



MLAI FSP

Imagine a new scientific method

Cheng Soon Ong

7 May 2019 - Data Driven Science from Earth to Space, C3DIS, Canberra

www.data61.csiro.au

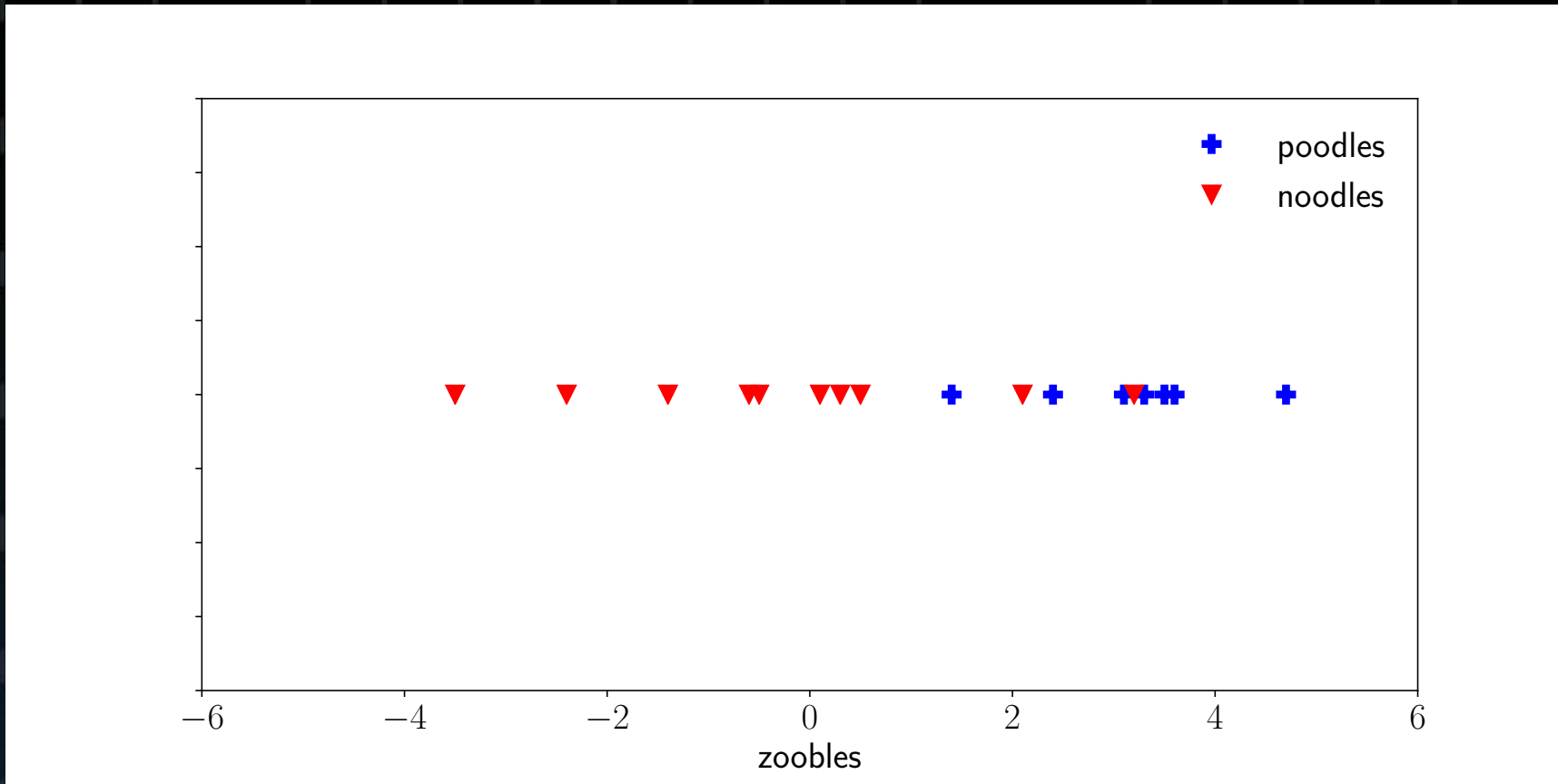
MLAI FSP - context

- *the FSP aims to develop capacity and platforms and position CSIRO at the forefront of key AI/ML areas so we can create new industries and support the transition of the Australian economy*
- *The ability to undertake science at scale, on real world problems, using new AI/ML algorithms, is of significant interest to all major research institutions. This proposal demonstrates the range of interest from research providers and industry partners.*



