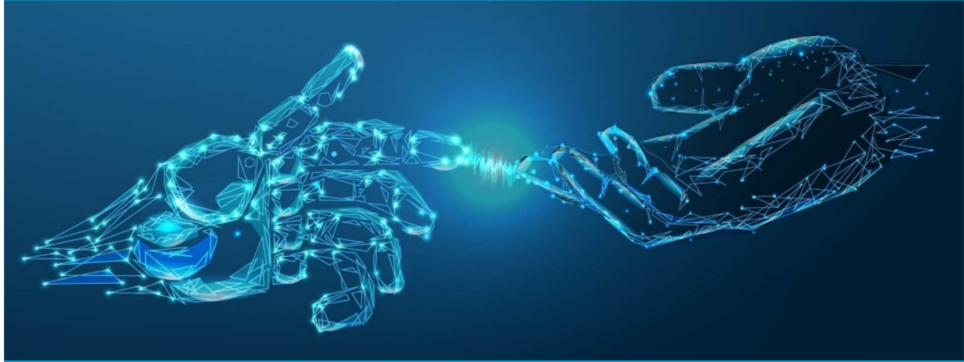
ABCDE: What to do with a predictor?



Cheng Soon Ong | 15 July 2020

Presented to The Adecco Group
Data & AI Conference

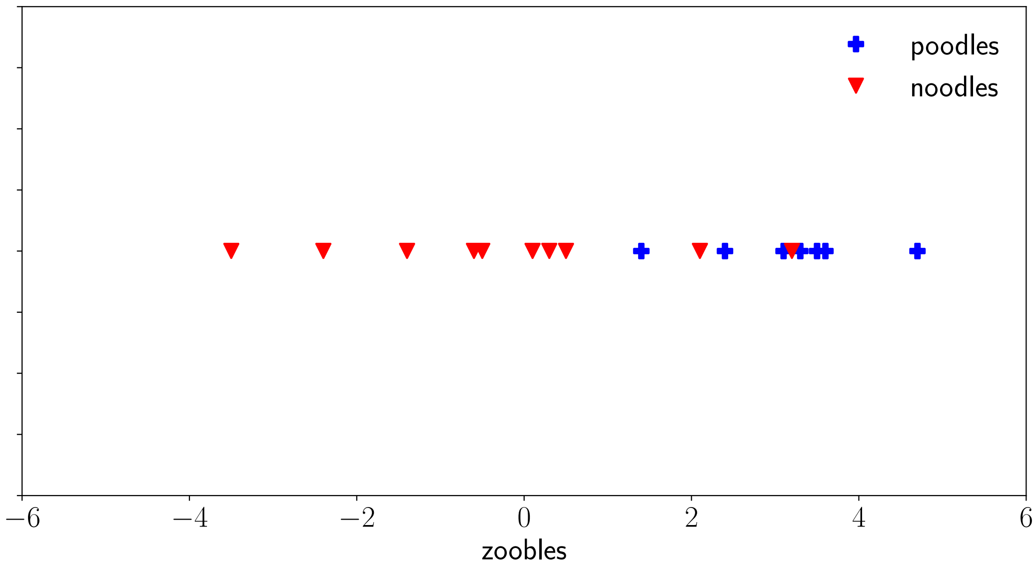
e cheng-soon.ong@data61.csiro.au
w ong-home.my

