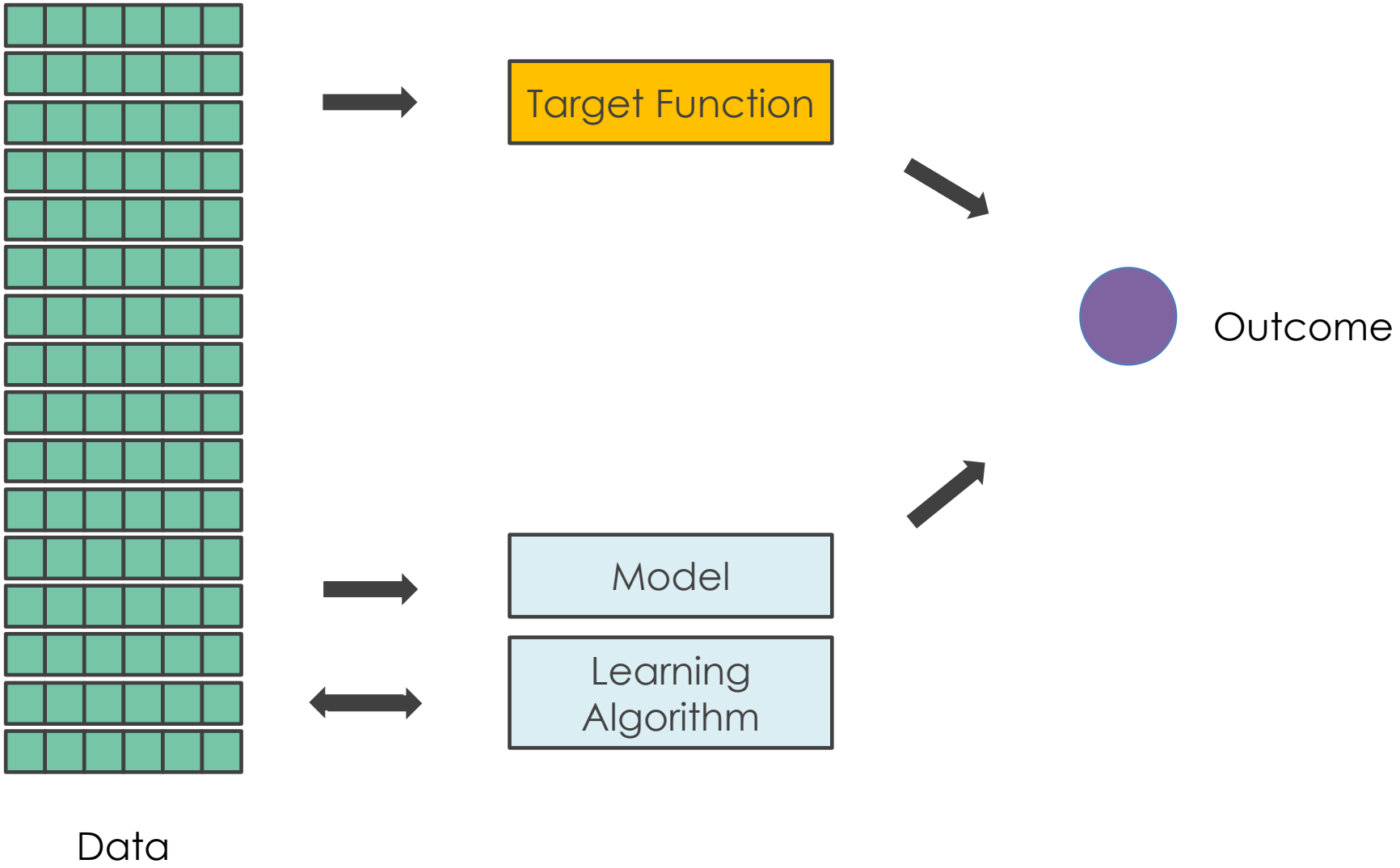
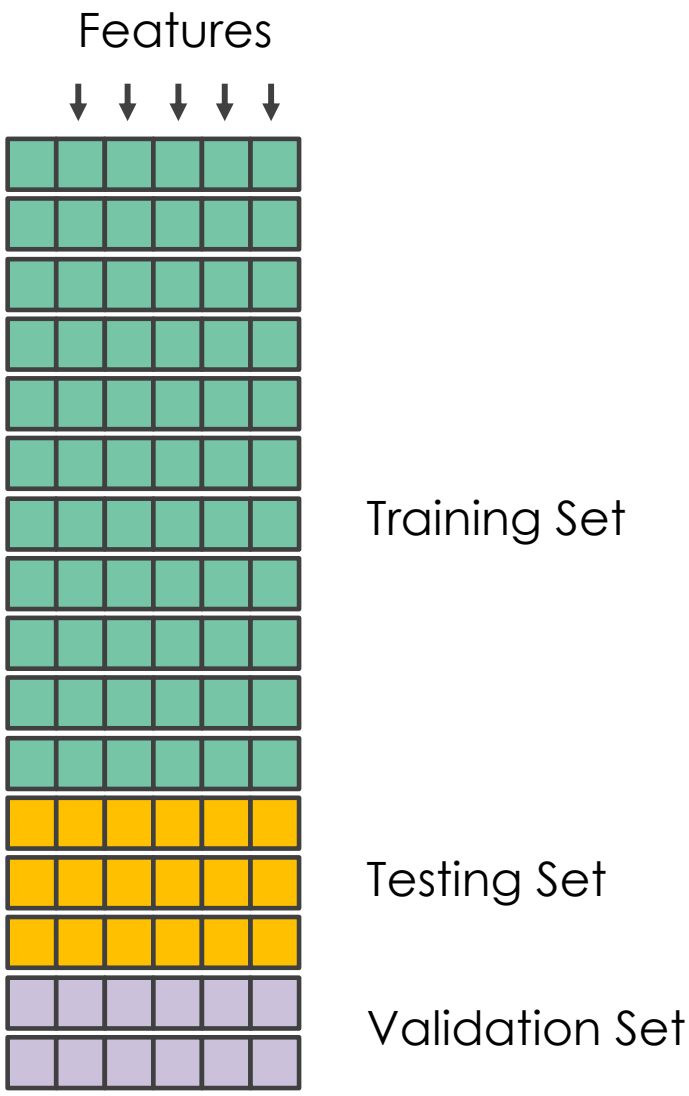