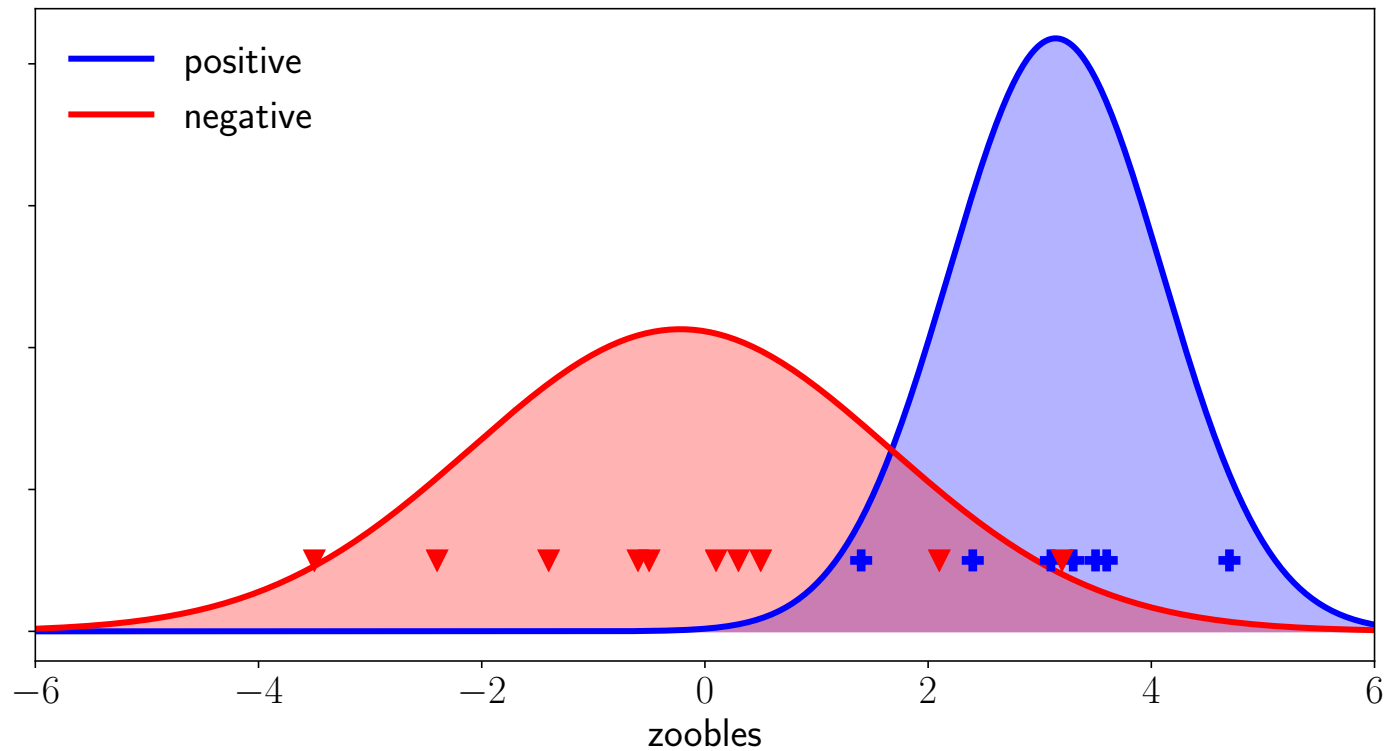
Given some data

Classify blue plus vs red triangles, based on features



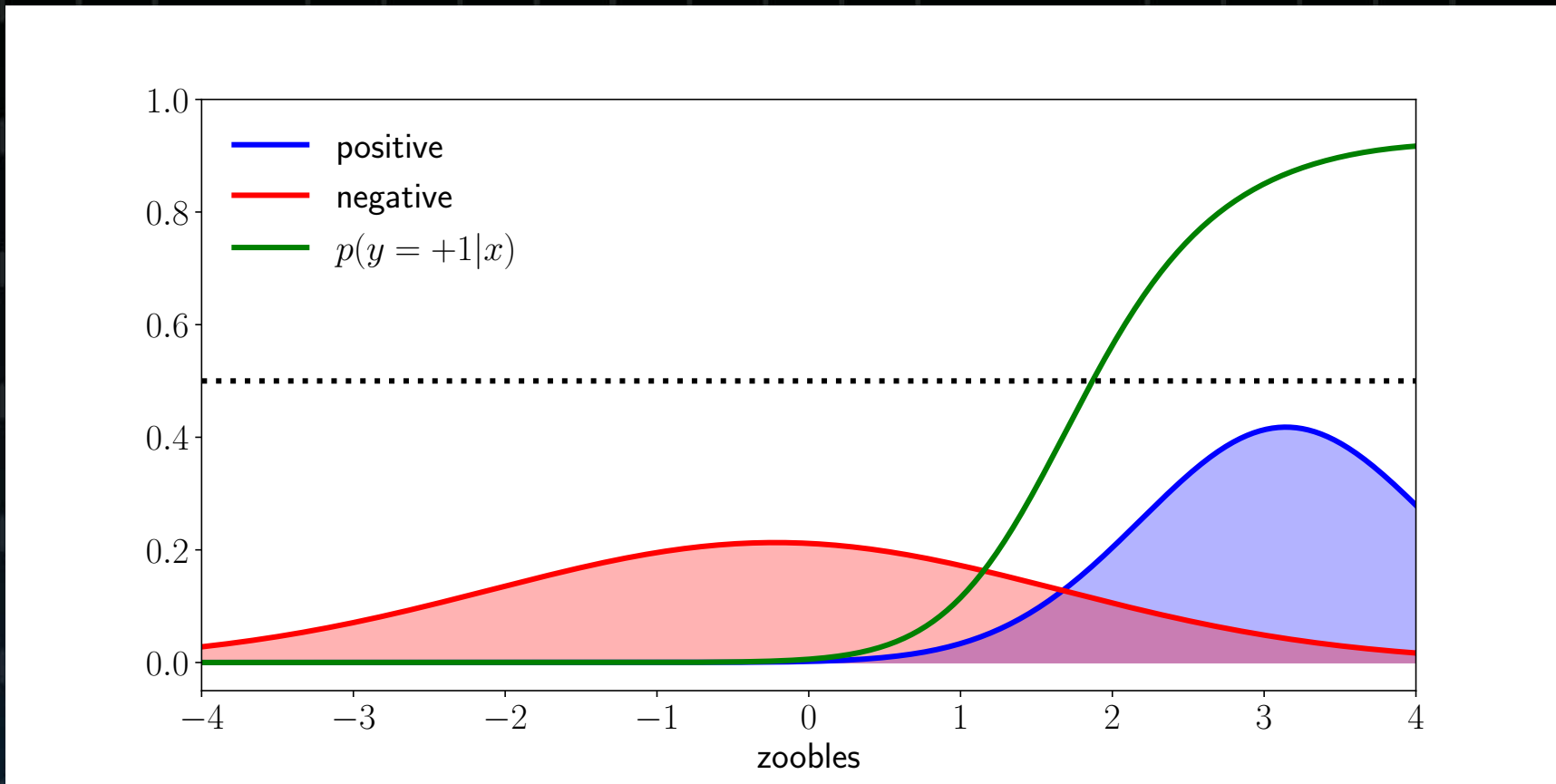
Fit a model to data

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

$$\mathbf{x}_1, \dots, \mathbf{x}_n \sim \mathcal{X}$$
$$\mathbf{y}_1, \dots, \mathbf{y}_n \sim \mathcal{Y}$$

Predictor

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

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

MLAI FSP - context

- *the FSP aims to develop capacity and platforms and position **CSIRO at the forefront of key AI/ML areas** so we can create new industries and support the transition of the Australian economy*
- *The ability to **undertake science at scale**, on real world problems, using new AI/ML algorithms, is of significant interest to all major research institutions. This proposal demonstrates the range of interest from research providers and industry partners.*



MLAI FSP - Goals

- **science**
The FSP is an investment that will deliver lasting impact to areas of strategic interest across CSIRO by exploiting and advancing Machine Learning and Artificial Intelligence (MLAI) research.
- **people**
The platform we create, and the people we train will become a capability that fundamentally changes the way CSIRO undertakes core research challenges.
- **technology**
We will deliver new MLAI solutions to age-old problems, novel solutions for new data, and platforms for emerging challenges in a data driven world.