https://research.csiro.au/mlai-fsp/

Australia's National Science Agency

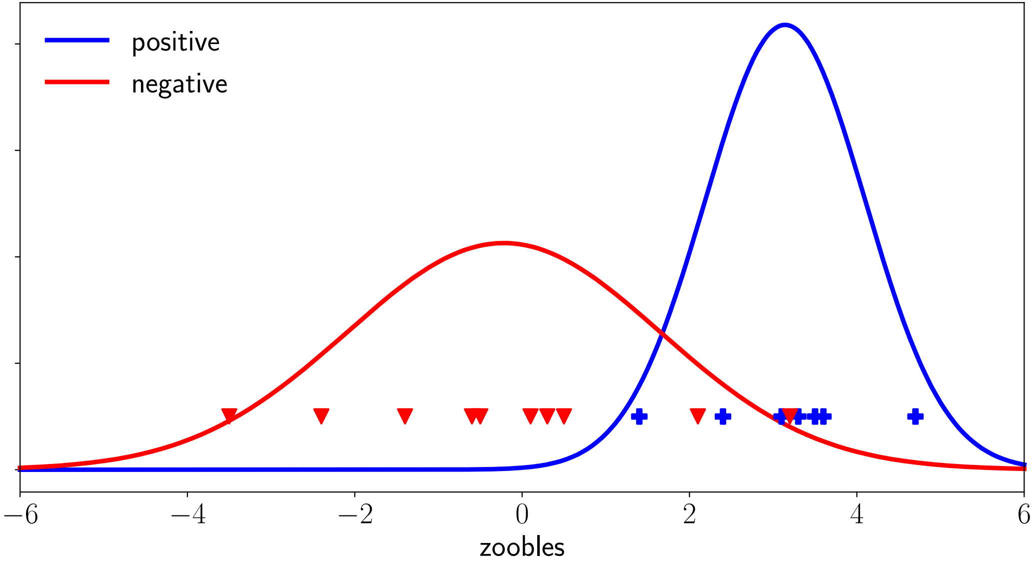
Given some data

Classify blue plus vs red triangles, based on features



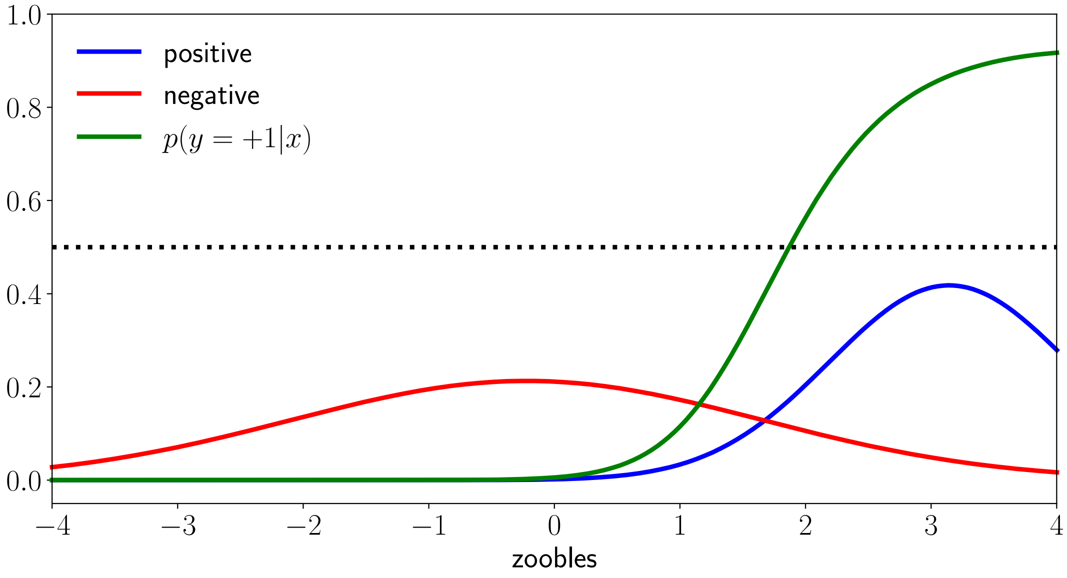
Fit a model to data

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

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

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

Who are we?

Commonwealth Scientific and Industrial Research Organisation, Australia

Our research and development

We are one of the largest and most diverse scientific research organisations in the world.
Our research focuses on providing solutions in nine core areas.

Key areas of research

Animals and plants
Astronomy and space
Climate
Environment
Farming and food production
Health
Information technology
Mining and manufacturing
Renewables and energy

Australia's innovation catalyst

Nurturing and enabling the national innovation and commercialisation ecosystem

697
Patent families

497
Active licences

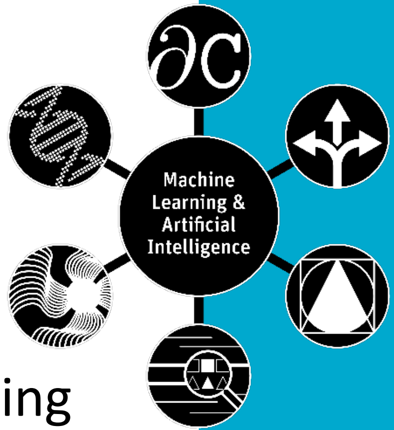
$1B+
Total market capitalisation of portfolio companies

170+
Start-up companies from CSIRO science and technology

2,400 partners
Turning science into solutions with industry, government and research collaborators

150k school students
Delivering STEM education programs to equip Australia's future workforce

