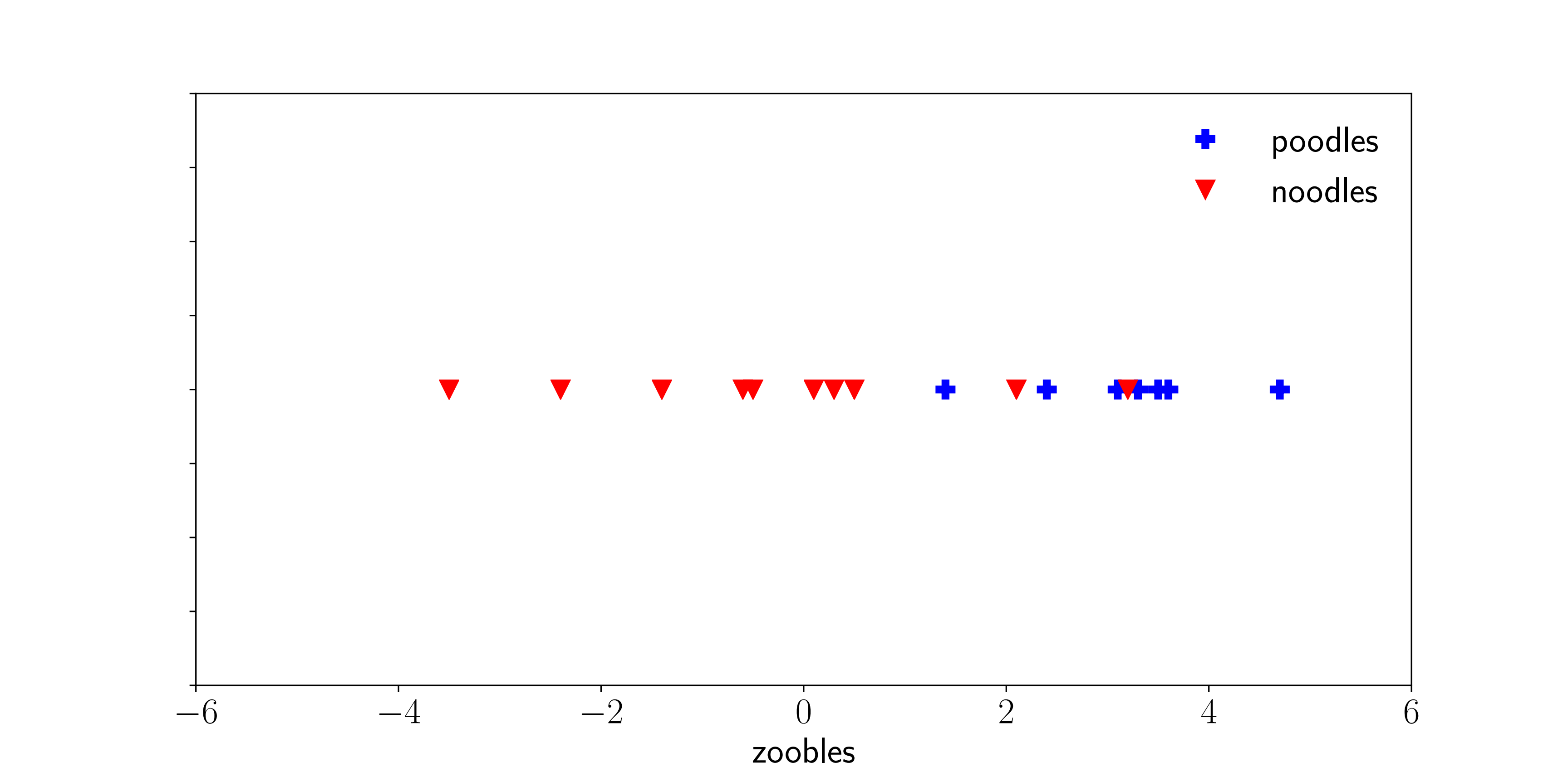
Human in the Loop Machine Learning

Cheng Soon Ong

9 July 2019 – Bukalapak, Jakarta

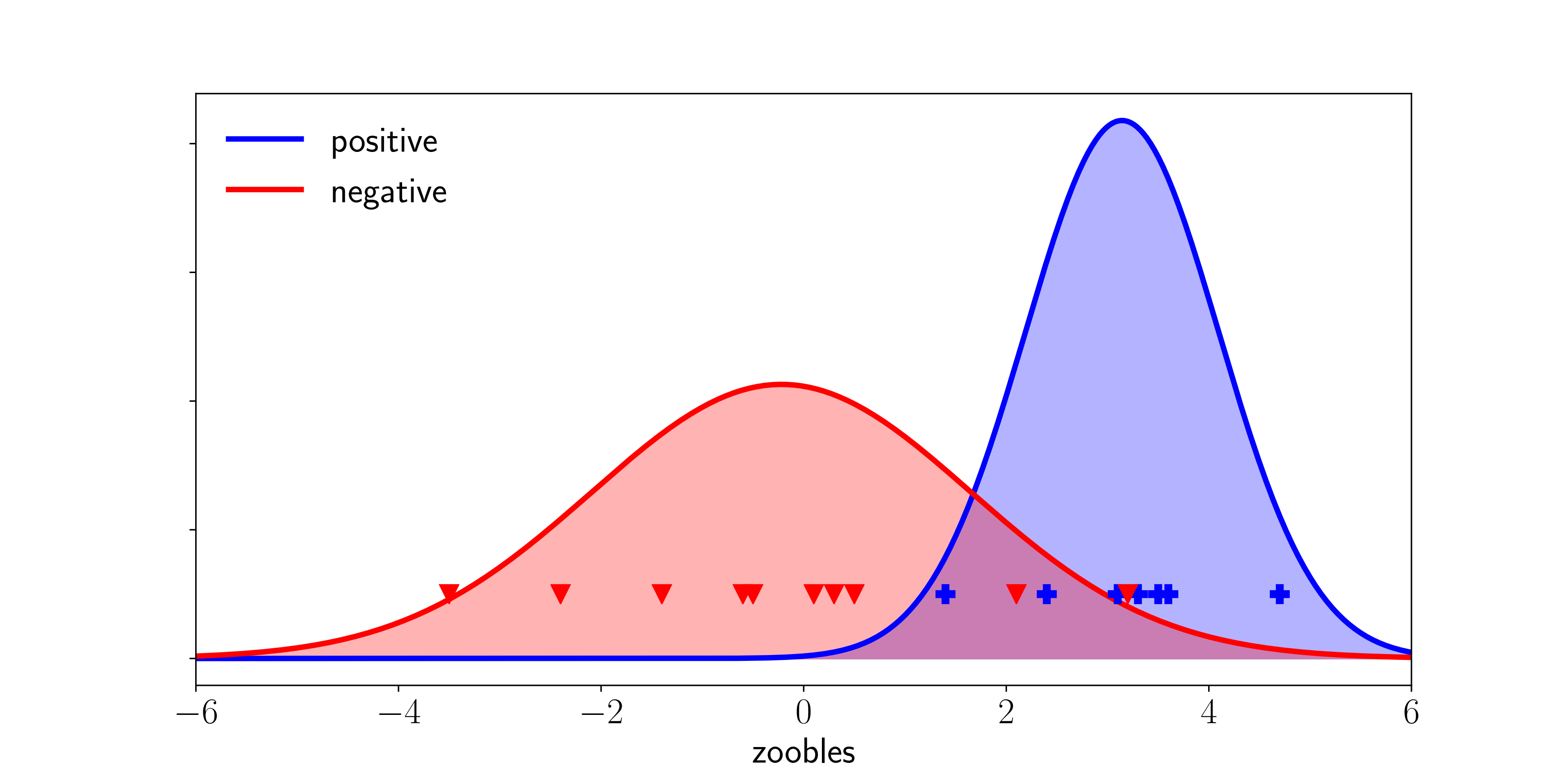
Given some data

Classify blue plus vs red triangles, based on features



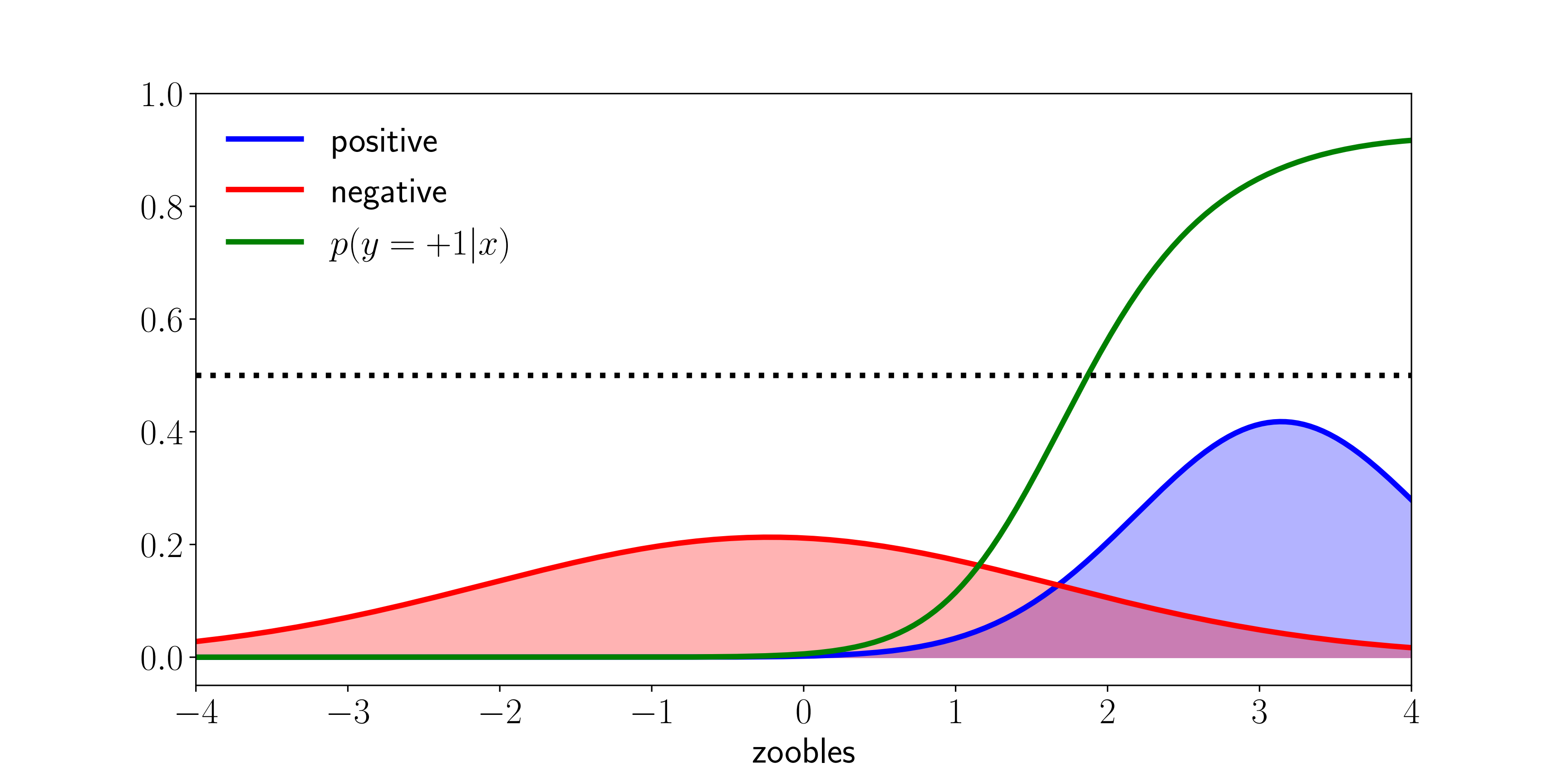
Fit a model to data

Estimate a Gaussian for each class conditional



Build a classifier

Compute the posterior probability of blue plus



What is Machine Learning?

- Machine Learning is about prediction
  - Examples/covariates/features
  - Labels/annotations/target variable

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \sim \mathcal{X}$$
$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \sim \mathcal{Y}$$

Predictor

$$f_{\boldsymbol{w}}(\boldsymbol{x}) : \mathcal{X} \rightarrow \mathcal{Y}$$

