MACHINE LEARNING

Dr Yun Sing Koh

School of Computer Science

WHAT IS MACHINE LEARNING?

Field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Tom Mitchell, 1998

Machine learning algorithms can figure out how to perform important tasks by generalizing from examples.

Pedro Domingos, 2012

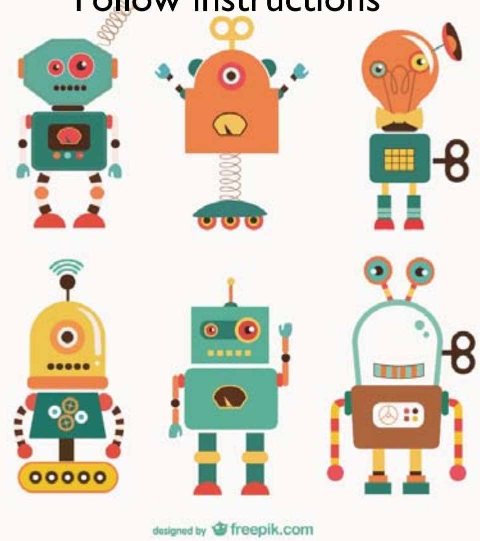
It can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

Ethem Alpaydin, 2020

WHAT IS MACHINE LEARNING?

Learn from experience

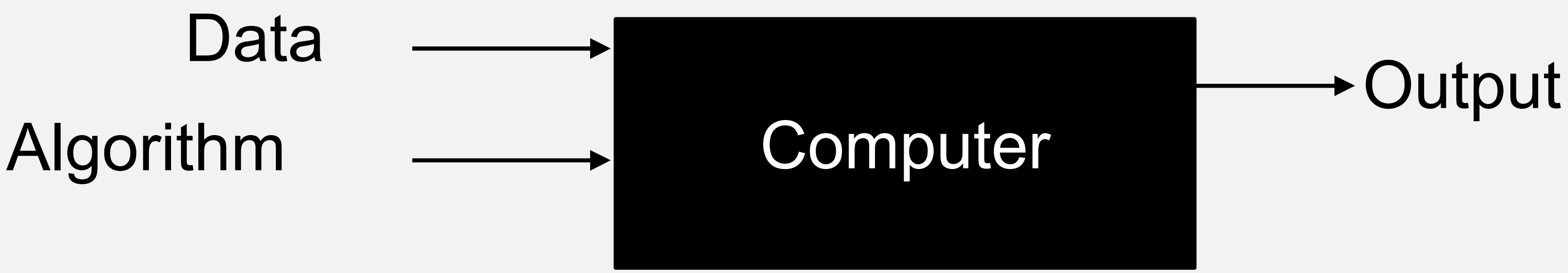
Follow instructions



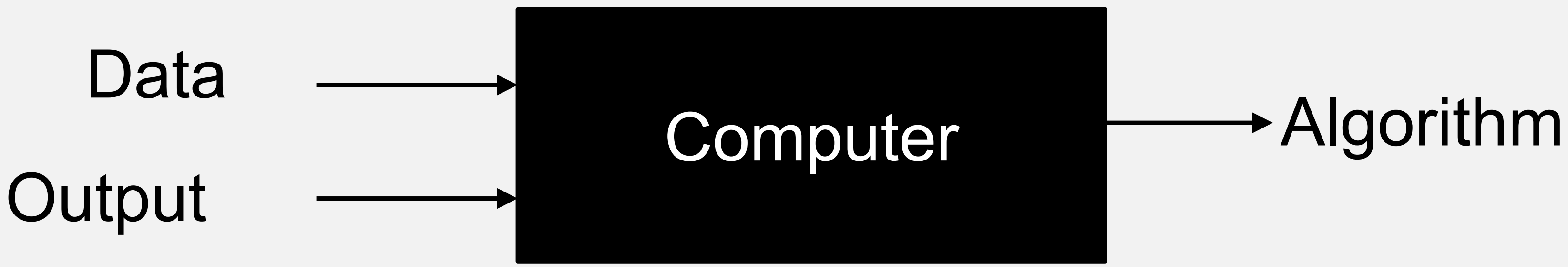
data

Learn from ~~experience~~

Traditional Programming



Machine Learning



Most of the knowledge in the world in the future is going to be extracted by machines and will reside in machines.

