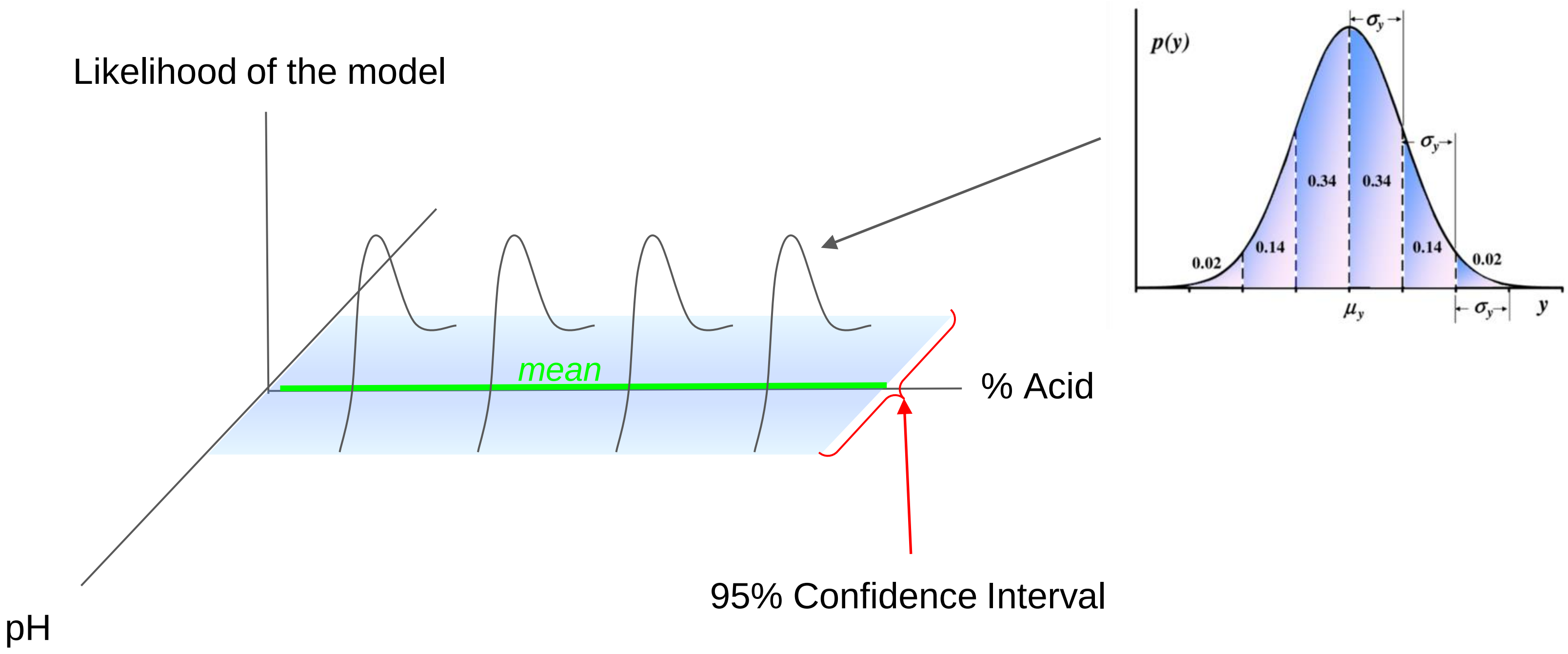


Link Provided in Chat

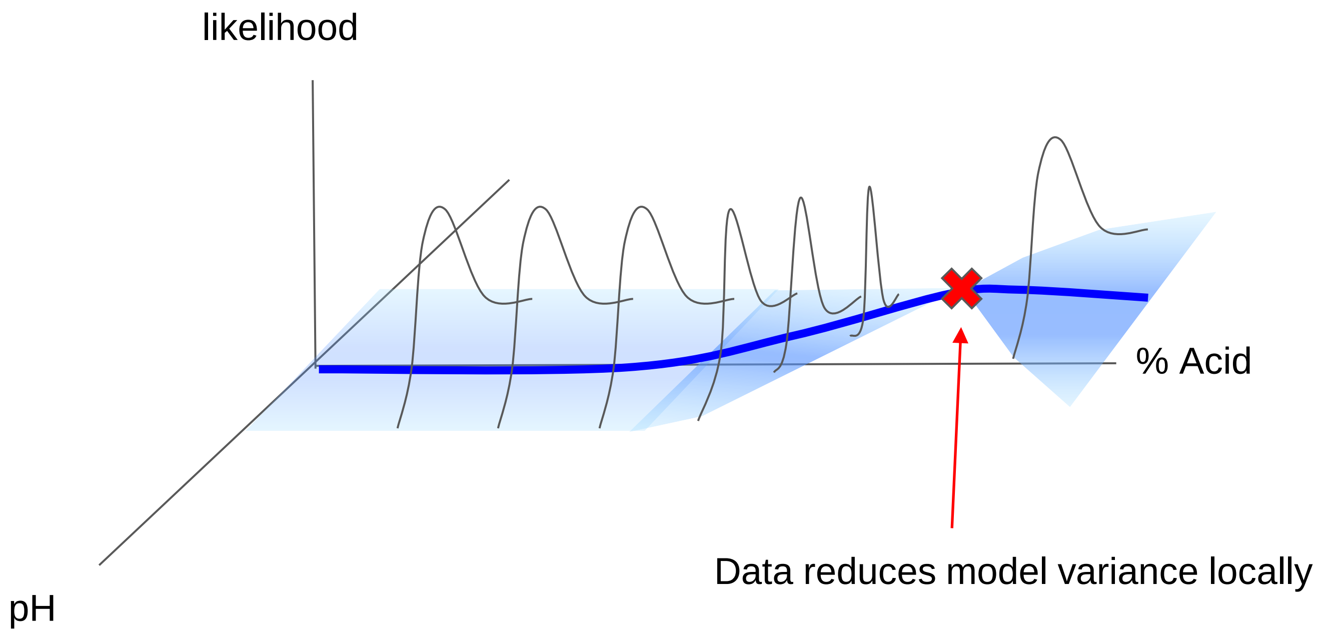
Statistical Methods and Results

Exploration Initiative - **(Gaussian Process)**

Normal distribution of predicted pH for each potential % Acid composition