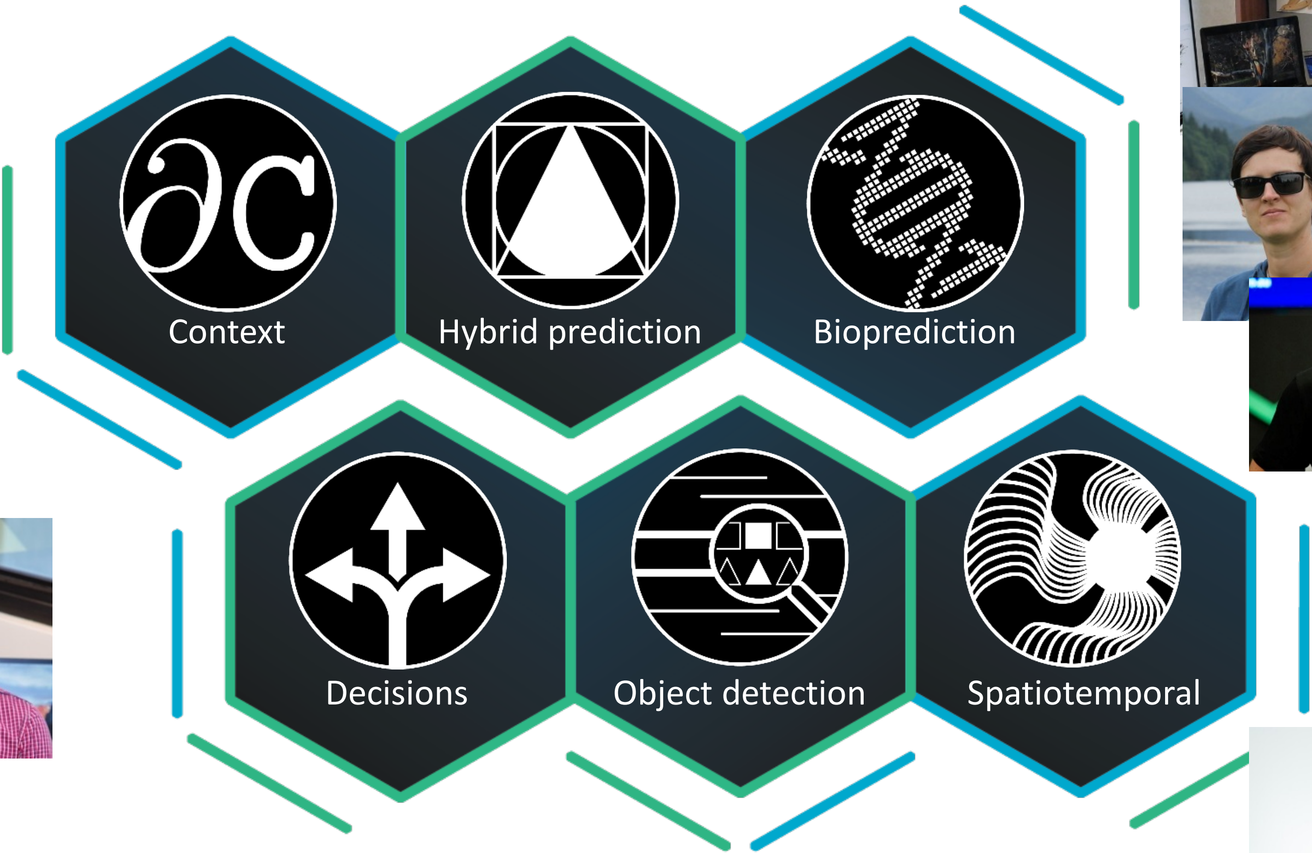
MLAI
Future Science Platform

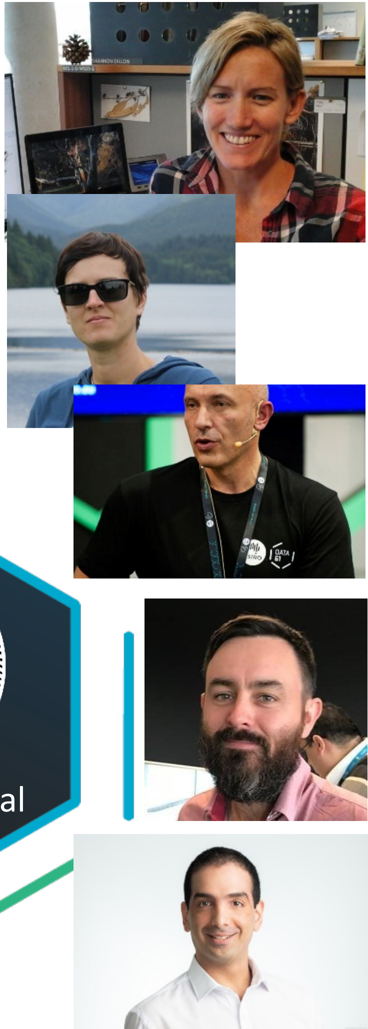


- science
  Demonstrate machine learning for scientific discovery
- people
  Lead a network of machine learning and science experts (create critical mass in Australia)
- technology
  Create languages or systems to specify machine learning problems

https://research.csiro.au/mlai-fsp/

MLAI FSP Activities





What to do with a predictor?