MLAI FSP - Timeline

- May-June 2019: Internal CSIRO workshops to scope activities
 - Activities are a collection of projects
 - Aligns to pillars of FSP proposal
 - Potential domain impact
 - Open machine learning problem
- Leadership team:
Simon Barry, Ben Trevaskis, Tara Martin, Peter Fitch, Brent Henderson, Cheng Soon Ong



What is Machine Learning?

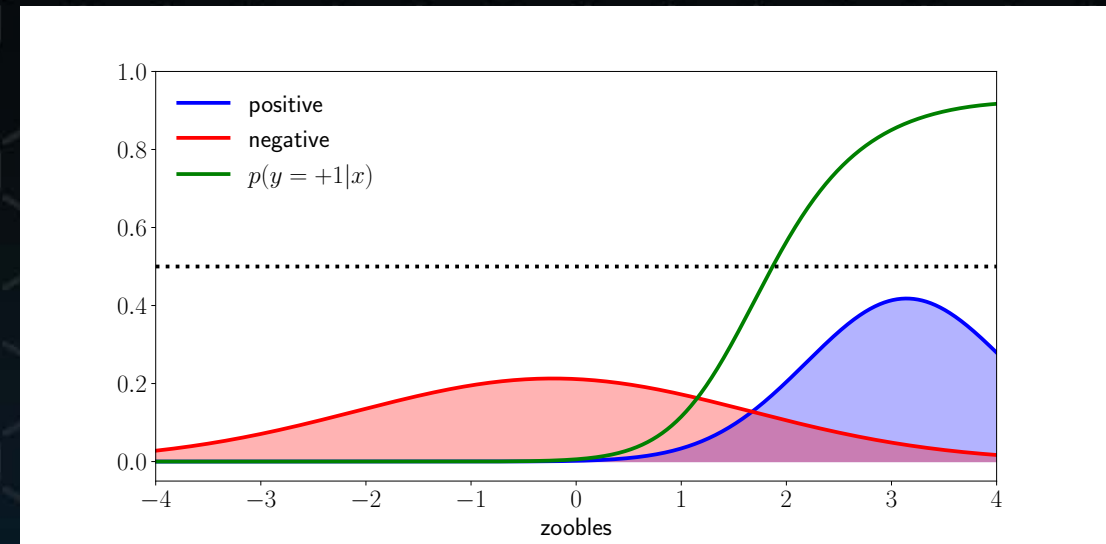
- Machine Learning is about prediction
 - Examples/covariates/features
 - Labels/annotations/target variable

Predictor

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

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

$$\mathbf{x}_1, \dots, \mathbf{x}_n \sim \mathcal{X}$$
$$\mathbf{y}_1, \dots, \mathbf{y}_n \sim \mathcal{Y}$$



