UNIVERSITY OF Southampton

Active Learning for Discovery in the Laboratory

Characterising Biomolecular Systems

Chris Lovell

UNIVERSITY OF Southampton

Active Learning and Real World Problems

Image: Mars Rover
Courtesy NASA/JPL-Caltech

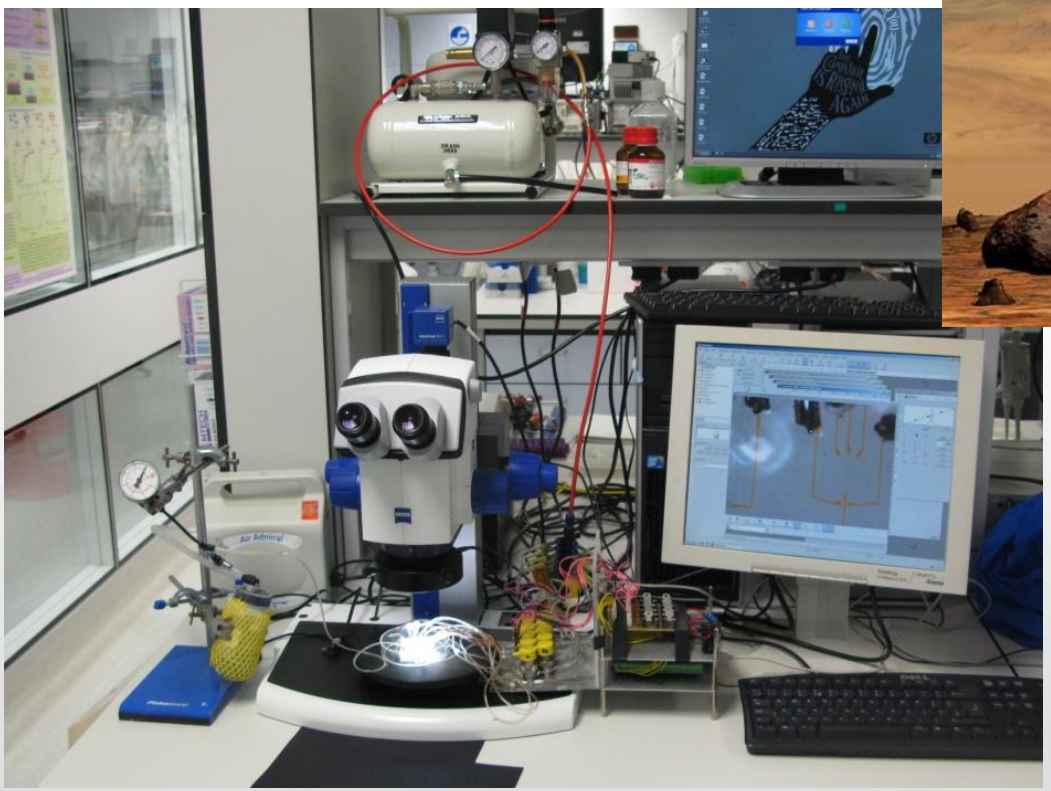


Image: Laboratory workbench by Gareth Jones

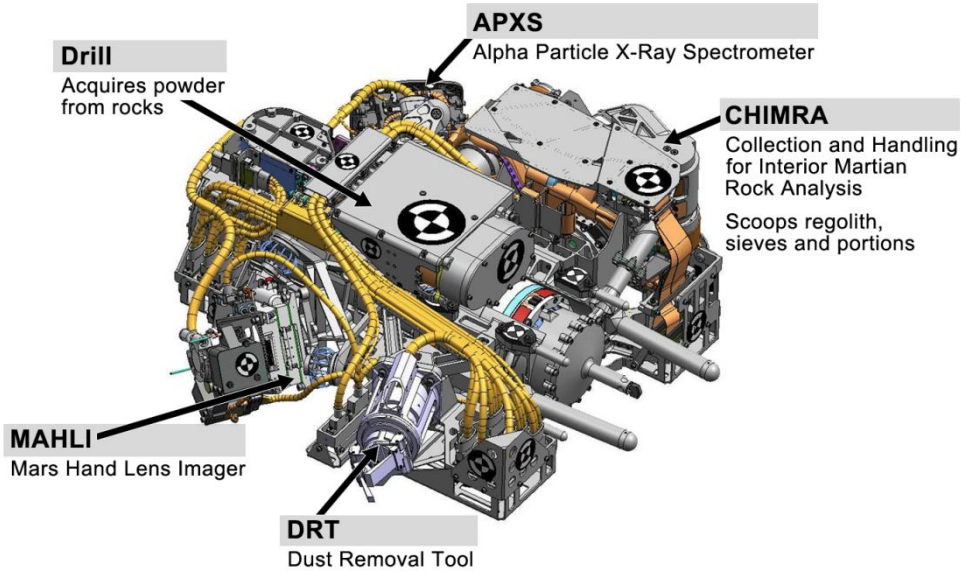
Active Learning

- Allow the learner to select the data to learn from
  - Eg. design the experiment to perform
- Enables the learner to focus information gathering on the areas they are uncertain about
- But once we start attempting real problems, there are some considerations that have to be addressed…

Real World Problems – Parameter spaces

Where to start collecting data from?
- large search spaces

What sort of data should I collect?
- high parameter dimensionality



Images courtesy NASA/JPL-Caltech

Real World Problems - Resources

"There's so much here for me to tell you about."

Resources limit learning. Possible resources include: money, time, communications bandwidth (robotic exploration), battery life, chemical.

"That's nice, but could you fit everything you want to tell me onto a postcard?"

Images courtesy NASA/JPL-Caltech

Real World Problems - Noise

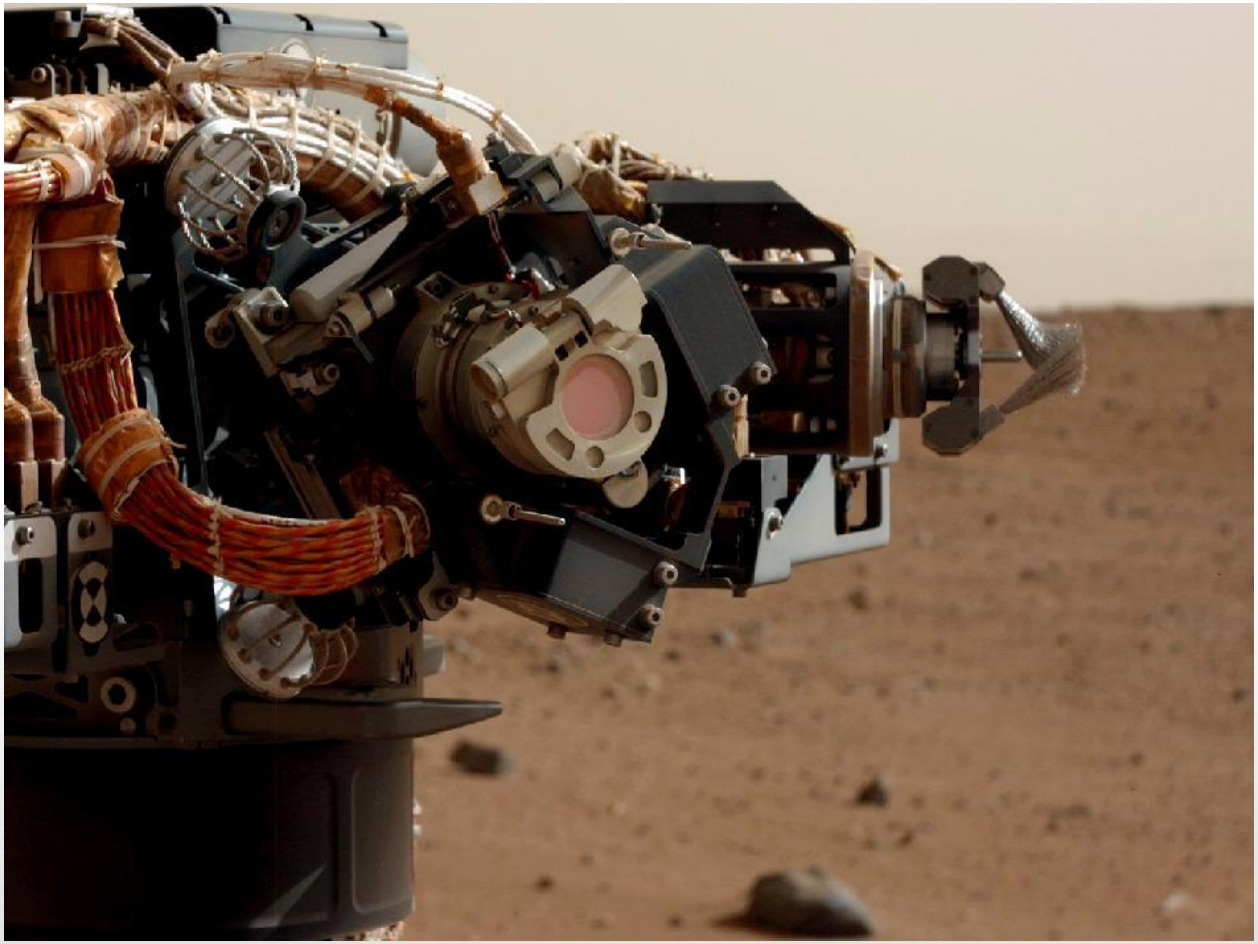
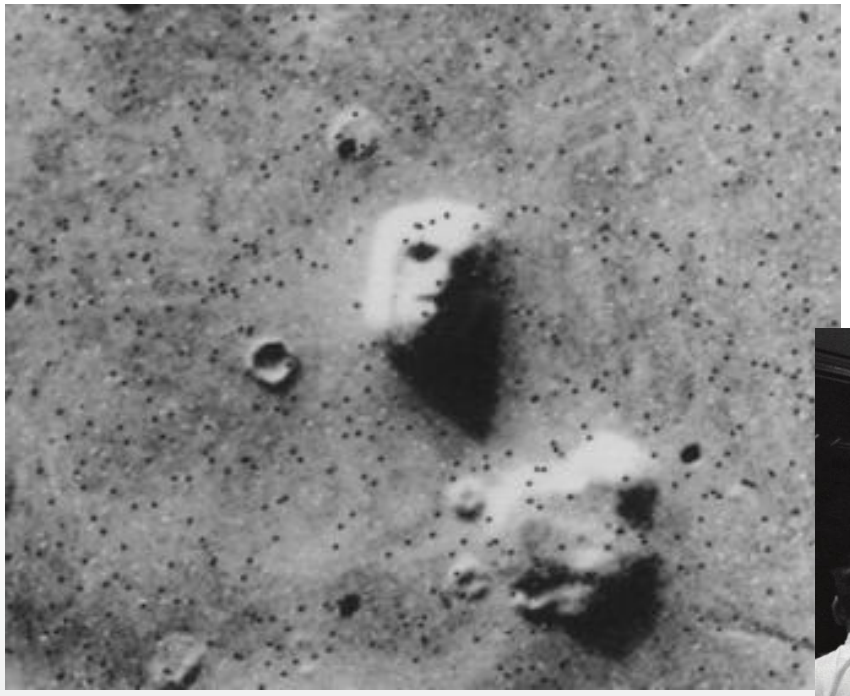


Image courtesy NASA/JPL-Caltech

I measured this rock 10 times and now have 10 different sizes for it.

As soon as you have real data, you have noise. Ordinarily would want to take multiple samples, however this will reduce the resources available.

Real World Problems – Errors



"I found life on Mars"

Mission Complete!

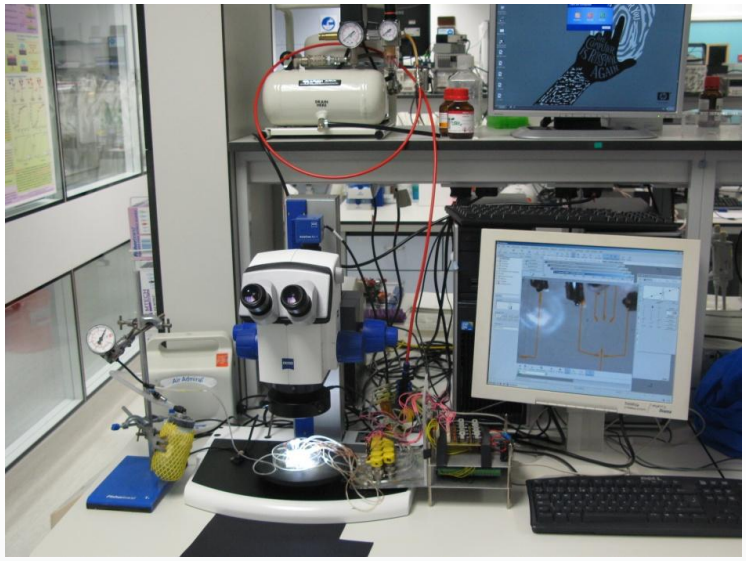
Errors can happen when learner does not understand much about what it is studying (eg. limited prior information) and may make incorrect assumptions (incorrect prior information).

Images courtesy NASA

Active Learning and Real World Problems

- Learning by selecting the training data
- Often limited by cost
  - Monetary, time, available resources, bandwidth
  - Very large or multi-dimensional parameter spaces
  - May mean very limited amounts of data available
- Sometimes things go wrong

Real World Problems



- But people deal with these problems daily and routinely
  - We've been discovering things for centuries
  - Successful scientists working in a lab
  - What can we learn by studying how they go about doing things?

Image credits: left Gareth Jones, right courtesy NASA/JPL-Caltech

Case Study: Biological Response Characterisation

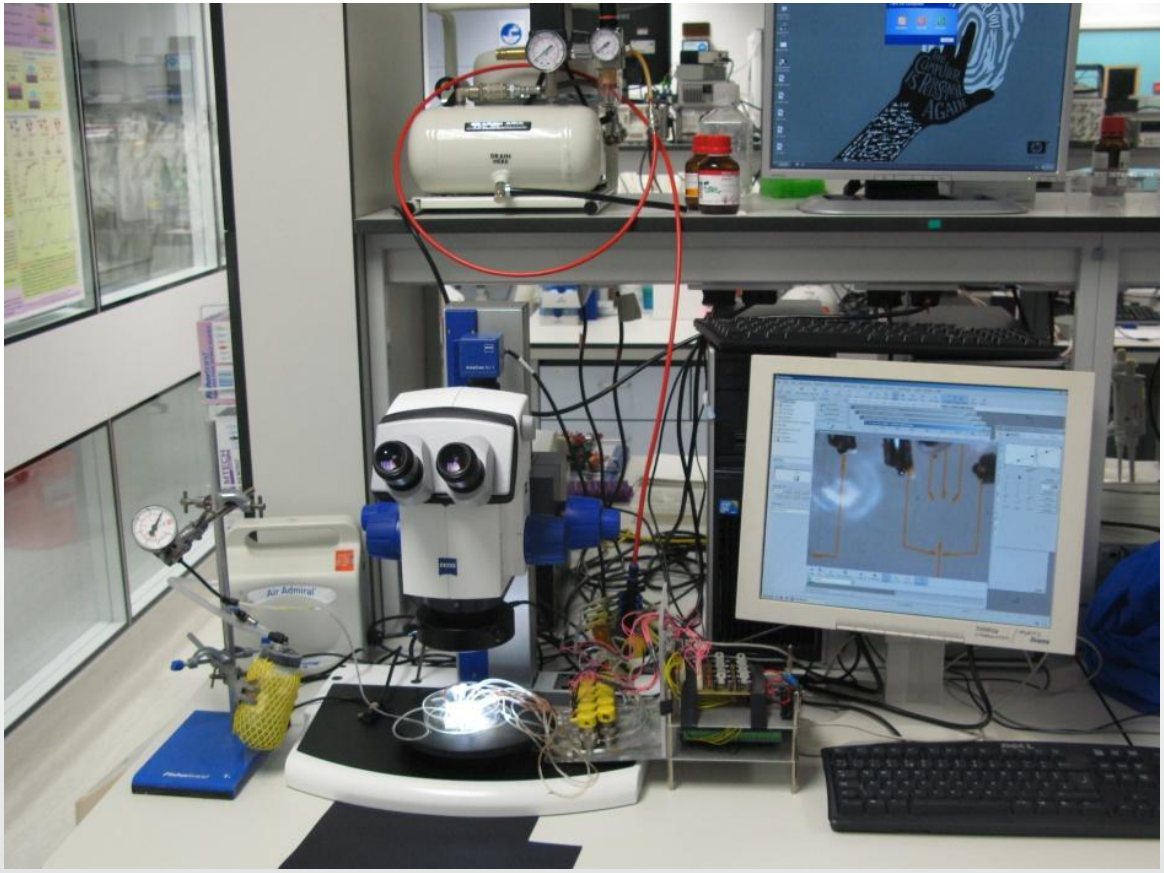


Image: Laboratory workbench by Gareth Jones

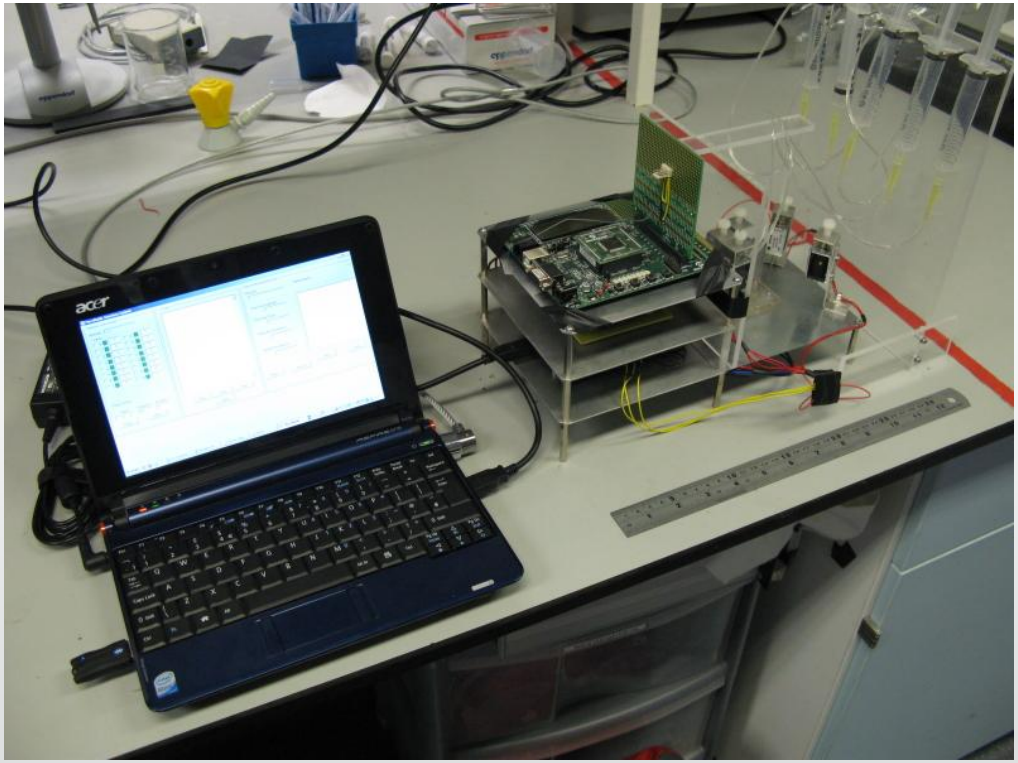
UNIVERSITY OF Southampton

Case Study: Biological Response Characterisation

- Goal:

  A machine capable of closed-loop autonomous discovery

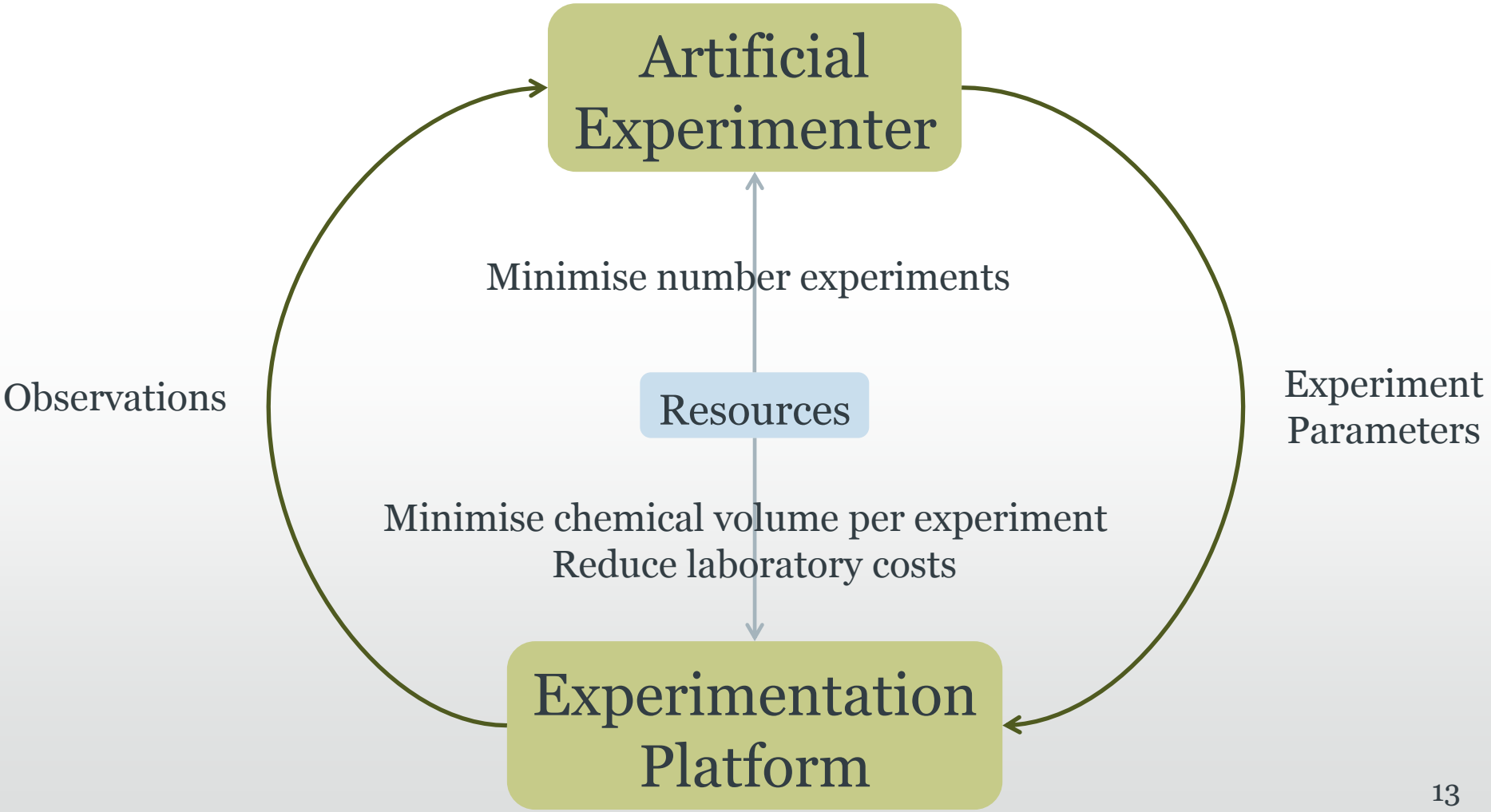
Information



Chemicals

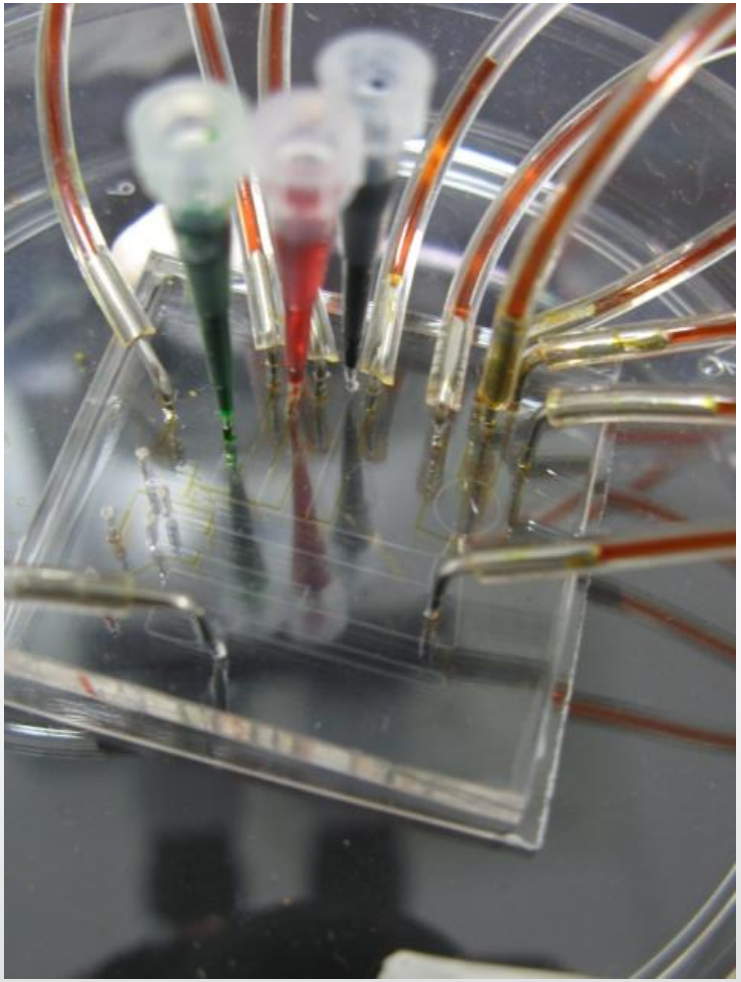
Image: closed-loop experimentation platform by Gareth Jones

Autonomous Experimentation



Lab-on-chip Platform

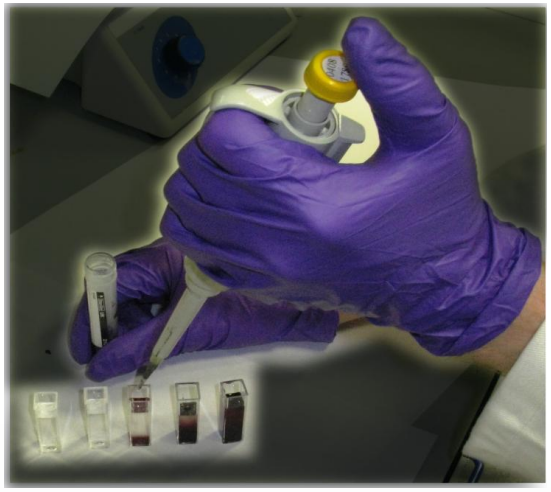
- Uses microfluidic technology
  - Microscale chemicals
  - Less equipment
  - Low initial cost
  - Then on, devices from roughly £2
  - Minimising per experiment costs

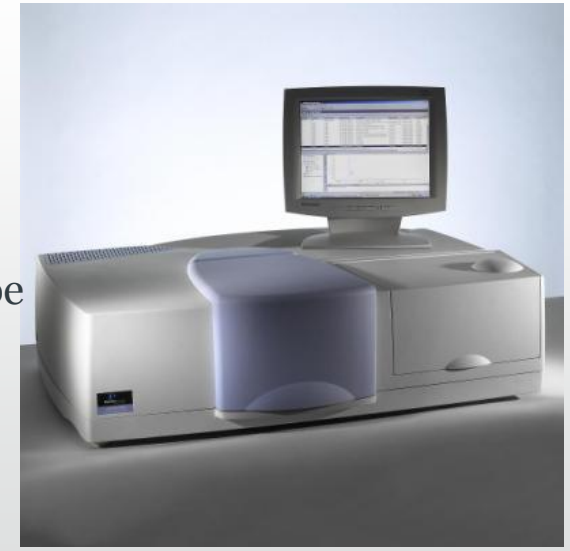


Microfluidic experimentation platform by Gareth Jones.

Discovery Issues

- Experiment parameter spaces very large
  - High dimensionality from many different possible
  - reactants to trial
- Problem of cost
  - Resources restrict number of experiments available
- Variability of the reactants
  - Eg. Chemical contamination – unexpected reactants
  - Eg. Unstable compounds – not doing what they should be
  - Observations unrepresentative of behaviours present
- Very little information currently known
  - Most information concerned with physiological conditions





Automated Discovery

- More than just selecting the x to perform in some $f(x)$

- Need to:

  Interpret the data obtained (not a lot per dimension)

  Ensure the data obtained is representative (good)

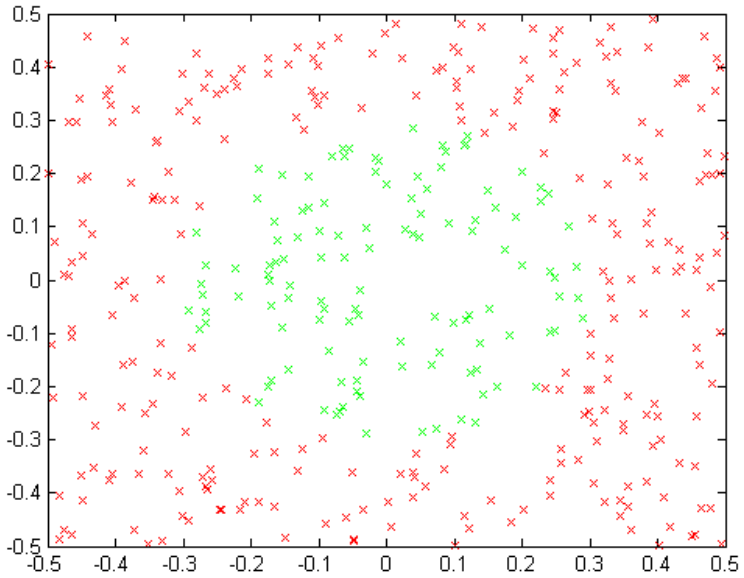
  Maximise the accuracy of our predicting capability

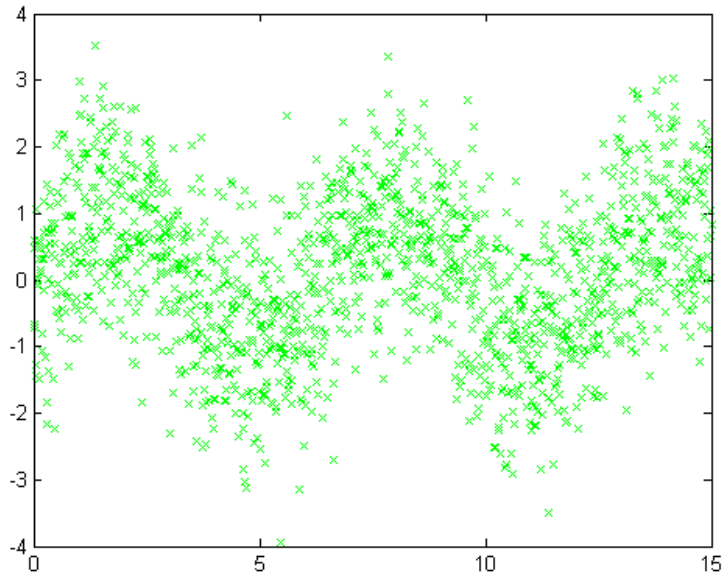
UNIVERSITY OF
Southampton
School of Electronics
and Computer Science

Managing the Data

Machine Learning

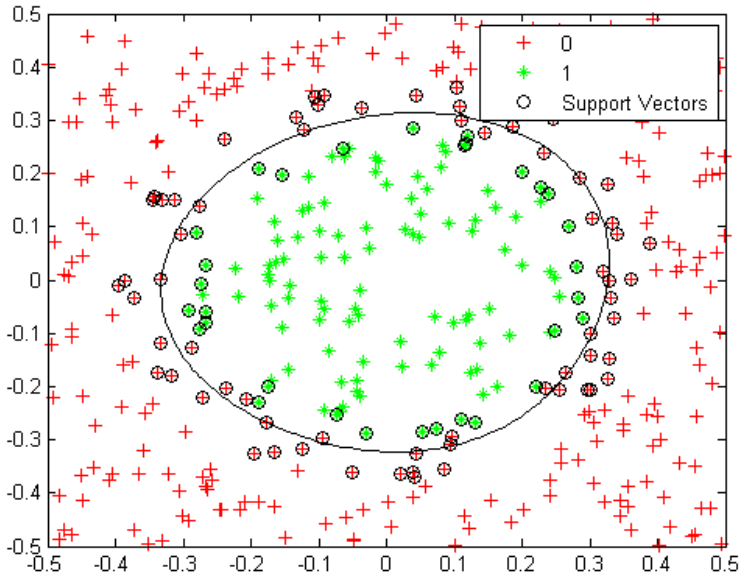
- Find patterns within a set of given data

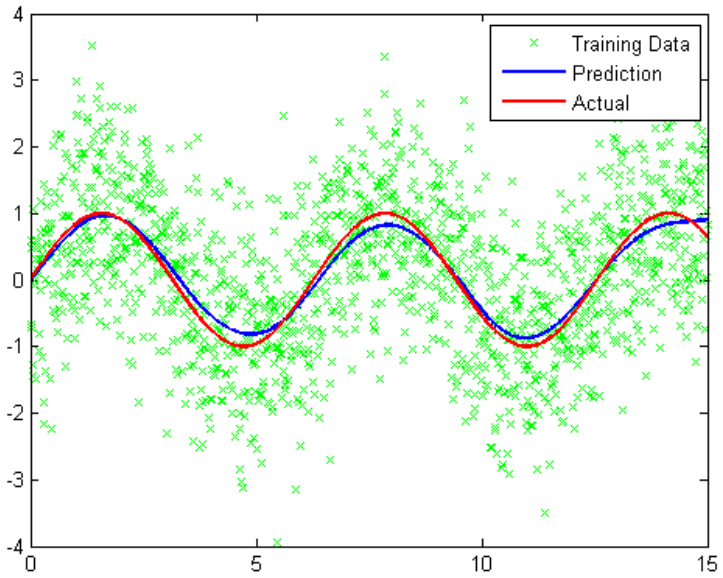




Machine Learning

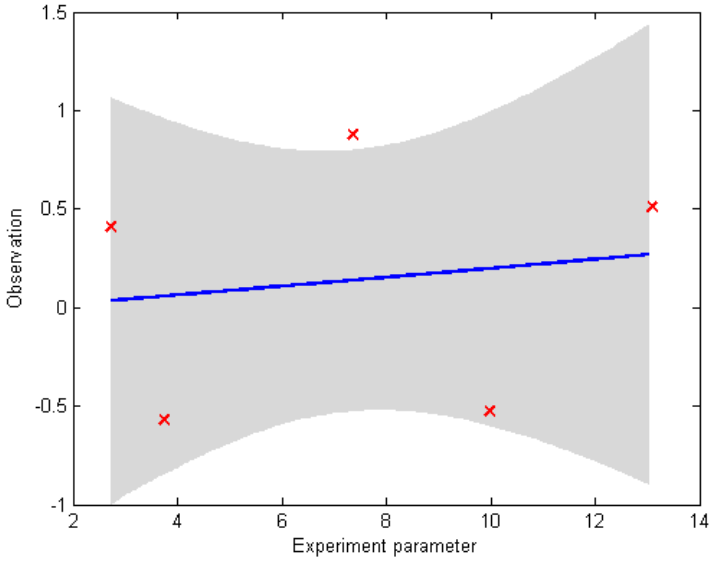
- Find patterns within a set of given data





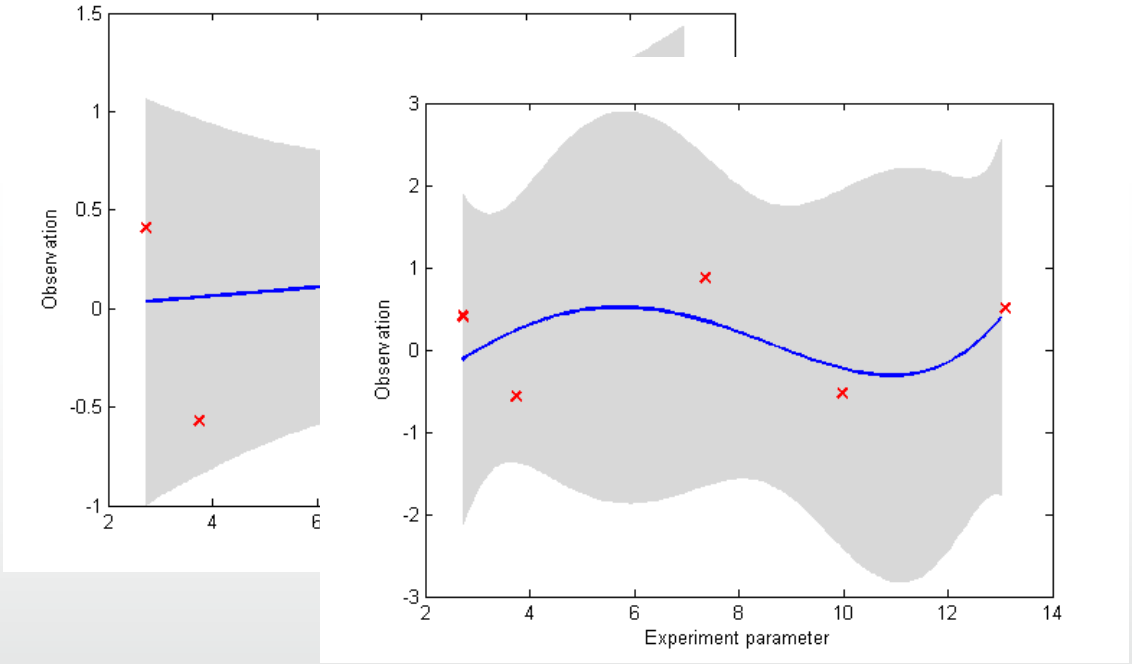
Hypothesis Management

- With less data – which one is correct?



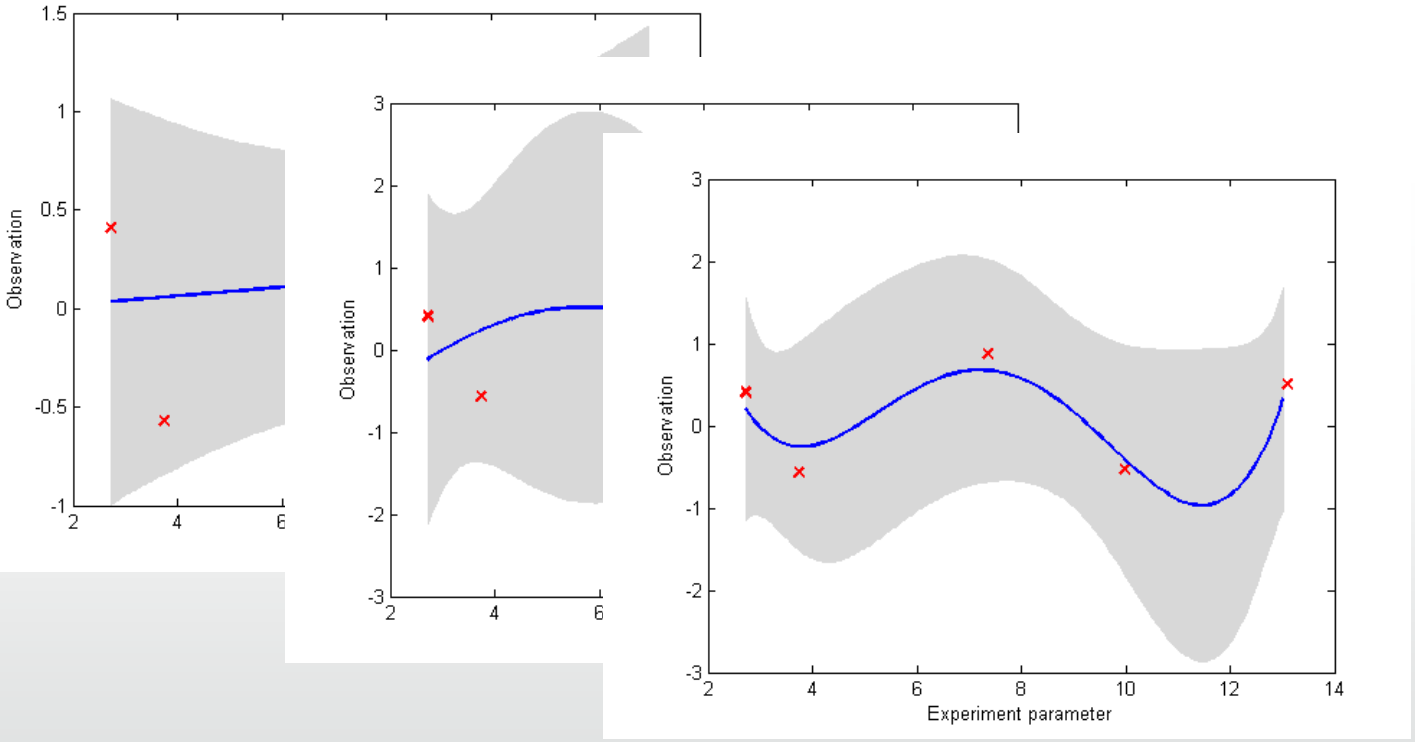
Hypothesis Management

- With less data – which one is correct?