- Estimate the best predictor = training
  - No mechanistic model of the phenomenon
  - There are many examples
  - The outcomes (labels) are well defined (usually binary)

CSIRO DATA 61

Data lifecycle

1. Can I load your data using `pandas` or `numpy`?
2. Confounders, missing values, scale, units, encoding
3. Define the problem you want to answer:
   - The business/scientific problem
   - The performance metric
   - The model for the predictor
4. Run `sklearn` or `statsmodels` (**machine learning part**)
   Do not train on the test set.
5. Convert predictions into human friendly form for decision making

Prediction ≠ understanding ≠ taking action

- How can we use prediction to help humans perform discovery?

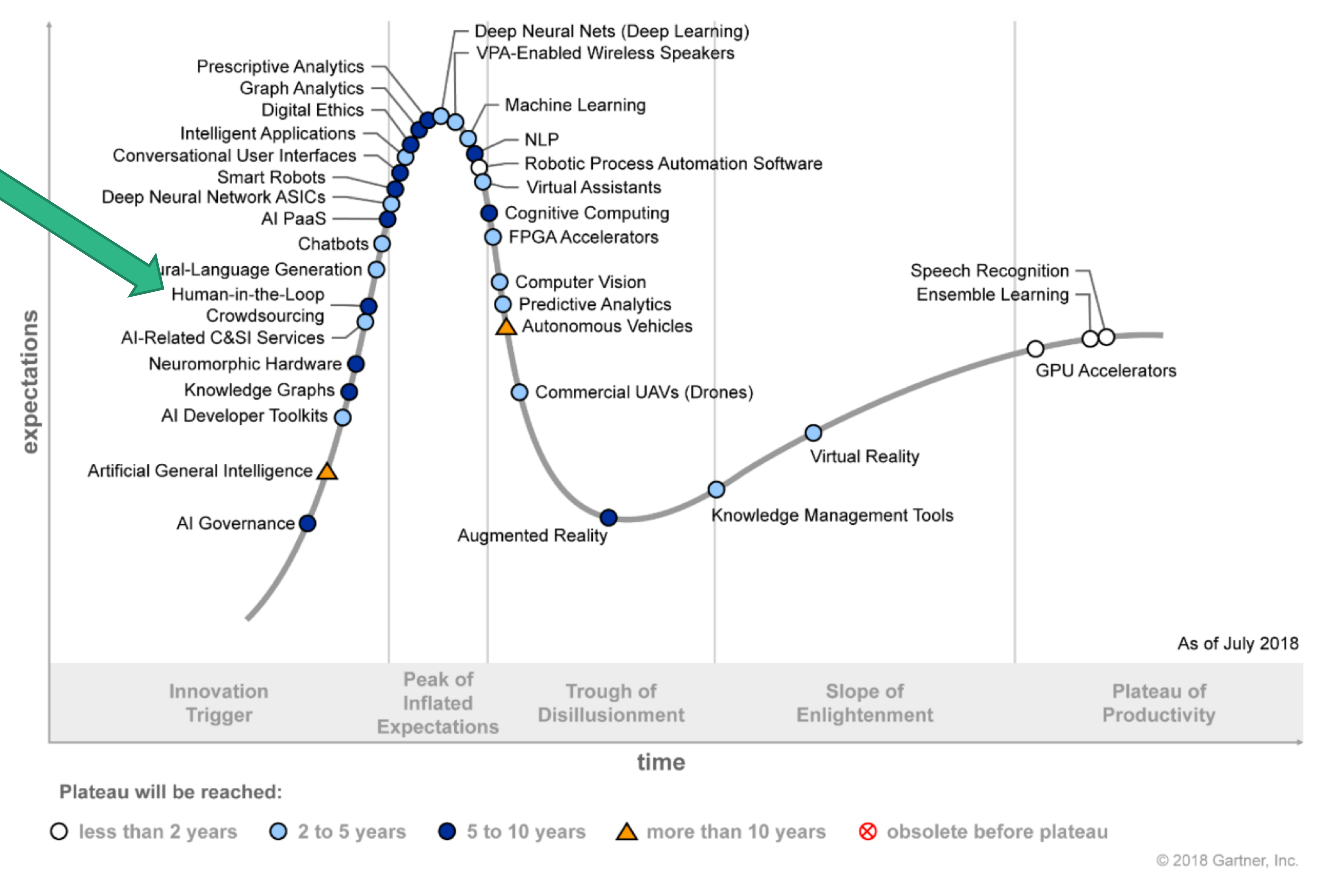
$$f_{\boldsymbol{w}}(\boldsymbol{x}) : \mathcal{X} \to \mathcal{Y}$$