Prediction \neq understanding \neq taking action

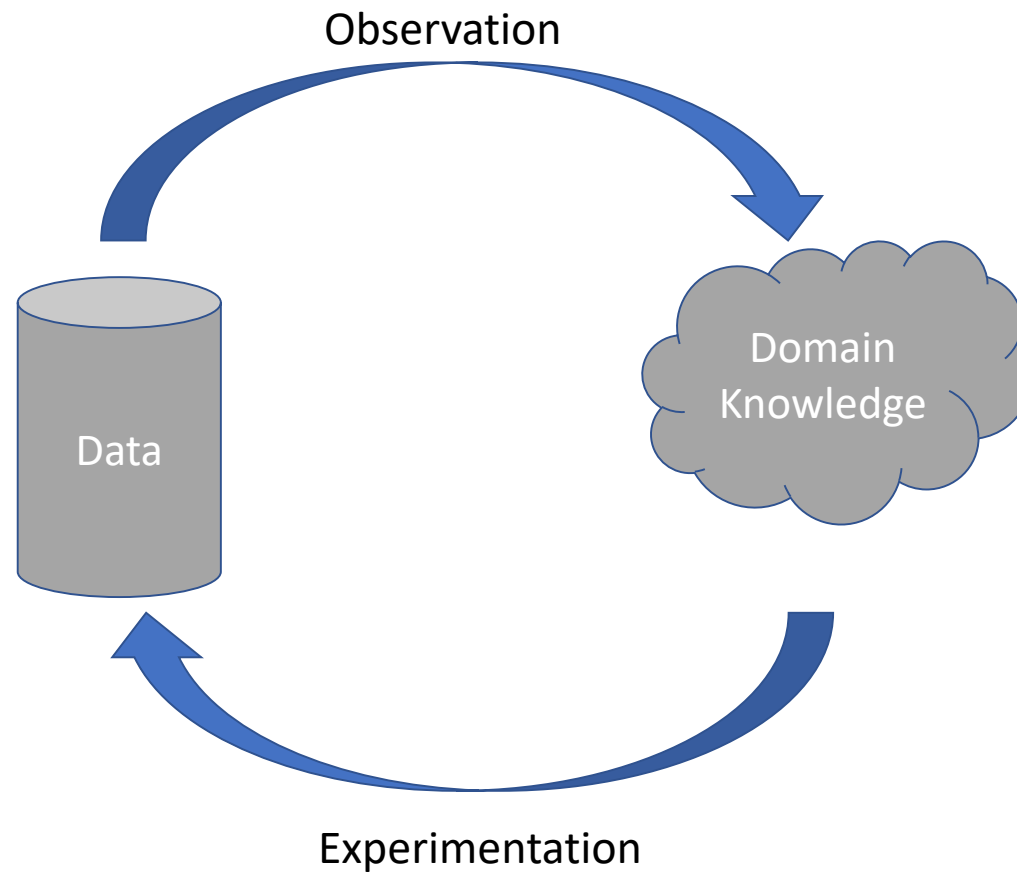
- How can we use prediction to help with scientific discovery?

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

What is scientific discovery?

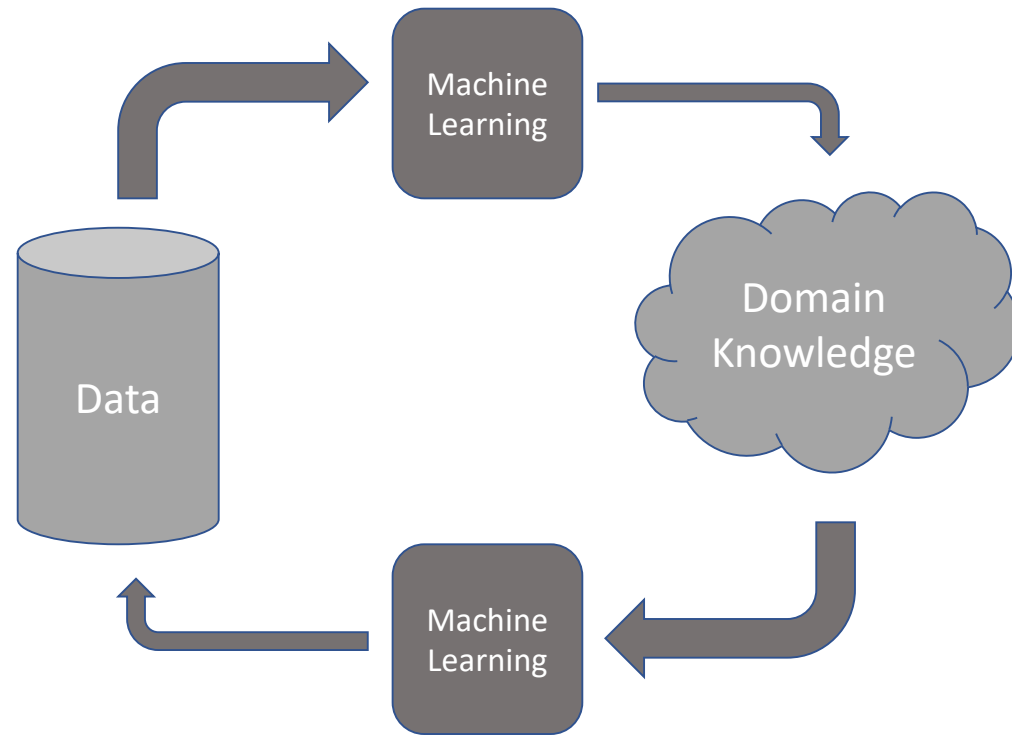


What is scientific discovery?