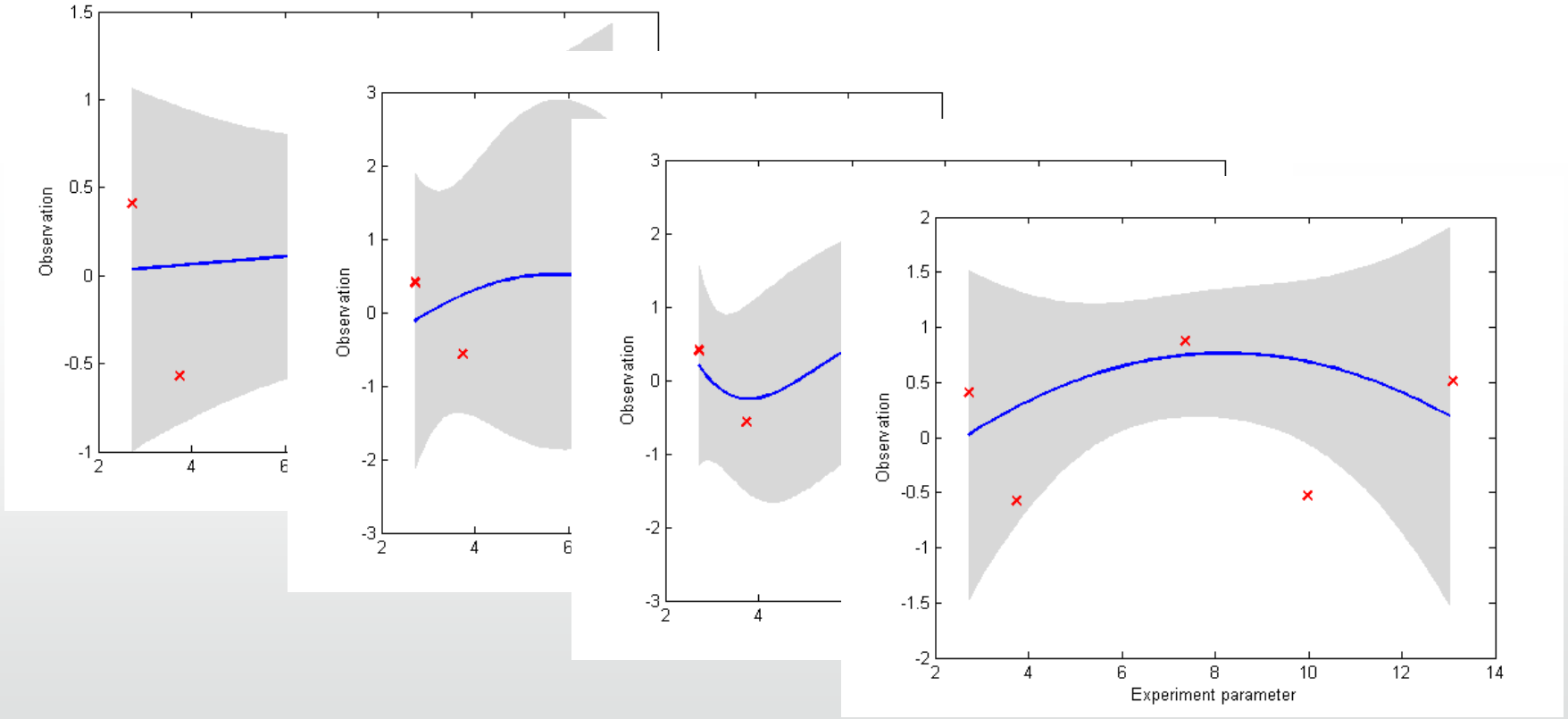
Hypothesis Management

- With less data – which one is correct?



Hypothesis Management

- With less data – which one is correct?



Making Sense of the Data

- Have:
  - Limited data
  - Erroneous data (questionable reliability)
  - Noisy data
- Want:
  - Some mechanism for prediction of response (hypothesis)
  - A single hypothesis does not seem sensible

Making Sense of the Data

- How a scientist does it:
  - Multiple competing hypotheses
  - Philosophy of Science: Popper
  - New hypotheses should explain deficiencies in existing

Making Sense of the Data

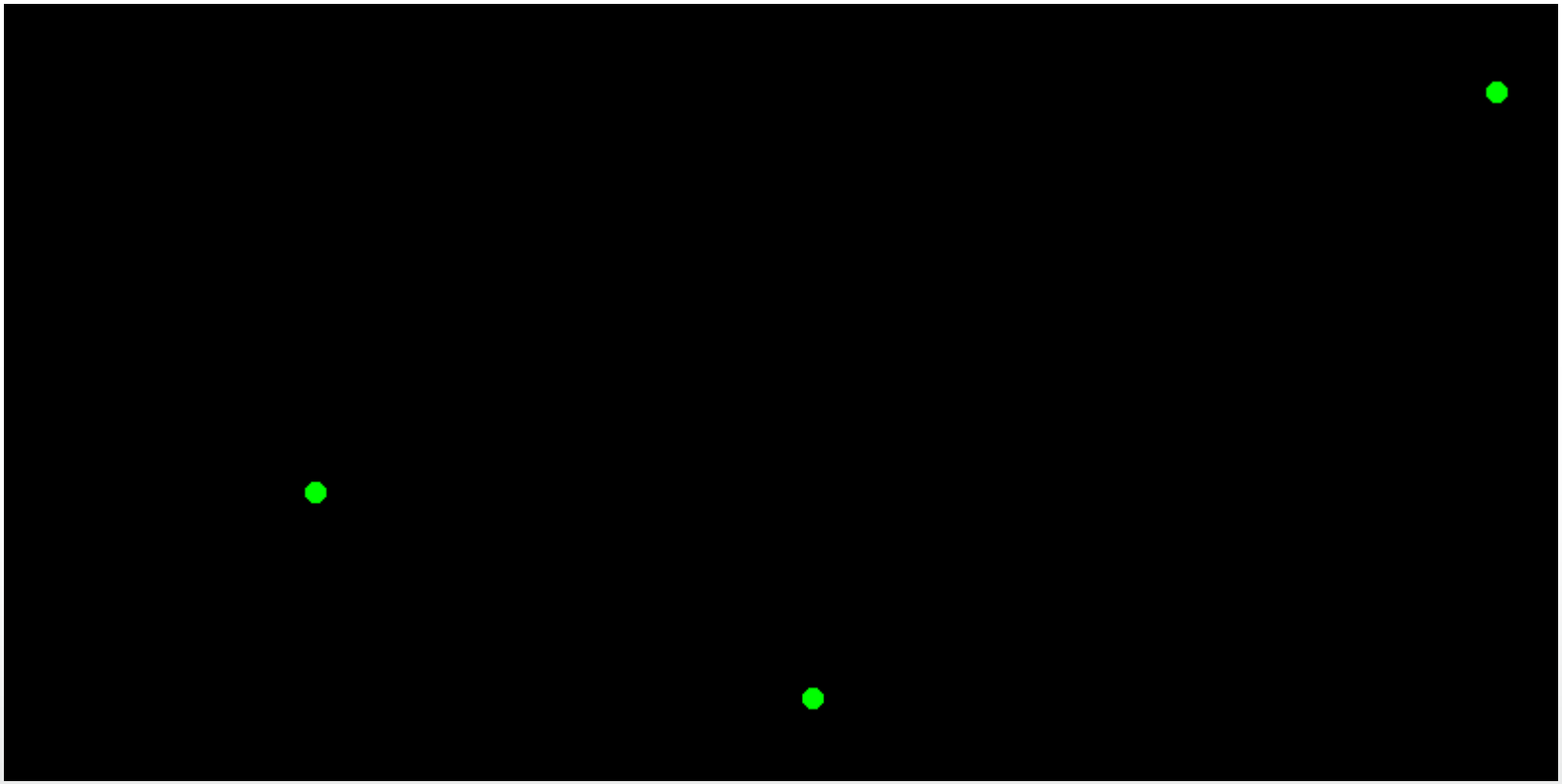
- How a scientist does it:
  - Multiple competing hypotheses
  - Philosophy of Science: Popper
  - New hypotheses should explain deficiencies in existing

- Machine learning:
  - Ensemble based methods, query by committee
  - But, often data subsets chosen at random

Multiple Hypotheses

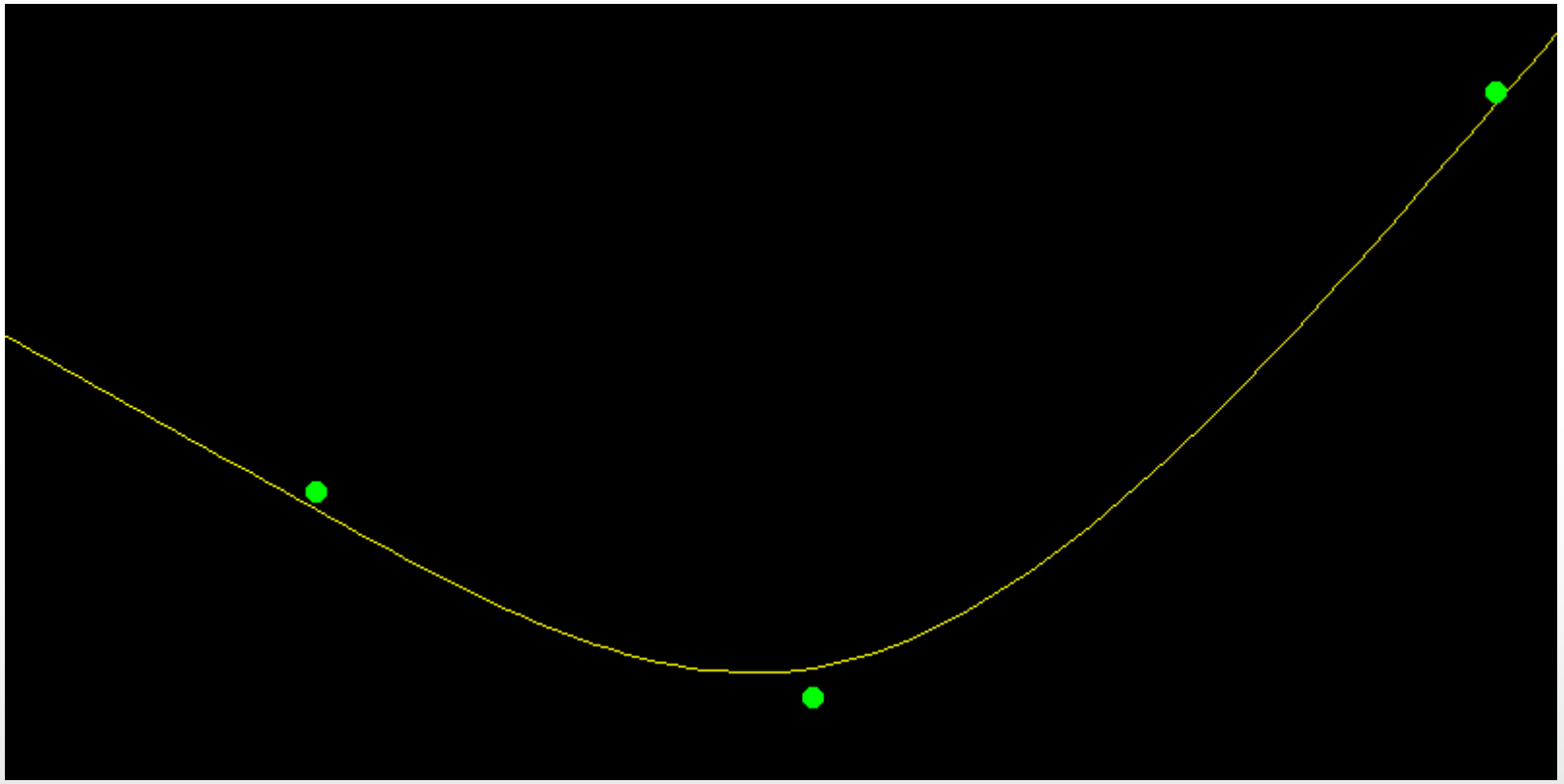
- Many thousands of hypotheses can be considered

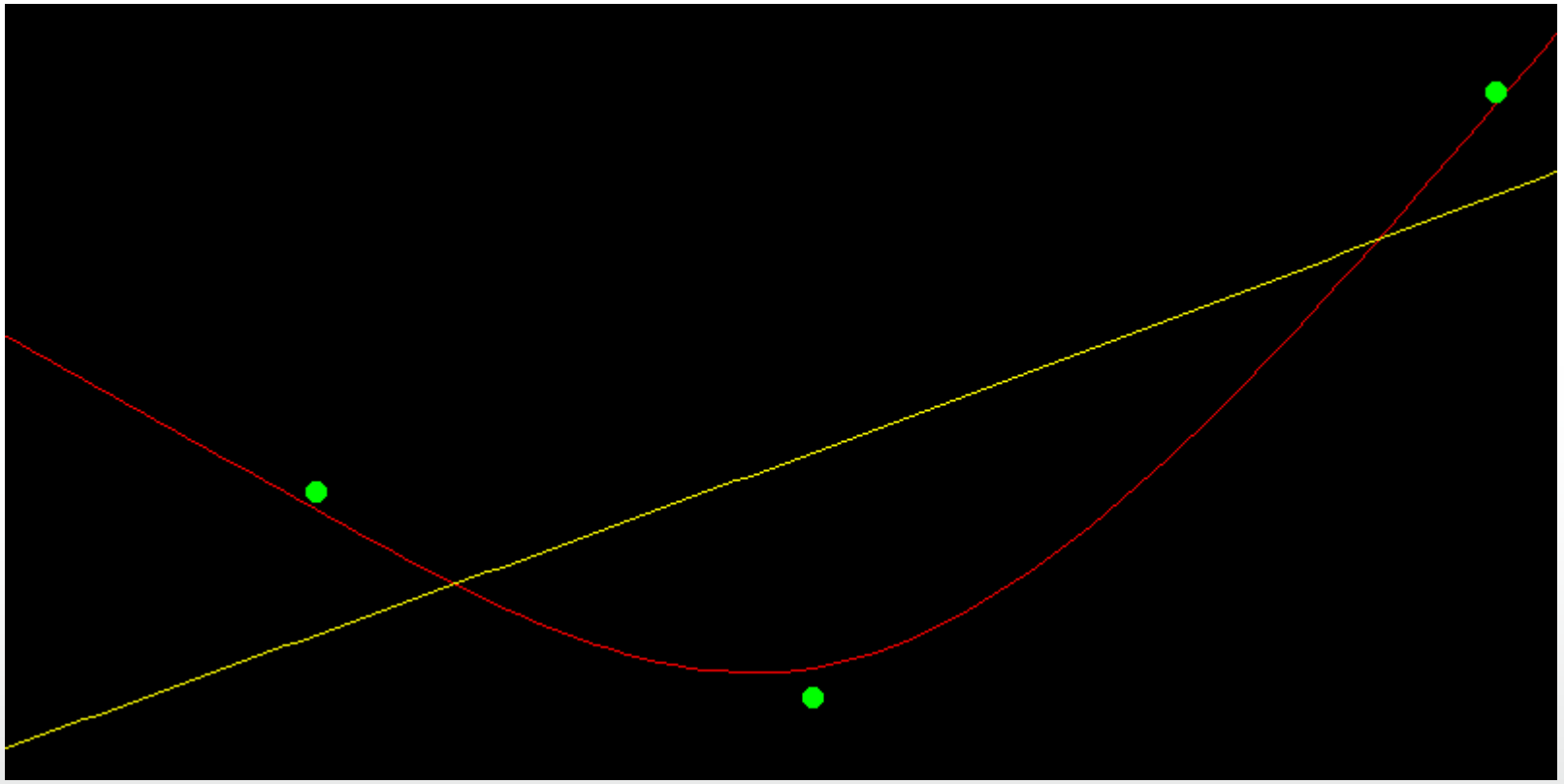
- Allows decisions about uncertainty to be made later
  - Is the observation erroneous or not?
  - Different hypotheses with competing views

