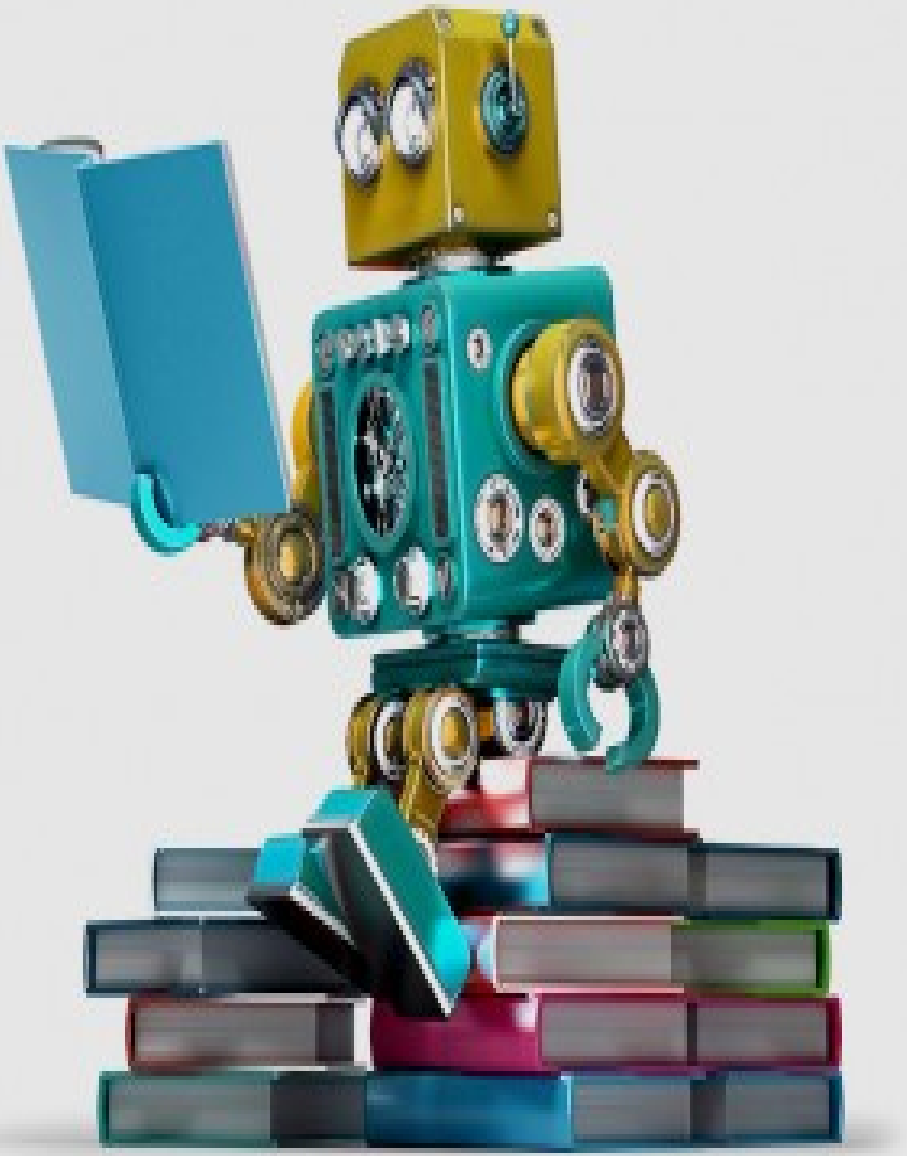
PWL ❤ BSB # 11:
"A Few Useful Things to Know about Machine Learning"
Pedro Domingos

Papers We Love ❤ Brasília

Presenter: Alessandro Leite

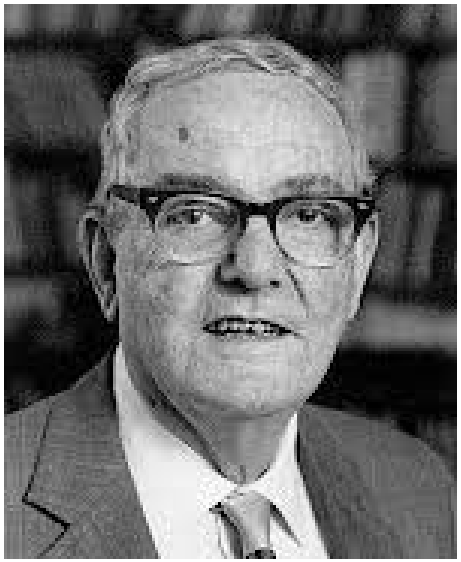
October 11, 2018



What is Machine Learning?

"Learning is any process by which a system improves performance from experience".

"Machine Learning is concerned with computer programs that automatically improve their performance through experience".



Herbert Simon
Turing Award (1975)
Nobel Prize in Economics (1978)

What is Machine Learning? (2)

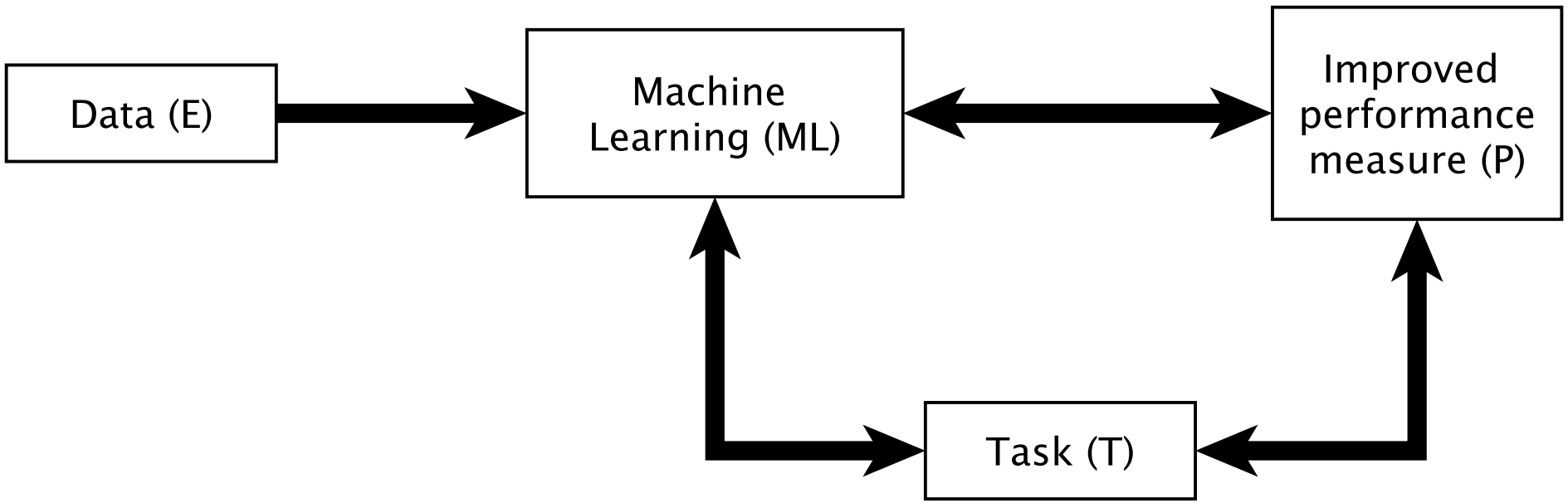
Definition

Machine learning can be defined as the process of an **algorithm extracts patterns from data** and to **make predictions without** being **explicitly programmed** to do so.