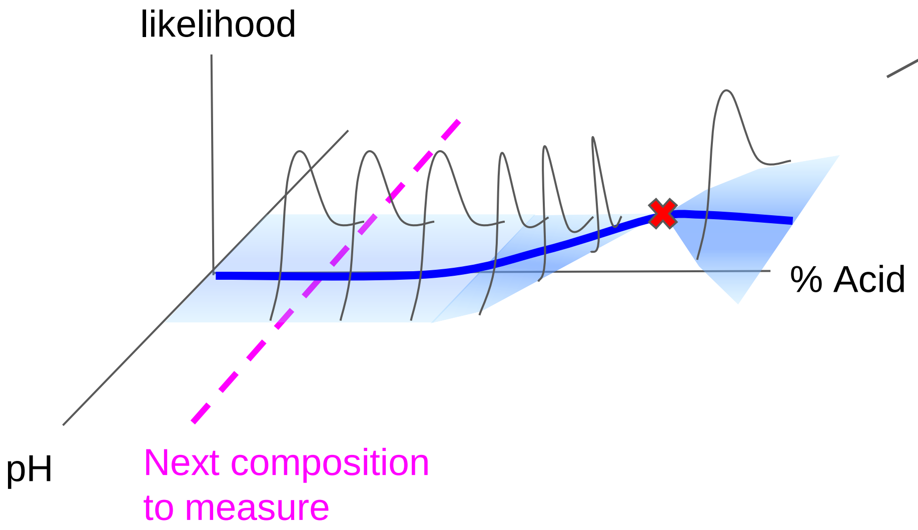
Exploration Initiative - **(Gaussian Process)**

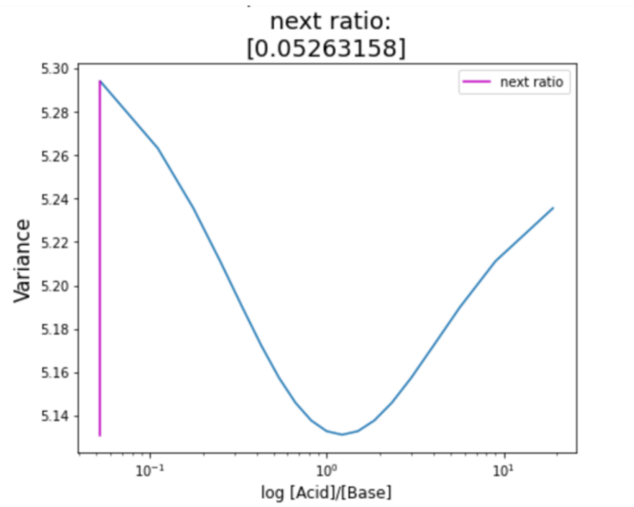


Exploration Initiative - **(Gaussian Process)**

Active Learning:

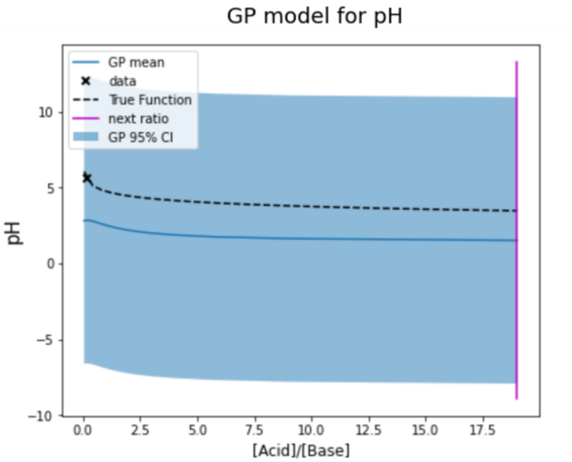
- ➔ Acquisition Function
- ➔ *argmax (variance)*