Human in the Loop Machine Learning

- Use a combination of human experts and machine learning predictions
- This talk:

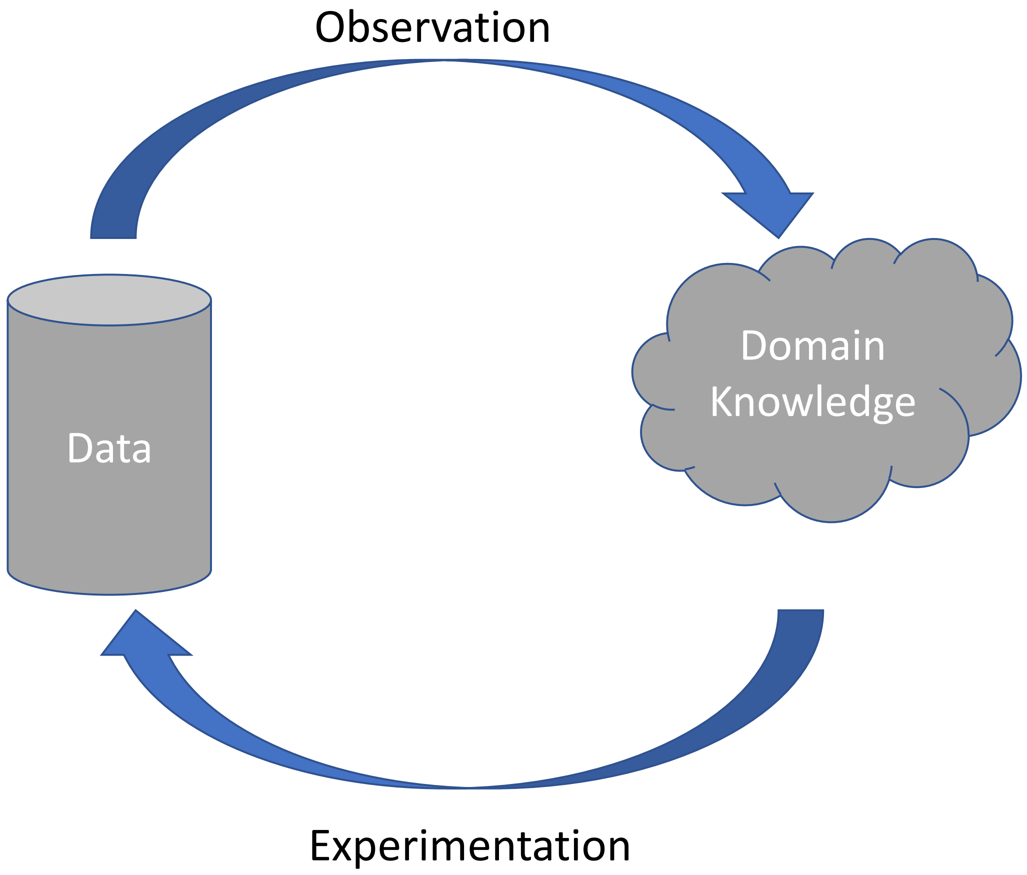
  Where should the machine ask for help?

Figure 1. Hype Cycle for Artificial Intelligence, 2018



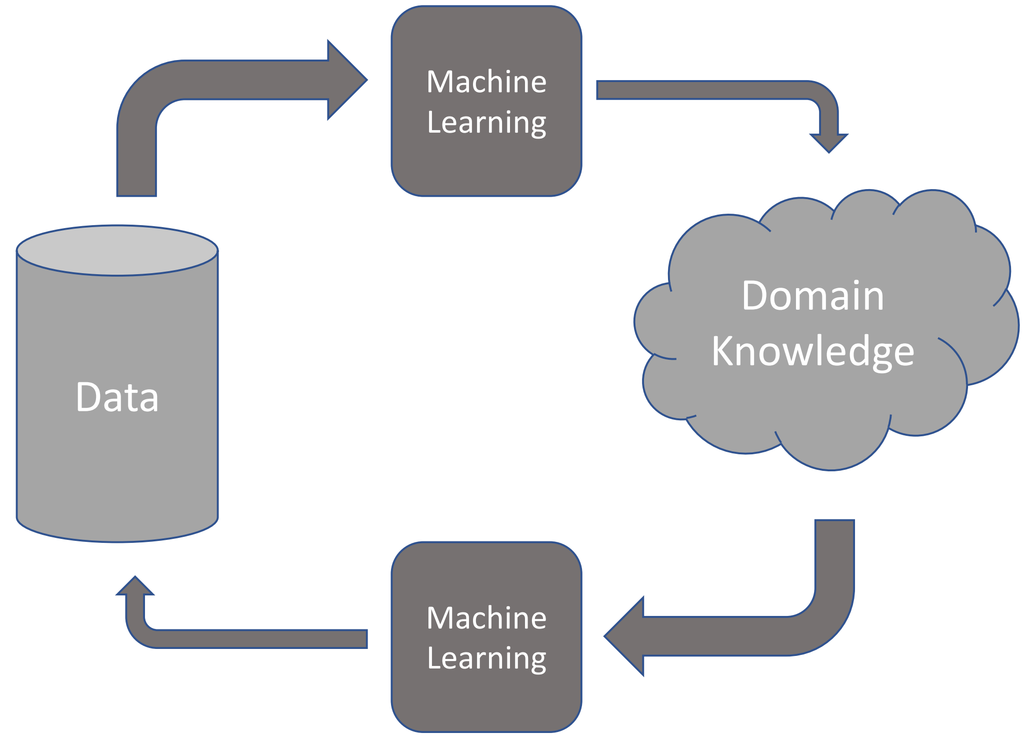
Source: Gartner (July 2018)

What is scientific discovery?