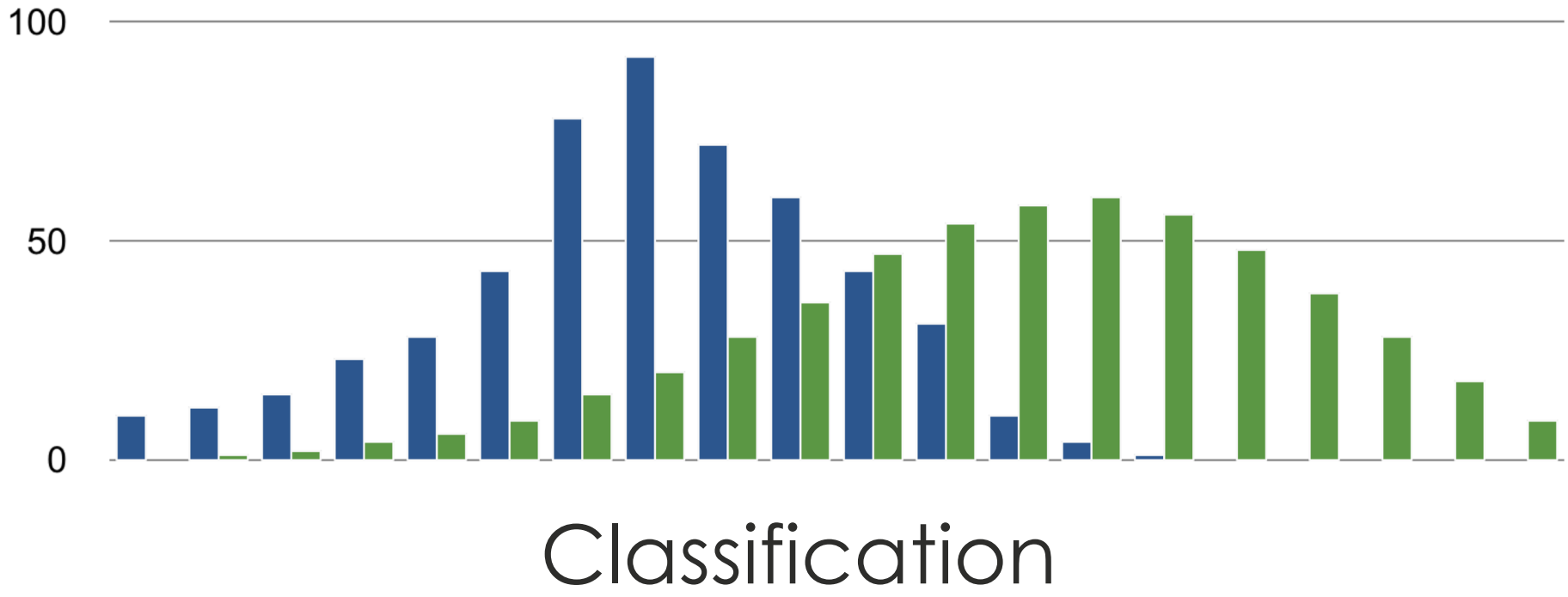
Machine Learning 101