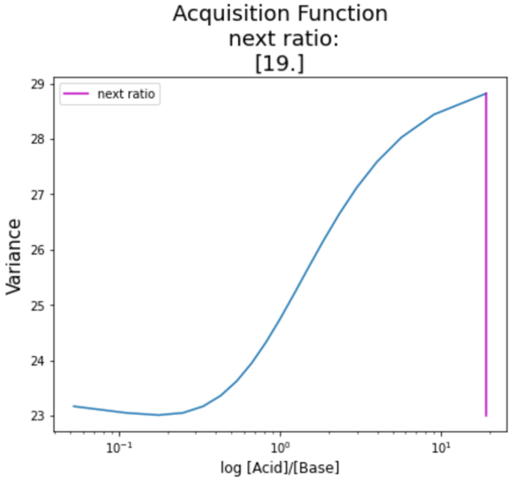




Autonomous Results - **(Gaussian Process)**

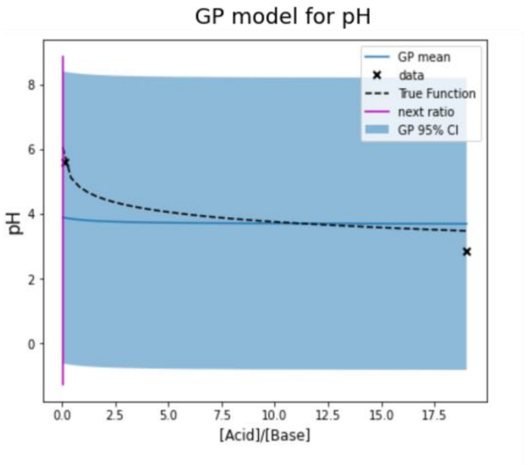
1 data point

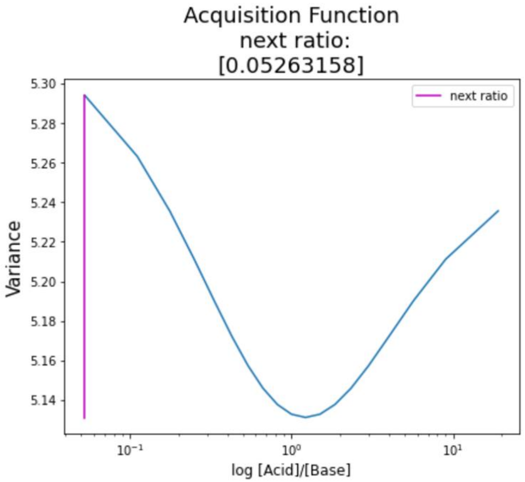




Autonomous Results - **(Gaussian Process)**

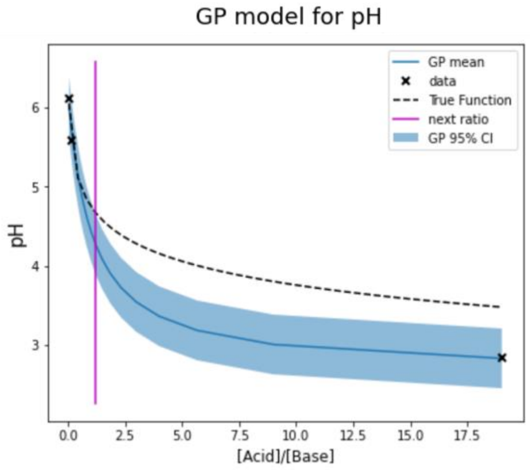
2 data points

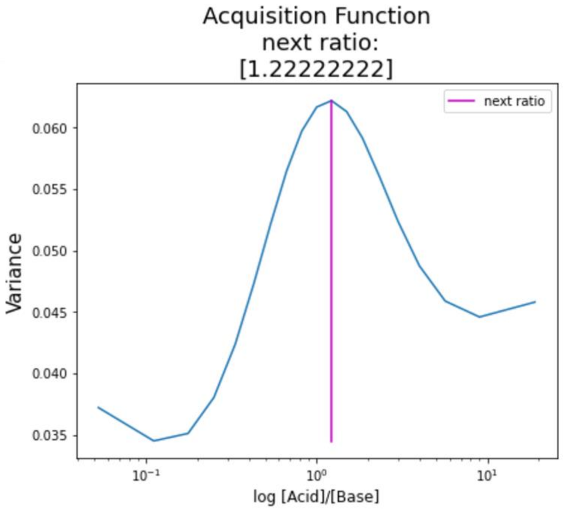




Autonomous Results - (Gaussian Process)