What is Machine Learning? (3)

Definition

A computer program is considered to learn from experience **E** with respect to some class of tasks **T** and performance measure **P** if its performance tasks in **T**, as measured by **P**, improves with experience **E**.



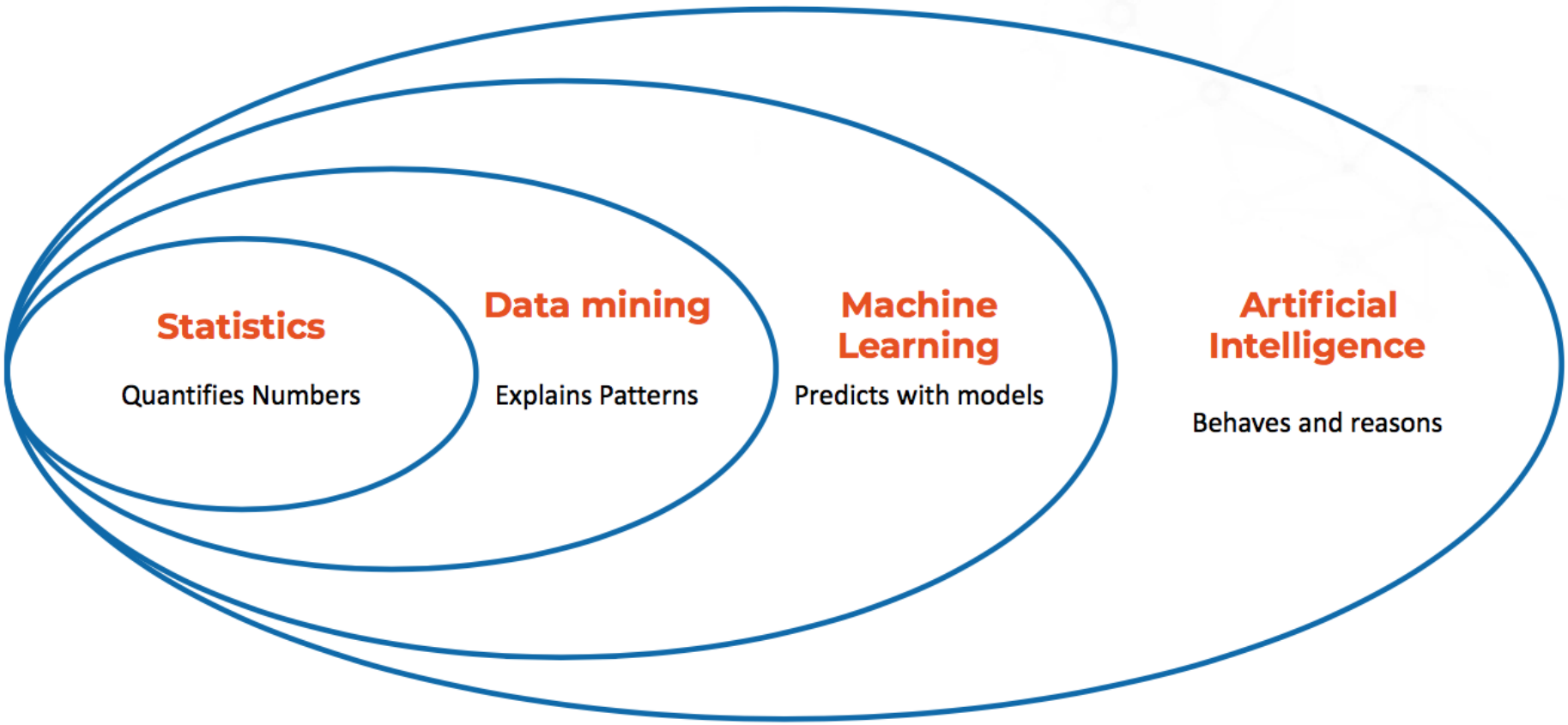
What are the reasons to rely on Machine Learning?

- Machine Learning can be used as an alternative approach to build complex systems
- Examples of use case scenarios include:
  - When human cannot build program a system manually
  - When human cannot provide a solution easily
  - When needing rapid decisions that human cannot make
  - When needing to be user-oriented in a massive scale

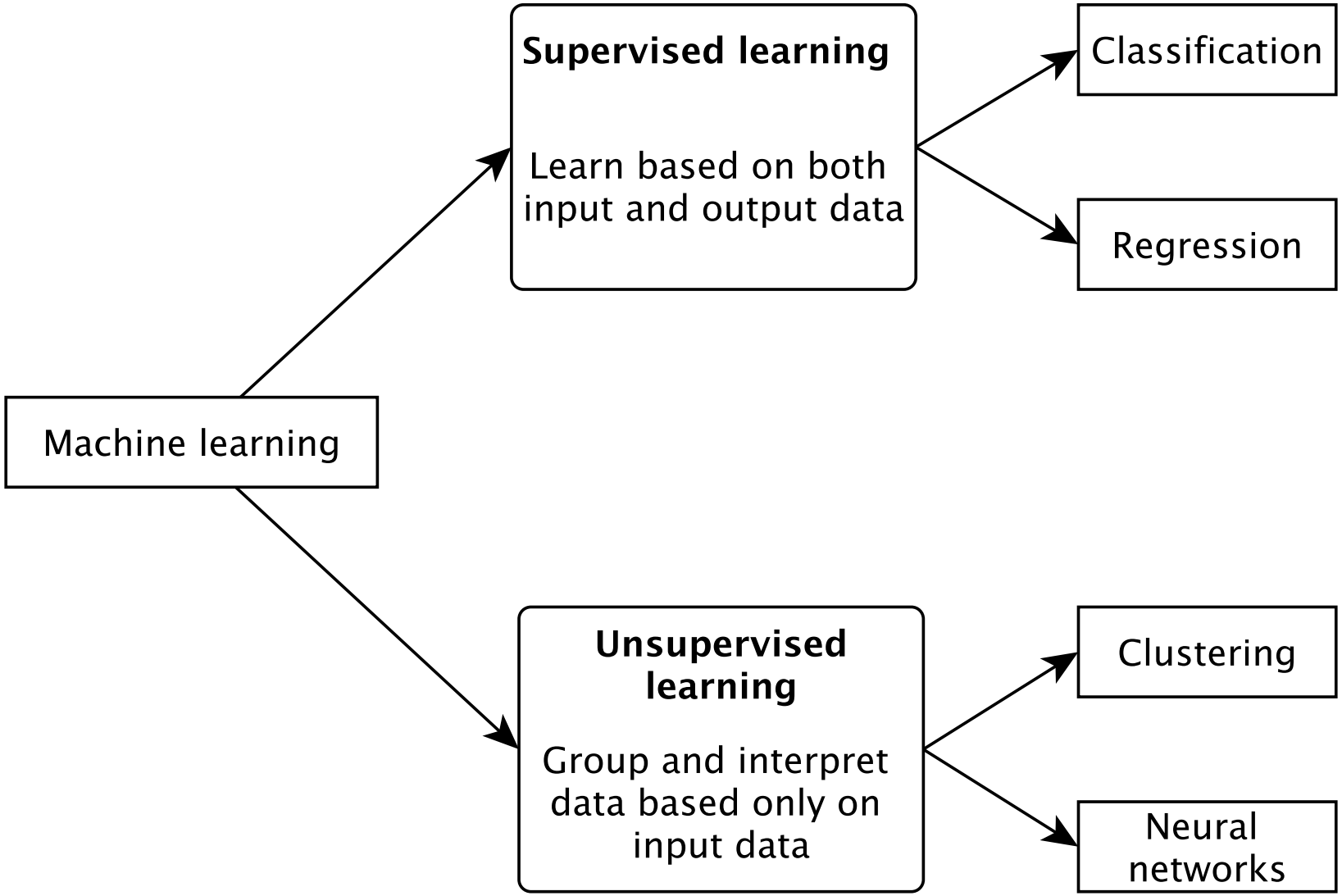
What are the essence elements to decide whether to use Machine Learning?