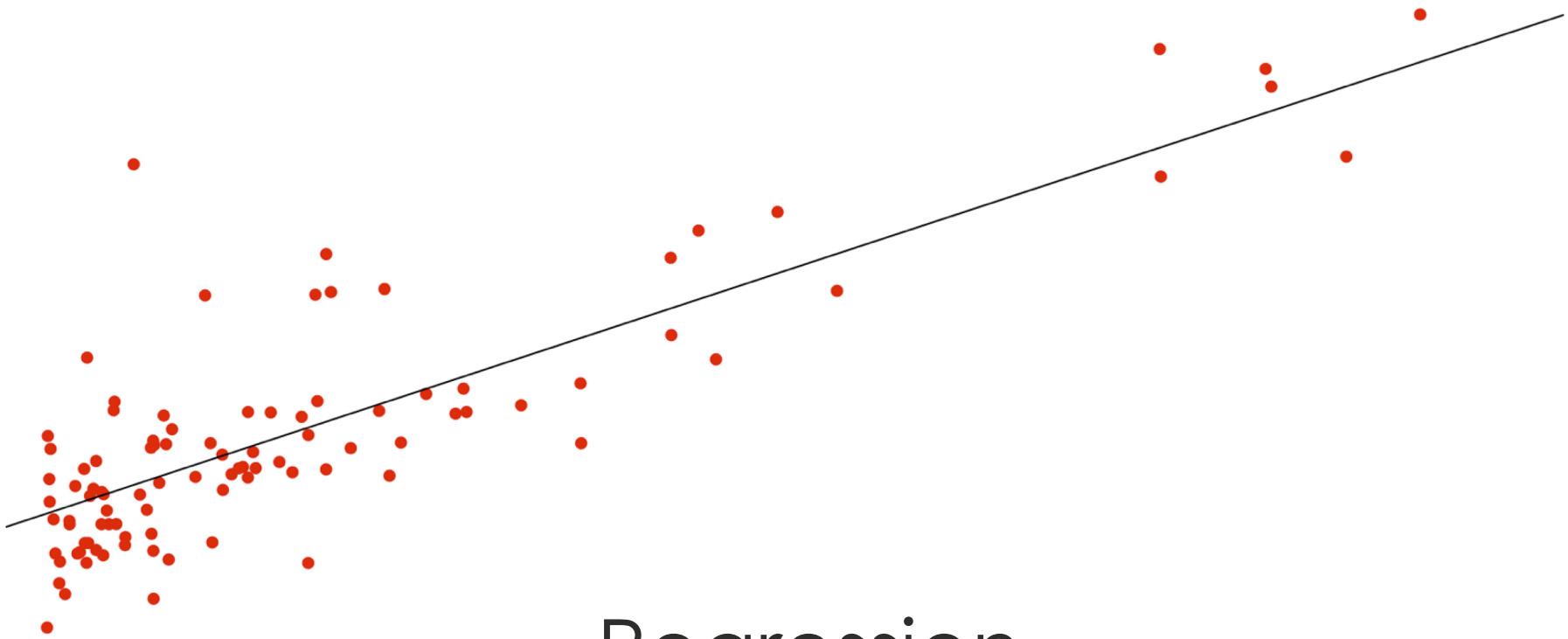
What is Machine Learning?