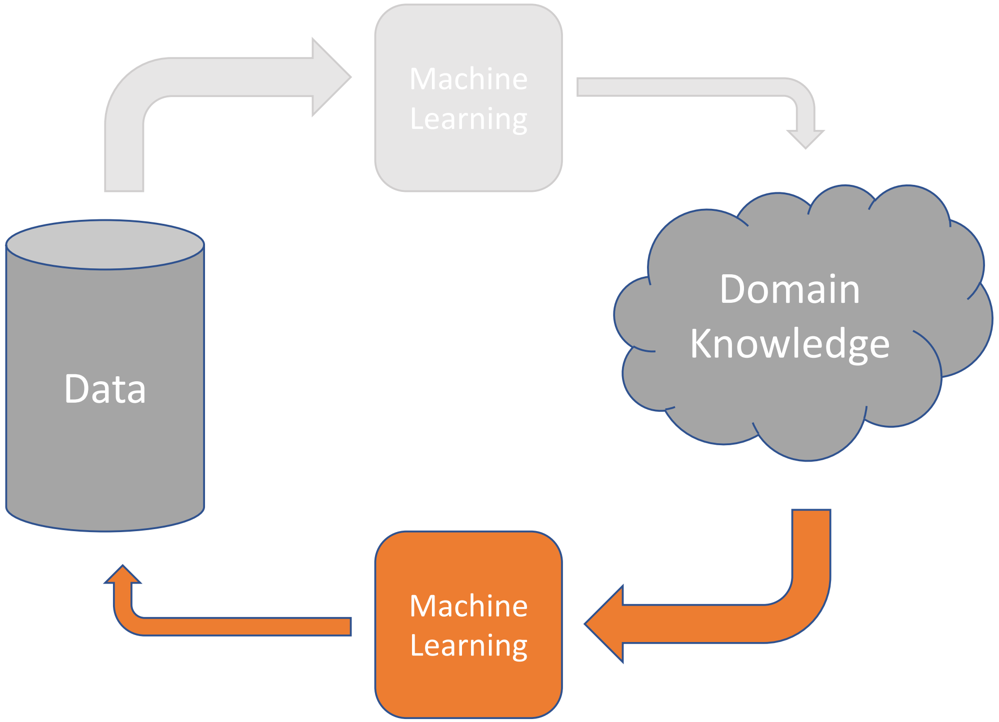


Francis Bacon,
credited with the modern
scientific method

Scientific discovery with machine learning



$$f_{\boldsymbol{w}}(\boldsymbol{x}) : \mathcal{X} \to \mathcal{Y}$$



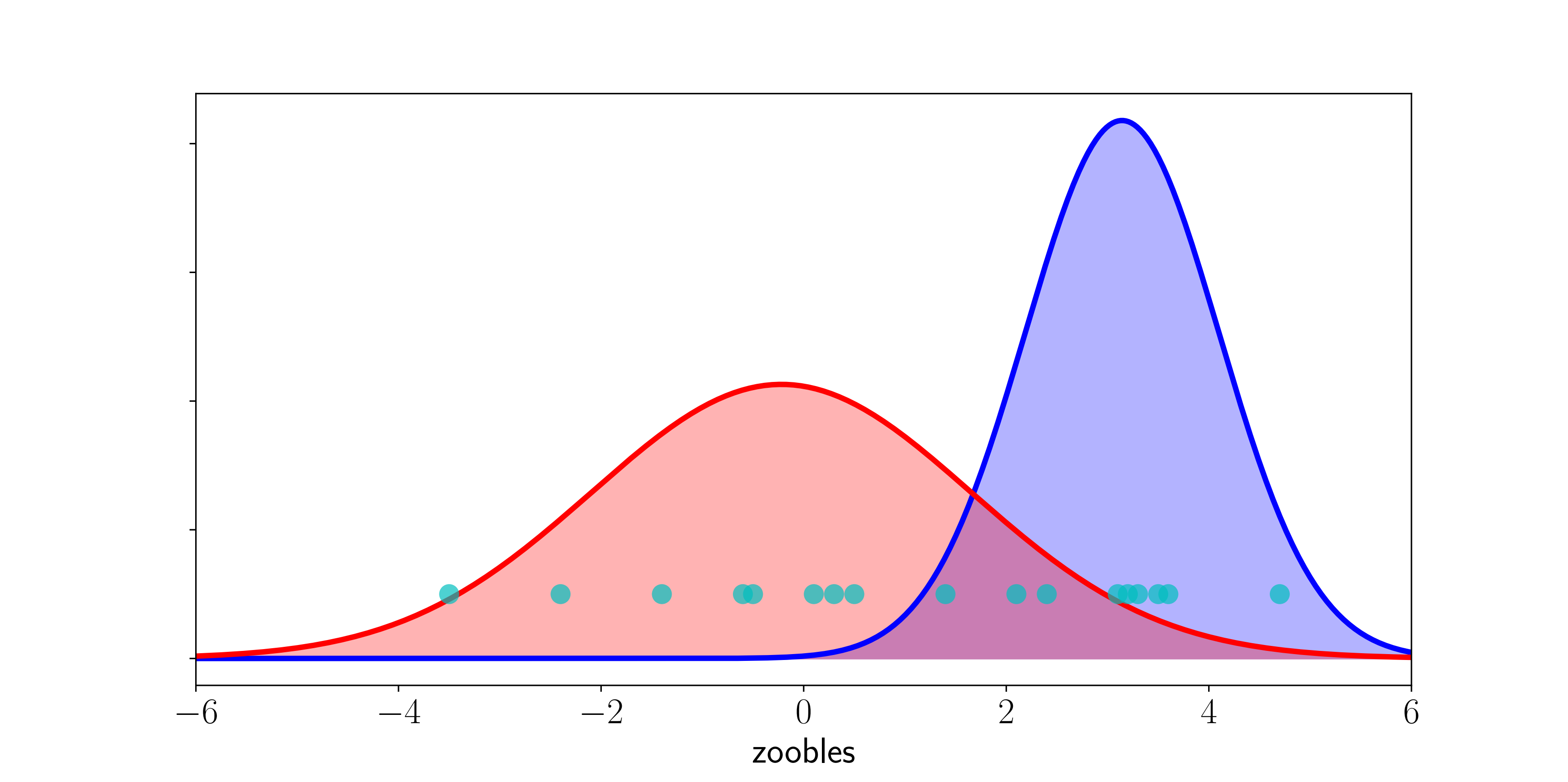
$$f_w(x) : \mathcal{X} \to \mathcal{Y}$$

- Assume that domain knowledge is captured by a predictor

- Use predictor to decide where to measure (ABCDE)
  - (A) Active Learning
  - (B) Bandits / Bayesian Optimisation
  - (C) Choice Theory
  - (DE) Design of Experiments