- When exists some **underlying pattern** to be **learned**
  - Thus, the performance measure can be improved
- When **lacks** easily **programmable** definition
  - Machine learning is needed
- There is **data** about the pattern
  - As a result, Machine Learning has an **input** to learn from

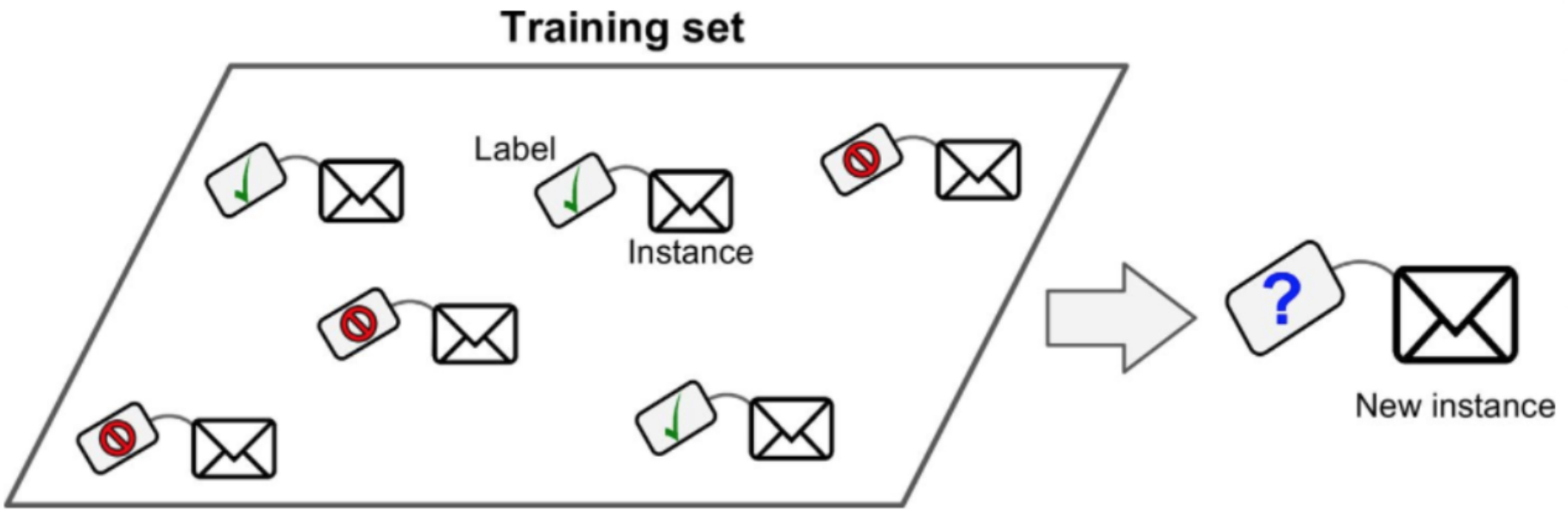
The relationship of Machine Learning with other fields



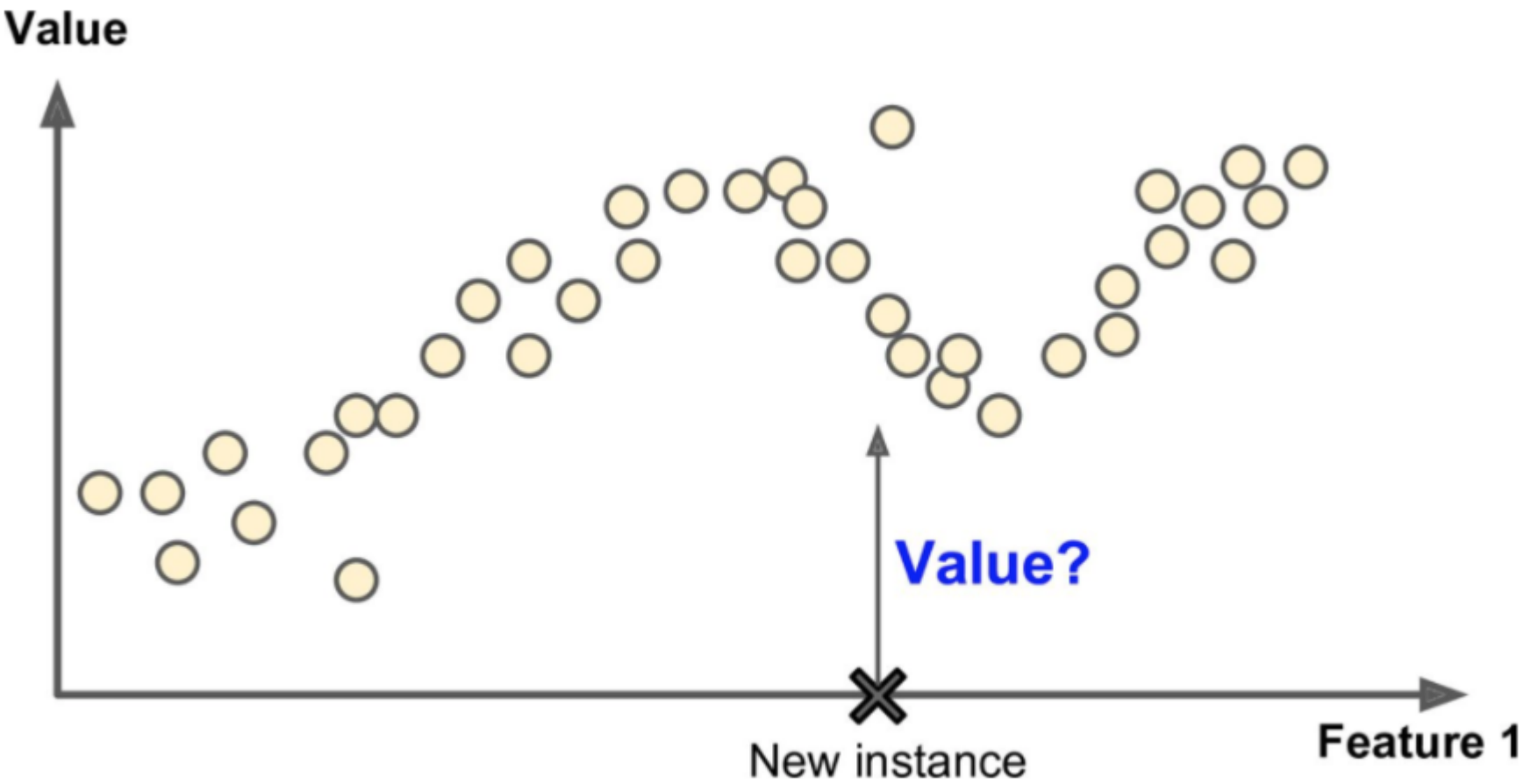
There are two major Machine Learning techniques