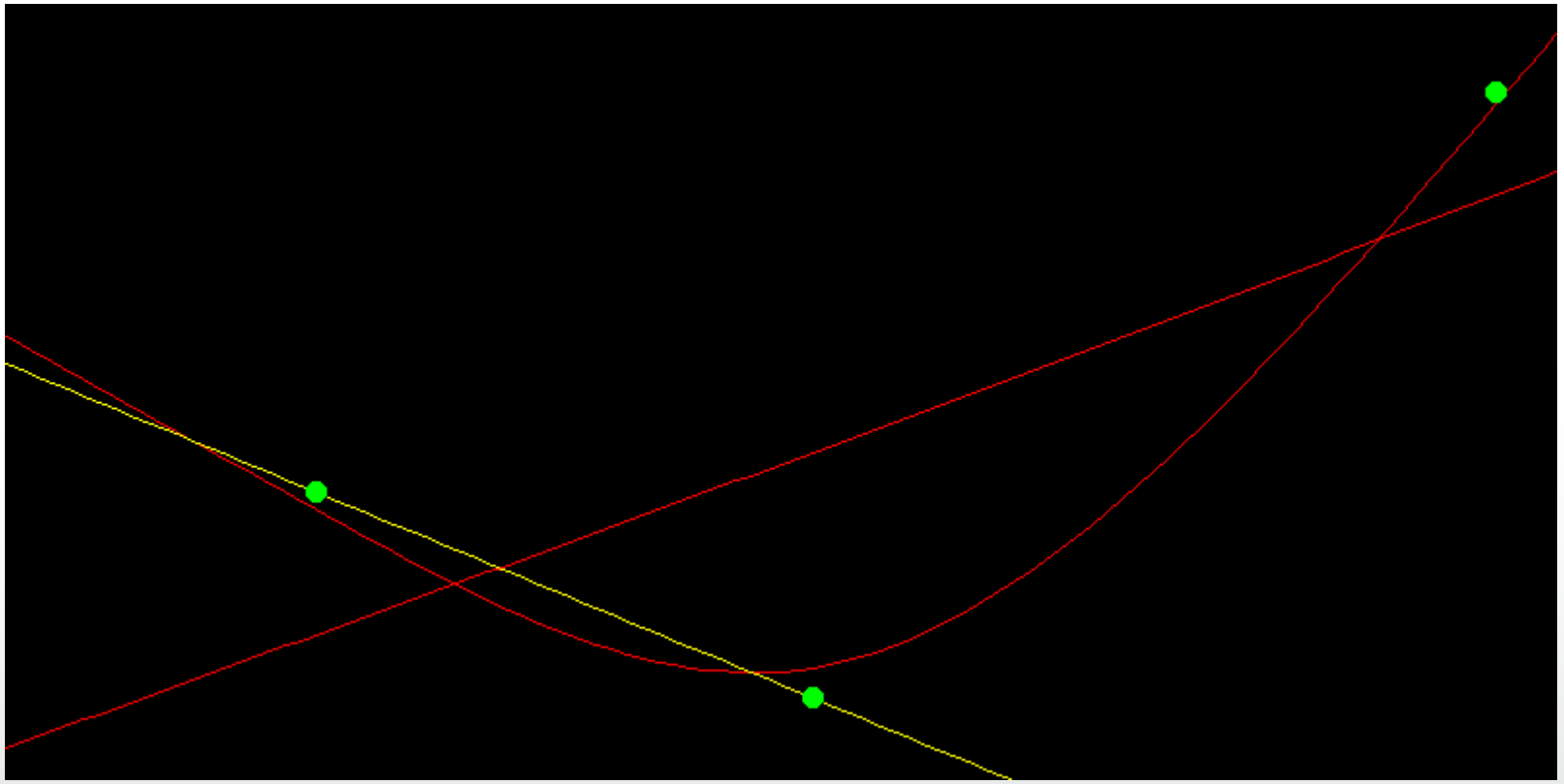


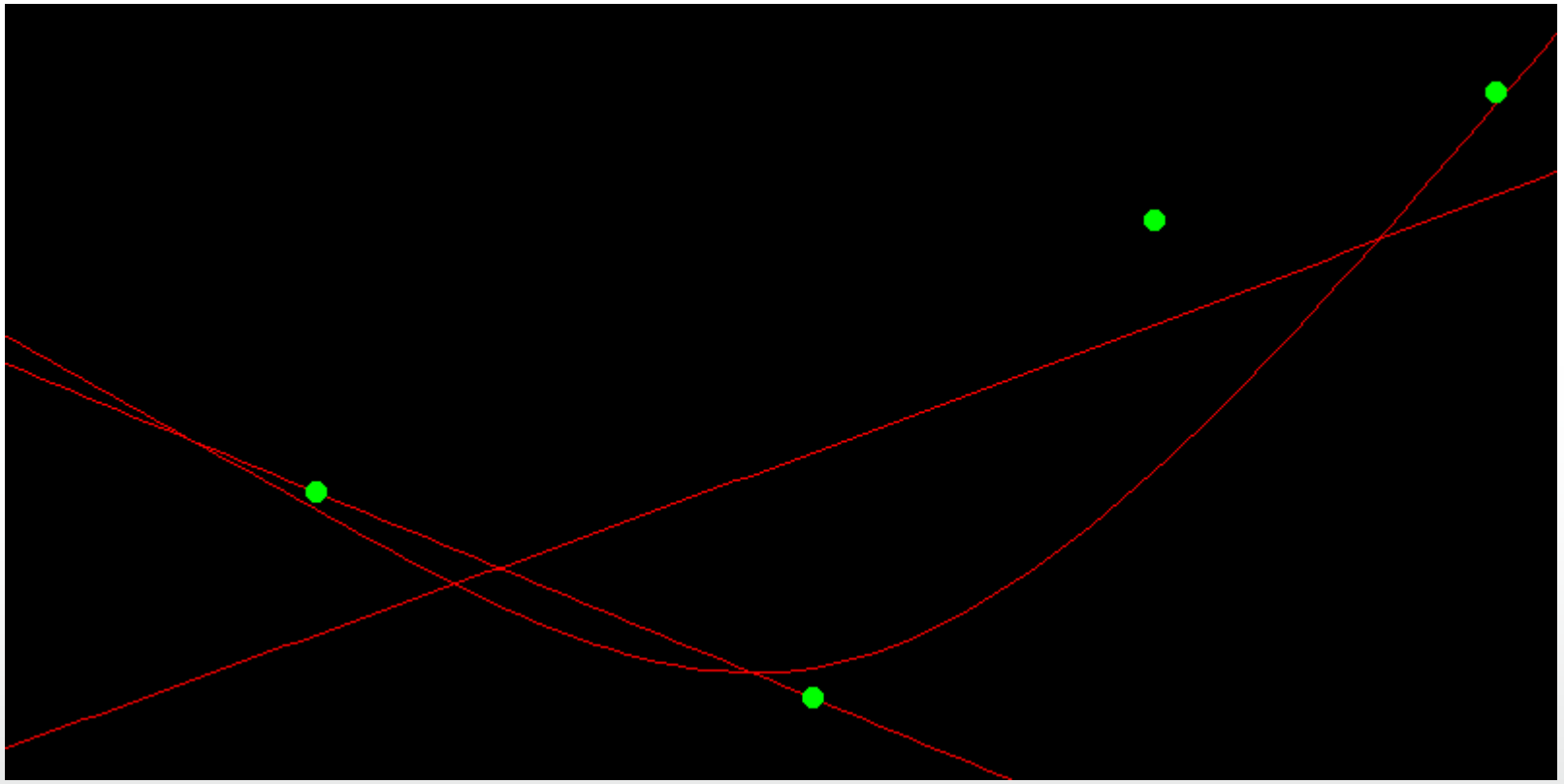
Observation

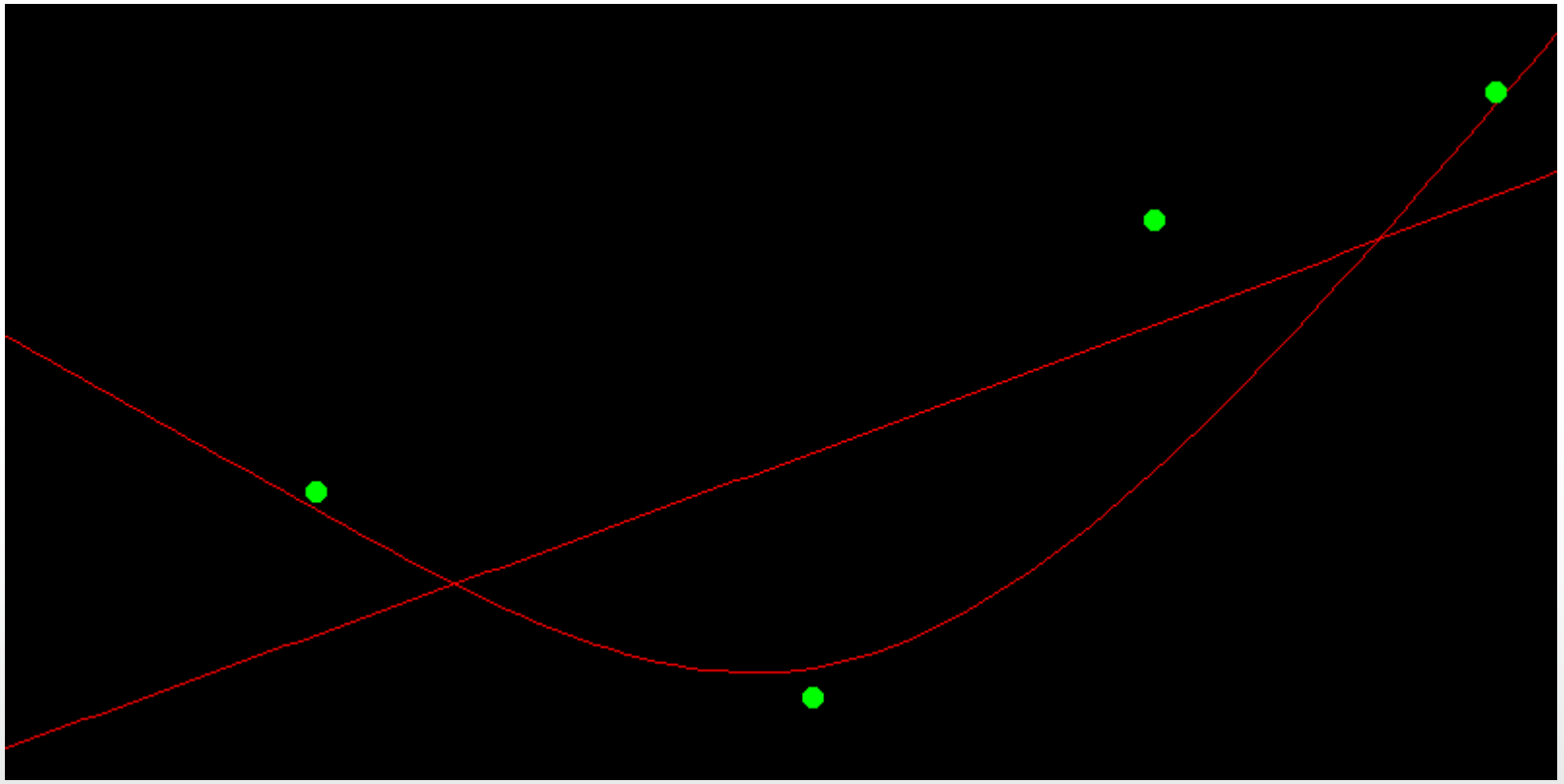
Experiment parameter

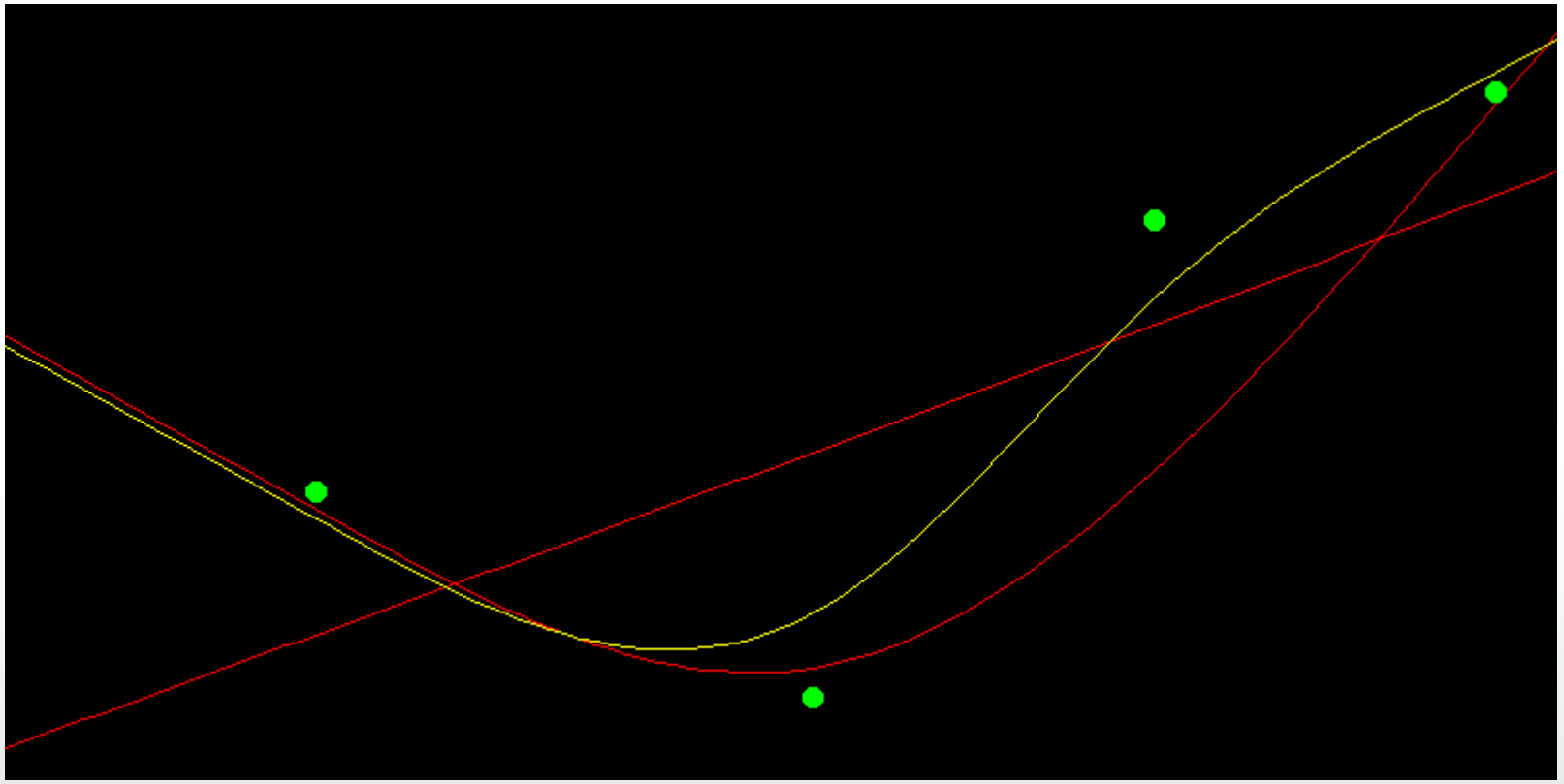

Observation
Experiment parameter

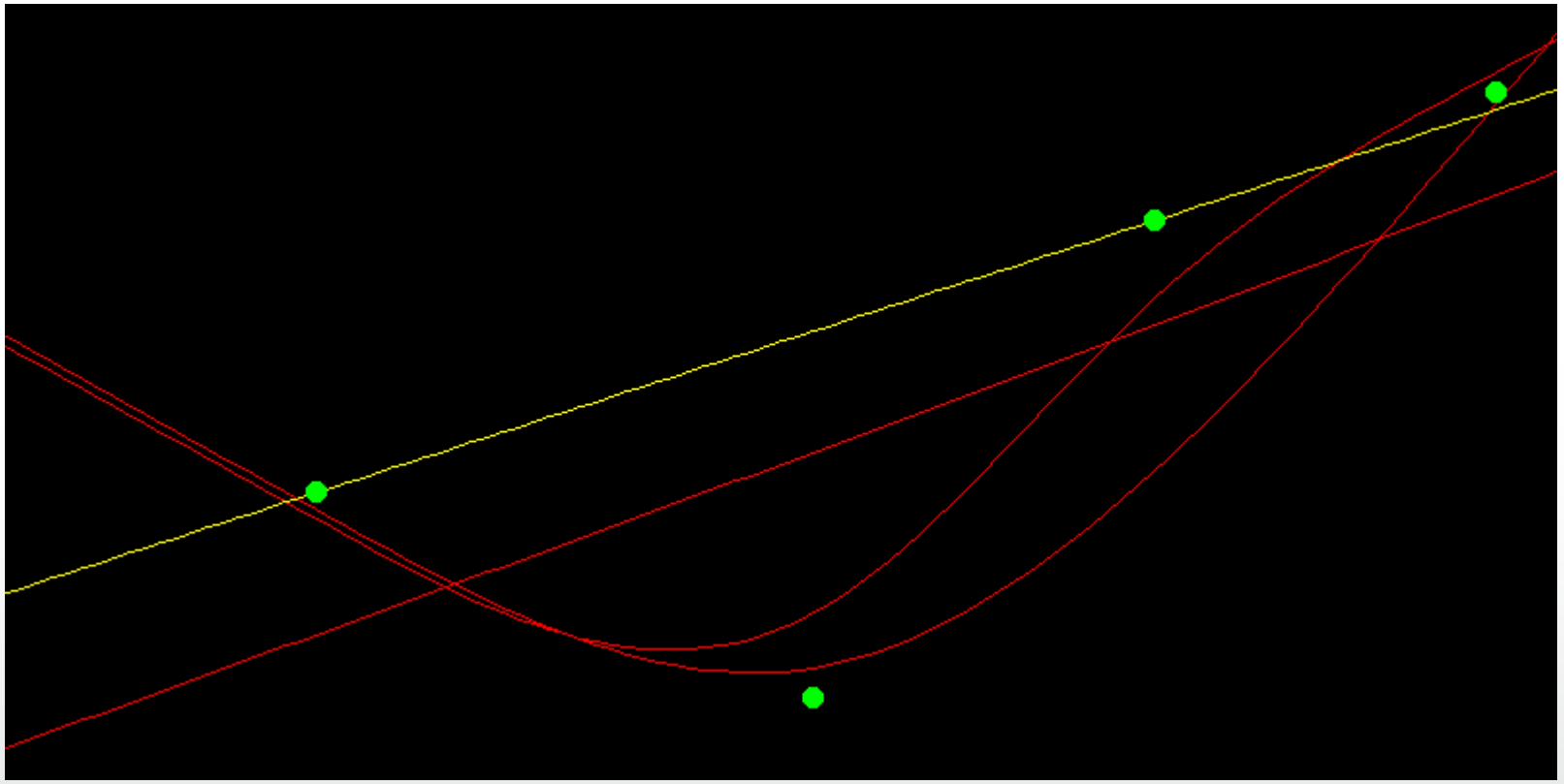

Observation
Experiment parameter


Observation
Experiment parameter


Observation
Experiment parameter

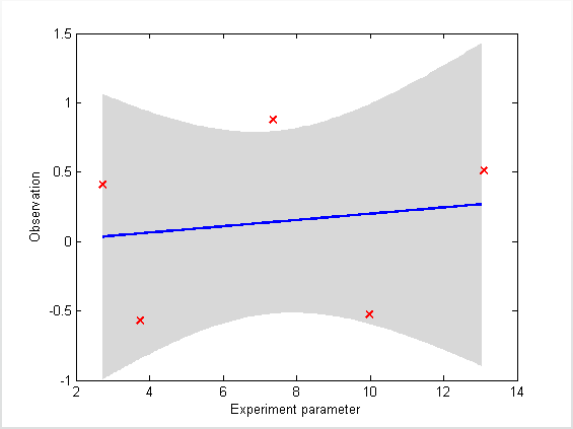

Observation
Experiment parameter

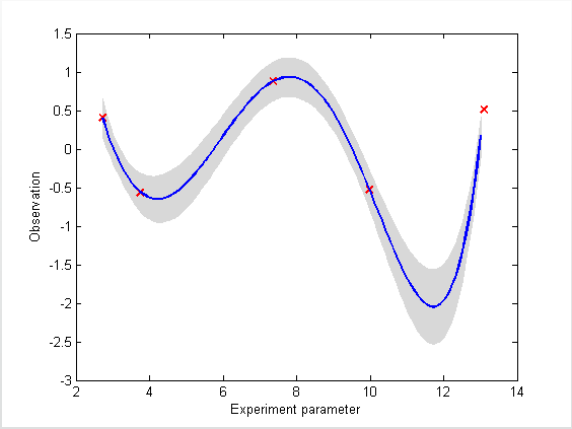

Observation
Experiment parameter

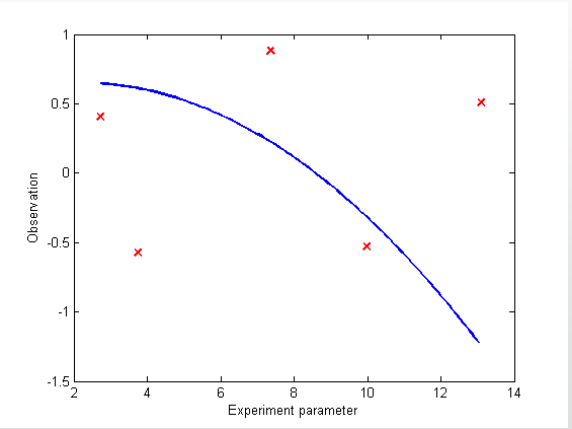

Observation
Experiment parameter

Building Multiple Hypotheses: *Diversity*

- No good if all your hypotheses are essentially the same
- Initially create random hypotheses
  - Random training sets
  - Random regularisation parameters (how smooth the line is)
  - This gives different initial views of the data

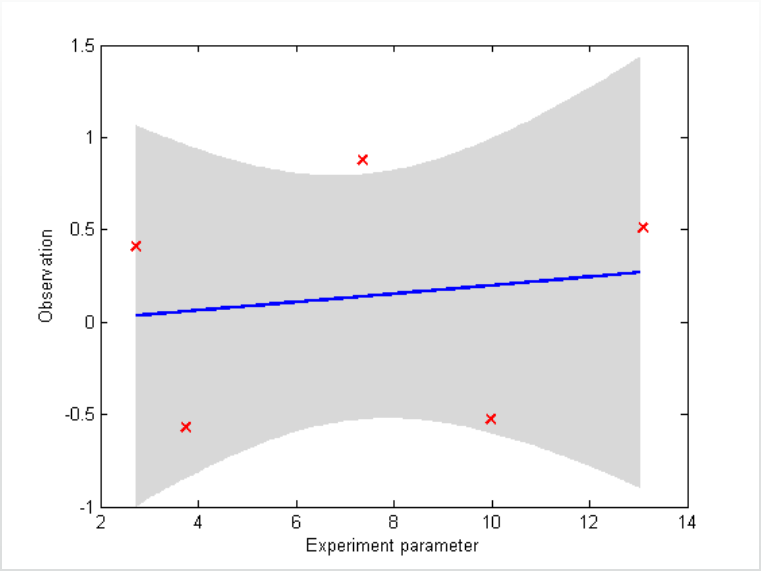






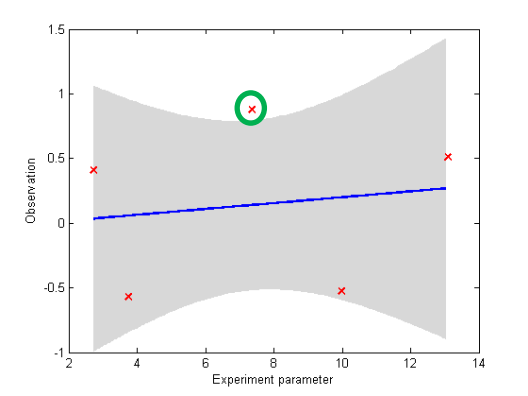
Building the Hypotheses: *Learning*

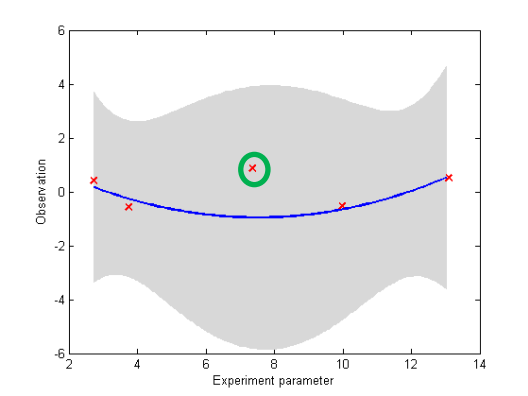
- Don't want to rely on just luck
- Is it the hypothesis or the observation that is erroneous?
  - Repeat experiments, but then resources being taken away from discovery
  - Real exploration-exploitation trade-off problem