A – Active Learning

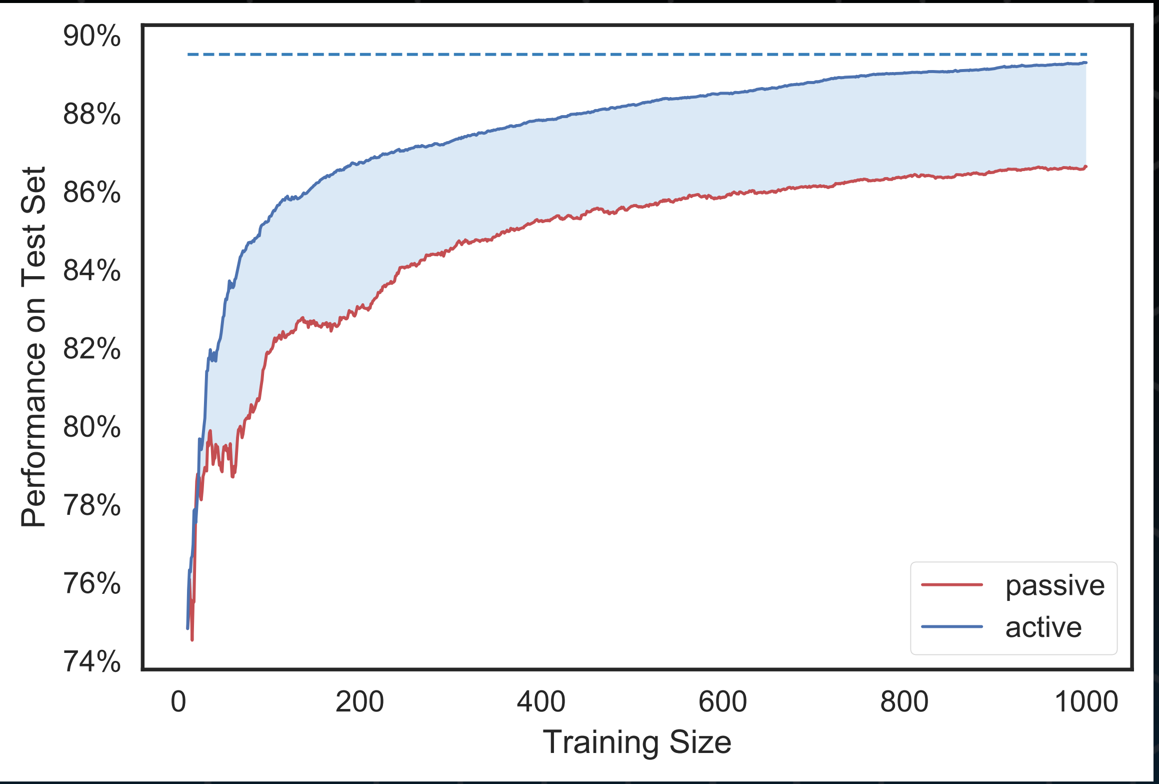
Want to build a classifier without paying for a lot of labels



A - Active Learning

- Choose a particular example to label using heuristics
- Annotator assumed to provide ground truth

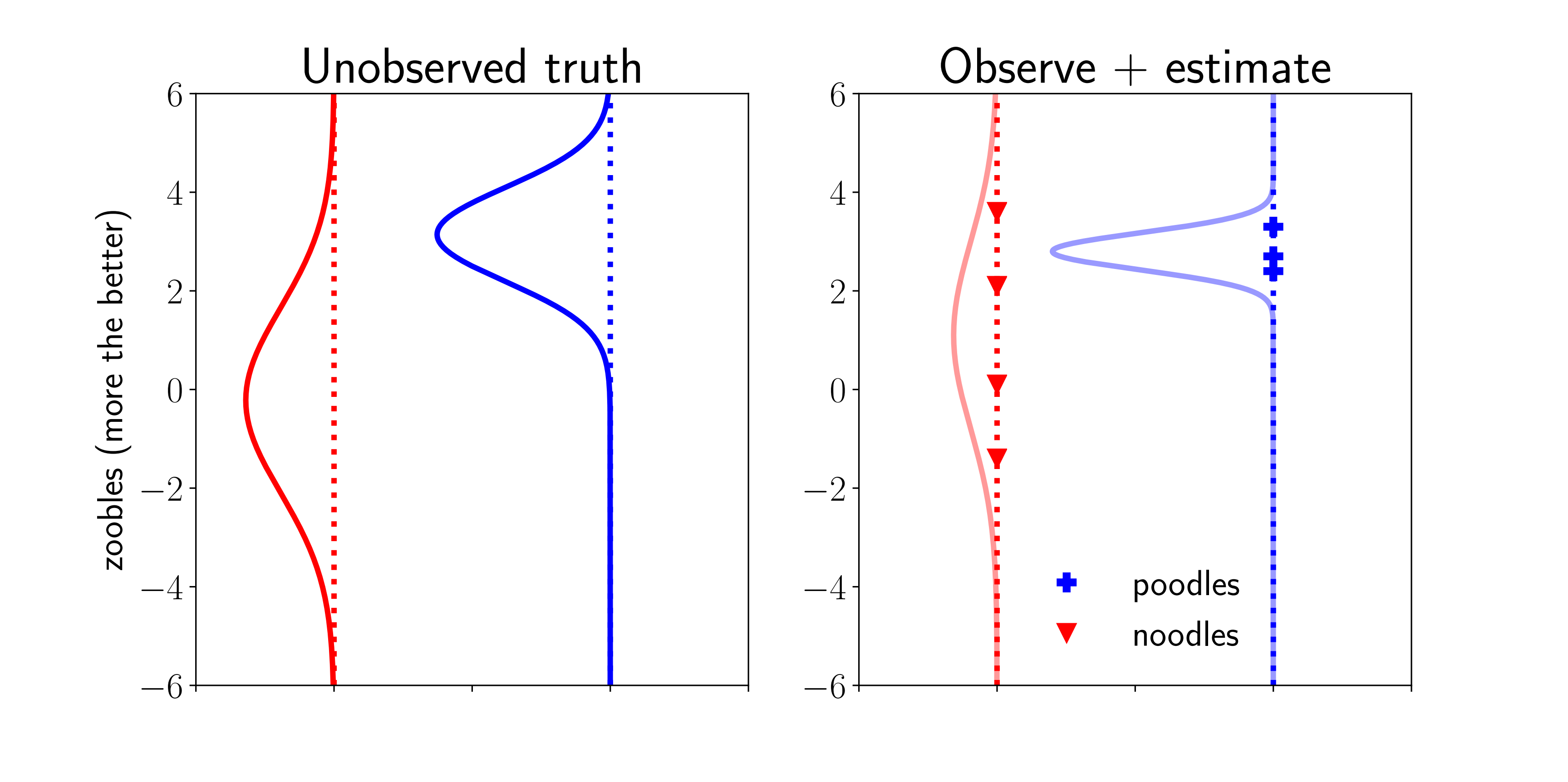
- Examples:
  - Uncertainty sampling
    (sample near the decision boundary,
    or maximal variance)
  - Committee of classifiers
    (where they disagree)



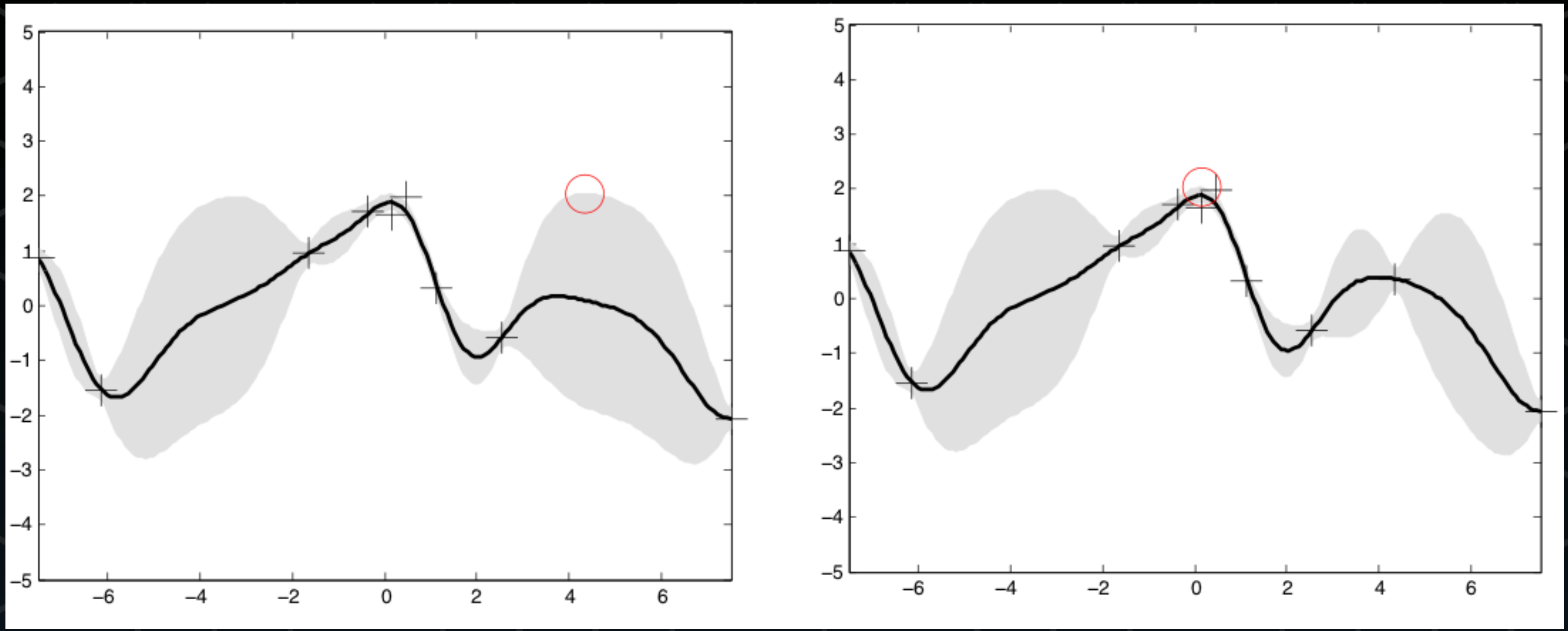
Tran, Ong, Wolf, Combining active learning suggestions, PeerJ, 2018