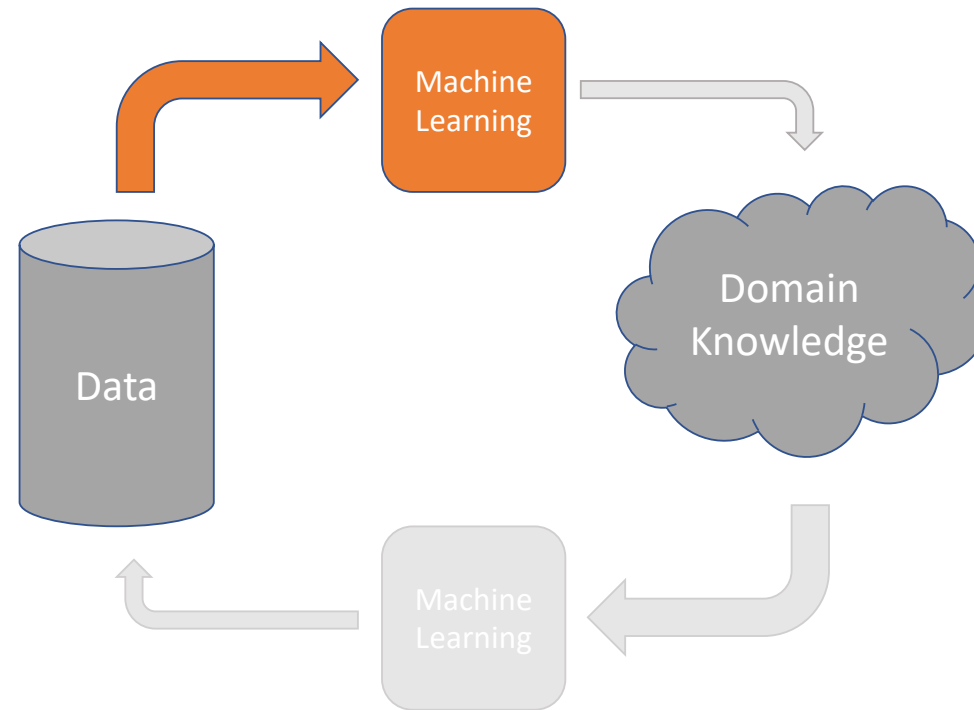


Francis Bacon,
credited with the modern
scientific method

Scientific discovery with machine learning



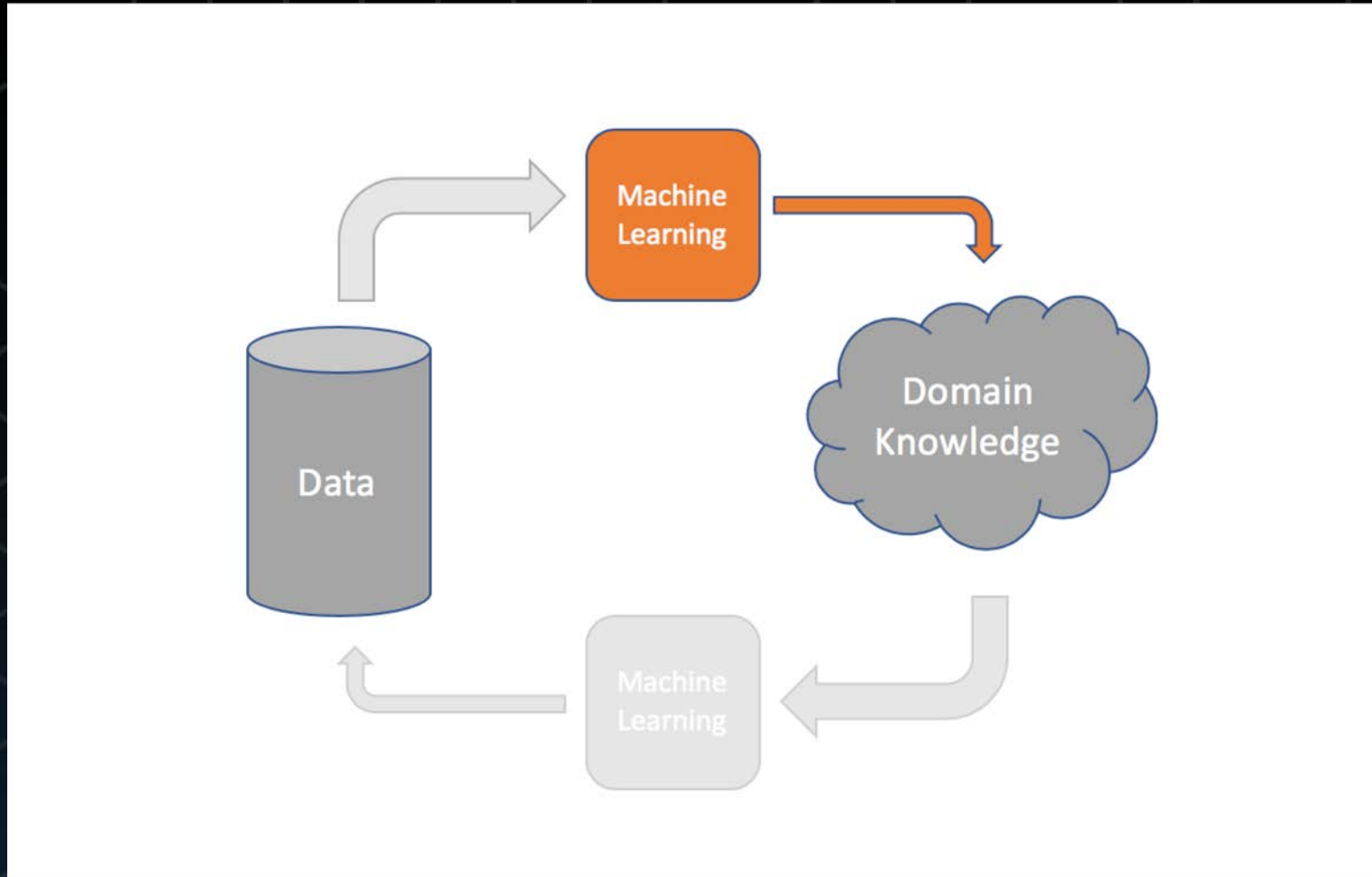
$$f_w(x) : \mathcal{X} \rightarrow \mathcal{Y}$$



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

- Structured data, e.g. graphs, trees, sequences