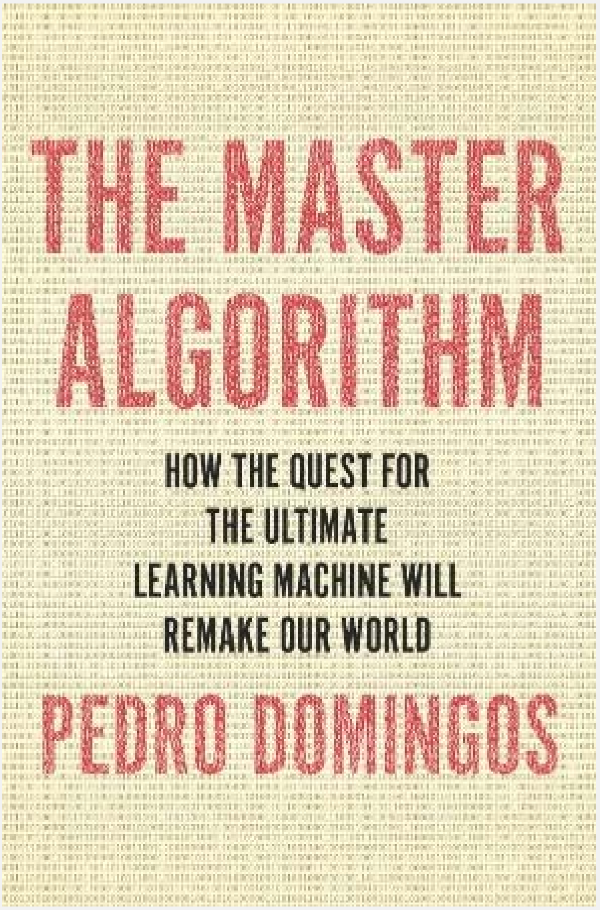
–Yann LeCun, Director of AI Research, Facebook

SO HOW DO MACHINES LEARN?

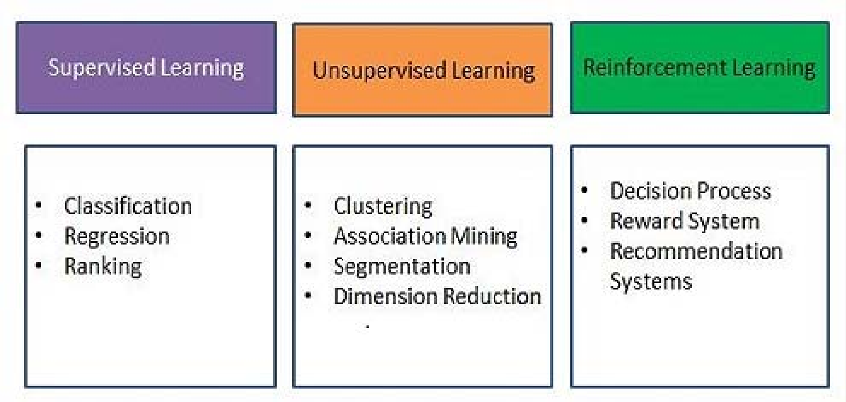
Fill in	Fill in gaps in existing knowledge
Emulate	Emulate the brain
Simulate	Simulate evolution
Reduce	Systematically reduce uncertainty
Notice	Notice similarities between old and new