B – Bandits / Bayesian Optimisation

Want to maximise the outcome of different choices



B – Bandits / Bayesian Optimisation



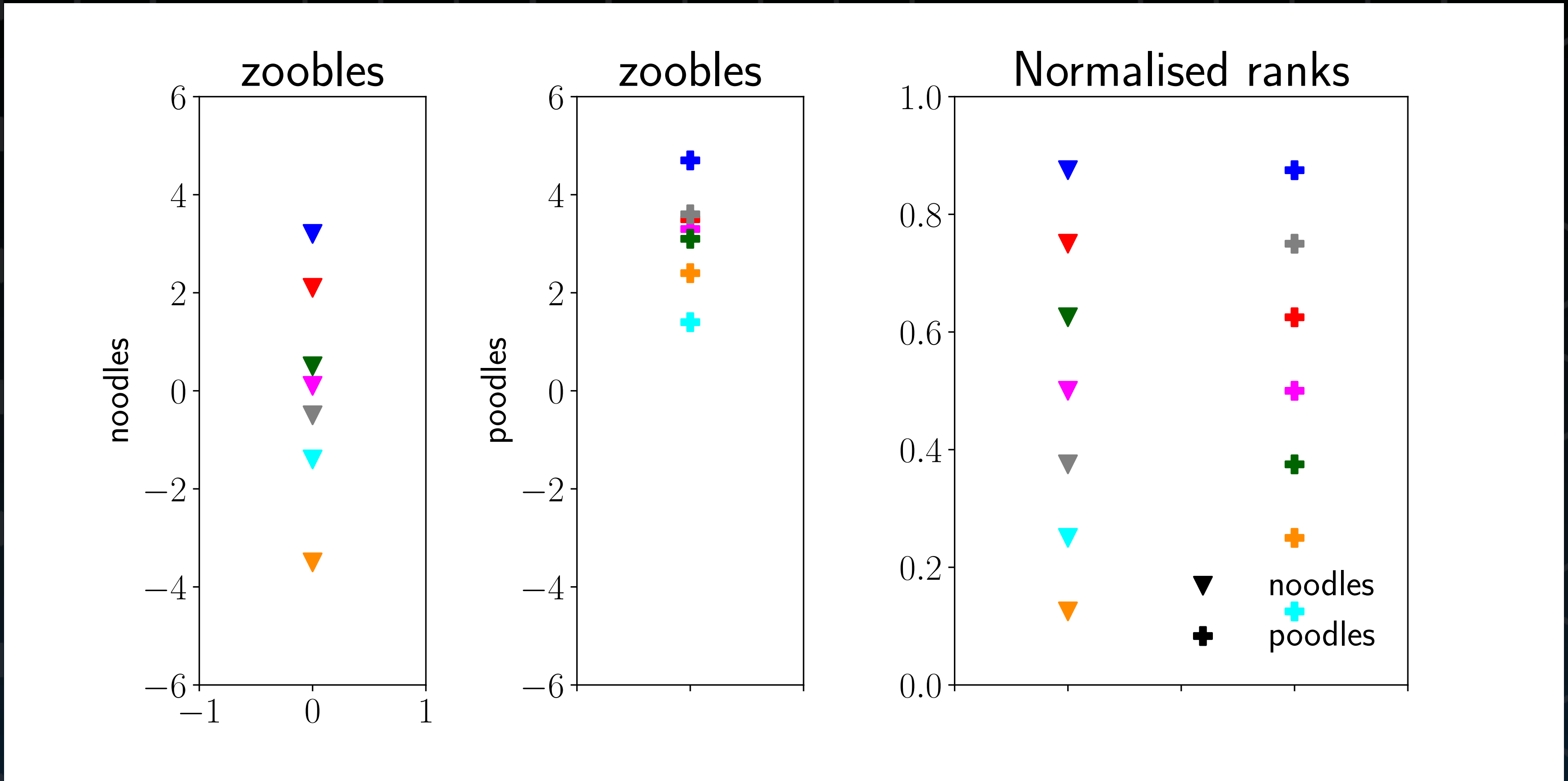
- Select a choice from a set of actions. Maximise reward/payoff from each action
- Simple algorithms with theoretical guarantees
- Manage uncertainty with repeated sampling

Krause, Ong, Contextual Gaussian Process Bandit Optimization, NIPS 2011

CSIRO DATA 61

C – Choice Theory

Want to integrate different sources of information



C – Choice Theory

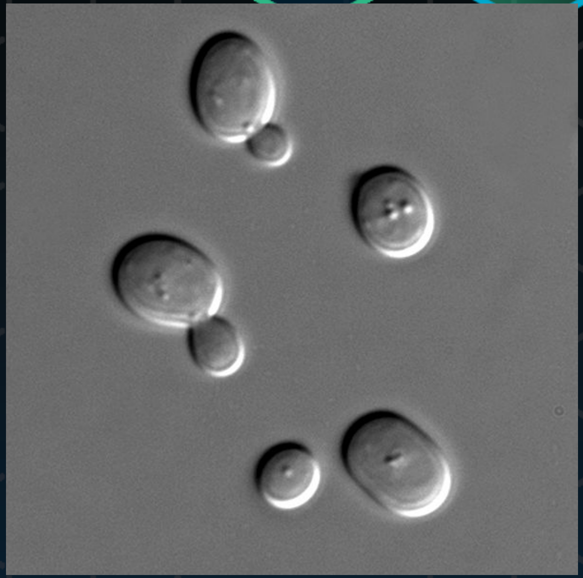
- **Main idea**
  - Aggregate set of ranks into one ordering (combine predictions)
  - Economics and social science, impossiblity theorems
- **Equivalent representation of ranks**
  - Ordered list of n objects selected from Ω
  - List of values [1, . . . , n] (can be normalised to ∈ (0, 1))
  - Permutation mapping $R : \Omega \rightarrow (0, 1)$
- Combine by using the geometric mean

Justin Bedő, Cheng Soon Ong, Multivariate Spearman's rho for Aggregating Ranks Using Copulas, *JMLR 2016*

DE - Design of Experiments

- **Glucose metabolism in Yeast**
  - Multiple possible models
  - Design biological experiments that maximise information gain

Busetto, Hauser, Krummenacher, Sunnåker, Dimopoulos, Ong, Stelling and Buhmann.
Near-optimal experimental design for model selection in systems biology , Bioinformatics 2013



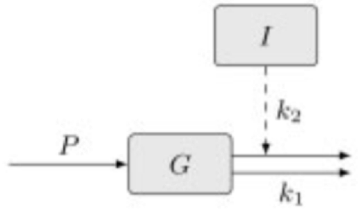
What is a model?