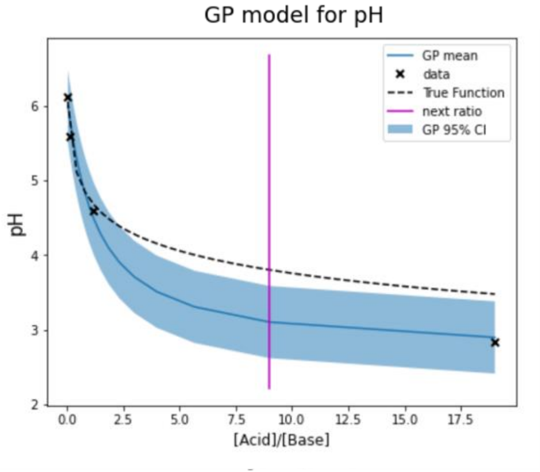
3 data points

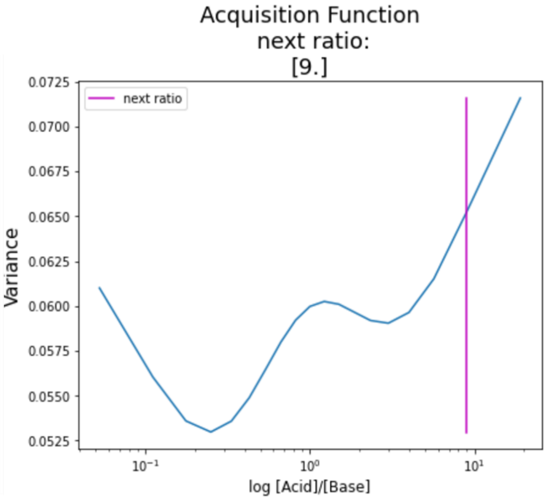




Autonomous Results - **(Gaussian Process)**

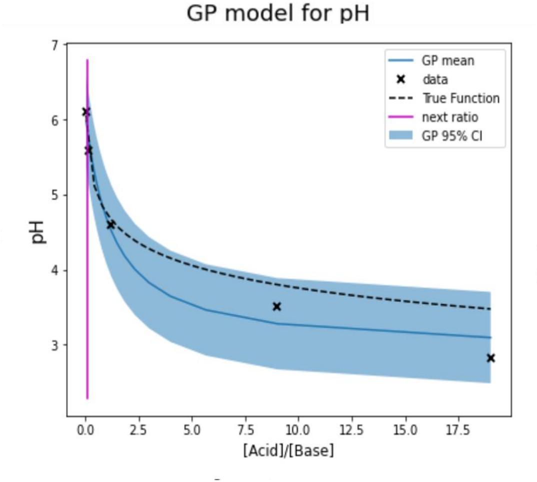
4 data points

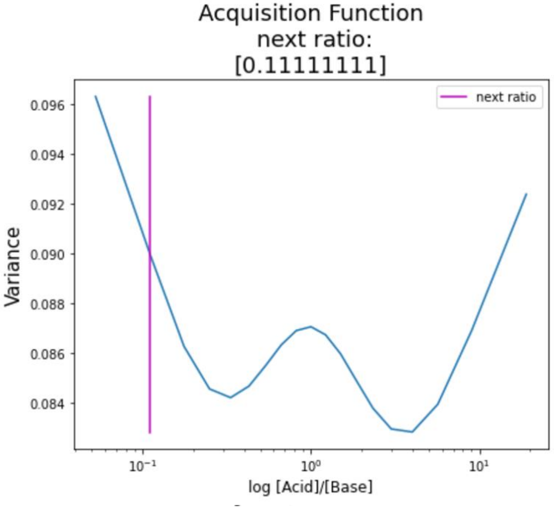




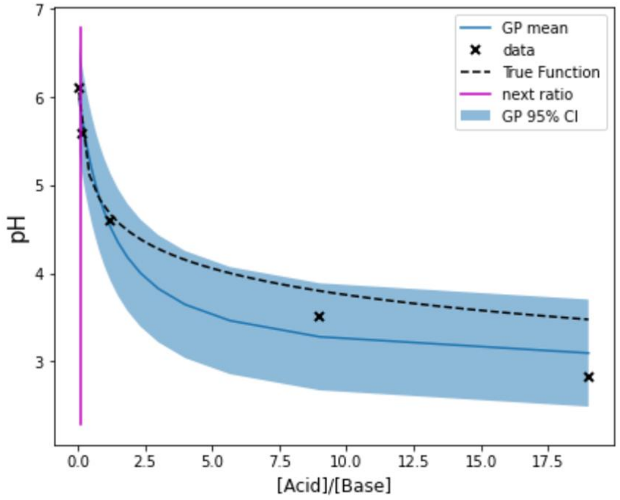
Autonomous Results - <u>**(Gaussian Process)**</u>