PROLOGUE TO THE MASTER ALGORITHM - PEDRO DOMINGOS

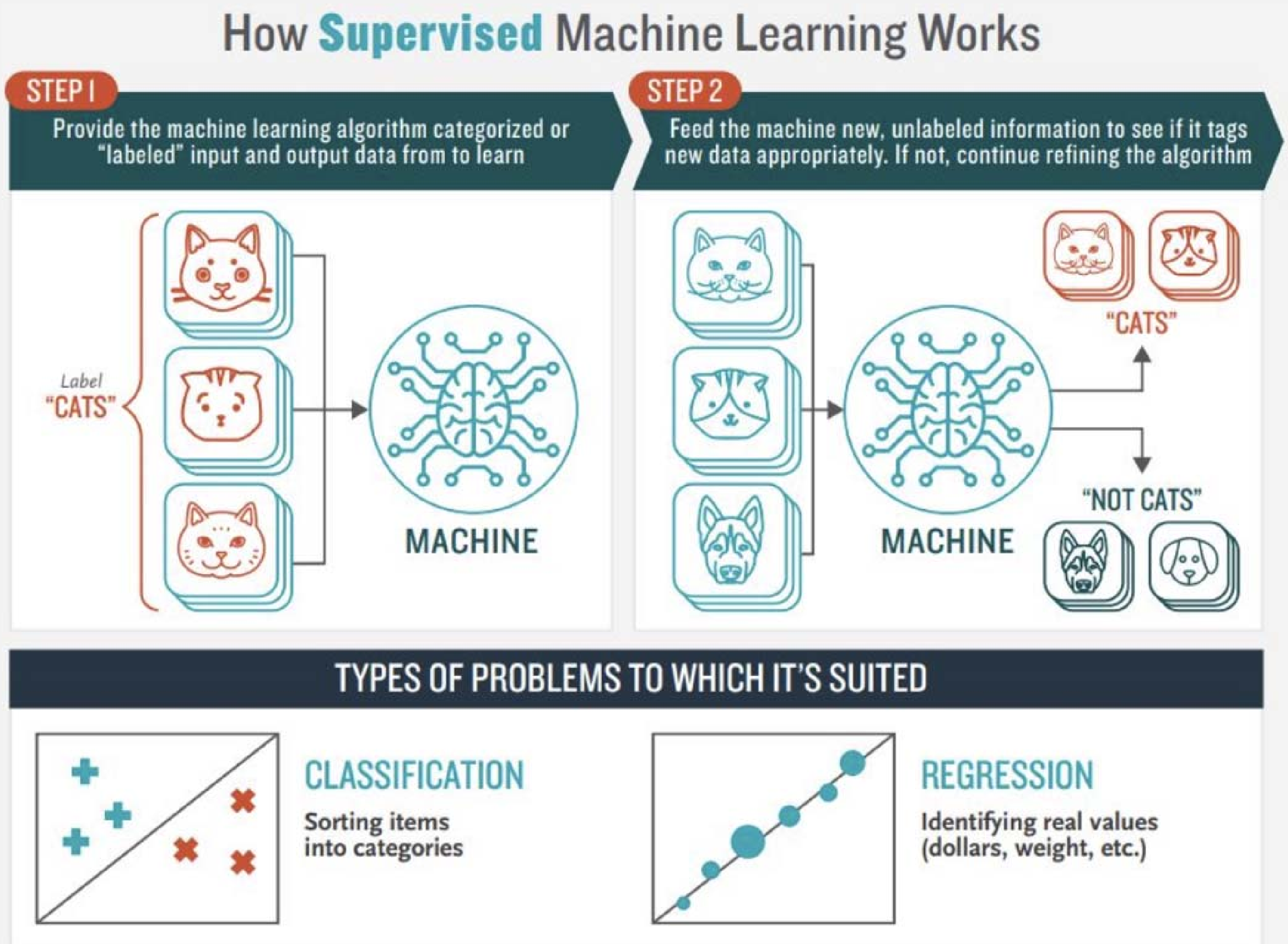
"Traditionally, the only way to get a computer to do something—from adding two numbers to flying an airplane—was to write down an algorithm explaining how, in painstaking detail. But machine learning algorithms, also known as learners, are different: they figure it out on their own, by making inferences from data. And the more data they have, the better they get. Now we don't have to program computers; they program themselves. It's not just in cyberspace, either: your whole day, from the moment you wake up to the moment you fall asleep, is suffused with machine learning."