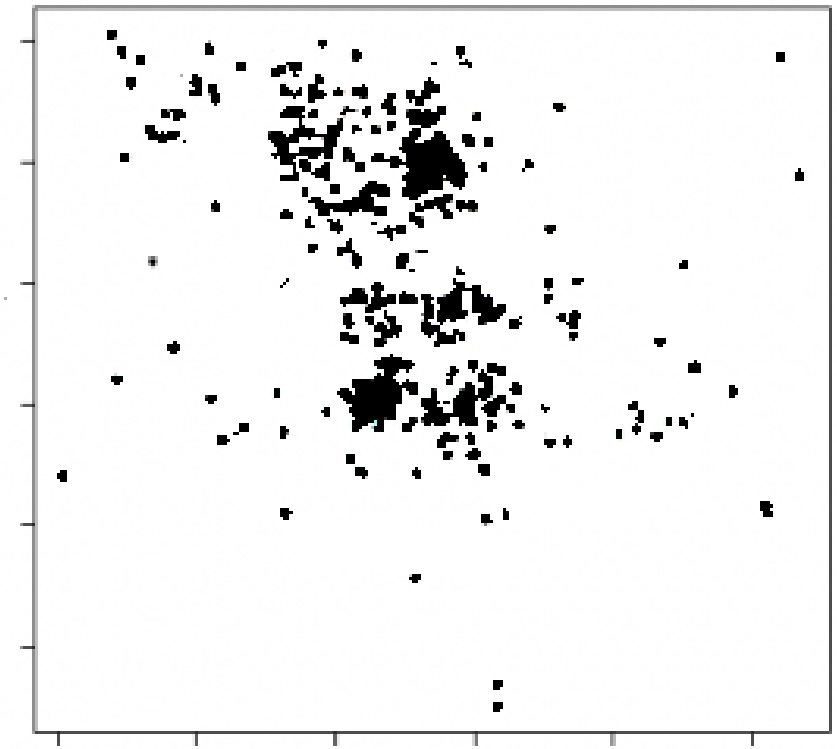
Example of supervised Machine Learning use case: classification



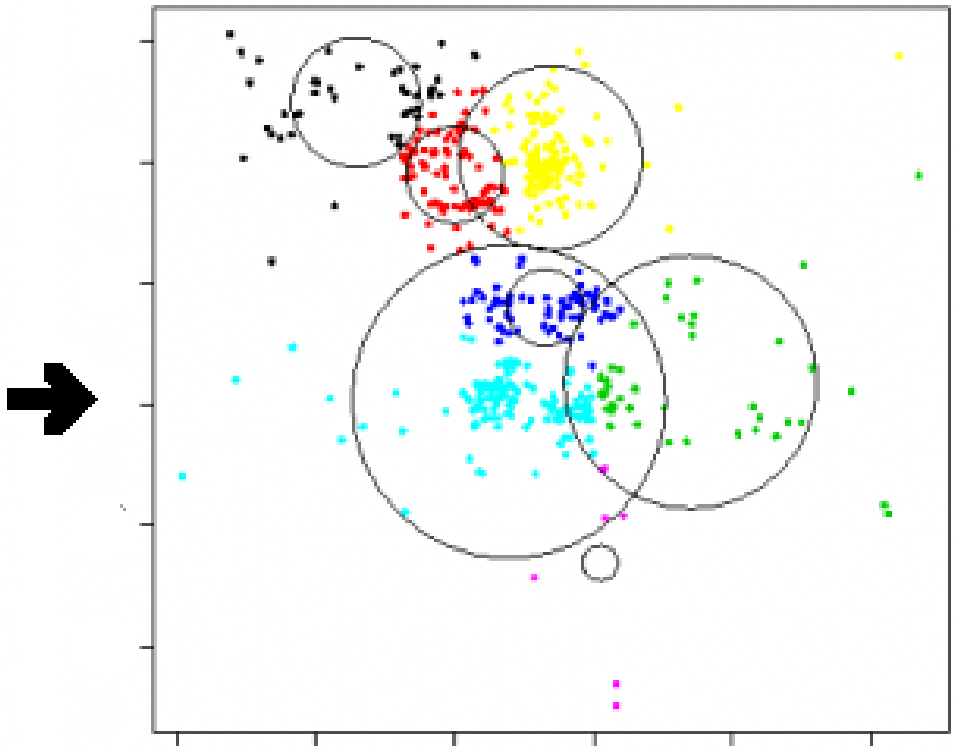
Example of supervised Machine Learning use case: regression