5 data points





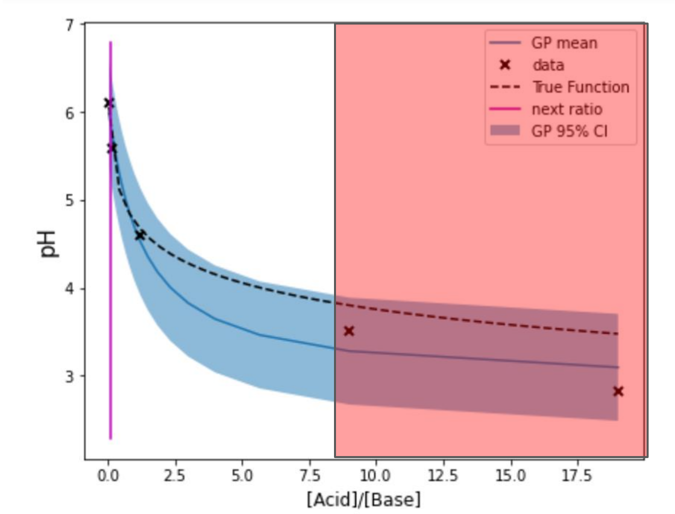
GP flexibility - **(Gaussian Process)**



HH equation relies on assumptions

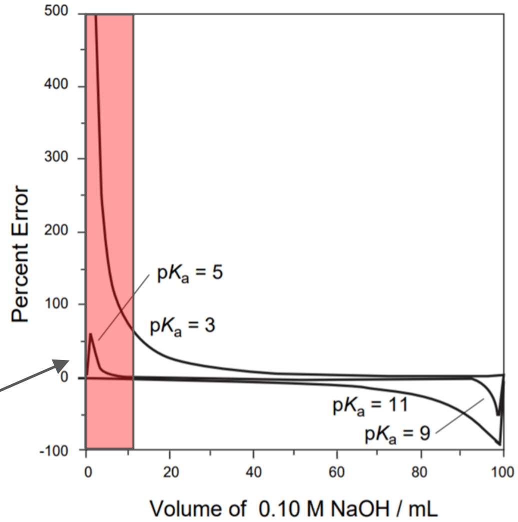
- ➔ No self-ionization of water
- ➔ Valid only in certain composition range
- ➔ Our pKa ~ 4.7

GP flexibility - <u>**(Gaussian Process)**</u>



HH equation relies on assumptions

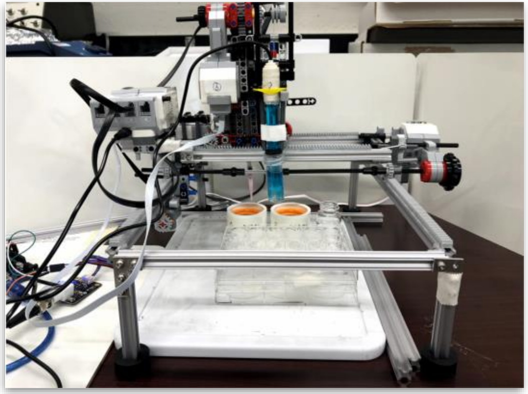
- ➔ No self-ionization of water
- ➔ Valid only in certain composition range
- ➔ pKa ~ 4.7

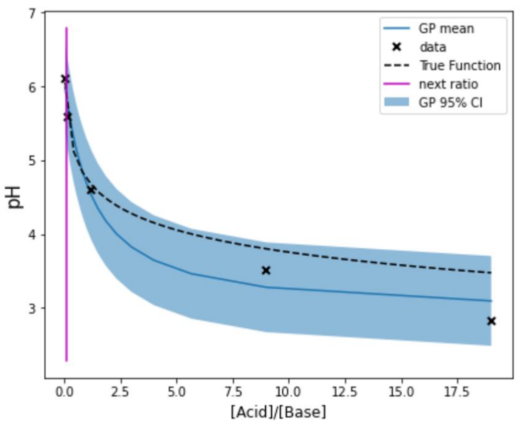


↑↑ % error in HH simplification ...

Summary

Closed Loop Autonomous Science System
- ➔ Educational Tool
- ➔ Low-cost
- ➔ Modular
- ➔ Materials Exploration





Used Gaussian Processes to explore pH as function of composition
- ➔ Flexible model
- ➔ Explore other active learning methods ...

Acknowledgments

Dr. Gilad Kusne, PhD

Dr. Ichiro Takeuchi, PhD

Dr. Austin McDannald, PhD

Alex Wang

Haotong Liang

Questions

References

Burger, B., Maffettone, P.M., Gusev, V.V. *et al.* A mobile robotic chemist. *Nature* 583, 237–241 (2020). https://doi.org/10.1038/s41586-020-2442-2

De Levie, R. (2003). The Henderson-hasselbalch equation: Its history and limitations. Journal of Chemical Education, 80(2), 146. https://doi.org/10.1021/ed080p146

Gerber, L. C., Calasanz-Kaiser, A., Hyman, L., Voitiuk, K., Patil, U., & Riedel-Kruse, I. H. (2017). Liquid-handling Lego robots and experiments for STEM education and research. PLOS Biology, 15(3), e2001413. https://doi.org/10.1371/journal.pbio.2001413

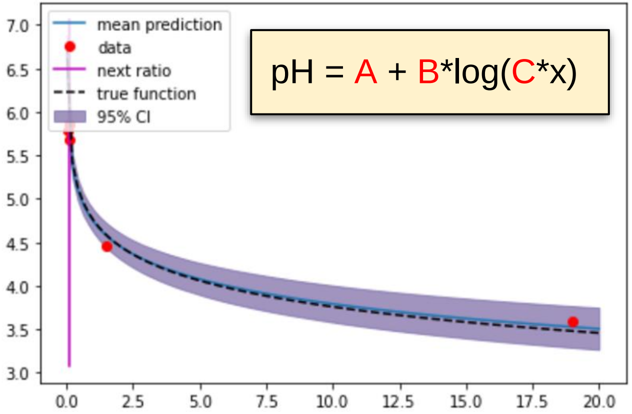
Appendix

The Next Steps:

For pH Measurement Setup

- ➔ Parameter Refinement

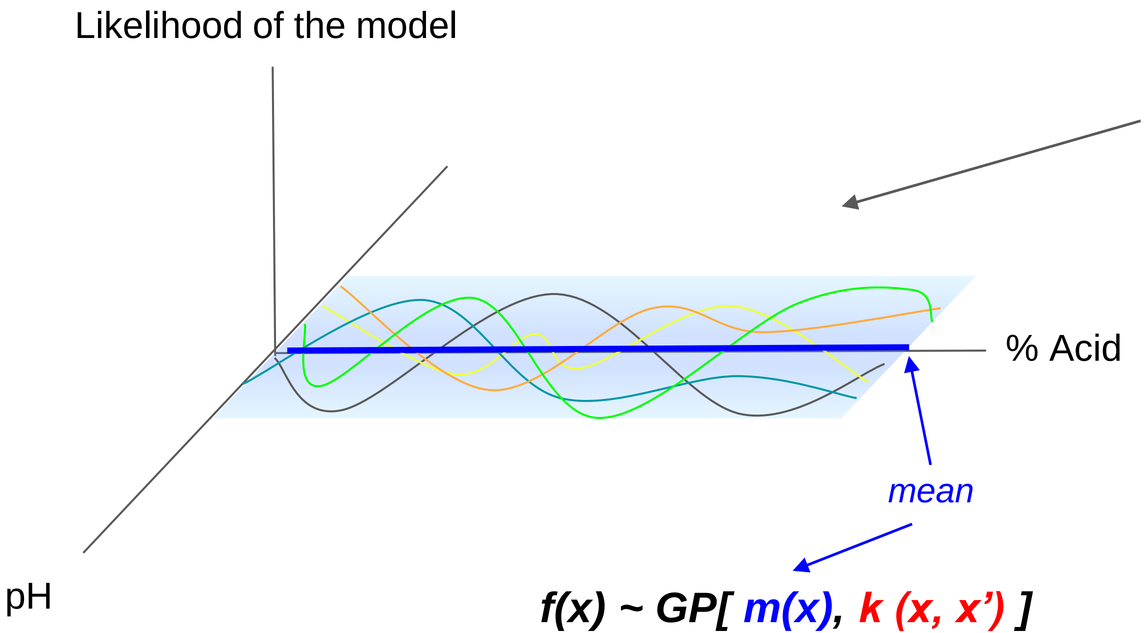
- ➔ Hypothesis Testing
  - ➔ Bayesian methods
  - ➔ Filter between candidate functions



Other Applications (Educational Tool)

- ➔ Camera attachment
  - ◆ Learn color mixing trends

Exploration Initiative - **(Gaussian Process)**



Many models generated:

- ➔ Non-parametric
- ➔ Flexible
- ➔ Averaged
- ➔ Kernel determines "smoothness"
- ➔ Noise tolerant

f(x) ~ GP[m(x), k (x, x')]

Kernel is a covariance function

Brief Overview - Bayesian Machine Learning

Probabilistic interpretation … quantifying **uncertainty** (how confident are we?)

Bayes Theorem

$$\mathbf{P\,(model \mid data)} = \frac{P(\text{data} \mid \text{model})\, P(\text{model})}{P(\text{data})}$$

Brief Overview - Bayesian Machine Learning

Probabilistic interpretation … quantifying **uncertainty** (how confident are we?)

Bayes Theorem

$$\mathbf{P\,(model \mid data)} = \frac{P(\text{data} \mid \text{model})\,\mathbf{P(model)}}{\cancel{P(\text{data})}}$$

“Posterior”

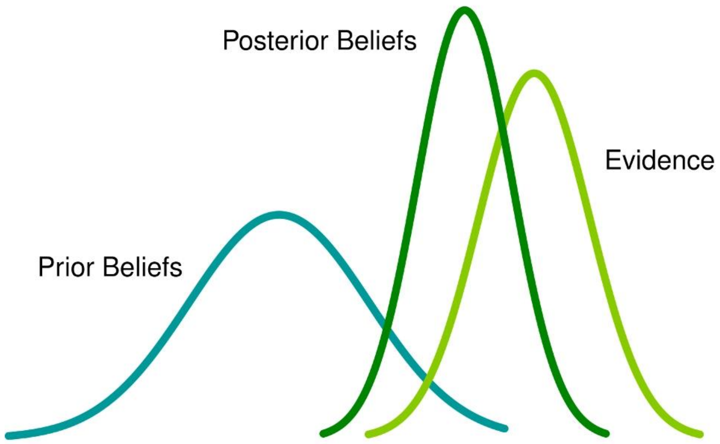
Our confidence in this model being “correct” given the data (what we want to know)

“Prior”

Our confidence in this model being “correct” before getting data (assumption)