What is Machine Learning?

Assume we have managed to train a sensible predictor

- 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)

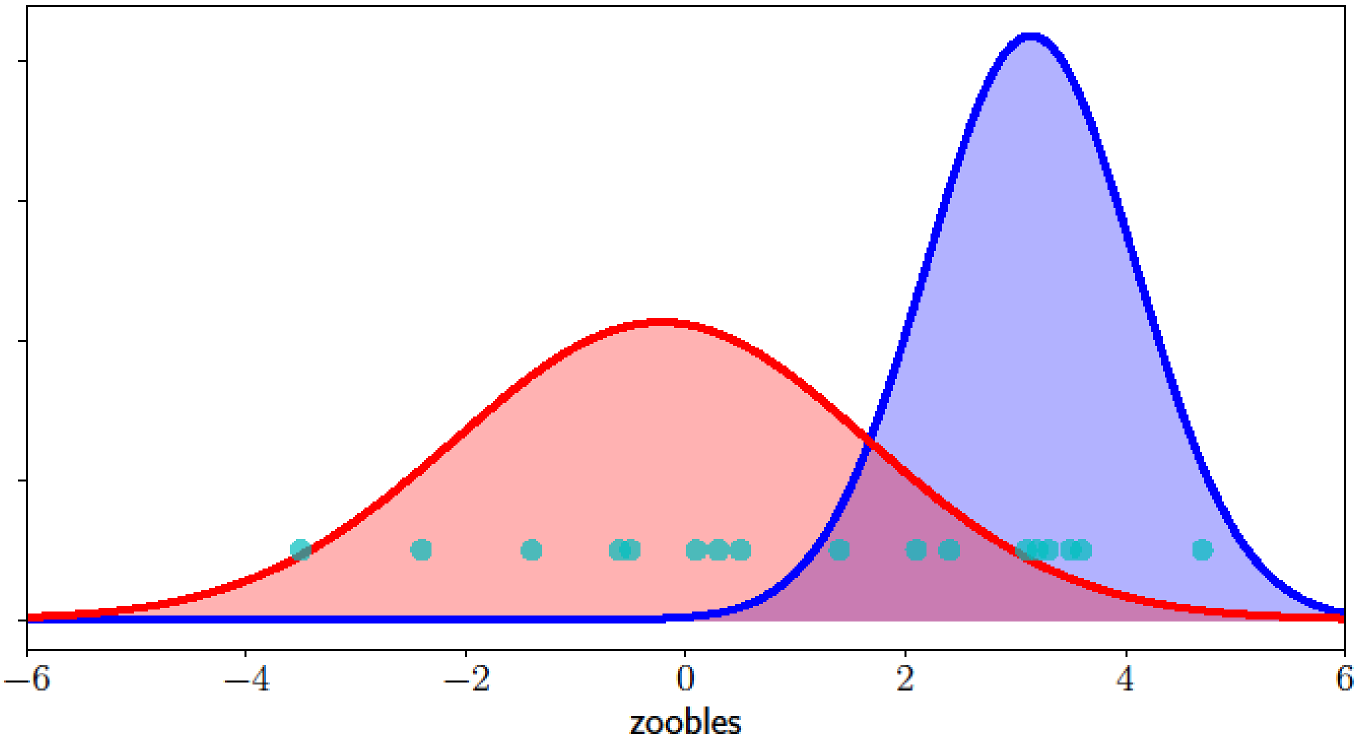
$$f_w(\boldsymbol{x}) : \mathcal{X} \to \mathcal{Y}$$