- Feature learning, e.g. computer vision
- Deep data, more features than samples
- Dependent features, e.g. spatio temporal data
- Knowledge about data, e.g. compositional data
- Privacy, encrypted data
- Non-representative data (biased sample)
- Ethics and Fairness

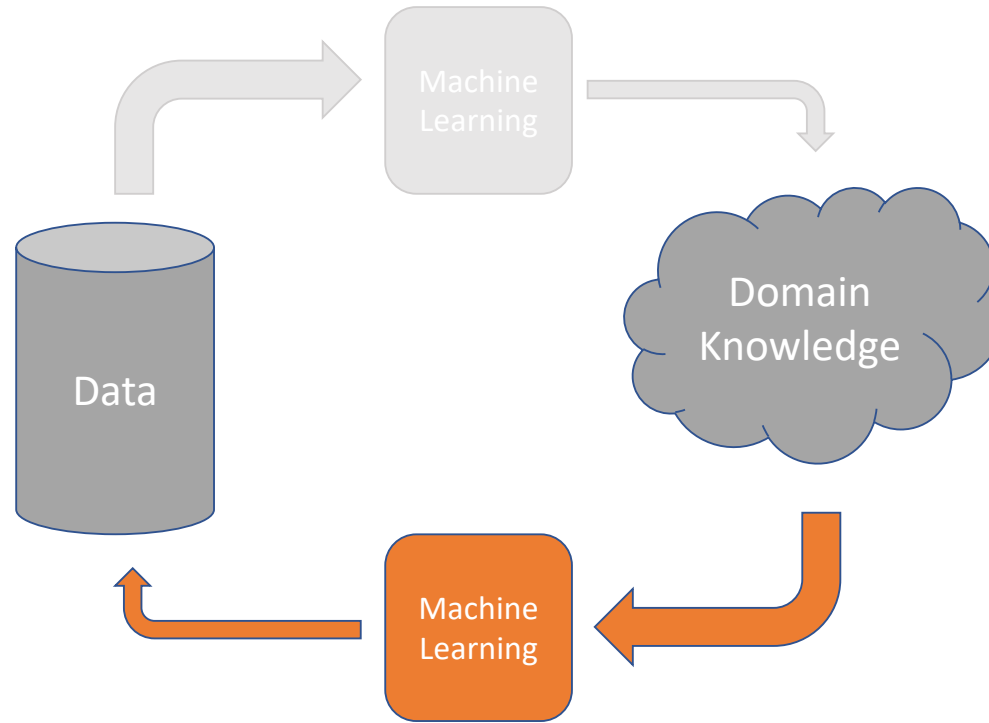
$$f_w(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{Y}$$