LEARNING PARADIGM OF MACHINE LEARNING



Source: https://www.kdnuggets.com/2015/09/questions-data-science-can-answer.html/2



https://medium.com/@jorgesleonel/supervised-learning-c16823b00c13

THE FIVE TRIBES OF MACHINE LEARNING

Tribe	Origins	Master Algorithm
Symbolists	Logic, philosophy	Inverse deduction
Connectionists	Neuroscience	Backpropagation
Evolutionaries	Evolutionary biology	Genetic programming
Bayesians	Statistics	Probabilistic inference
Analogizers	Psychology	Kernel machines

If it exists, the master algorithm can derive all knowledge in the world — past, present, and future — from data. Inventing it would be one of the greatest advances in the history of science. - Pedro Domingos

THE SYMBOLISTS

- The Symbolises work on the premise of inverse deduction.
- Instead of starting with the premise and looking for the conclusions, inverse deduction starts with some premises and conclusions, and essentially works backward to fill in the gaps.
- The system has to ask itself "what is the knowledge that is missing?" and acquire that knowledge through analysis of existing data sets.
- "It's an ever-growing virtual circle of knowledge."
- Examples of techniques: Decision Trees (C4.5), Random Forest, Logistic Regression, etc.

http://www.dataversity.net/pedro-domingos-on-five-machine-learning-tribes/

CONNECTIONISTS

- "Connectionists" want to reverse engineer the brain.
- Create artificial neurons and connect them in a neural network.
- Neurons work on a weighted value of inputs, and how binary results can be enhanced into a "continuous value" with methods like back propagation. All of this leads the computer to be able to learn more about a given set of information criteria – in this case, about what is and is not a cat, to be able to more correctly label random sets of images.
- Examples of techniques: Neural Networks, CNN, RNN.

http://www.dataversity.net/pedro-domingos-on-five-machine-learning-tribes/

THE EVOLUTIONARIES

- "Evolution made your brain and everything else,"
- Evolutionaries apply the idea of genomes and DNA in the evolutionary process to data structures.
- The survival and offspring of units in an evolutionary model are the performance data. An algorithm for an evolutionary learning project would mimic those processes in key ways.
- Examples of techniques: Genetic Algorithm, Particle Swarm Optimization.

http://www.dataversity.net/pedro-domingos-on-five-machine-learning-tribes/

THE BAYESIANS

- Bayesian deal in uncertainty and solutions.
- Their master algorithm solution is called probabilistic inference.
- Take a hypothesis and apply a type of "a priori" thinking, believing that there will be some outcomes that are more likely.
- Update a hypothesis as you see more data to make some hypotheses become more likely than others.
- As a sort of scientific process, the probabilistic models do bring a certain concrete result to Machine Learning.

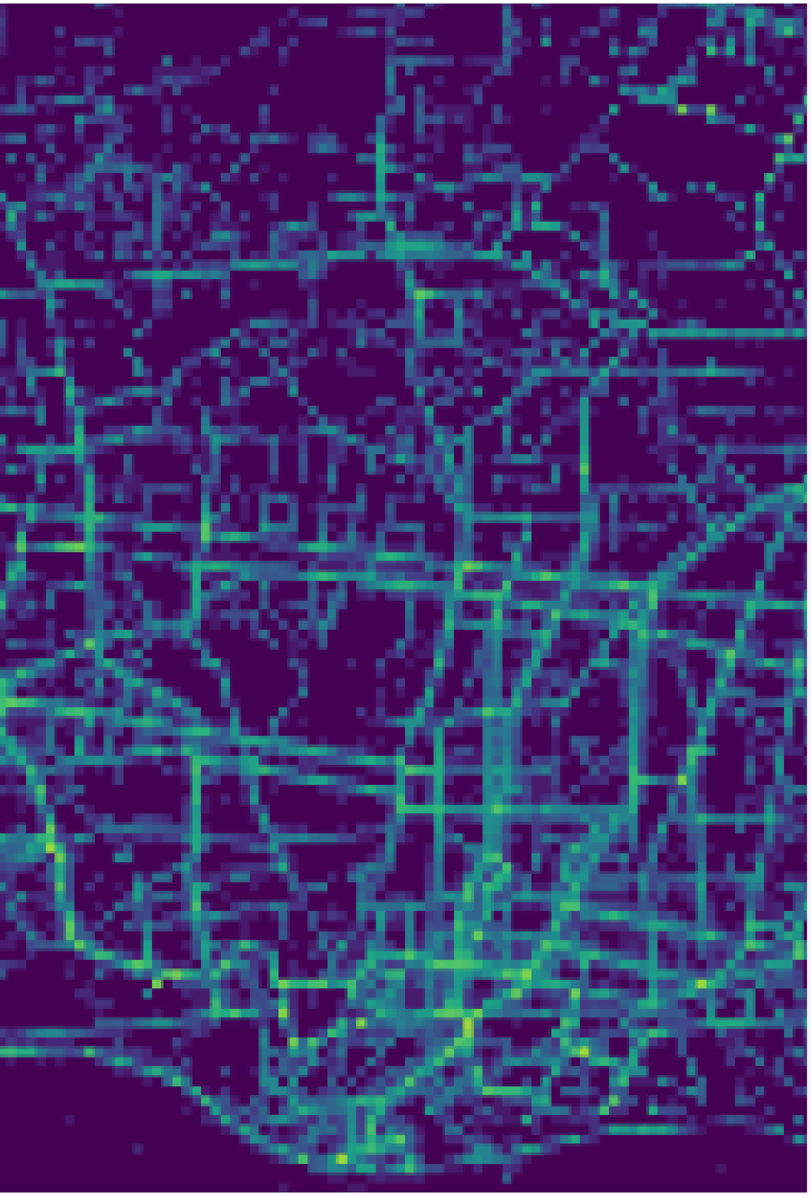
- Examples of techniques: Naïve Bayes, Bayesian Graphical Models.