- 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

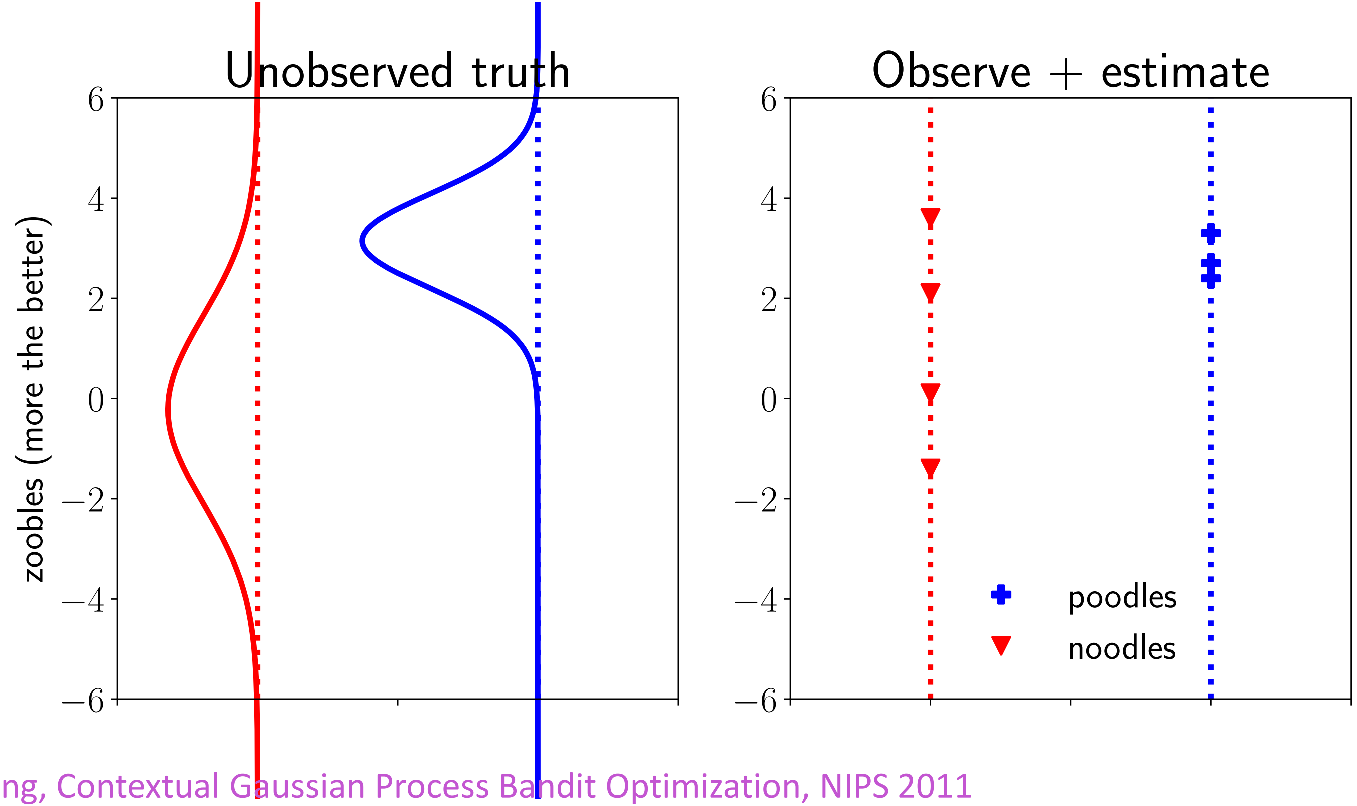
Want to build a classifier without paying for a lot of labels



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

B – Bandits / Bayesian Optimisation

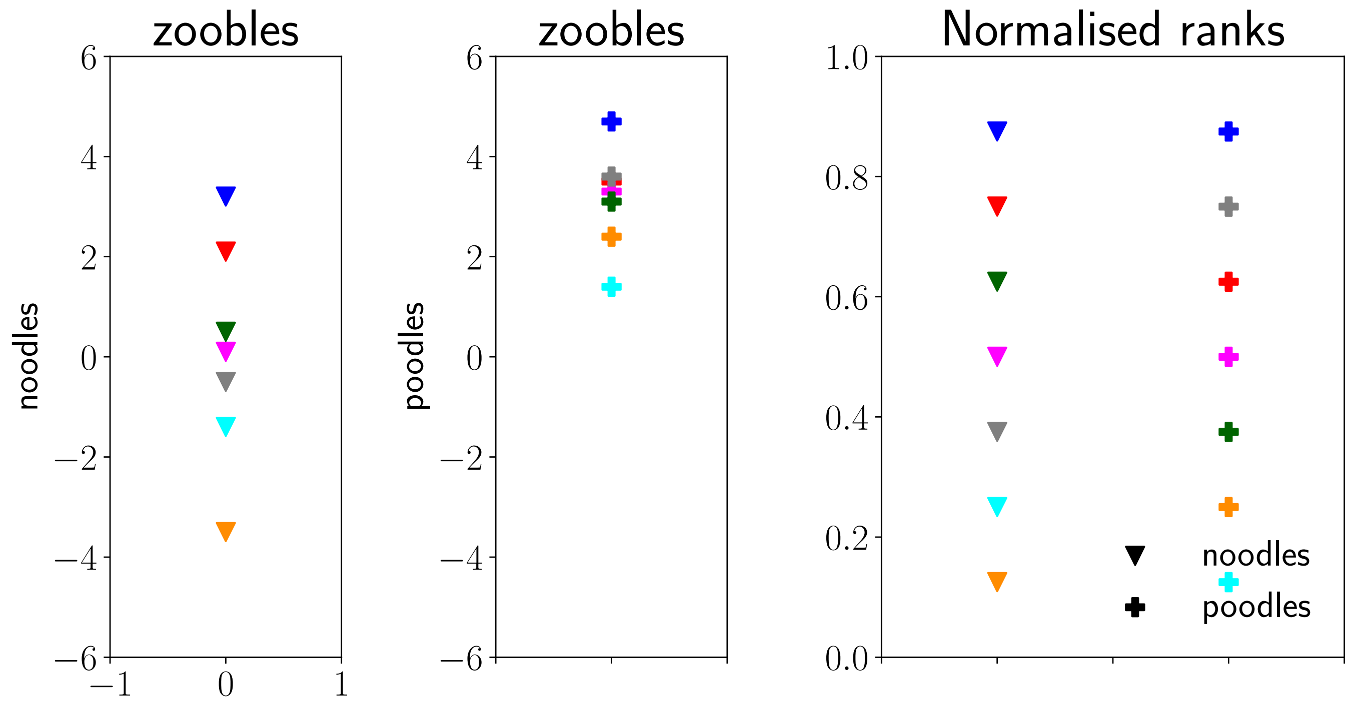
Want to maximise the outcome of different choices