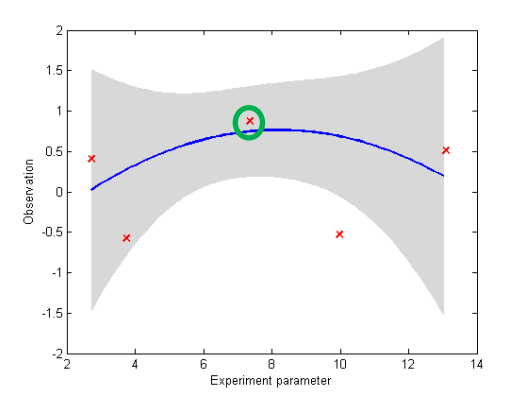


Building Multiple Hypotheses: *Learning*

- If an observation disagrees with a hypothesis
  - Refine the hypothesis with 2 new hypotheses
  - One hypothesis declares the observation valid
  - One hypothesis declares the observation invalid
  - Keep all 3 hypotheses







Procedure

On each data point obtained:

1. Add new random hypotheses to the set of hypotheses in consideration
2. Refine hypotheses based on data available
3. Evaluate and discard worst ones
4. Wait for the next experiment

Hypothesis Management

- How many hypotheses to keep?
  - Ideally all hypotheses – but will become computationally infeasible very quickly
  - Practically some hypotheses have to be removed
- How to evaluate a hypothesis?
  - Some form of mean squared error
    - Difference between seen data and hypothesis prediction
  - Danger of forcing a hypothesis to evaluate itself on erroneous observations

UNIVERSITY OF Southampton

Choosing Experiments

Now we have these hypotheses, what do we do with them?

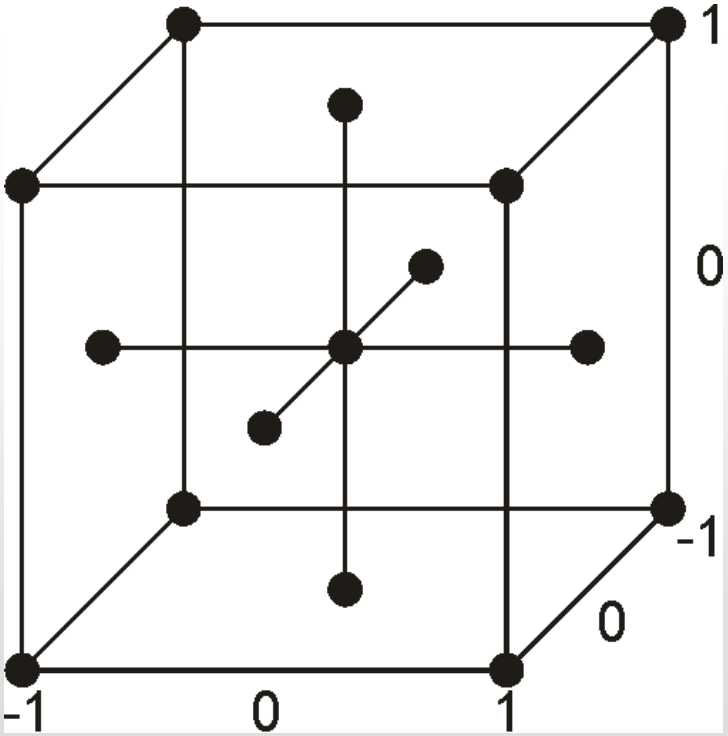
Experiment Manager

- Use all information available to choose next experiment
  - Minimise the number of experiments
  - Maximise the information obtained

- Fields doing this:
  - Design of Experiments
  - Active Learning
  - Computational Scientific Discovery
  - Space or deep sea exploration

Design of Experiments

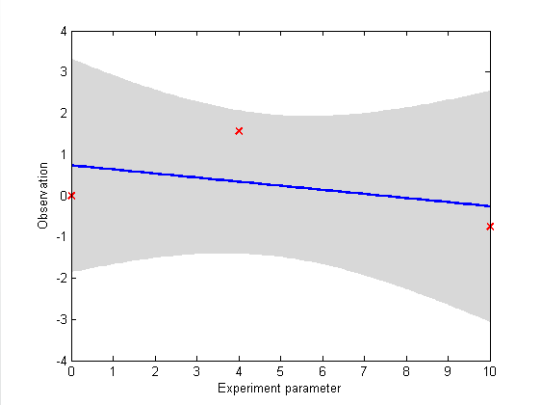
- Statistically tested techniques to choose best parameters
  - But normally do not adapt to observations obtained



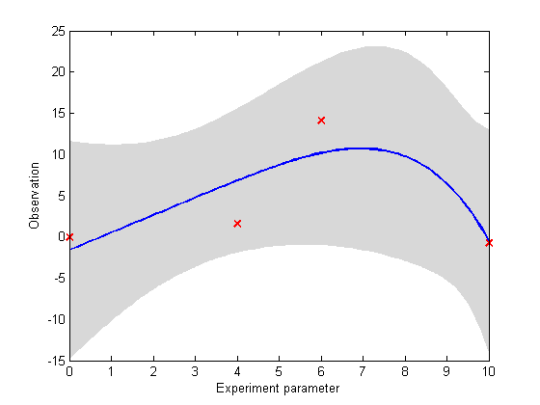
UNIVERSITY OF Southampton

Active Learning

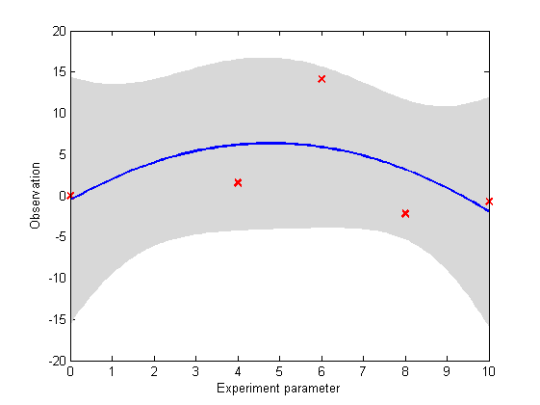
- Data is not given straight away
  - Instead must choose the data to learn from
  - Eg. select the experiments to be performed
- Adaptive based on the data it receives



1



2



3

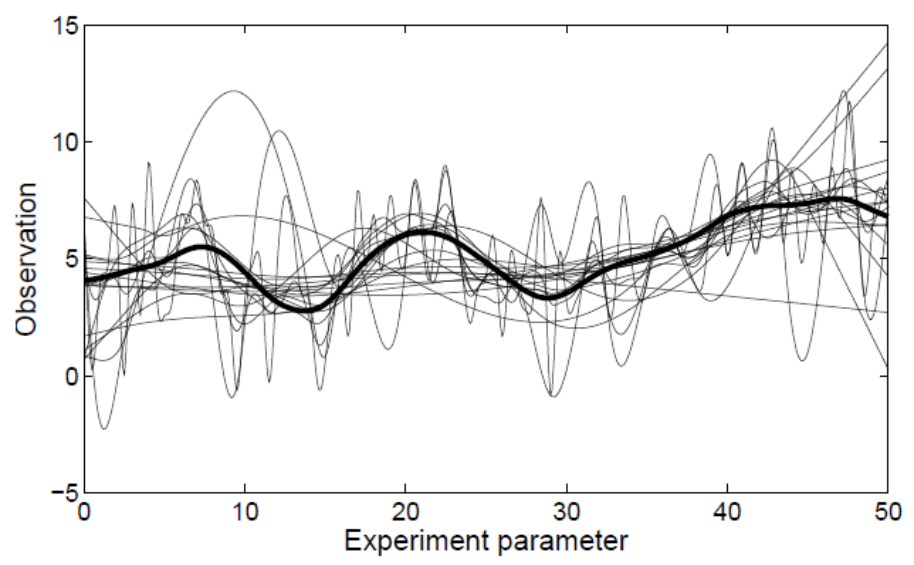
What do Scientists do? (part 1)

- Have lots of hypotheses
  - Which one of them is the most representative one?

- Part of experiment selection will be trying to answer this question

Separating Hypotheses

- Consider a toy problem
  - Set of competing hypotheses
  - One of them is the *'true'* hypothesis
  - Experiments get noise adjusted values from *'true'* hypothesis
  - How long until the most confident hypothesis is the *'true'* hypothesis



Separating Hypotheses

- Consider a toy problem

Hypothesis Similarity (increasing order)	Strategy				
	Random	Variance	Max Discrepancy	Surprise	KL Divergence
$N(0, 4^2)$	3	**2**	**2**	3	**2**
$N(0, 2^2)$	8	4	**3**	7	4
$N(0, 1^2)$	18	**7**	**7**	13	11

Variance

$$x^*_{\text{Var}} = \arg\max_x \; k \sum_{i=1}^{|\mathcal{H}|} C(h_i) \left(\hat{h}_i(x) - \mu^*\right)^2$$

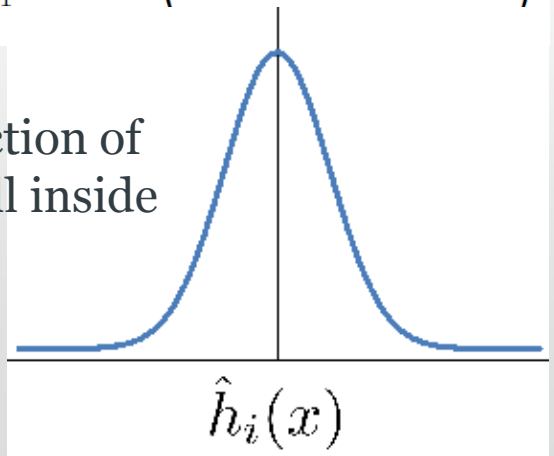
Max Discrepancy

$$D(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} 1 - \exp\left(\frac{-\left(\hat{h}_i(x) - \hat{h}_j(x)\right)^2}{2\sigma_i^2}\right)$$

$\hat{h}_i(x)$ Prediction of hypothesis i for experiment x

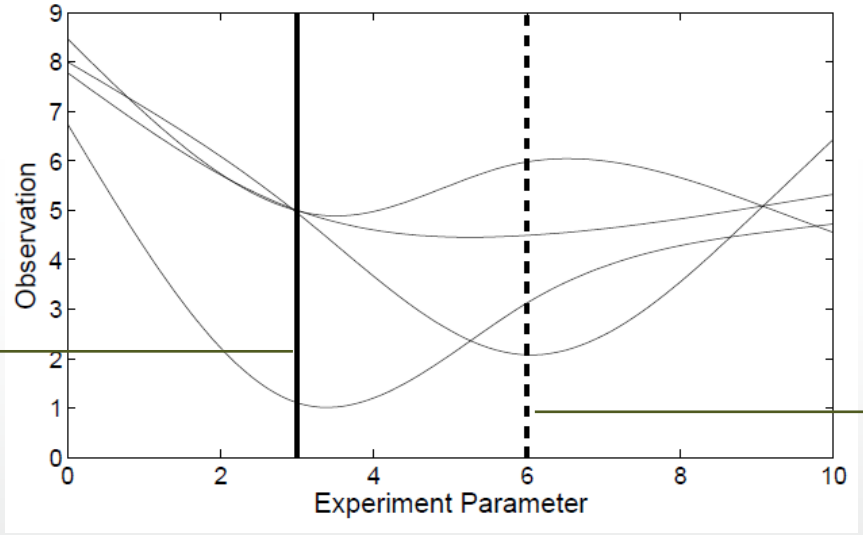
$2\sigma_i^2$ Uncertainty in the prediction

Does the prediction of hypothesis j, fall inside the normal distribution of hypothesis i?



Separating Hypotheses

Variance



Discrepancy

$$x^*_{\text{Var}} = \arg\max_x \; k \sum_{i=1}^{|\mathcal{H}|} C\left(h_i\right) \left(\hat{h}_i(x) - \mu^*\right)^2$$

$$D(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} 1 - \exp\left(\frac{-\left(\hat{h}_i(x) - \hat{h}_j(x)\right)^2}{2\sigma_i^2} \right)$$

What do Scientists do? (part 2)

- Separating hypotheses is only part of the solution

- How do we know any of the hypotheses are actually any good?
  - Could be finding best of a bad bunch
  - May have missed features of the behaviour investigating

- Part of experiment selection will be trying to answer this question

Exploration or Exploitation

- Experiments must explore
  - Find new behaviours
  - Provide data to allow different hypotheses
  - Maximally distant to each other

- Experiments must exploit (the hypotheses)
  - Test the hypotheses
  - Test the observations
  - Disprove the hypotheses (good scientific method!)

Exploration or Exploitation?

- How to decide whether to:
  - Search for something new (exploration)
  - Test our views of what is going on (exploitation)

- Various methods exist
  - Simple ones like: 90% of the time exploit
  - Others make assumptions about the problem which are not valid here

How do Scientists manage the trade-off?

- Scientists often talking about surprising results
  - Things they didn't expect
  - A lot of things found by accident…

- If you find something surprising
  - You would want to find out more about it (exploitation)
- If nothing currently surprising
  - Look for something surprising

Bayesian Surprise (Itti and Baldi, NIPS 2006)

- First used to find surprising occurrences in videos
- Follows a KL-divergence
- C(h) – the prior confidence of hypothesis h
  - How much we believed h before the experiment
- C'(h) – the posterior confidence of h
  - How much we believed it after the experiment

$$S = \sum_i C(h_i) \log \frac{C(h_i)}{C'(h_i)}$$

Bayesian Surprise

- Calculate C(h), perform experiment, calculate C'(h)

$$C(h) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(\frac{-\left(\hat{h}(x_i) - y_i\right)^2}{2\sigma^2}\right)$$

y is the actual observation obtained for experiment x
n is the number of experiments performed so far

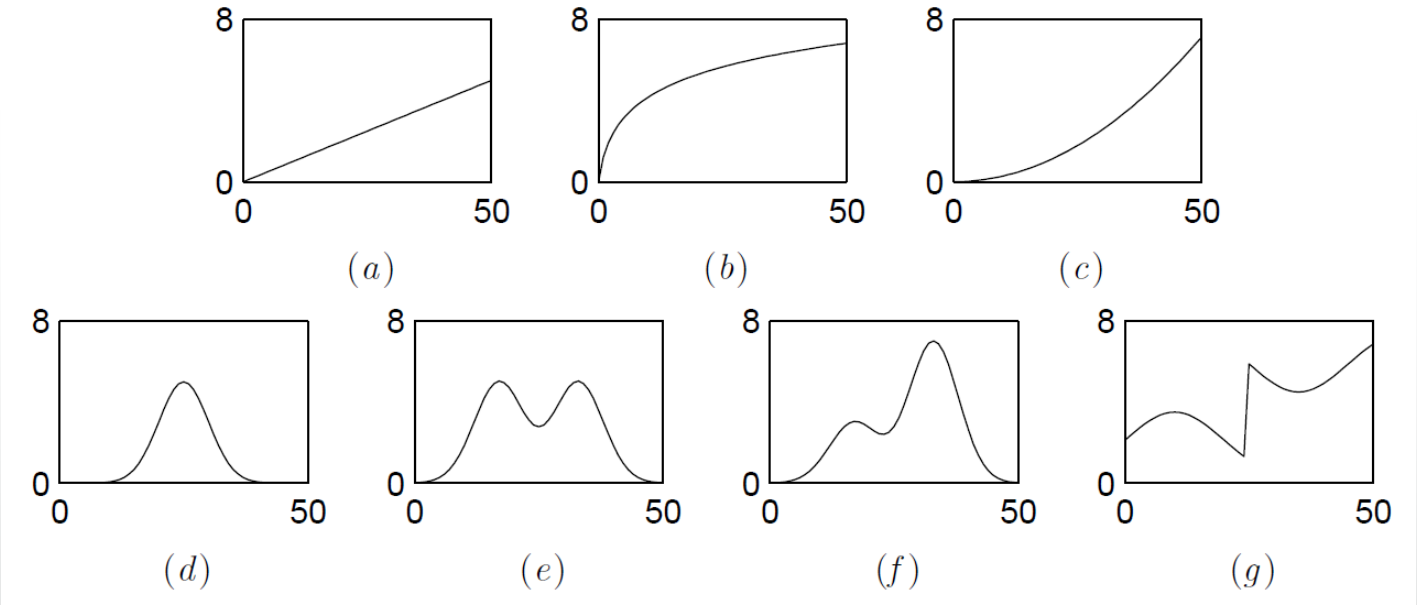
$$S = \sum_i C(h_i) \log \frac{C(h_i)}{C'(h_i)}$$

Bayesian Surprise

- In words:
  - A surprise is when a good hypothesis (high C(h)) becomes a bad hypothesis (low C(h)).
  - When this happens, we would like to know why it happened.
  - So when we get a surprise (S > O), we exploit to investigate the surprise
  - Otherwise we explore to learn something new

Evaluating the Approach

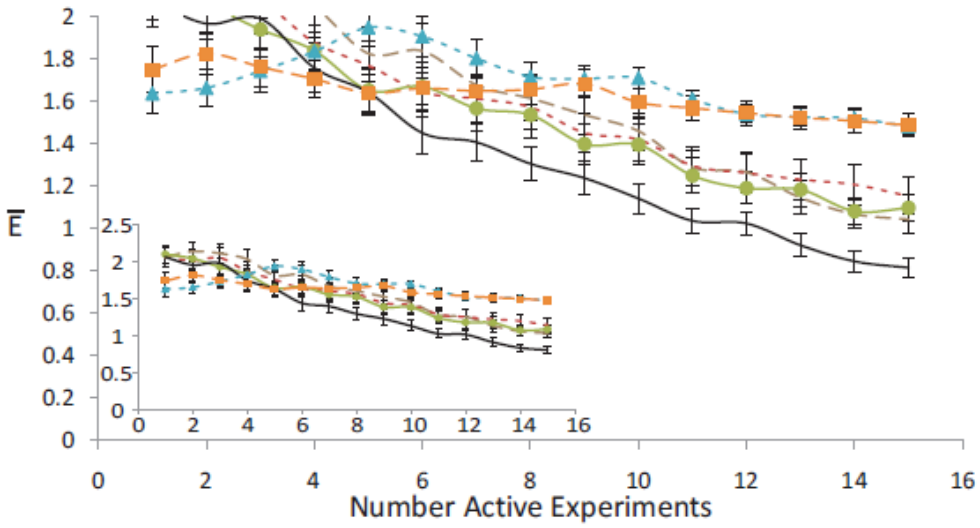
- Some possible behaviours that could be observed