Brief Overview - Bayesian Machine Learning

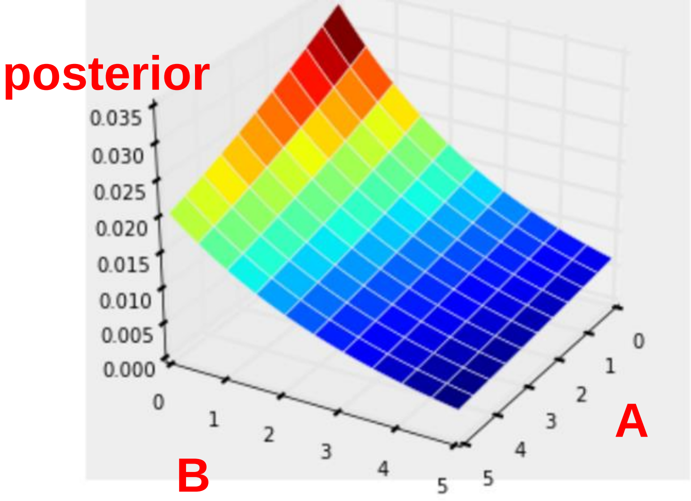
New data alters our prior beliefs → posterior beliefs



Parameter Refinement - (**Bayesian Inference**)

Assume the model has a certain form

Create parametric model with **model parameters …** (ex: model = **A** + **B** * x)



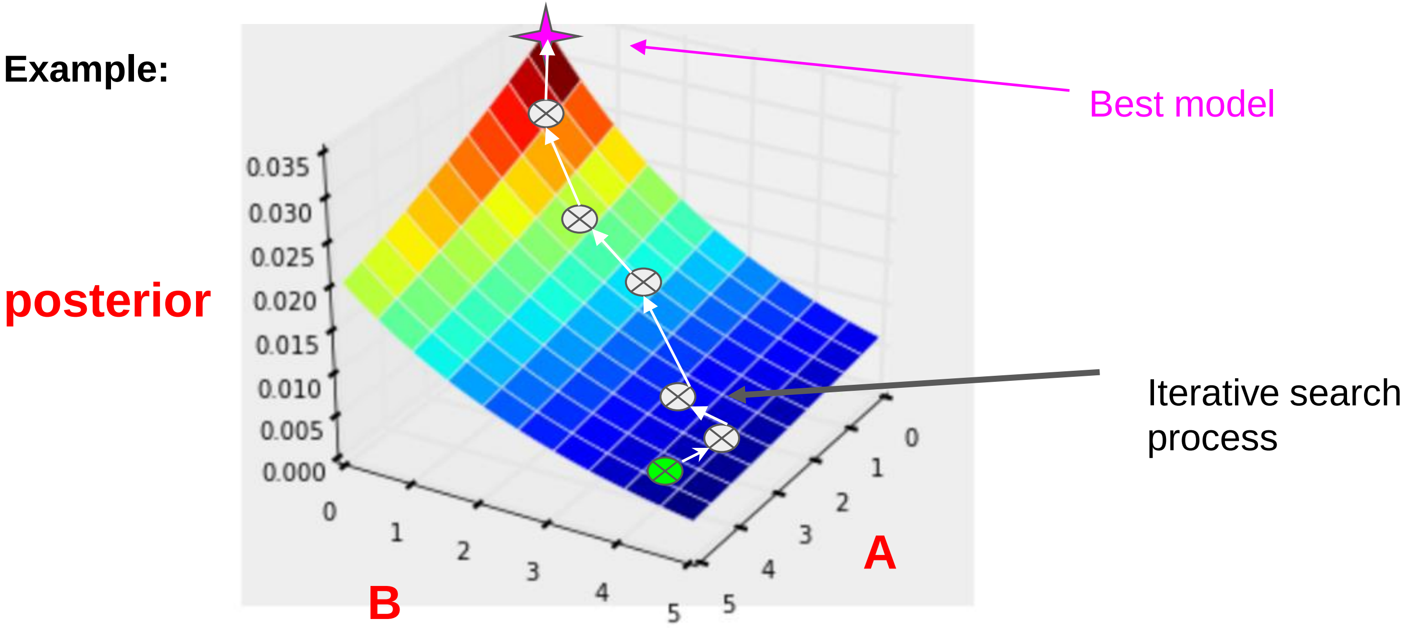
Problem: How to identify combination of parameters where posterior probability for model is greatest (i.e. best model?)

Solution: Sample posterior distribution in parameter space using Markov Chain Monte Carlo (**MCMC**) method

Image Credit: https://nbviewer.jupyter.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter3_MCMC/Ch3_IntroMCMC_PyMC3.ipynb

Parameter Refinement - (**<u>Bayesian Inference</u>**)

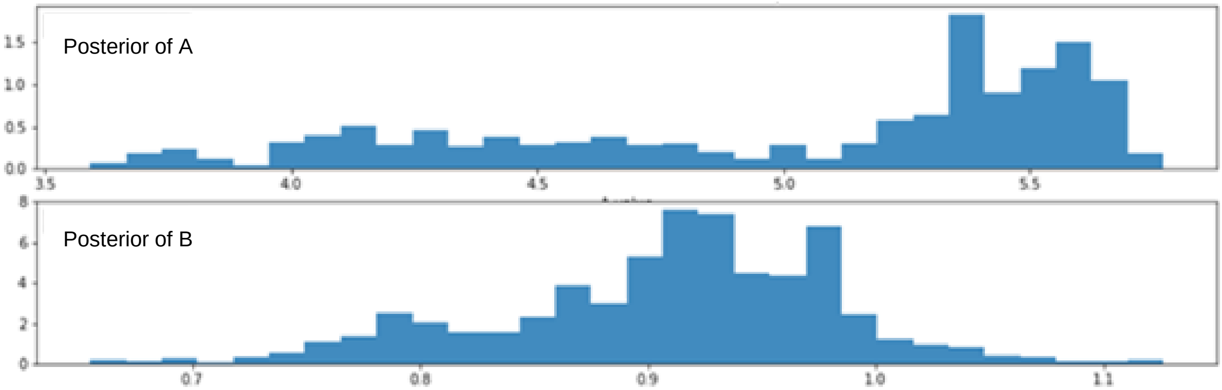
MCMC samples parameter space to find maxima in posterior (best model)