http://www.dataversity.net/pedro-domingos-on-five-machine-learning-tribes/

THE ANALOGIZERS

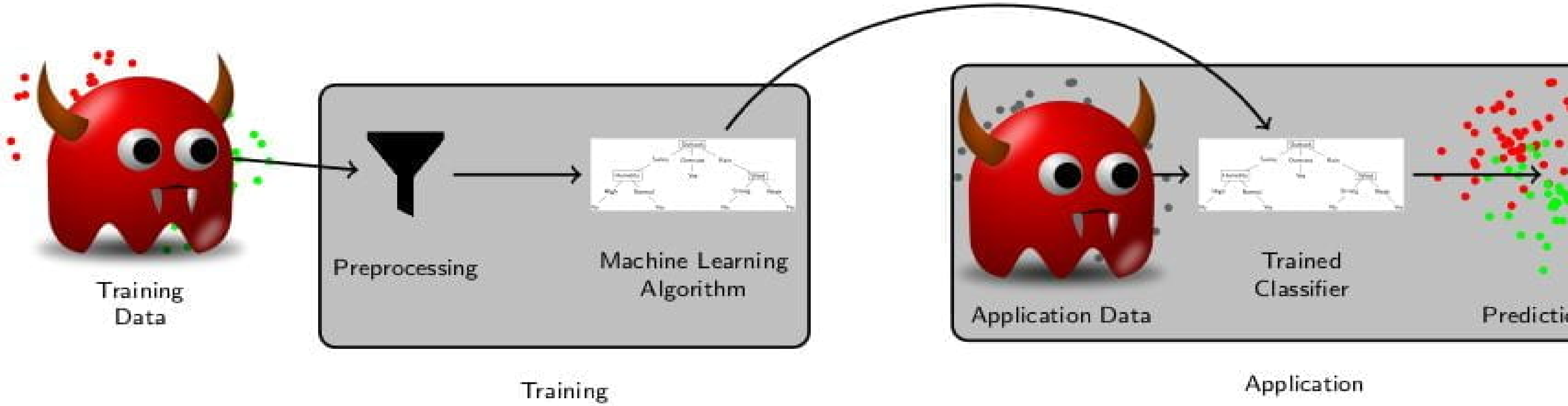
- The analogizers, or pioneers in the field of matching particular bits of data to each other.
- The master algorithm is the "nearest neighbour" principle "generalizing from similarity"
- It's a very nice type of similarity-based learning and useful:
- One third of Amazon sales are based on recommendations.
- Examples of techniques: K-NN, SVM

http://www.dataversity.net/pedro-domingos-on-five-machine-learning-tribes/



SPATIO-TEMPORAL DATA MINING

The prevalence of location positioning devices such as GPS have made huge amount of geo-tagged data available. Spatio-temporal data mining is to automatically uncover latent patterns from data associated with GPS coordinates over time. This research area links to various of applications including geographical topic analysis, local event detection, location recommendations and traffic prediction. (Kaiqi Zhao)