Example of unsupervised Machine Learning use case: clustering

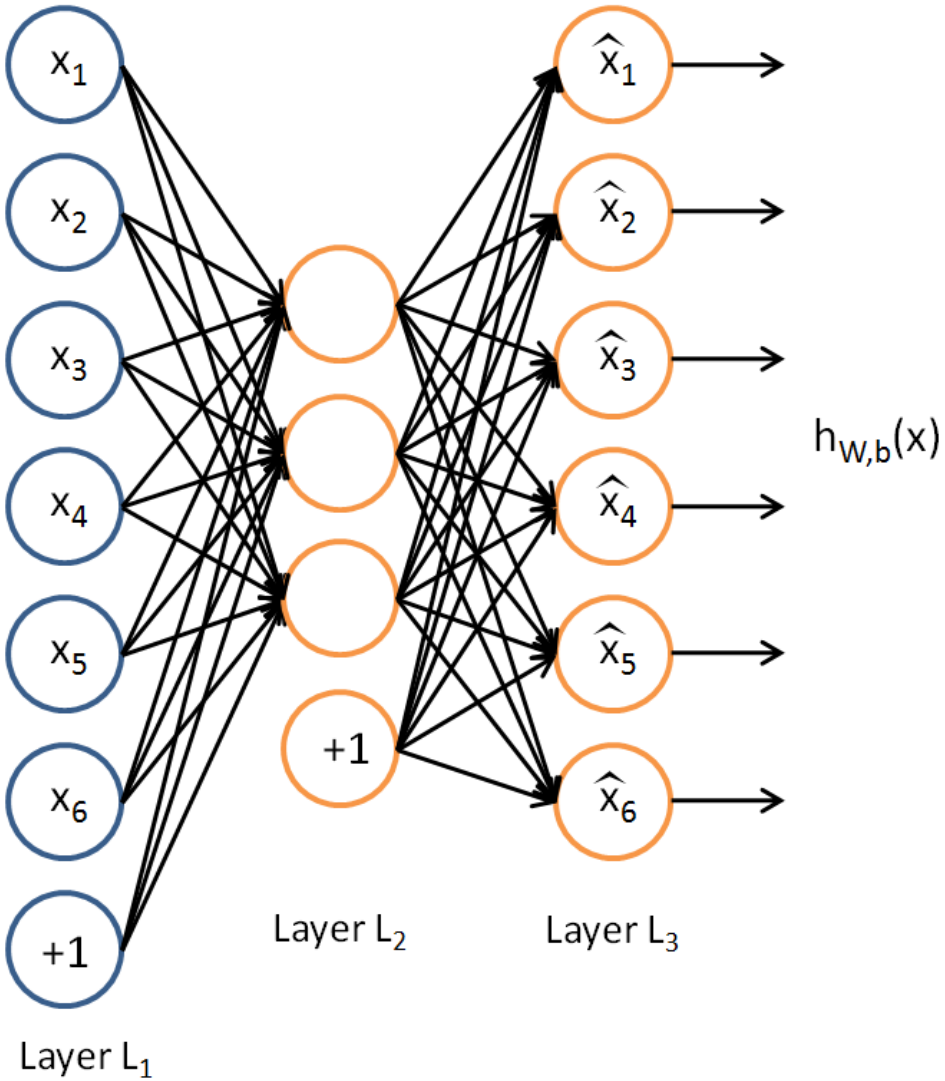


Raw Data



Clustered Data

Example of unsupervised Machine Learning use case: neural networks

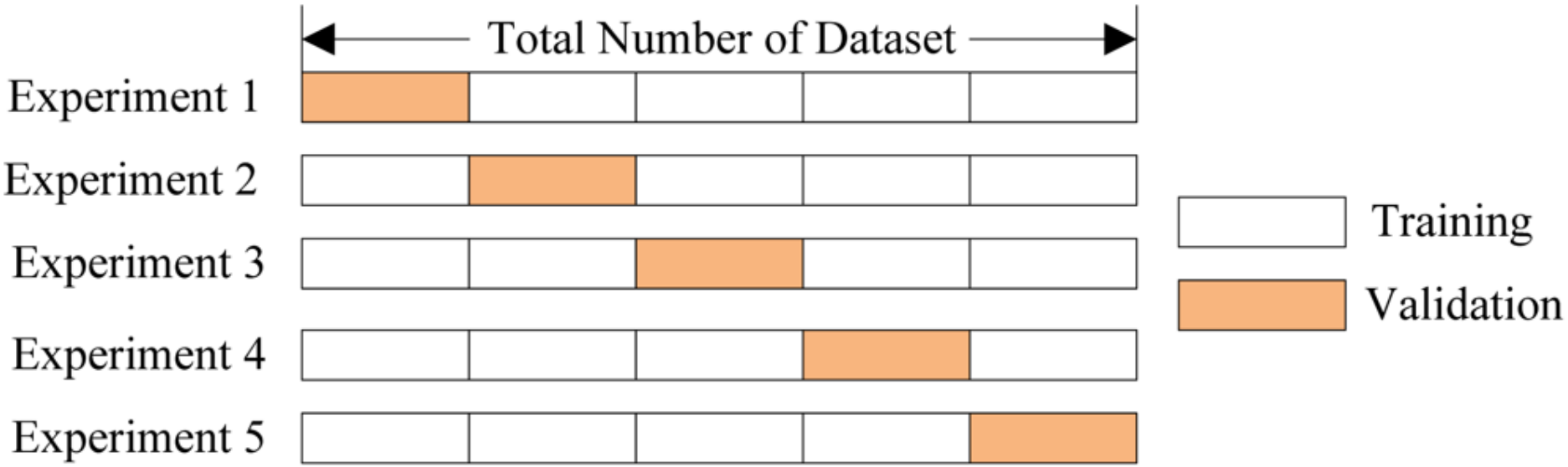


Learning = representation + evaluation + optimization

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

It's generalization that counts

- Separate your data set into training and validation set
- Use **cross-validation** to assess that your model can generalize to independent data set
- The fundamental goal of machine learning is to generalize beyond the examples in the training set



Data alone is not enough

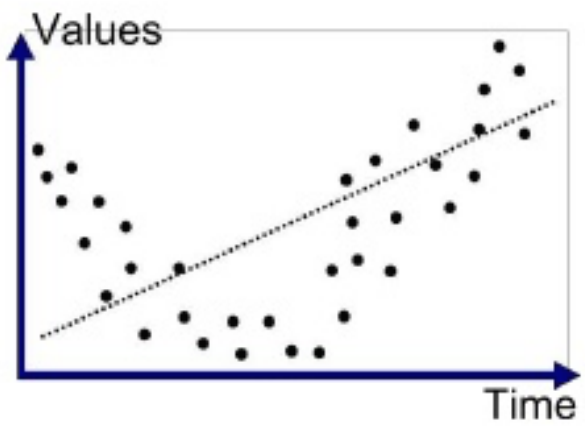
- No Free Lunch theorem (Wolpert, 2011)

> *"For any two learning algorithms, there are just as many situations (appropriately weighted) in which algorithm one is superior to algorithm two as vice versa, accordingly to any of the measures of superiority"*

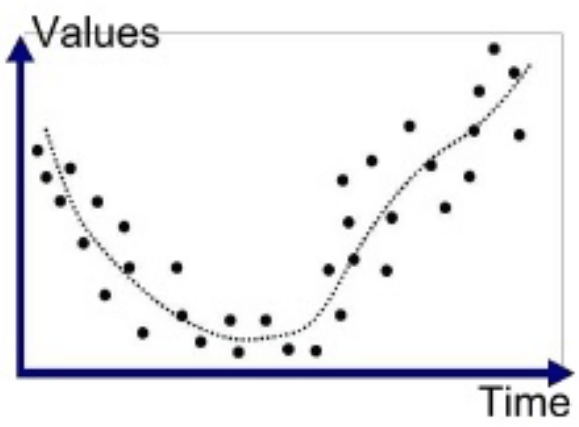
Data alone is not enough

- Knowledge representation is a key concept in machine learning
- Programming, like all engineering, is a lot of work: everything need to be built from scratch
- Learning is more like farming
- Farmers combine seeds with nutrients to grow crops
- Learners combine knowledge with data to grow program