Parameter Refinement - (**Bayesian Inference**)

Produces **posterior distributions** for each **model parameter**

Example:



Represent confidence in parameter values

Active learning: [Parameter Refinement] → argmax (variance in model)

Parameter Refinement - (**Bayesian Inference**)

Prior: Assume model has logarithmic form (pH = A + B*log(C*x))

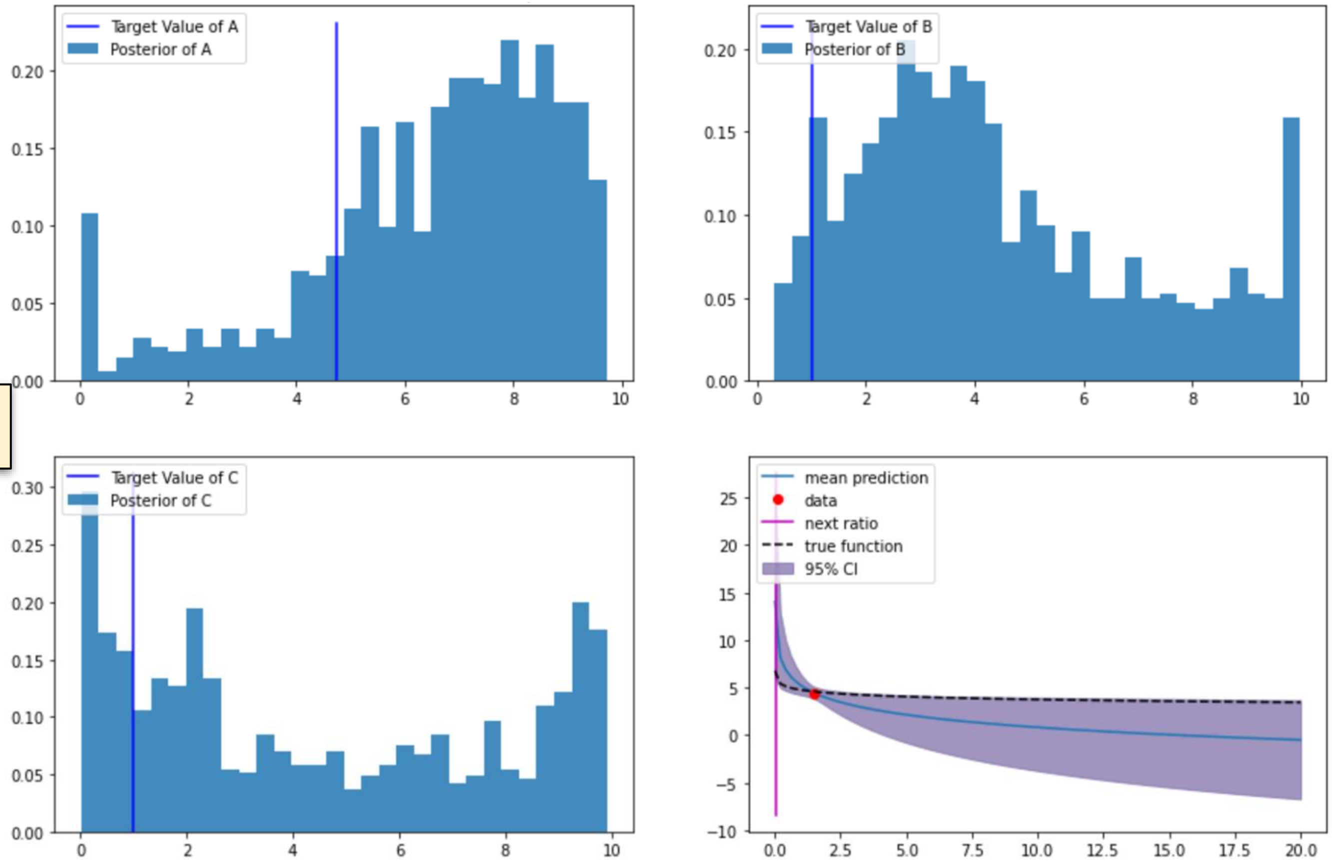
➔ A, B, C are our model parameters

Posterior: Probability of this model and its model parameters given the data

Autonomous Results - (**Bayesian Inference**)

1 data point

pH = A + B*log(C*x)



Autonomous Results - (**Bayesian Inference**)

5 data points

pH = A + B*log(C*x)