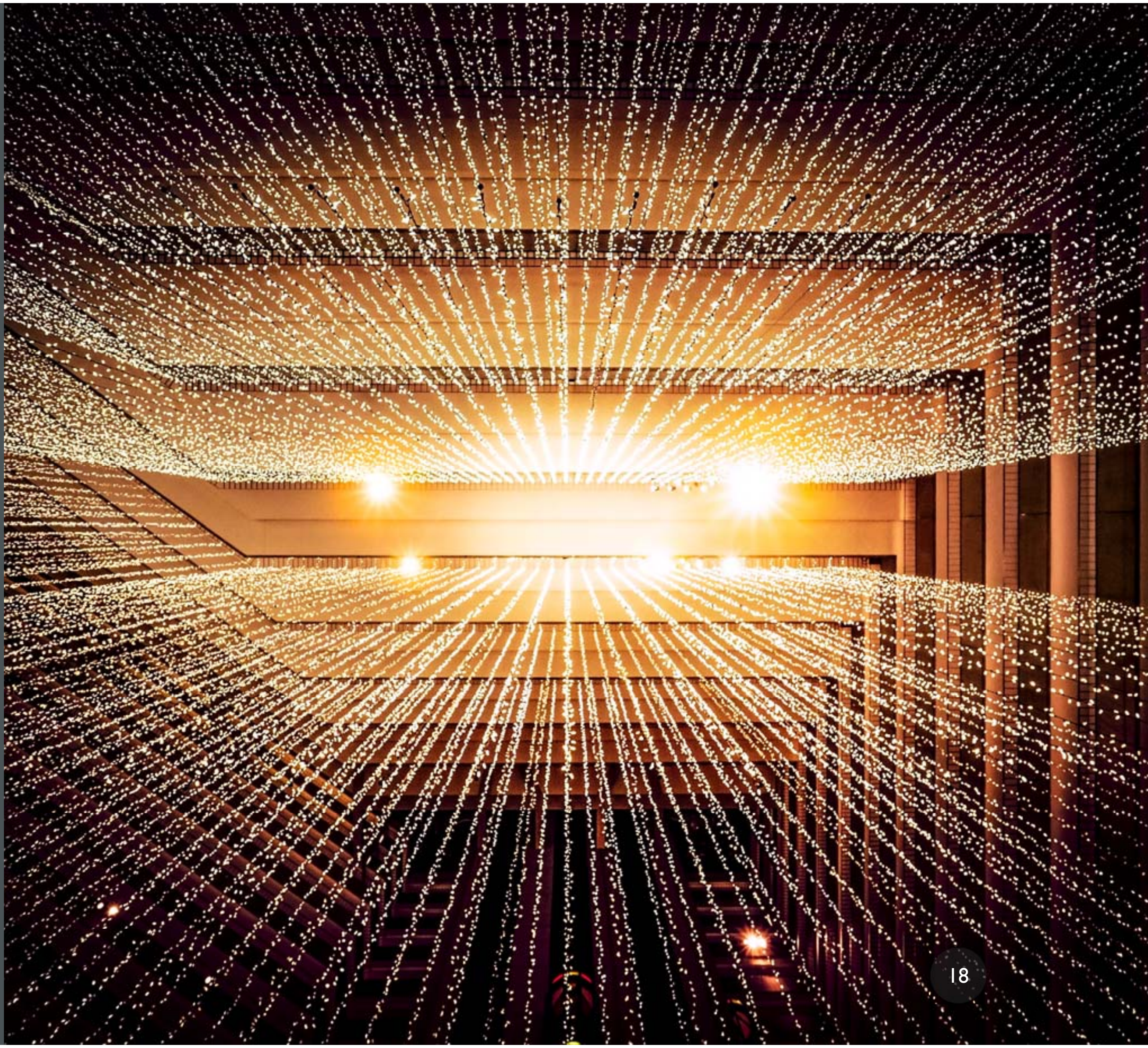
ADVERSARIAL LEARNING

Adversarial learning aims to identify weaknesses in machine learning models. The goal is to identify potential problems that cannot be found using traditional evaluation using test sets. It has been used successfully in a wide range of applications, typically focused on a specific model or domain. In image classification, methods have been developed to fool models that recognize traffic signs by rather simple modifications of pictures. Another direction of adversarial learning aims to identify examples that could break or improve the training of the model if that example would be added to the training. (Joerg Wicker, Gillian Dobbie)