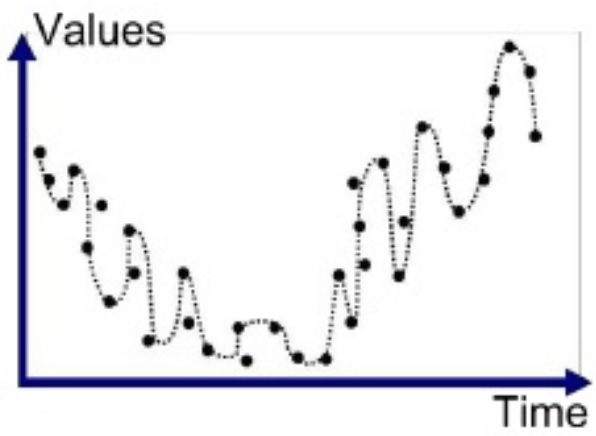
Overfitting has many faces



high bias
low variance



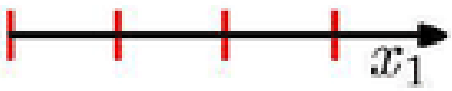
medium bias
medium variance



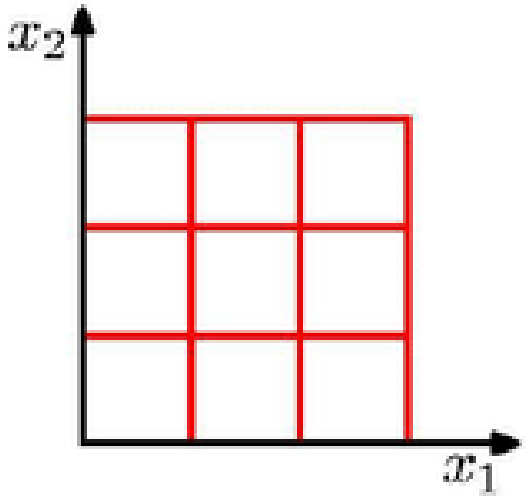
low bias
high variance

Intuition fails in high dimension

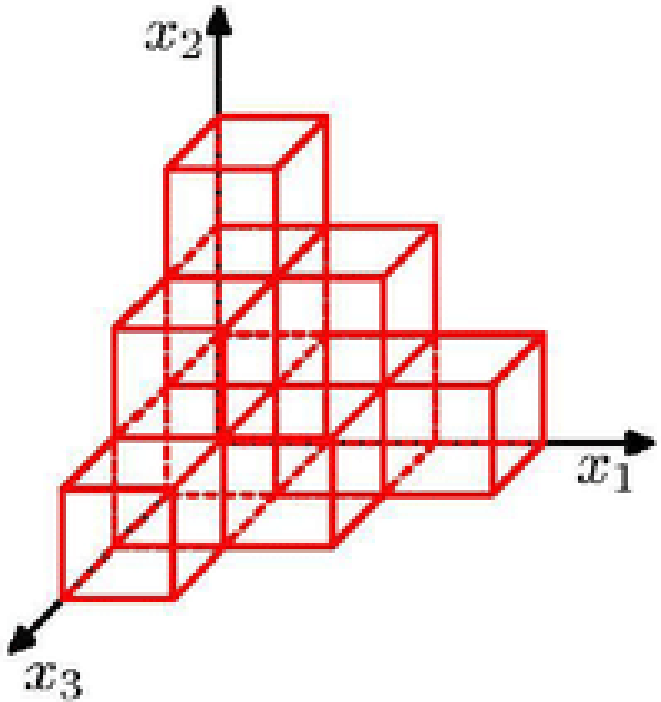
- We live in a three dimension world
- Our intuition usually fails when we try to image a high-dimension space



$D = 1$



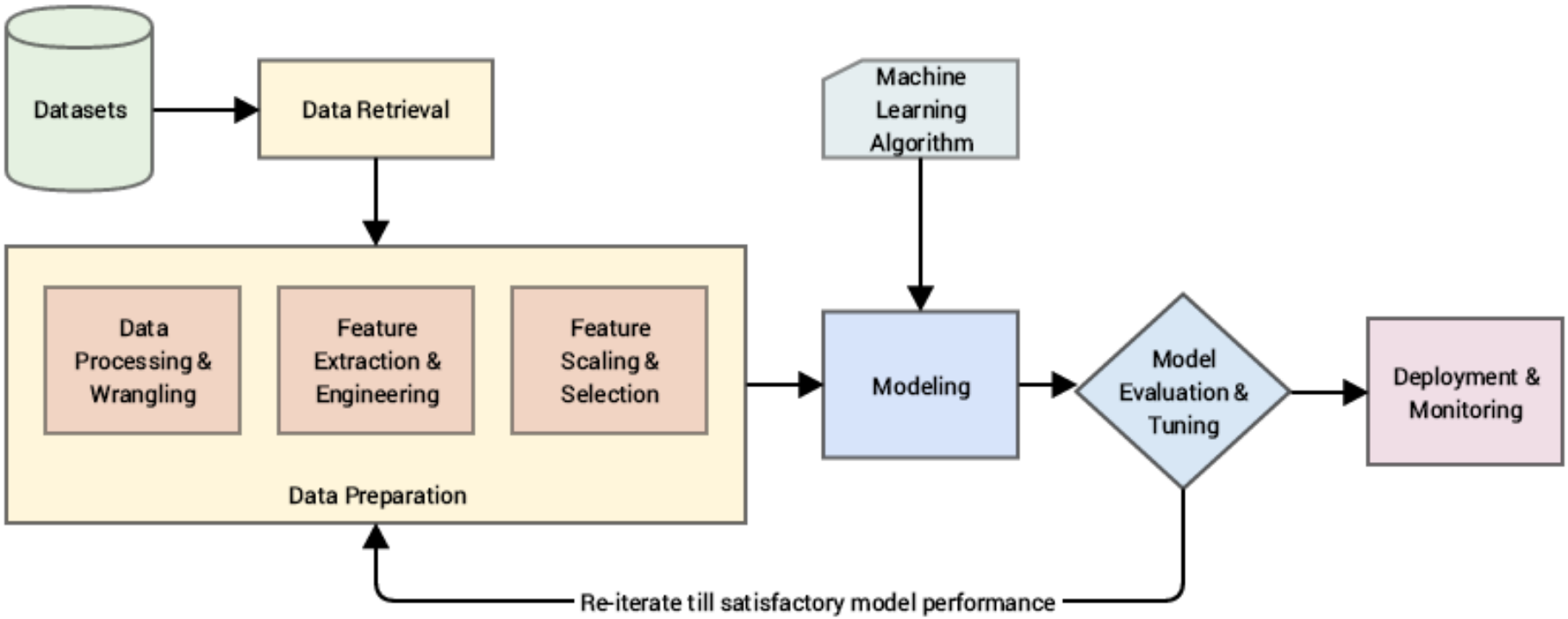
$D = 2$



$D = 3$

Feature engineering is the key

- Raw data is rarely in a form that is amenable to learning
- Therefore, we can build features from it that are.