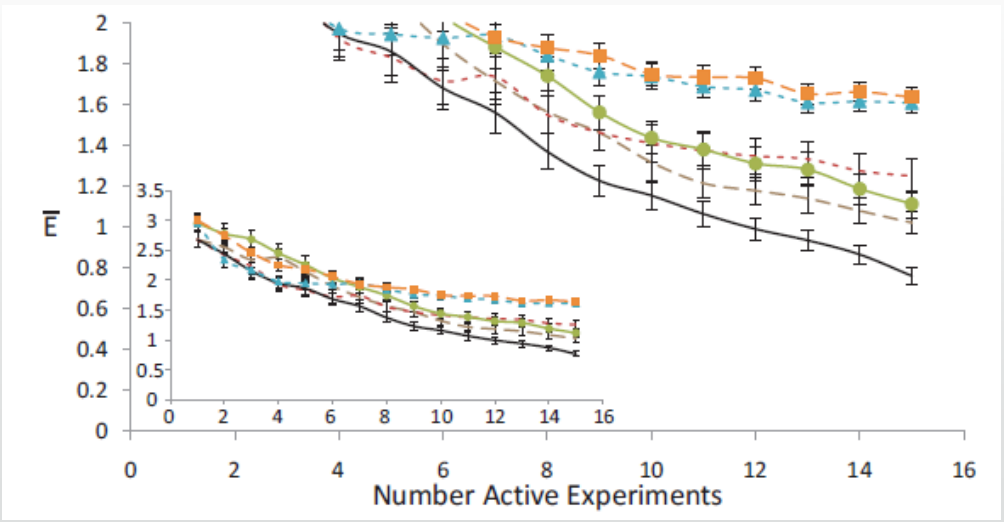


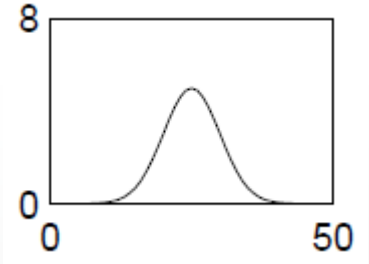
$$y = f(x + \delta) + \epsilon + \phi$$

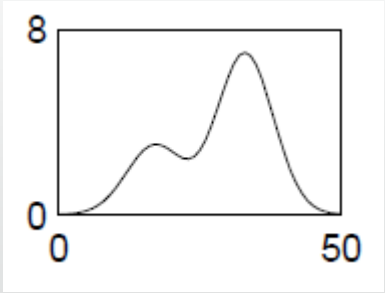
δ, ε Gaussian noise, φ shock noise

Simulation Results



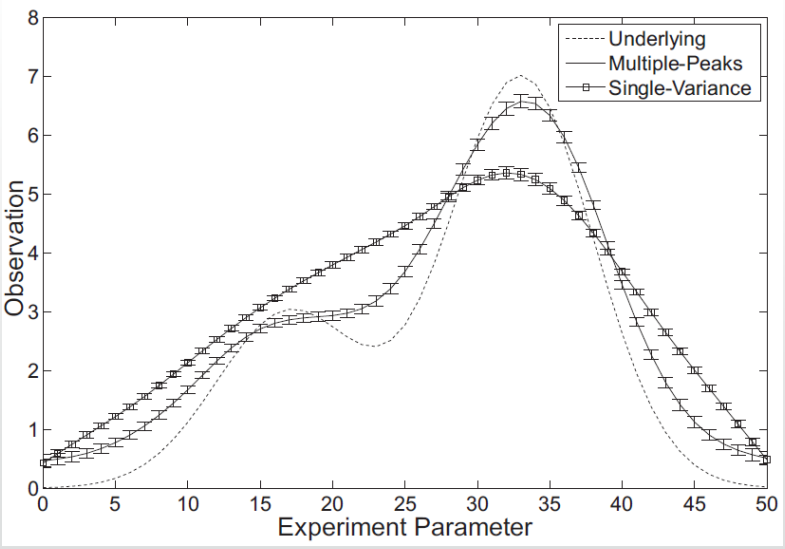






Simulation Results

- Single hypothesis works only in the monotonic behaviours
- Multiple hypotheses work for all the behaviours



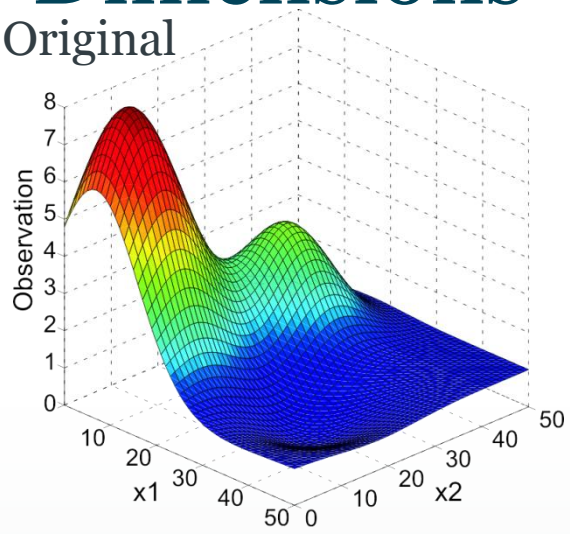
Mean prediction of most confident hypothesis over 100 trials for single and multiple hypotheses techniques.

Single hypothesis misses features of the behaviour.

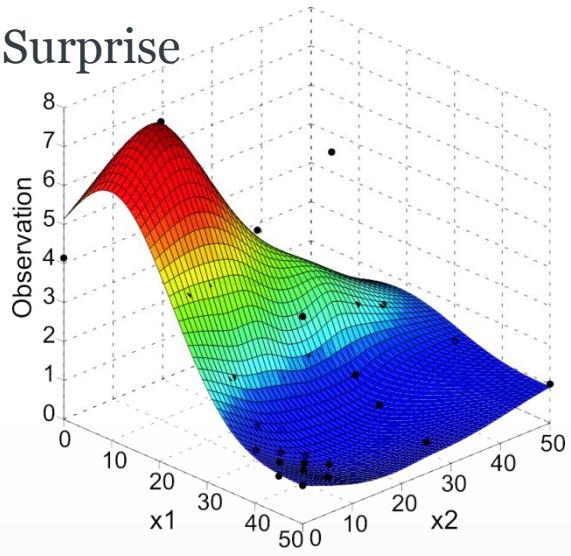
UNIVERSITY OF Southampton

Two Dimensions

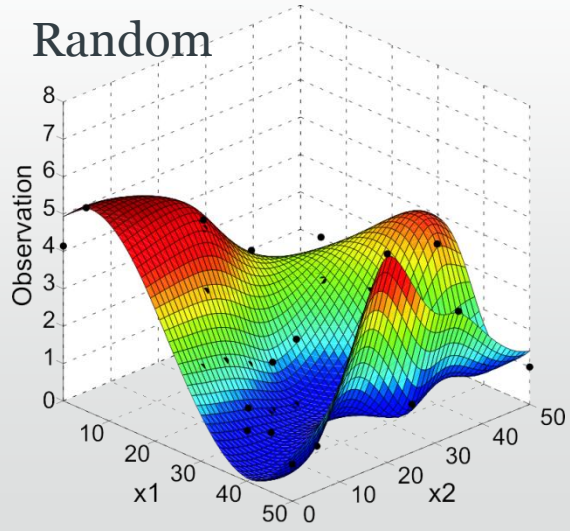
Original



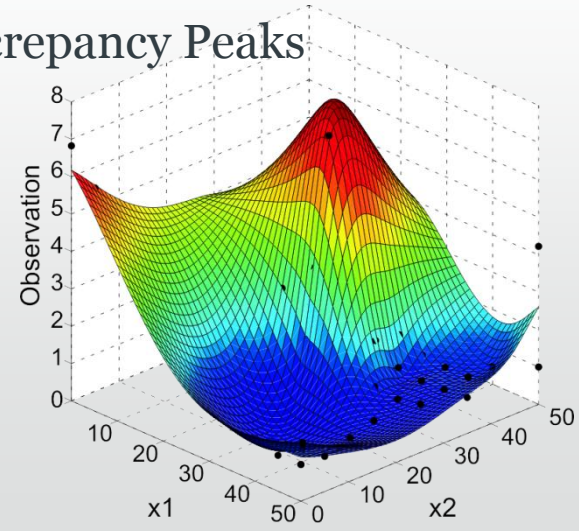
Surprise



Random



Discrepancy Peaks



Simulation Findings

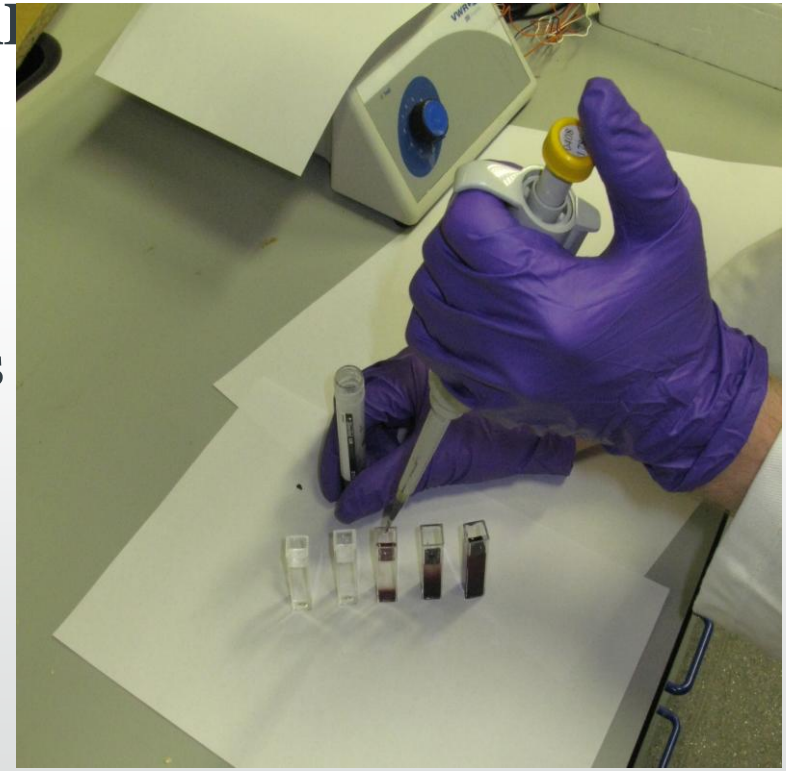
- Need multiple hypotheses
  - Single view of the data will not work
- Have an efficient way of identifying the most representative hypothesis in a set
- Have an effective way of managing the exploration – exploitation trade-off

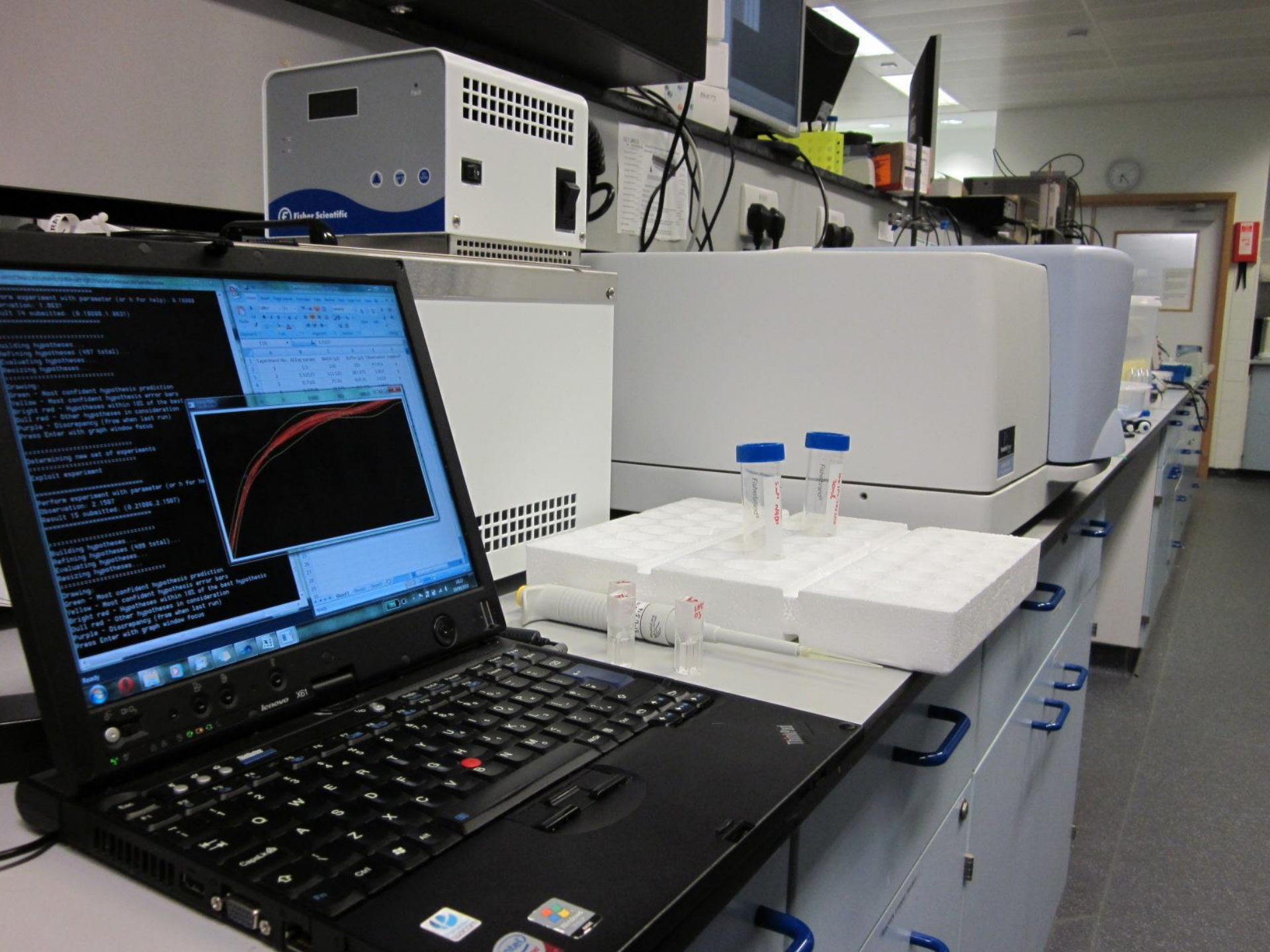
UNIVERSITY OF Southampton

Laboratory Evaluation

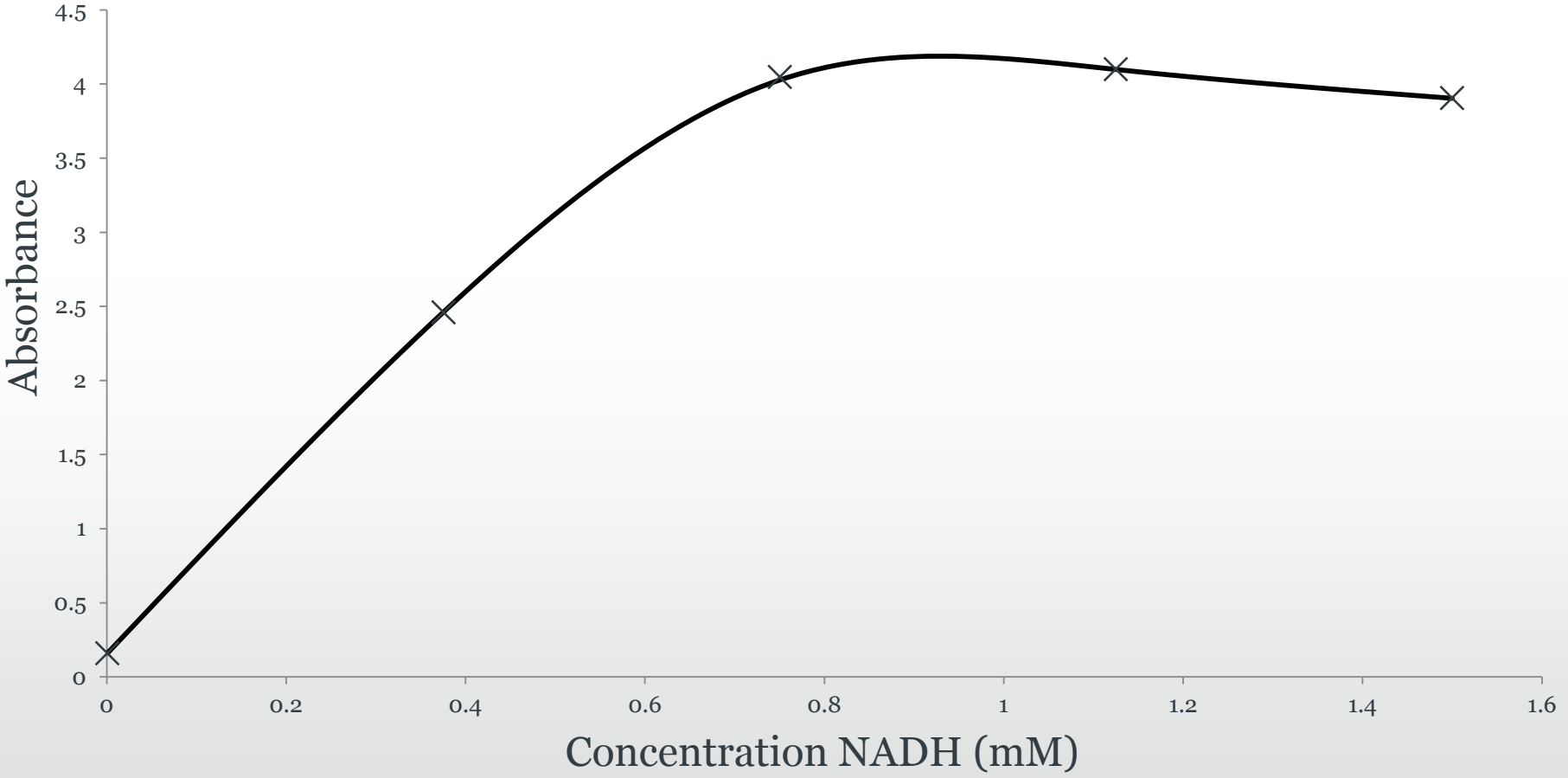
Testing in the laboratory

- Manual experiments directed by the algorithms
  - Characterisation of NADH
  - *x* parameter is concentration NAD
  - Observation is absorbance at a particular wavelength
  - 5 initial exploratory experiments
  - 9 actively chosen experiments
  - Compared to theoretical values from Beer-Lambert law