DATA STREAM MINING

Data Stream Mining extracts knowledge from continuous data which arrives into the system in a stream. Data stream sources consist of IoT sensors, through to air pollution sensors, health records. In this area of research, we are developing new and novel predictive techniques that can deal with data that is evolving and changing on a continuous basis. We are interested in continued adaptation of our model, to provide an accurate prediction in real time.

(Yun Sing Koh, Gillian Dobbie, Pat Riddle)



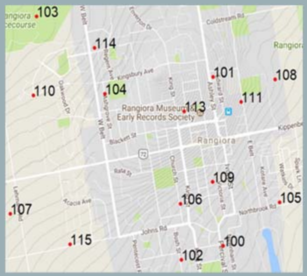


Figure 1: ODIN deployments in Rangiora overlayed on a Google Map. Locations are labeled by ODIN serial number.

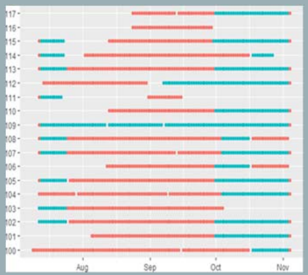


Figure 2: Timeline of ODIN locations. Blue denotes colocation at the ECan site and red denotes deployment.

PREDICTING AIR QUALITY FROM LOW-COST SENSOR MEASUREMENTS

Researchers: Hamish Huggard (UoA), Pat Riddle (UoA), Gustavo Olivares (NIWA), Yun Sing Koh (UoA)

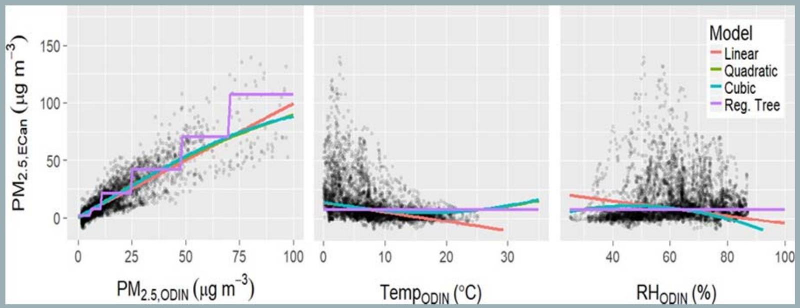


Figure (**MODEL SELECTION**): The result of different modelling techniques. On the x-axes are the measurements made by the ODIN, compared to the actual $PM_{2.5}$ values (as measured by the ECan TEOM-FDMS) on the y-axis. These data points come from the first third of the Single ODIN dataset.

Motivation

- PM annually causes 6.4 million years of life lost globally* and so understanding patterns in PM concentrations is an important global health problem.
- We investigate PM predictions from one such instrument, the ODIN-SD (low-cost, portable air quality sensors).

Research Questions

- What kind of calibration model produces the best accuracy?
- How much training data needs to be collected to get reasonable accuracy?
- Does it matter when the training data is collected?
- Is concept drift an issue?

Results

- Model error stabilises after about 30 days
- Calibration models of robust linear regression over PM, temperature, relative humidity seems to work well
- Concept drift is an issue for low cost sensors due to changes in PM sources, changes in the instrument, and changes in PM concentration distribution.

TOWARD A UNIVERSAL LEARNER

- Much remains to be done . . .
- We need your ideas
- What will a Universal Learner enable?

THANK YOU

- “Last but not least, if you have an appetite for wonder, machine learning is an intellectual feast, and you’re invited—R.S.V.P.!” – Pedro Domingos