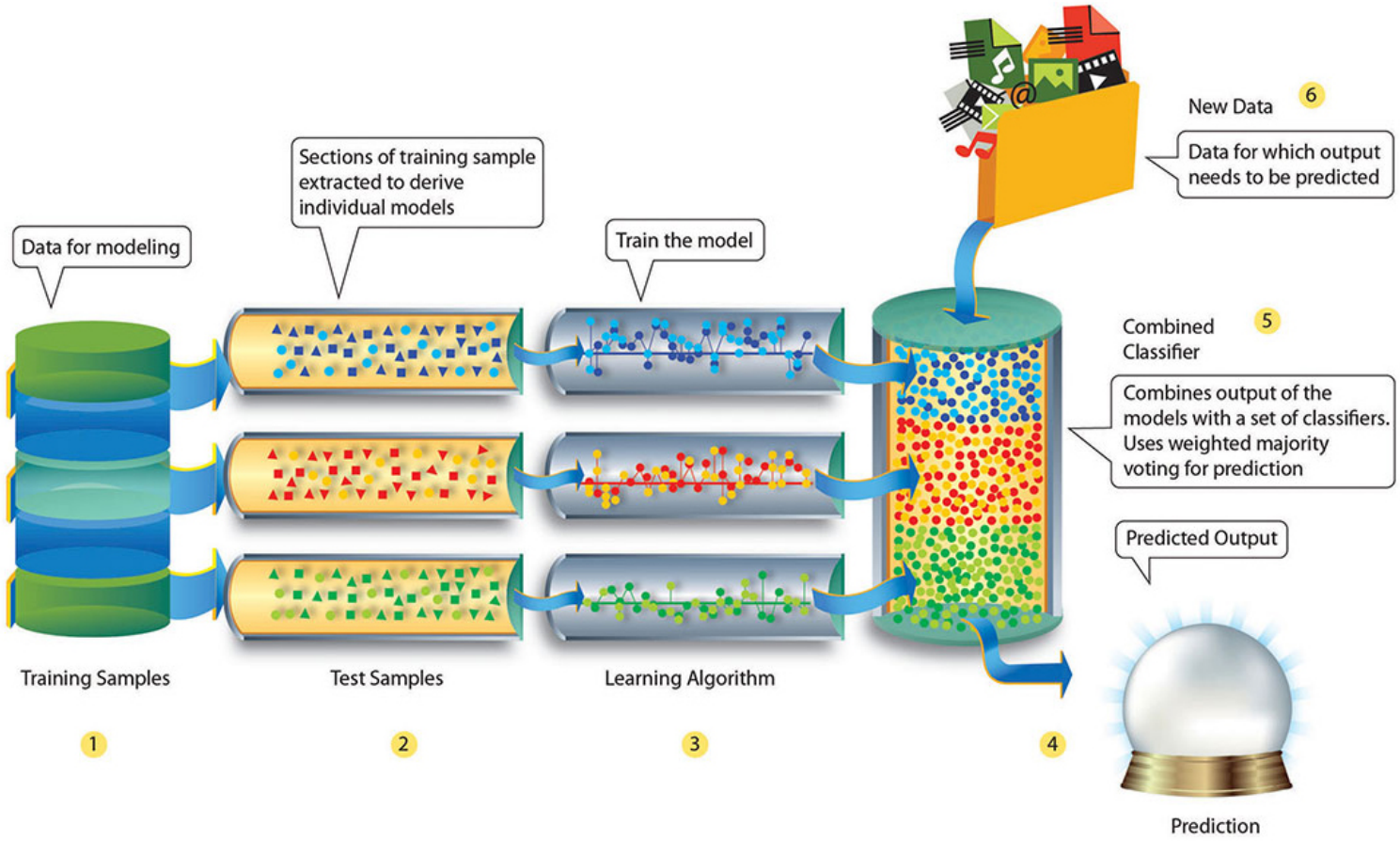


More data beats a cleverer algorithm

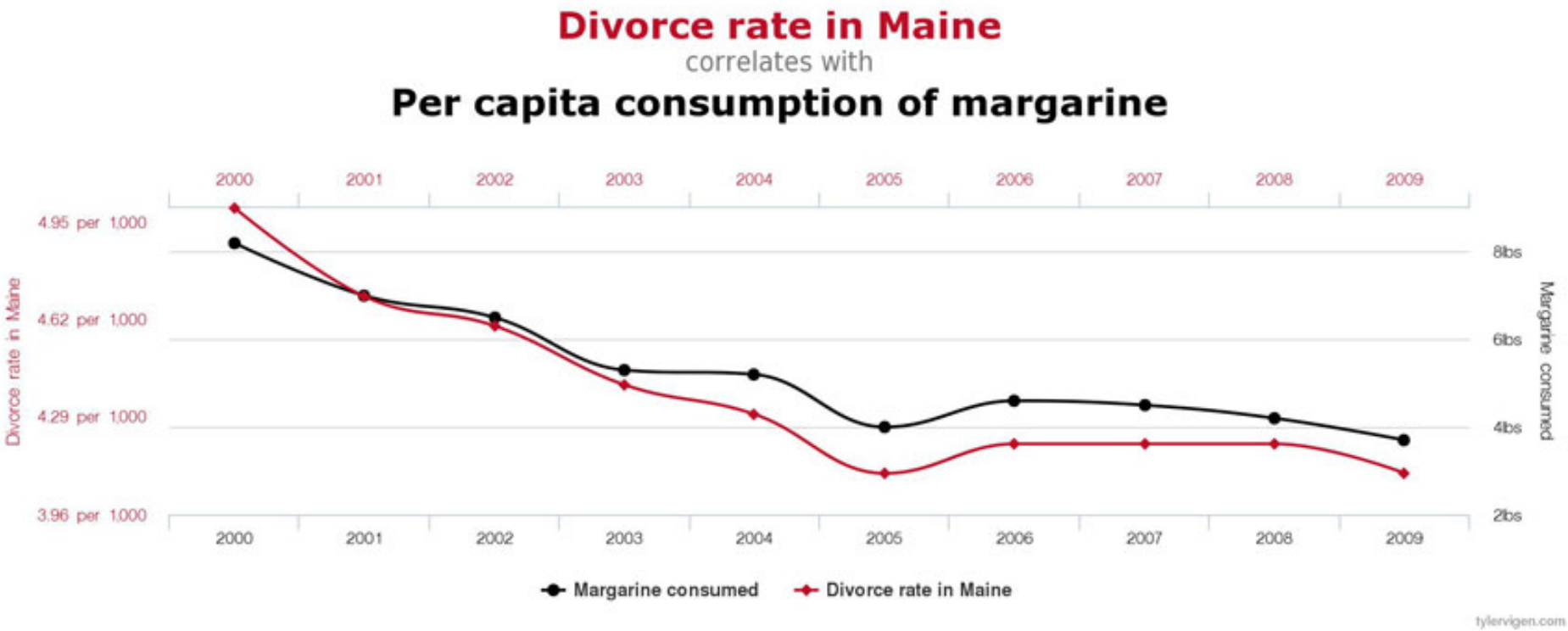
The *New* New Oil

Learn many models, not just one

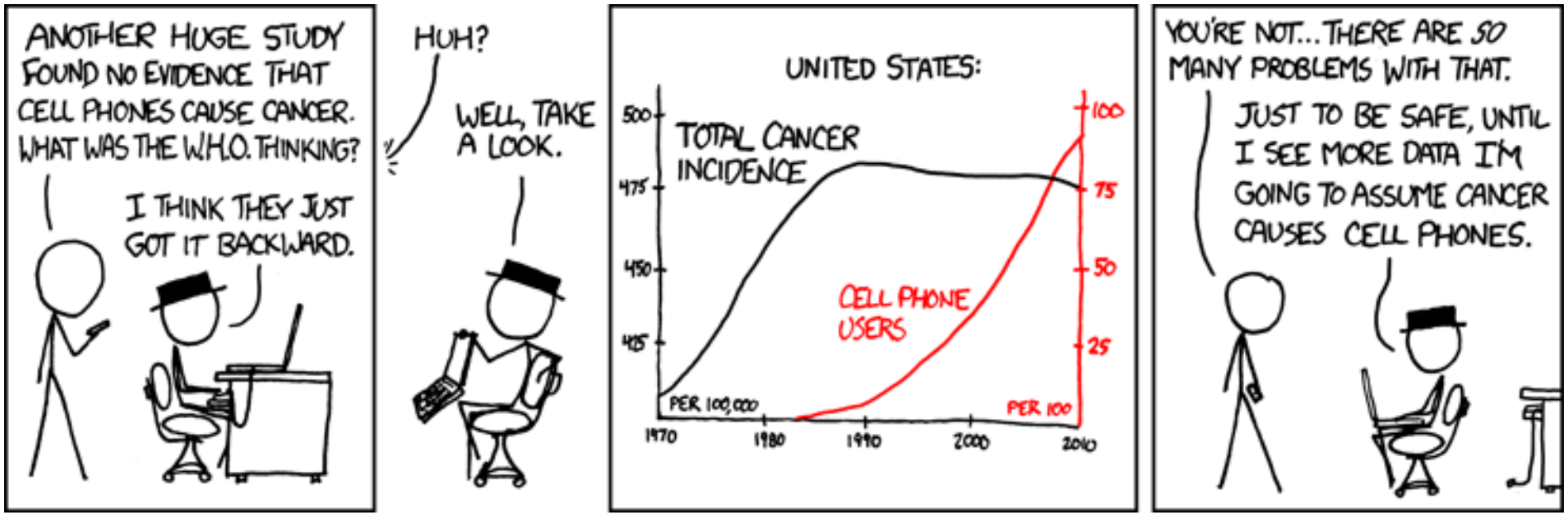