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

C – Choice Theory

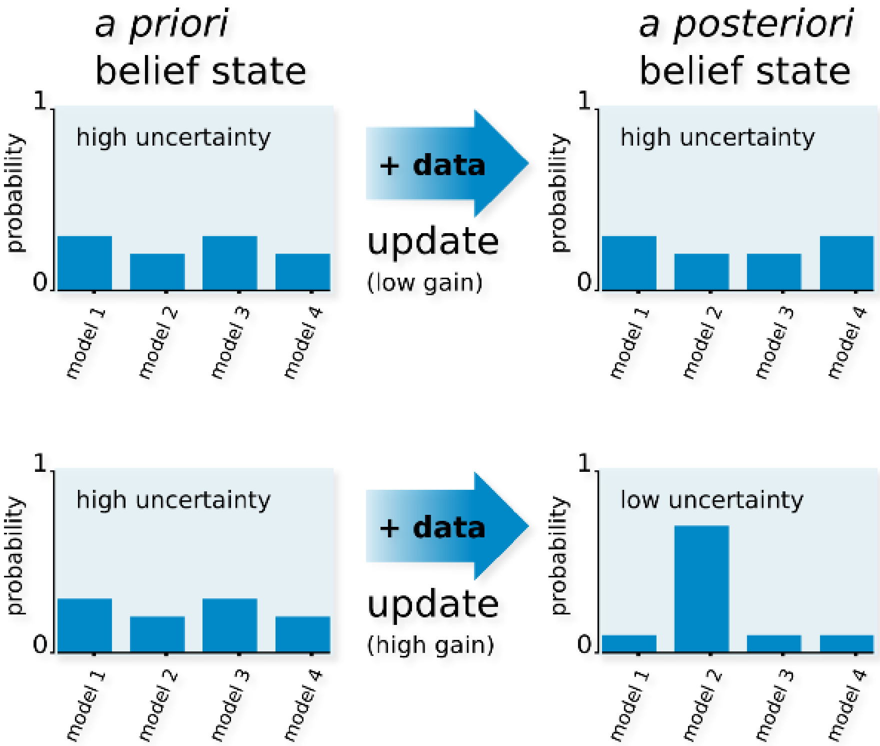
Want to integrate different sources of information



Bedo, Ong, Multivariate Spearman's rho for Aggregating Ranks Using Copulas, JMLR 2016

DE – Design of Experiments

Find good models by maximizing information gain

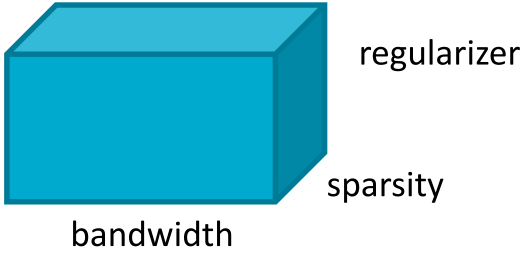


Busetto, et. al. Near-optimal experimental design for model selection in systems biology, Bioinformatics 2013

ABCDE: what are we sorting?