<table>
<tr><th>No.</th><th>Model</th><th>ODE and parameters</th></tr>
<tr><td>I</td><td>
</td><td>

$\dot{G} = \theta_1 G + \theta_2$

$\theta_1 = -k_1 \quad = (-1.09 \pm 0.11) \cdot 10^{-1}$

$\theta_2 = P \quad = 8.5 \pm 1.10$

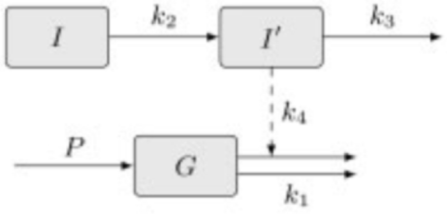
</td></tr>
<tr><td>IV</td><td>
</td><td>

$\dot{G} = \theta_1 G + \theta_2 I + \theta_3$

$\theta_1 = -k_1 \quad = (-1.44 \pm 0.35) \cdot 10^{-1}$

$\theta_2 = -k_2 \quad = (9.15 \pm 4.0) \cdot 10^{-2}$

$\theta_3 = P \quad = 11.3 \pm 3.1$

</td></tr>
<tr><td>V</td><td>
</td><td>

$\dot{G} = \theta_1 G + \theta_2 X + \theta_3$

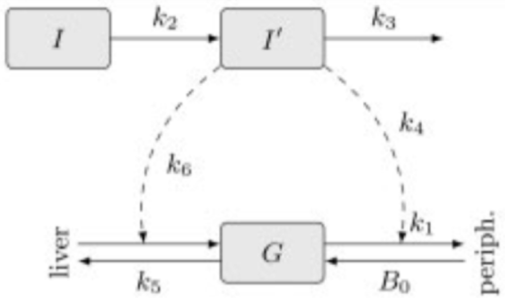
$\dot{X} = \theta_4 X + I$

$\theta_1 = -k_1 \quad = (6.50 \pm 0.73) \cdot 10^{-1}$

$\theta_2 = -k_2 k_4 \quad = (-9.10 \pm 1.73) \cdot 10^{-3}$

$\theta_3 = P \quad = 5.97 \pm 0.70$

$\theta_4 = -k_3 \quad = (-1.01 \pm 0.16) \cdot 10^{-1}$

</td></tr>
<tr><td>VI</td><td>
</td><td>

$\dot{G} = (\theta_1 - X)G + \theta_4$

$\dot{X} = \theta_2 X + \theta_3 I$

$X = I'/k_2$

$\theta_1 = -(k_1 + k_5) = (-4.90 \pm 0.97) \cdot 10^{-2}$

$\theta_2 = -k_3 \quad = (-9.10 \pm 1.20) \cdot 10^{-2}$

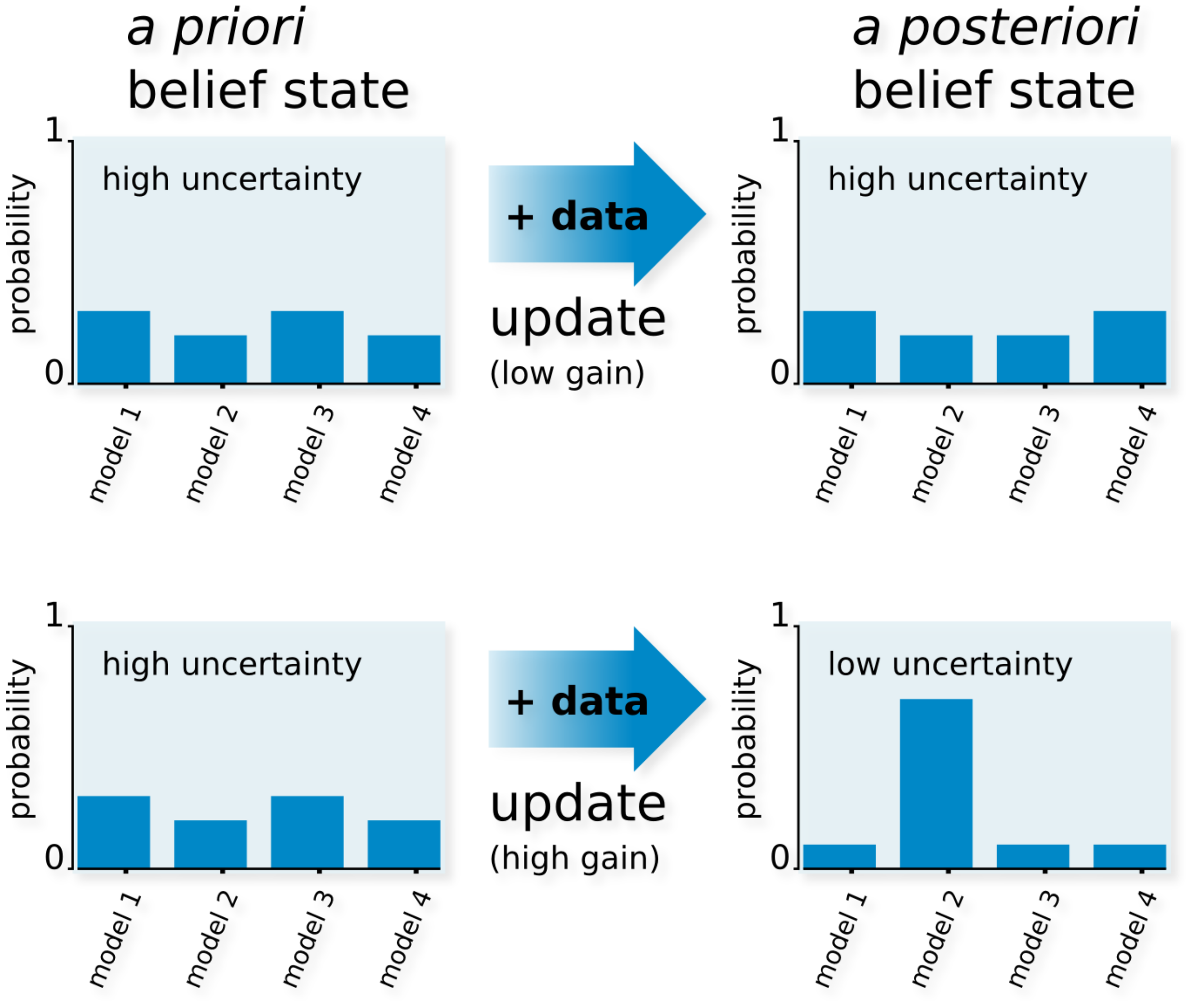
$\theta_3 = k_2(k_4 +_5) \quad = (8.96 \pm 1.88) \cdot 10^{-5}$

$\theta_4 = B_0 \quad = 4.42 \pm 0.74$

</td></tr>
</table>

Bergman insulin dependent glucose metabolism model.

Finding good models



Optimised experimental design (I)

- **Measurements**
  - Experiments produce readouts $y(t_i)$, grouped into datasets Y_π for an experiment π.