- The idea consists of combining multiple models to obtain better performance



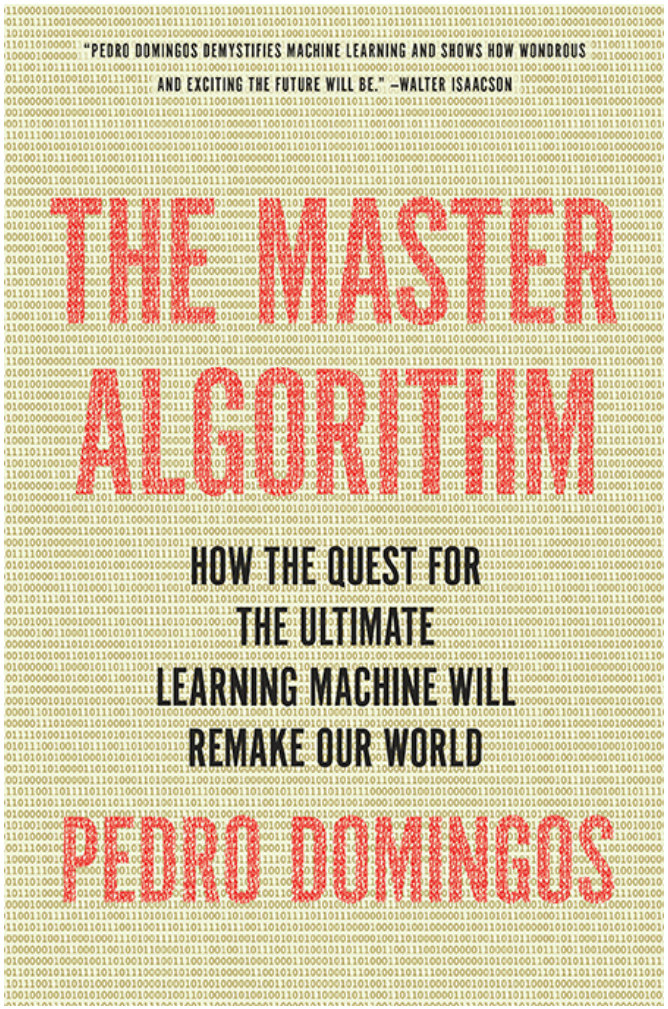
Correlation does not imply causation

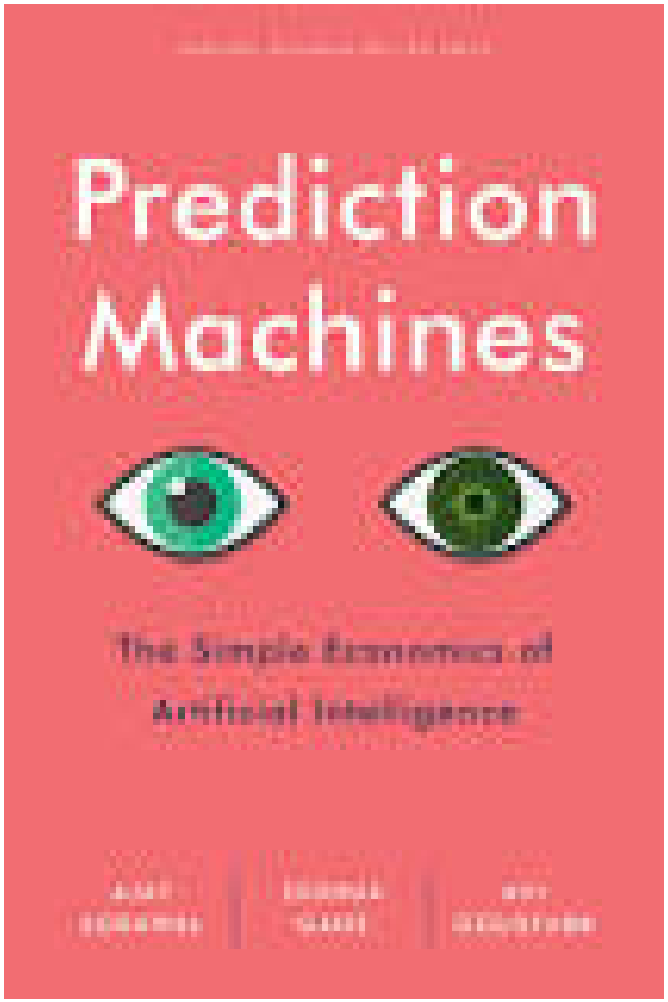


Correlation does not imply causation



Going further






That's all Folks!