A conceptual view of adaptive sampling

- Consider the set of all possible things to measure



- Think of the predictor output as a generator of features
  - Each generated features demonstrates the “importance” of a sample
  - Can get multiple features by a committee or ensemble of predictors

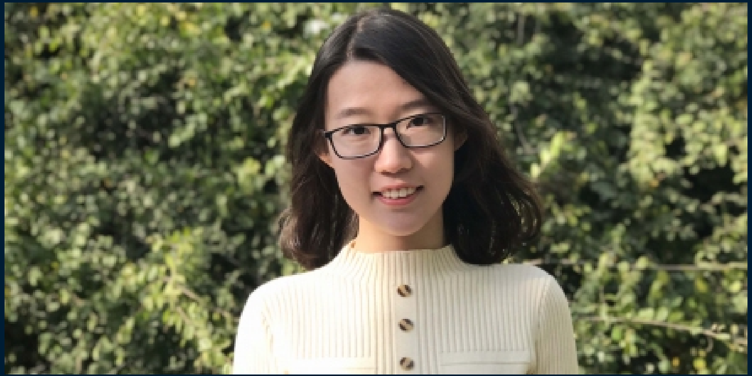
- Adaptively choose the next thing to measure by maximising an objective
  (machine learning is about defining good objective functions)

ABCDE – what are we sorting?

Illustration of the conceptual idea

- A – Active Learning
  1. Predictor generates a confidence that thing is positive
  2. Objective is to find the location where probability = 0.5
- B – Bandits / Bayesian Optimization
  1. Predictor generates a model of the reward
  2. Objective combines the summary statistic and uncertainty
- C – Choice Theory
  1. Predictor transforms scores into a comparable scale
  2. Objective maximises a multivariate copula score
- DE – Design of Experiments
  1. Predictor estimates the expectation over future experiments
  2. Objective identifies the notion of information gain

1. Predictor generates features
2. Define an objective function



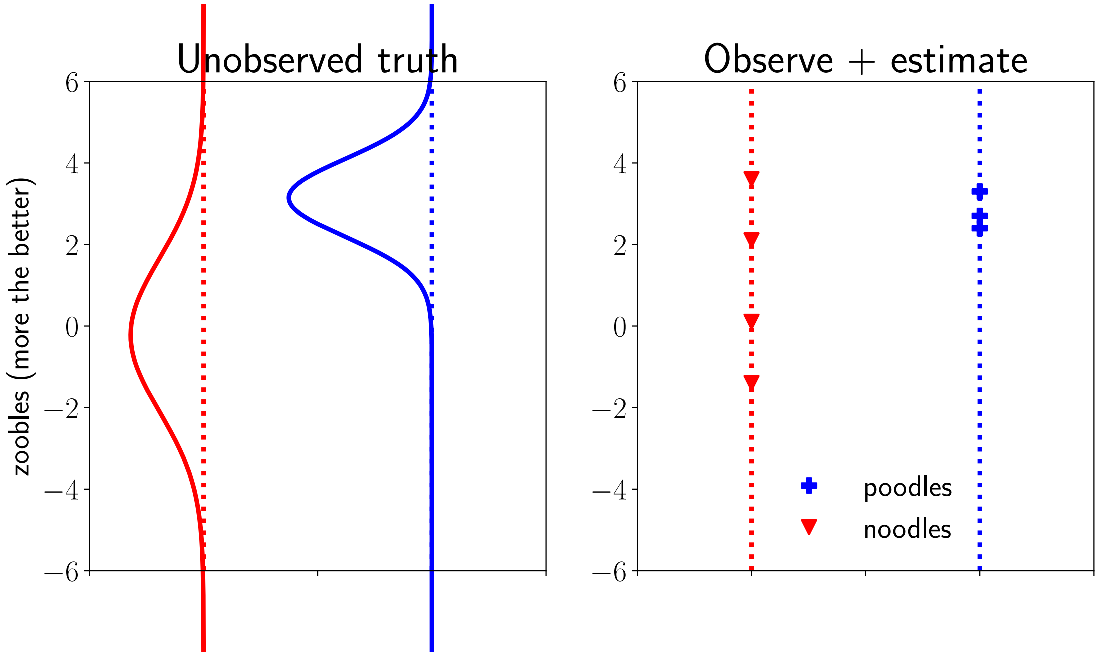
Mengyan Zhang, PhD candidate
Australian National University

Quantile Bandits

For the technical people in the audience…

B – Bandits / Bayesian Optimisation

Want to maximise the outcome of different choices

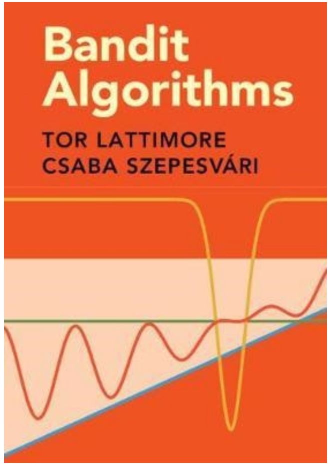


Anatomy of a Bandit Algorithm

Several design choices

- Given a set of arms, at each round:
  - Choose an arm (and get a reward)
    - depending on the task at hand