- **Bayes rule**
  - For a particular model f, (taking care of parameters)
    - $p(f|Y_\pi) = \frac{p(Y_\pi|f)p(f)}{p(Y_\pi)}$

Optimised experimental design (II)

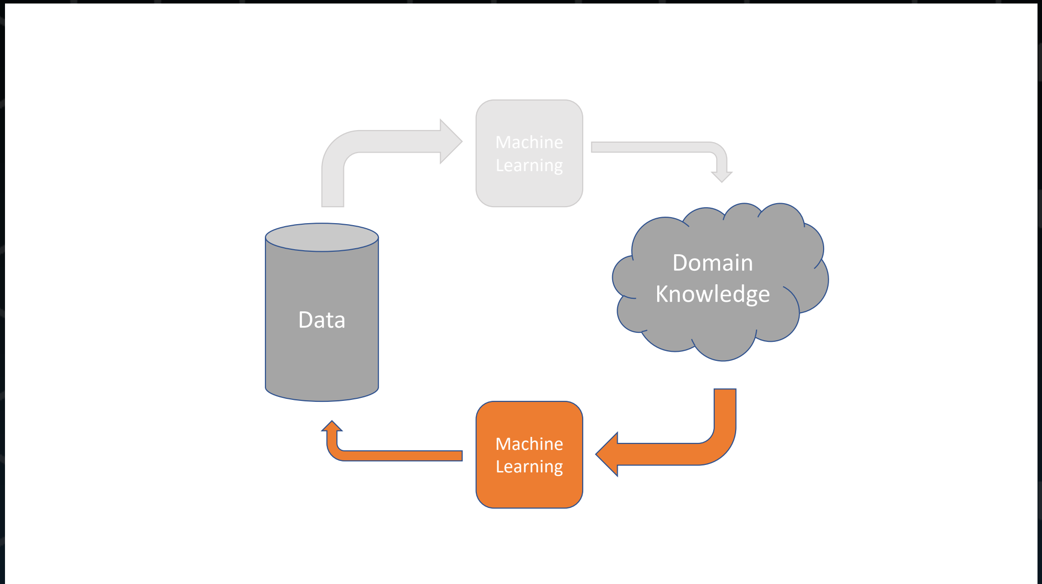
- **Information gain**
- We want to take measurements that change model probabilities
- $D_{KL}(p(f|Y_\pi)||p(f)) = \sum_f p(f|Y_\pi) \log p(f|Y_\pi)/p(f)$

- **Marginalise over possible outcomes**
- Maximise expected information gain (tough computational problem)
  - $argmax_\pi \mathbb{E}_{Y_\pi} D_{KL}(p(f|Y_\pi)||p(f))$

Scientific discovery with machine learning

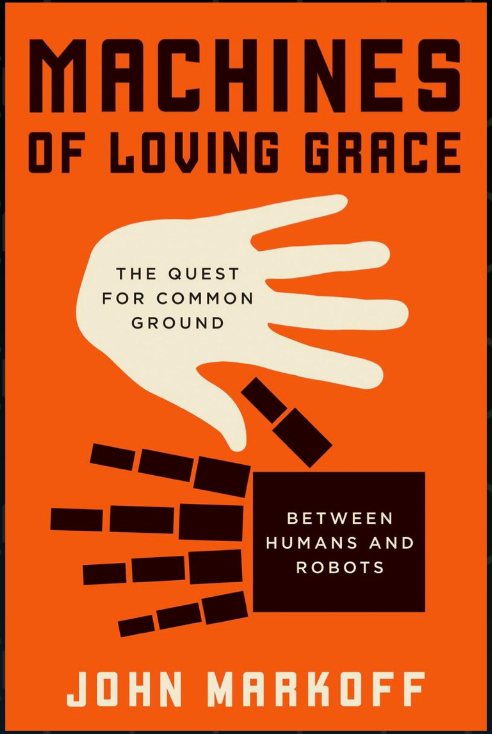
- How can we use prediction to help human experts perform discovery?

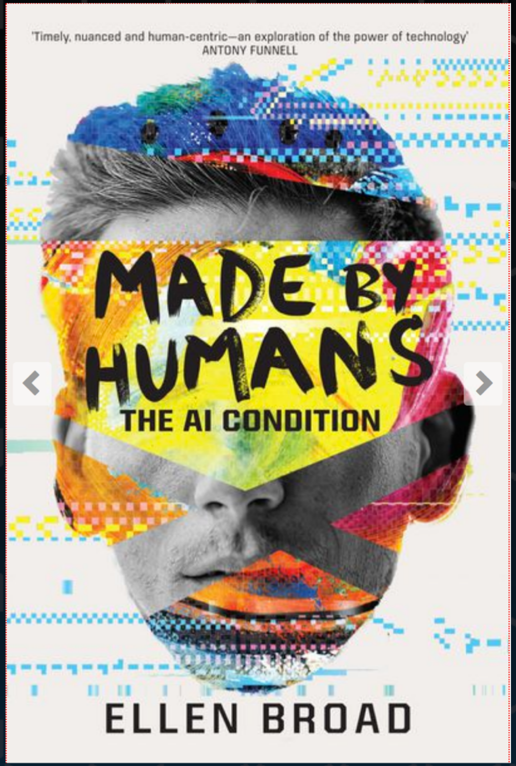
$$f_w(x) : \mathcal{X} \to \mathcal{Y}$$

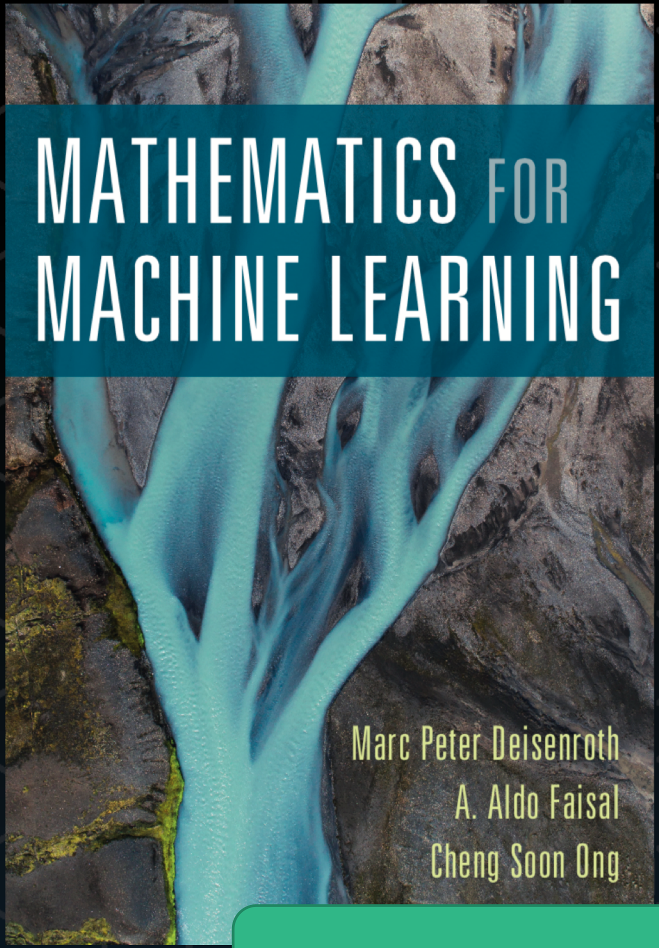


- Domain knowledge to Data
  - Human in the loop ML
  - Where to measure

- Use predictor to decide where to measure (ABCDE)
  - (A) Active Learning
  - (B) Bandits / Bayesian Optimisation
  - (C) Choice Theory
  - (DE) Design of Experiments







mml-book.com

The answers you get depend on the questions you ask.
– Thomas S. Kuhn

THANK YOU

Data61 CSIRO

Cheng Soon Ong

e cheng-soon.ong@data61.csiro.au
w ong-home.my

www.data61.csiro.au