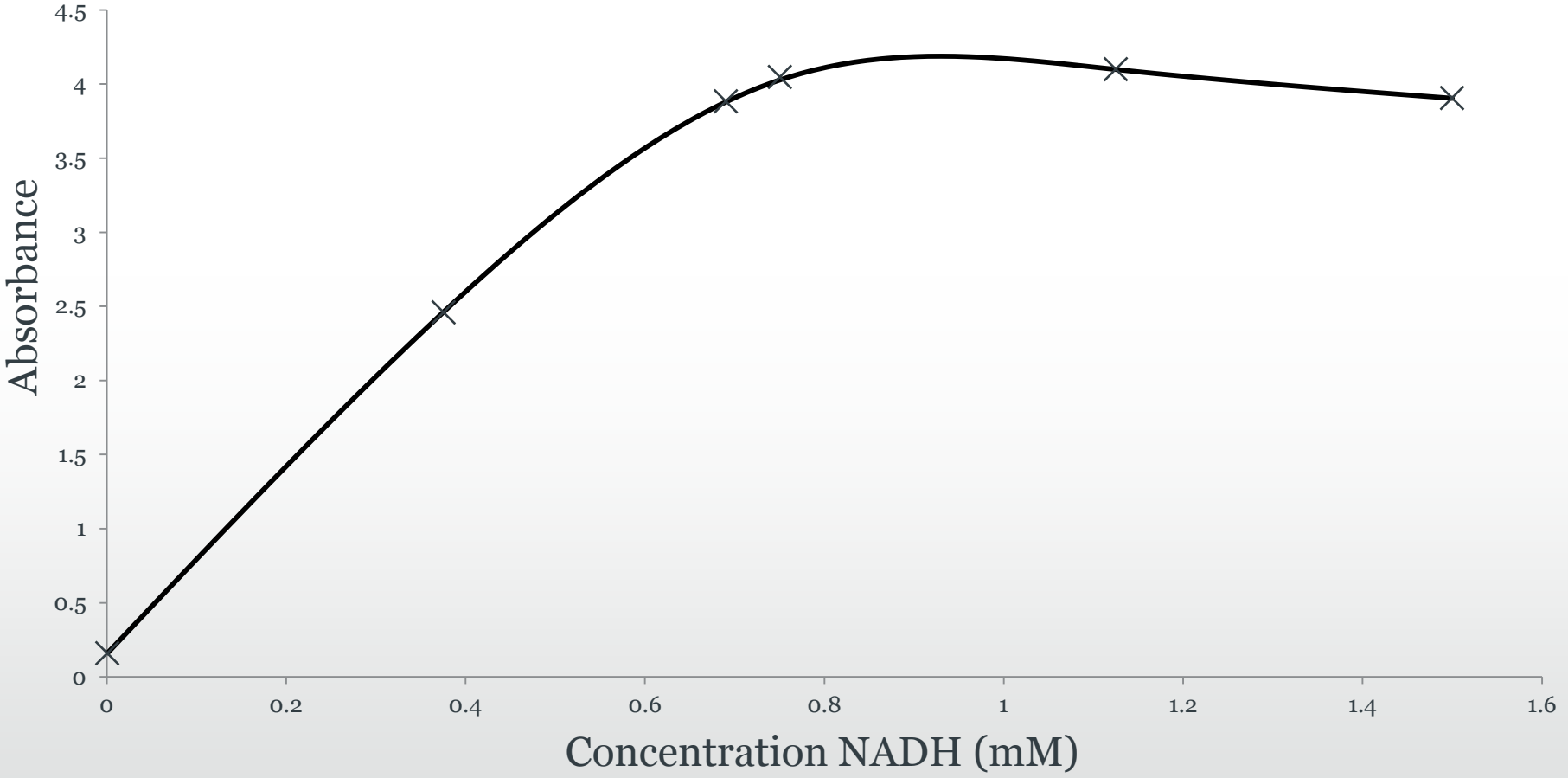



Fisher Scientific
Lenovo X61

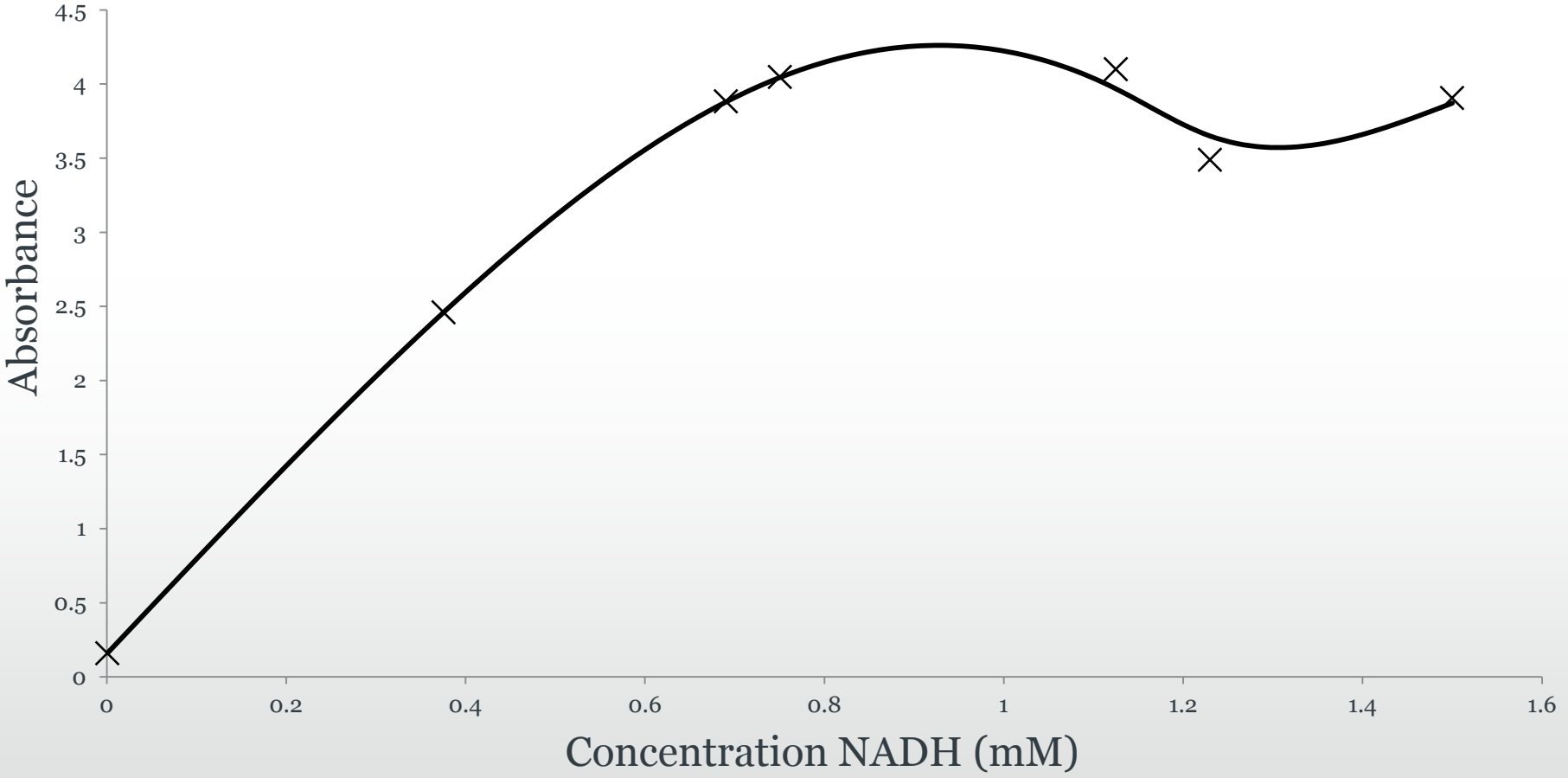
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

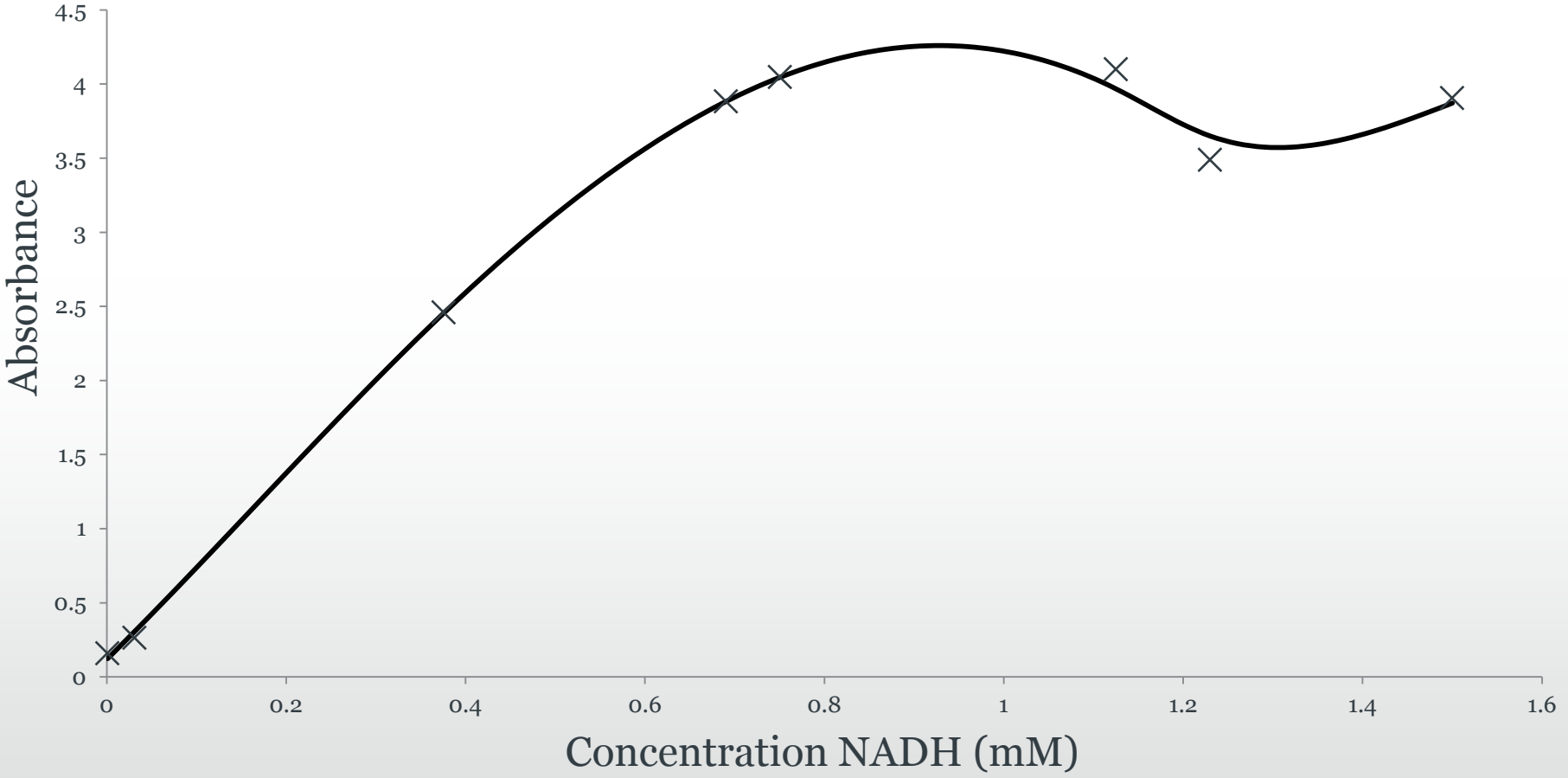
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

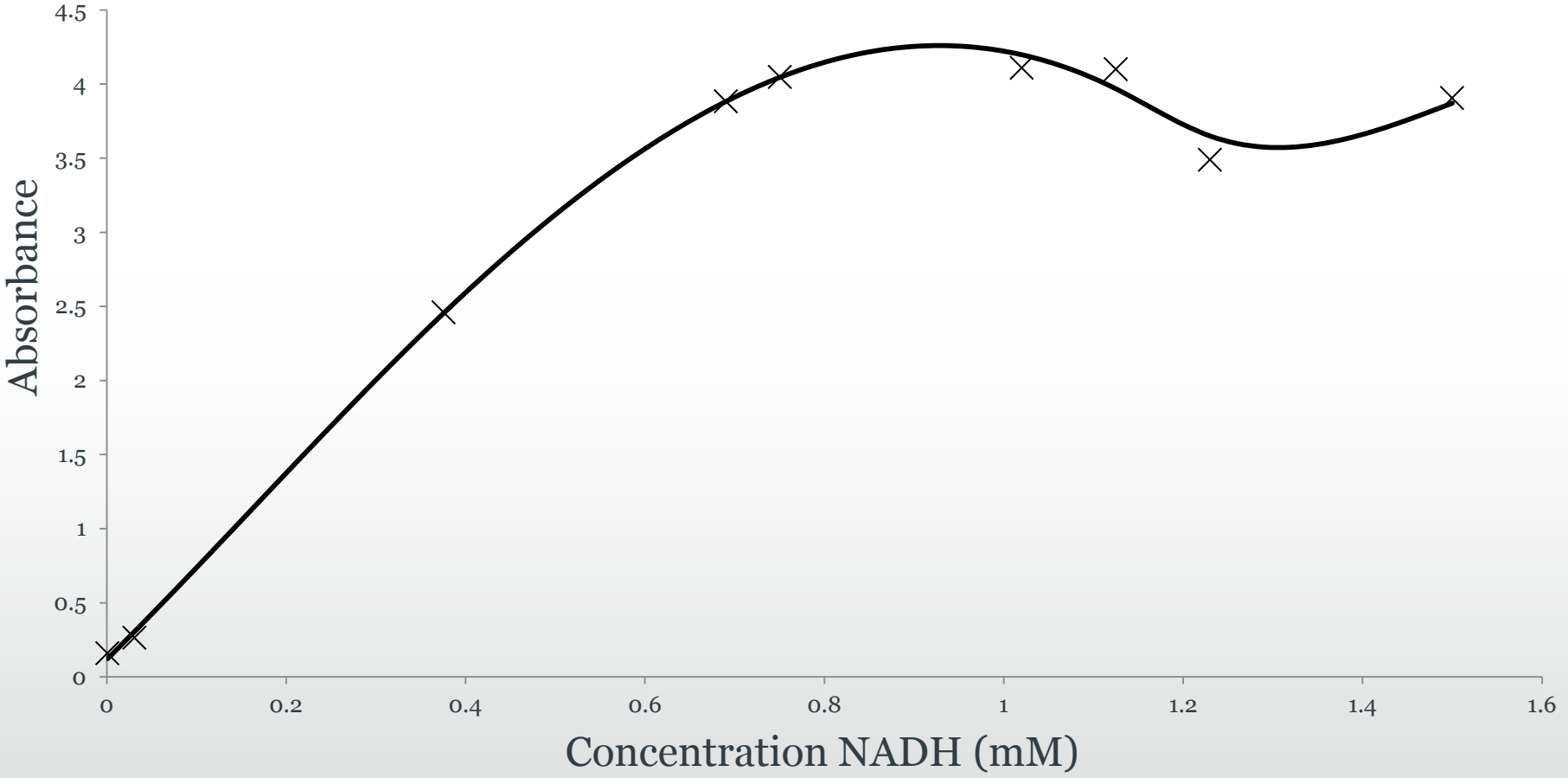
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

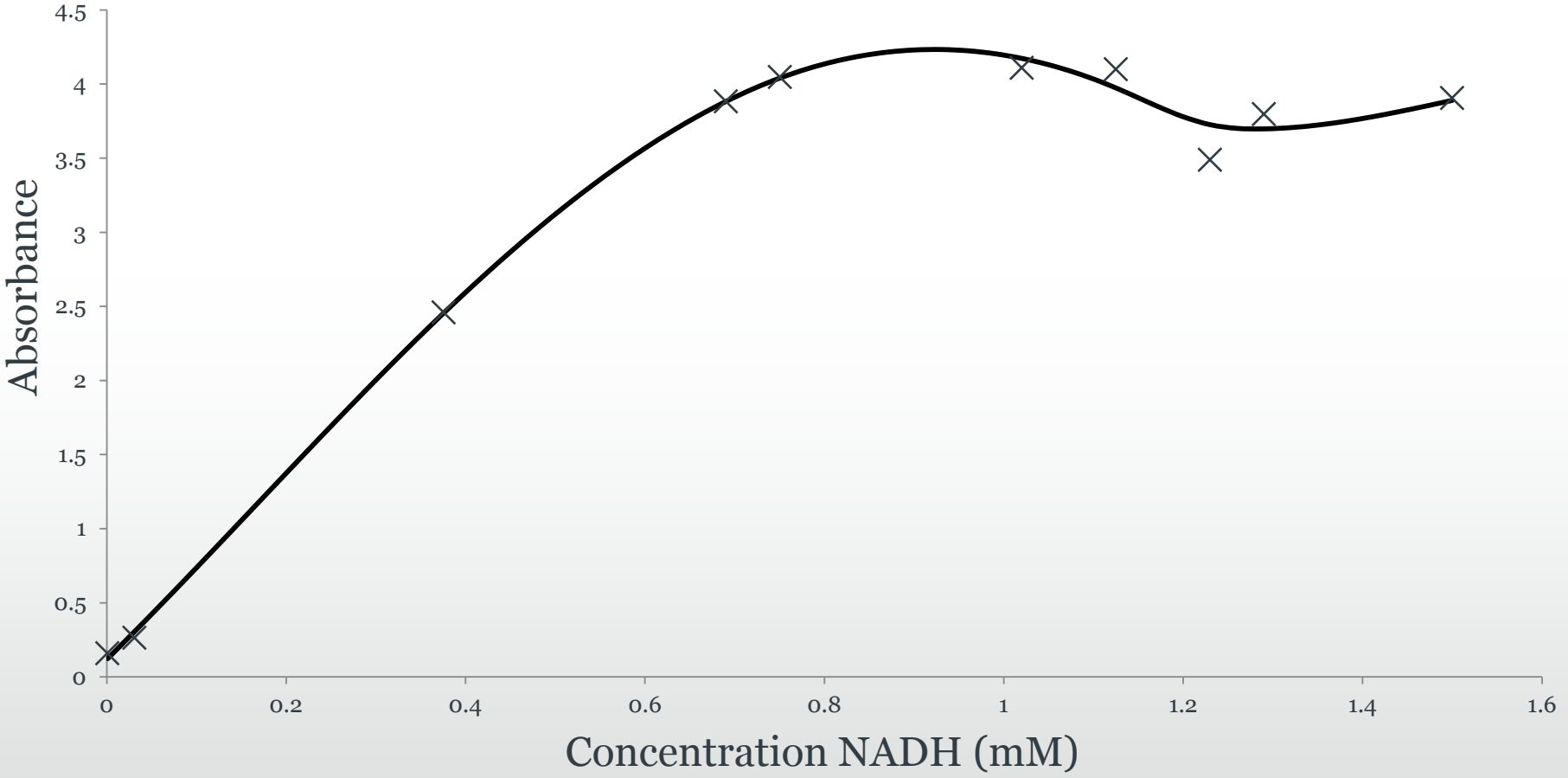
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

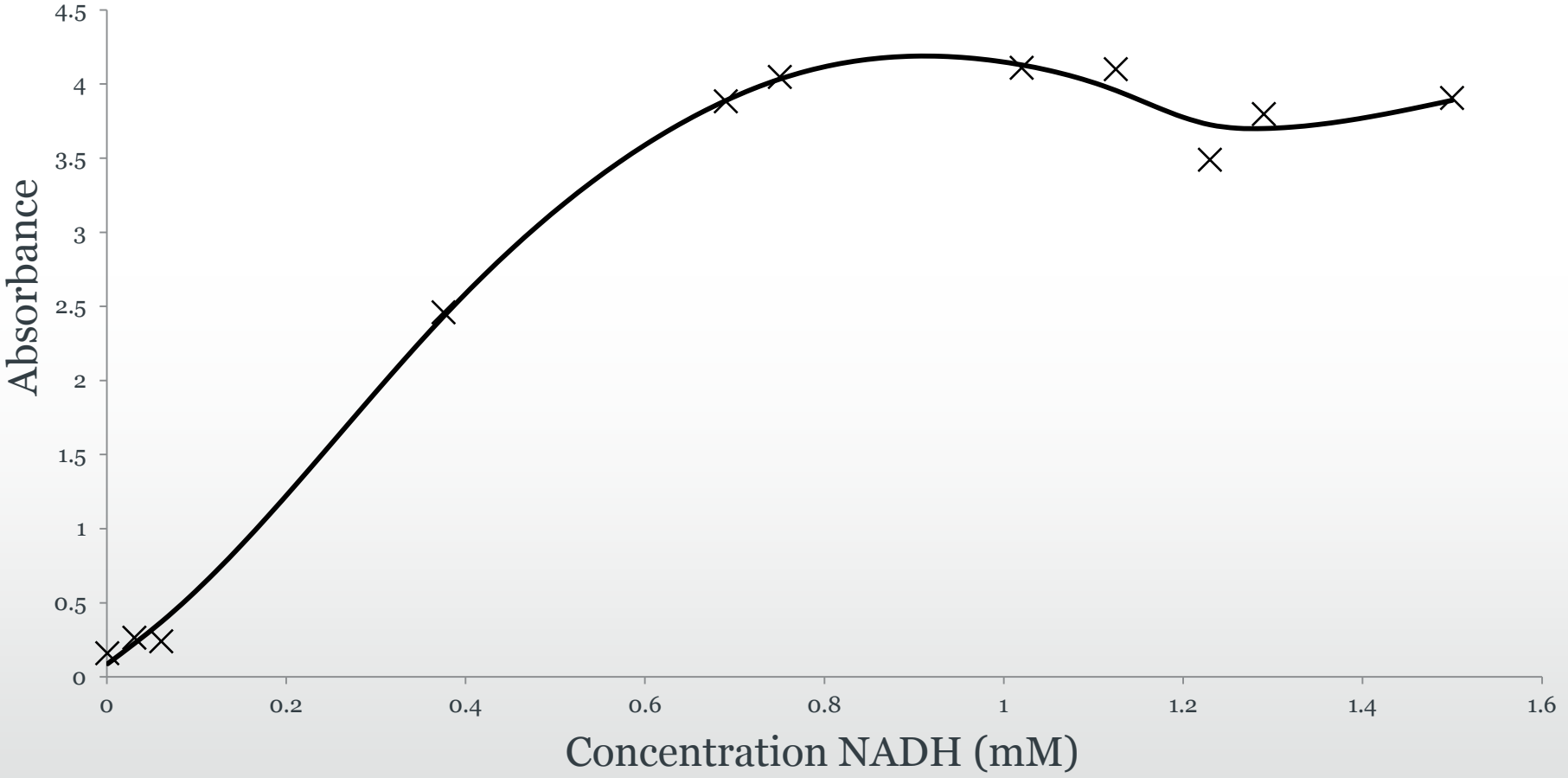
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