  - Estimate the distribution of the arm
    - Assumption needed for theoretical analysis
    - Usually skipped in the algorithm
  - Define a summary statistic for each distribution
    - usually the mean for risk neutral policy
      (Von Neumann-Morgenstern utility theory)



Anatomy of a Bandit Algorithm

Let's change from mean to quantile

- Given a set of arms, at each round:
  - Choose an arm (and get a reward)
    - depending on the task at hand
  - Estimate the distribution of the arm
    - Assumption needed for theoretical analysis
    - Usually skipped in the algorithm
  - Define a summary statistic for each distribution
    - usually the mean for risk neutral policy (Von Neumann-Morgenstern utility theory)
    - What if we are risk averse?

Replace mean with quantiles

Concentration of measure

Bound the distance from the empirical to the true quantiles

Theorem 2 (Two-side Concentration Inequality for Quantiles). *Denote the lower bound of hazard rate as L, the number of samples as n. For all quantile level $\tau \in (0,1)$, let $\widetilde{\tau} = \tau + 1/n$, with $\widetilde{k} = n(1-\widetilde{\tau})$, define $v = \frac{2}{\widetilde{k}L^2}$, $c = \frac{2}{\widetilde{k}L}$. Under Assumption 1, for $n > \frac{1}{1-\tau}$ and $\gamma > 0$, we have*

$$\mathbb{P}\left(\boxed{|\hat{Q}_n^\tau - Q^\tau|} \geq \sqrt{2v\gamma} + c\gamma\right) \leq 2\exp(-\gamma) \tag{6}$$

- Estimate the distribution of the arm
  - Assumption needed for theoretical analysis
  - Usually skipped in the algorithm
- Define a summary statistic for each distribution
  - usually the mean for risk neutral policy (Von Neumann-Morgenstern utility theory)
  - What if we are risk averse?

Assumption 1

Non-decreasing hazard rate

Lower bound of hazard rate L > 0

What do we want to optimize?

Machine learning is about defining objective functions

- Given a set of arms, at each round:
  - Choose an arm
    - depending on the task at hand
  - Estimate the distribution of the arm
    - Needed for theoretical analysis
    - Usually skipped in the algorithm
  - Define a summary statistic for each distribution
    - usually the mean for risk neutral policy (Von Neumann-Morgenstern utility theory)
    - What if we are risk averse?

Replace mean with quantiles

Objective Function
- Regret minimization
- Best Arm Identification
  - Fixed Budget
  - Fixed Confidence

Summary

A story with 3 levels …

- Machine learning is about prediction.
  - We can use predictions to help us make decisions
  - CSIRO is using ML and AI to reimagine scientific discovery

- ABCDE: What to do with a predictor?
  - (A) Active Learning
  - (B) Bandits / Bayesian Optimisation
  - (C) Choice Theory
  - (DE) Design of Experiments

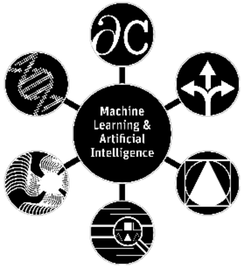
- Technical: For risk aware bandits, we can replace means with quantiles

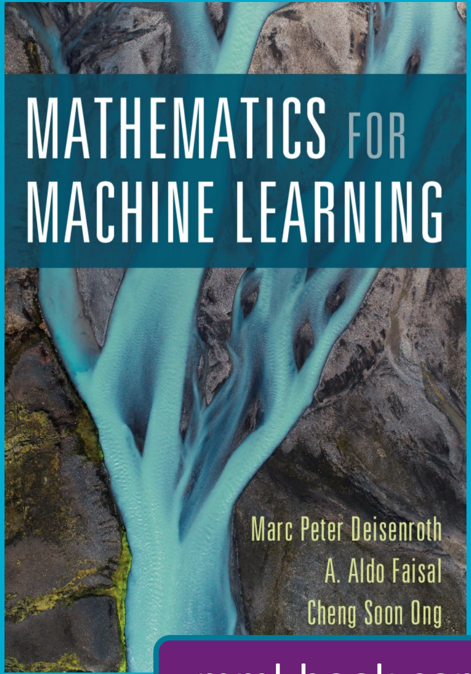
Thank you!

Cheng Soon Ong

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

https://research.csiro.au/mlai-fsp/





mml-book.com

Australia's National Science Agency