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

- Structured outputs, e.g. spatial predictions, DNA sequence
- Multi-task learning, domain adaptation
- Label noise, partial labels, multiple ground truths, weak supervision
- Uncertainty calibration and propagation
- Parameters of a dynamical system
- Causal decision making for intervention
- Dependent samples, e.g. pedigree information
- Justification of prediction, e.g. when predictions are used for taking action

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



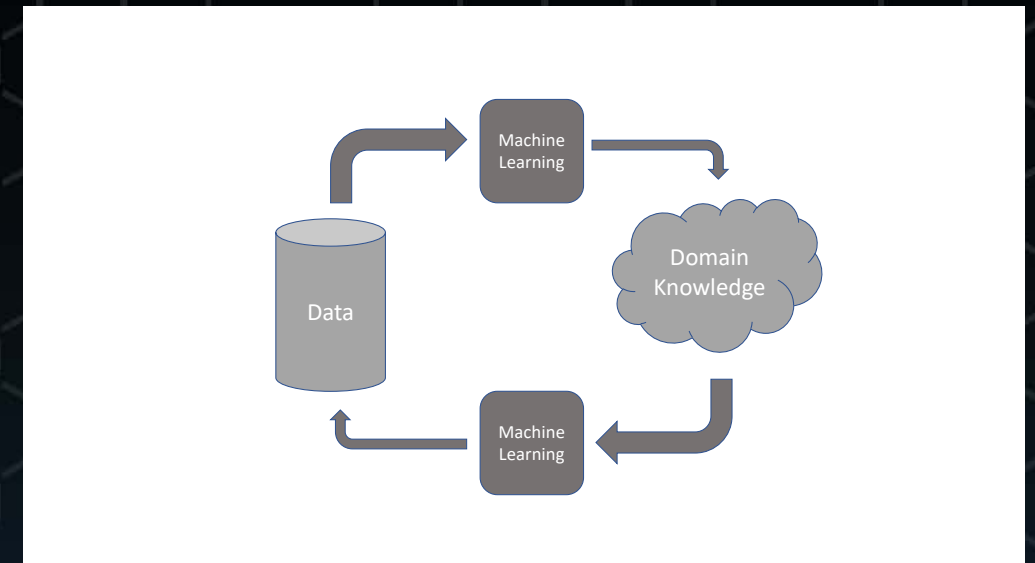
$$f_w(x) : \mathcal{X} \rightarrow \mathcal{Y}$$

- Human in the loop machine learning
- Use domain knowledge in modeling
- Infrastructure constraints, e.g. distributed, low power
- Qualification and Certification of predictors
- Use predictor to decide where to measure (ABCDE)
 - Active Learning
 - Bandits
 - Choice Theory
 - Design of Experiments

Scientific discovery with machine learning

- How can we use prediction to help with scientific discovery?
- Expand the types of prediction outputs \mathcal{Y}
- New constraints on inputs to ML \mathcal{X}
- Domain knowledge to Data
 - Human in the loop ML
 - Use predictor to decide where to measure (ABCDE)

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





THANK YOU

Data61 CSIRO

Cheng Soon Ong

e cheng-soon.ong@data61.csiro.au

w ong-home.my

www.data61.csiro.au