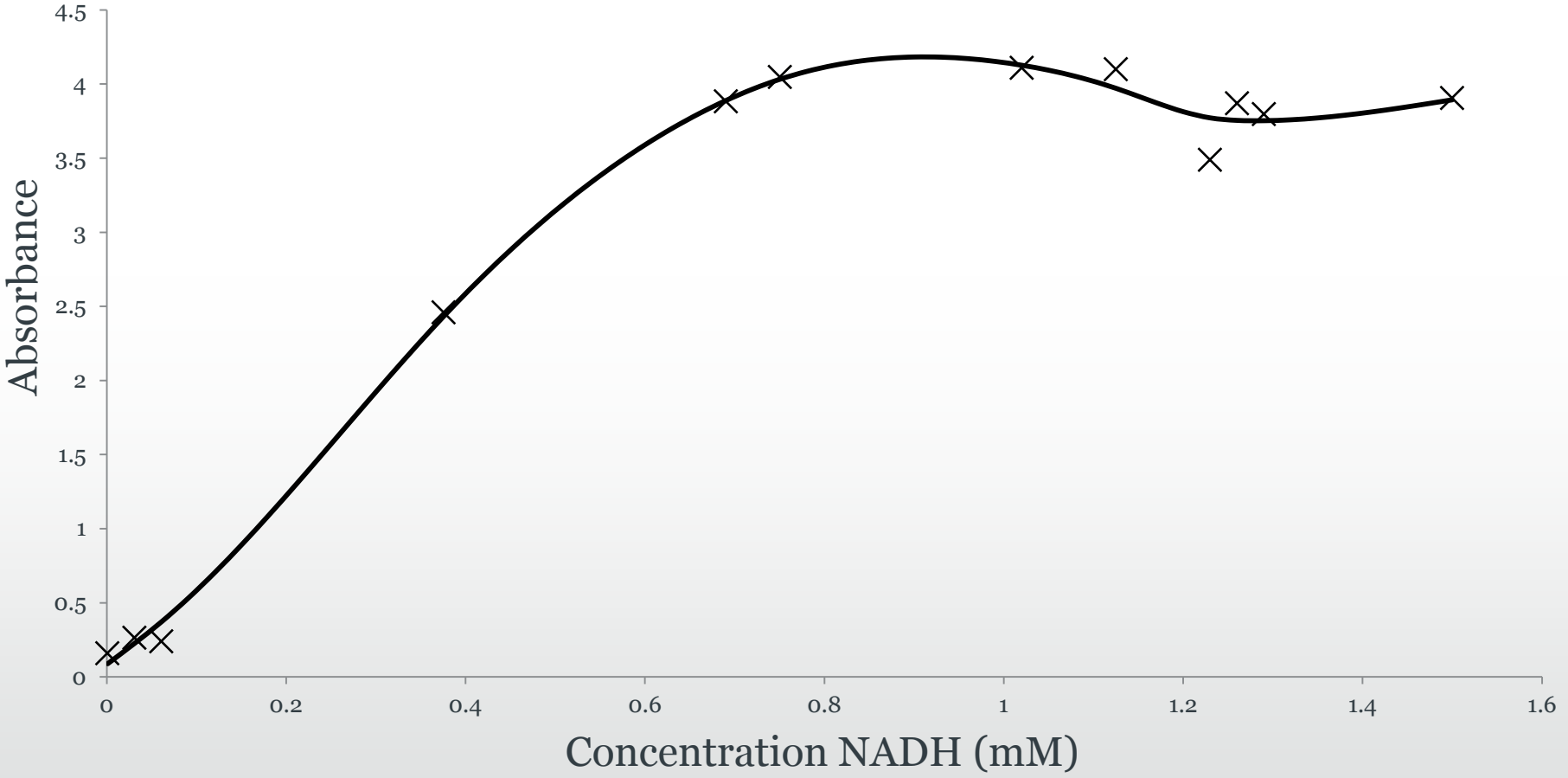
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

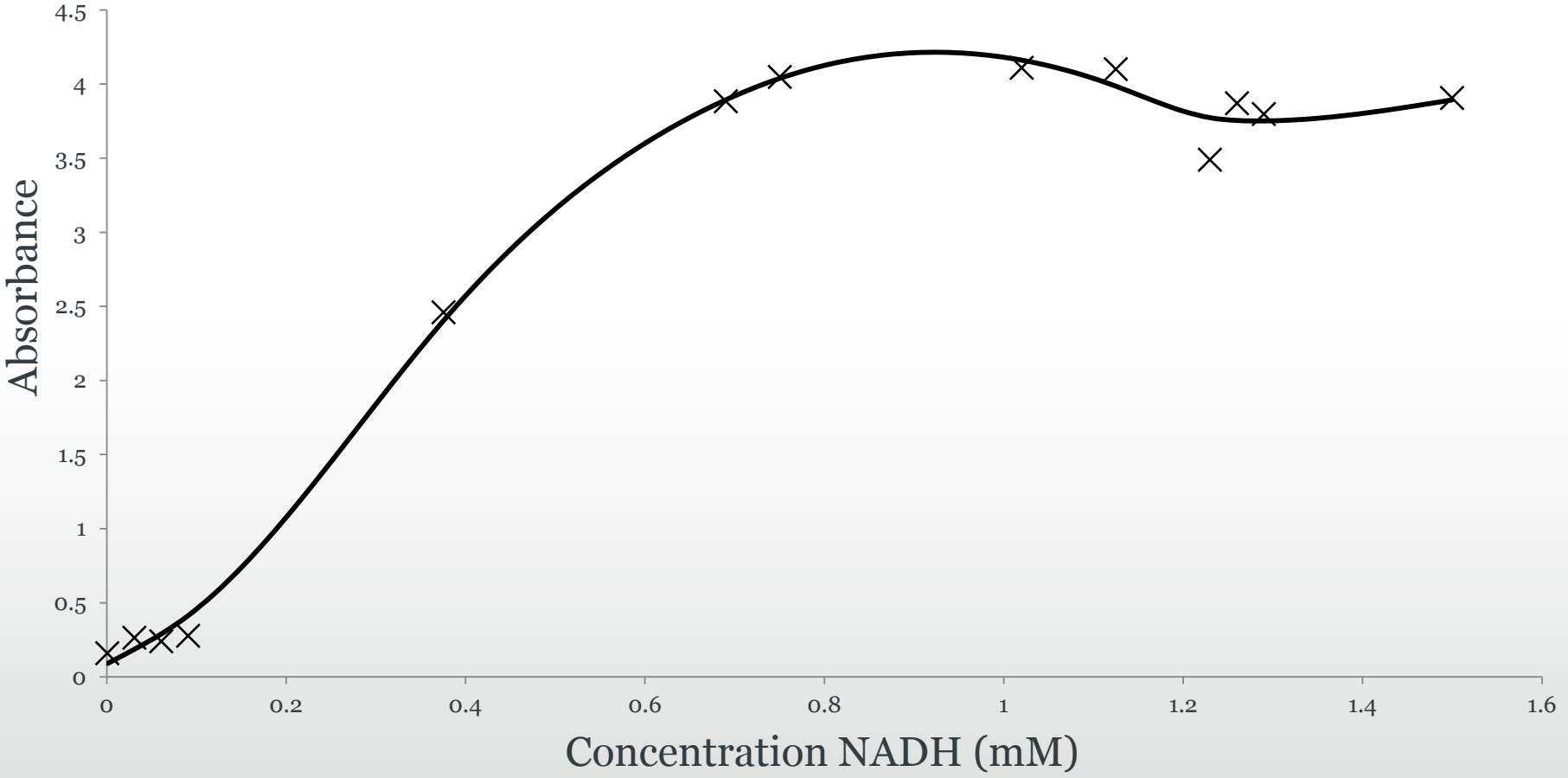
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

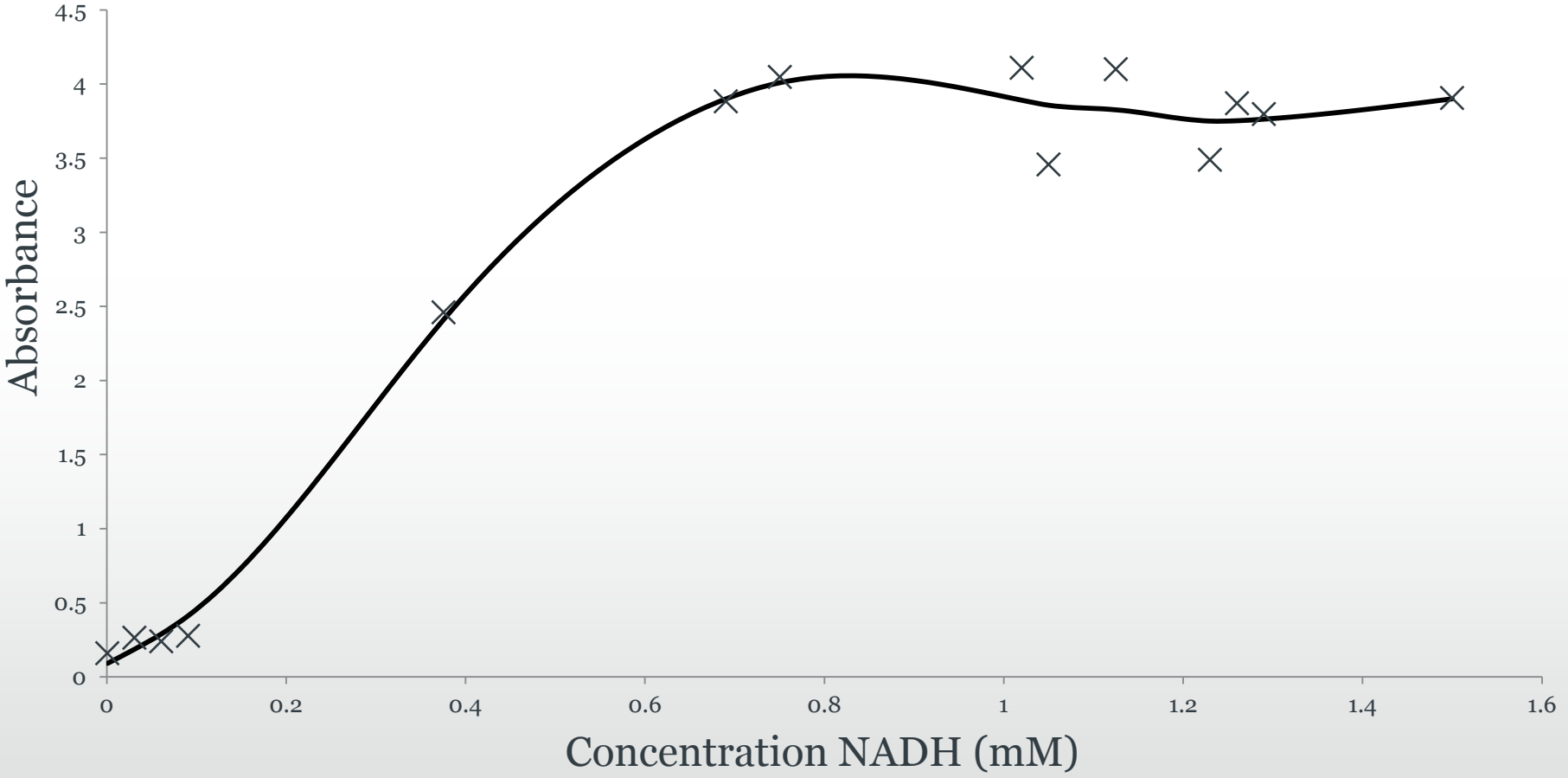
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

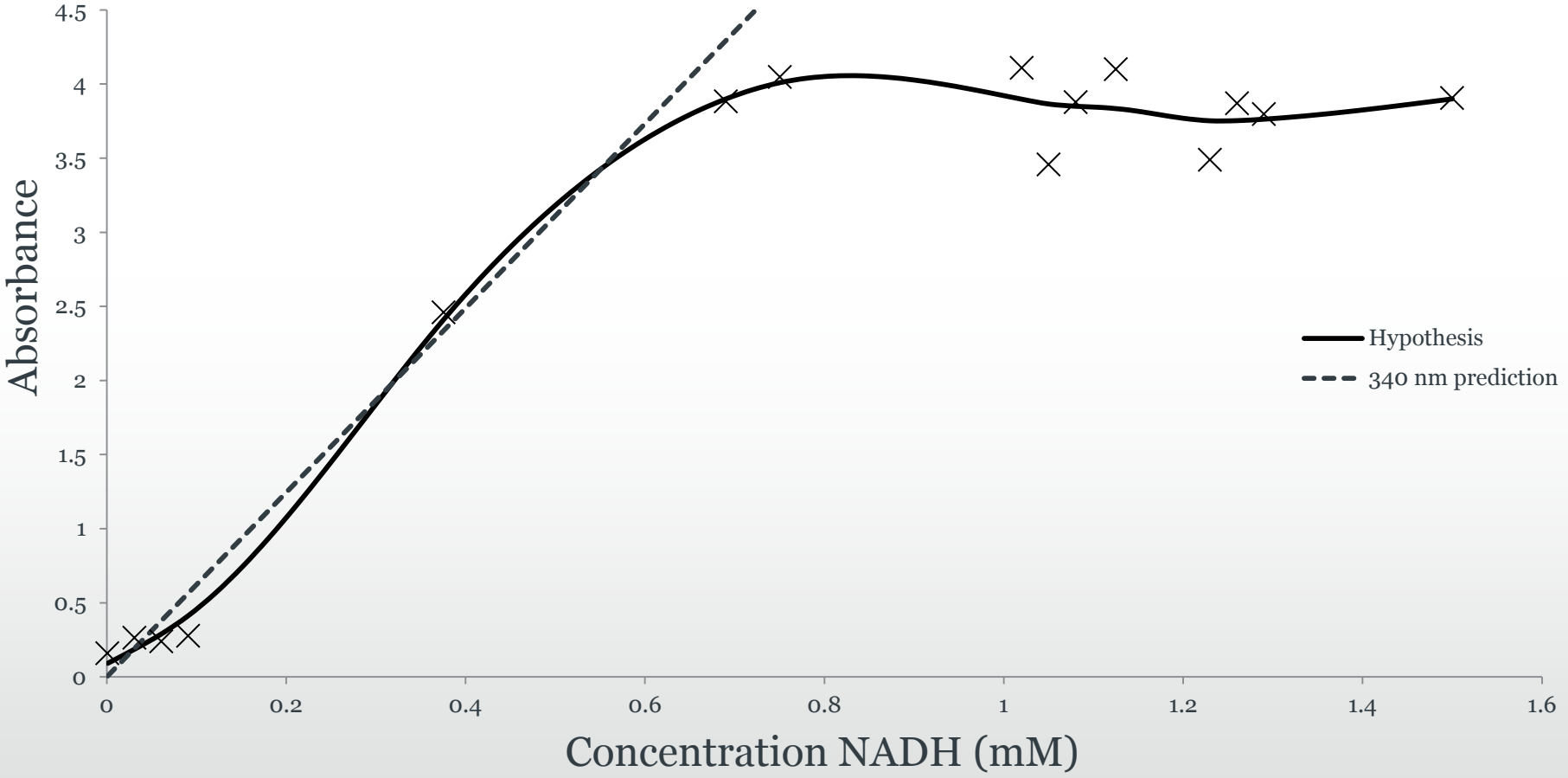
Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

Lab Run


Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)

Lab Run

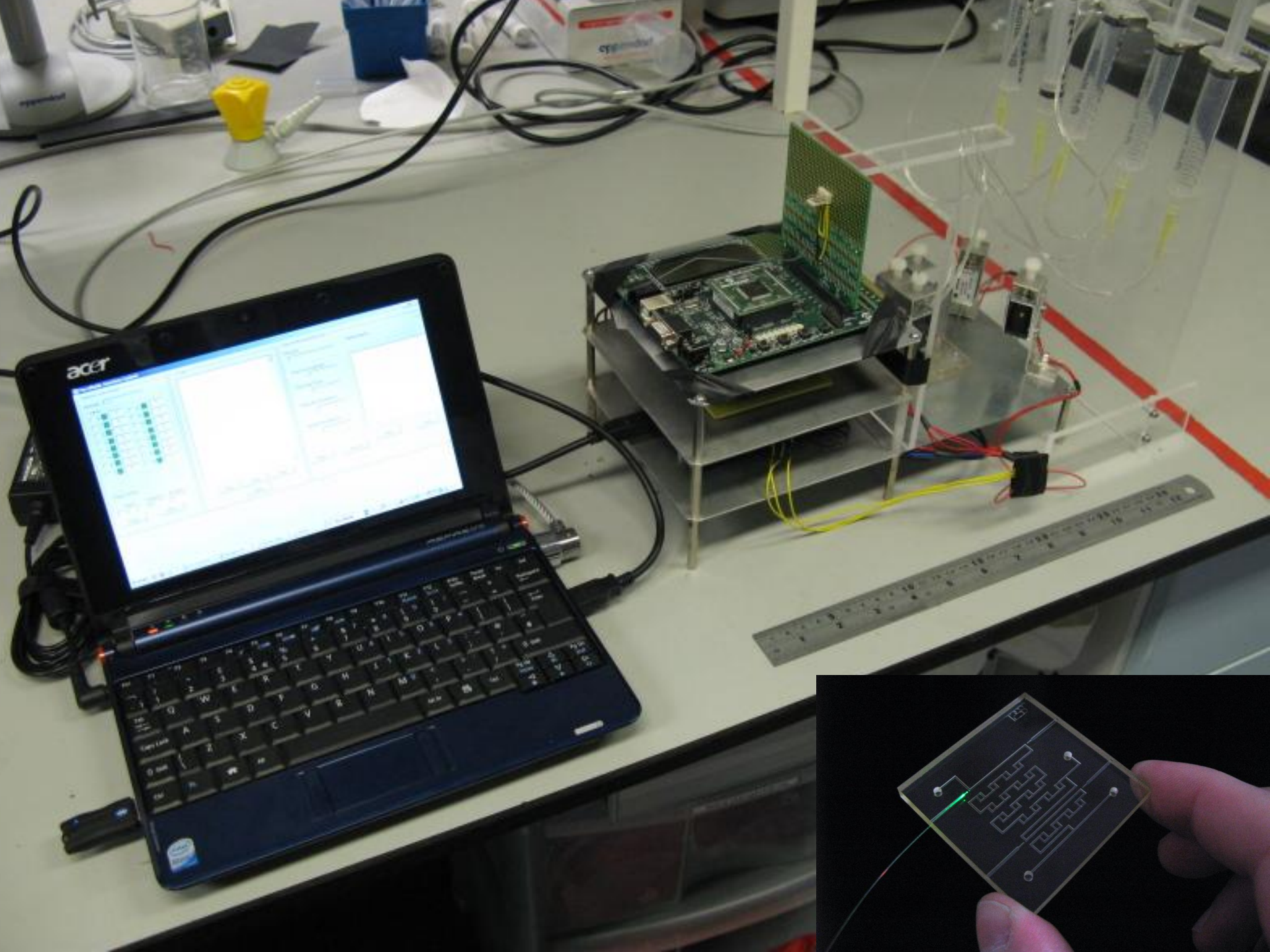

Absorbance
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Concentration NADH (mM)
Hypothesis
340 nm prediction

Conclusions

- Real-world active learning is more than just selecting the data points
  - Have to handle the data
  - Manage the issues surrounding the data
- Often resource costs are more limiting than you would expect
  - Less data around
- Scientists have been discovering for years
  - We can learn from them

Acknowledgements

- Gareth Jones
- Steve Gunn
- Klaus-Peter Zauner
- PASCAL2
- Microsoft Research


acer