Features
Training Set
Testing Set
Validation Set



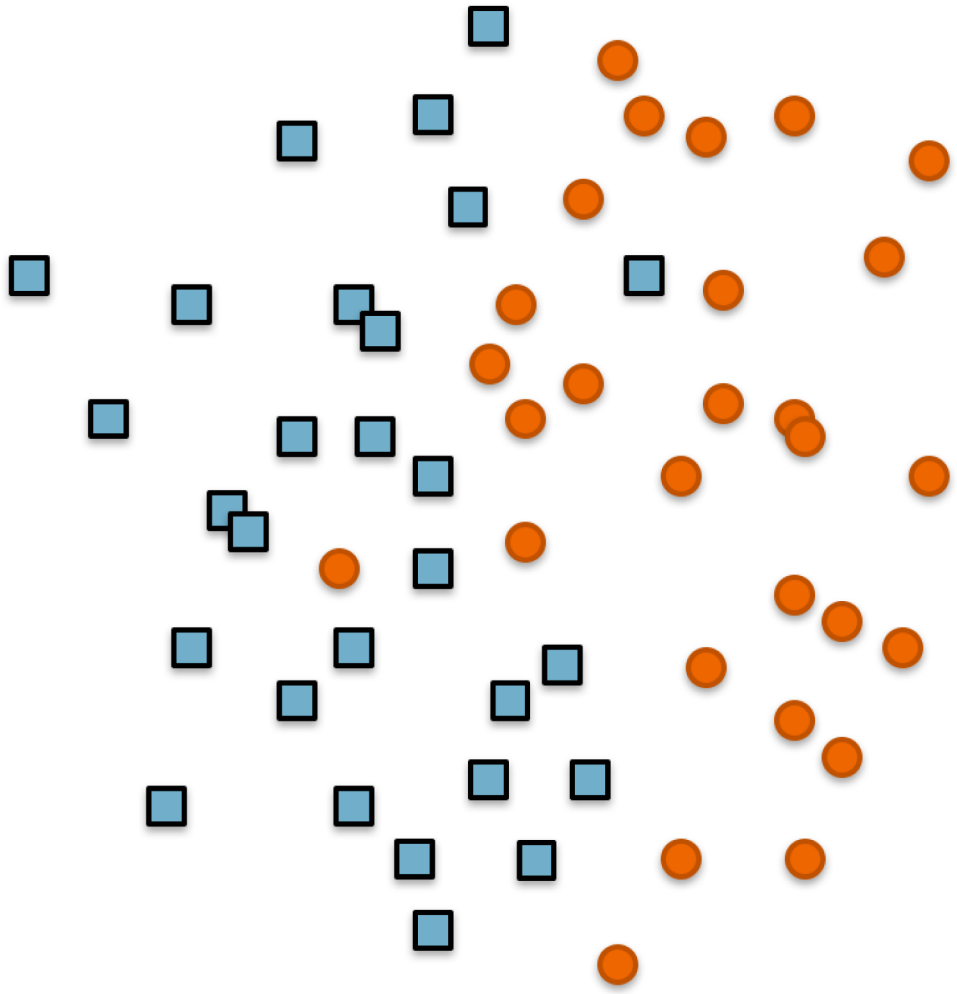
Source: H20.ai



Regression

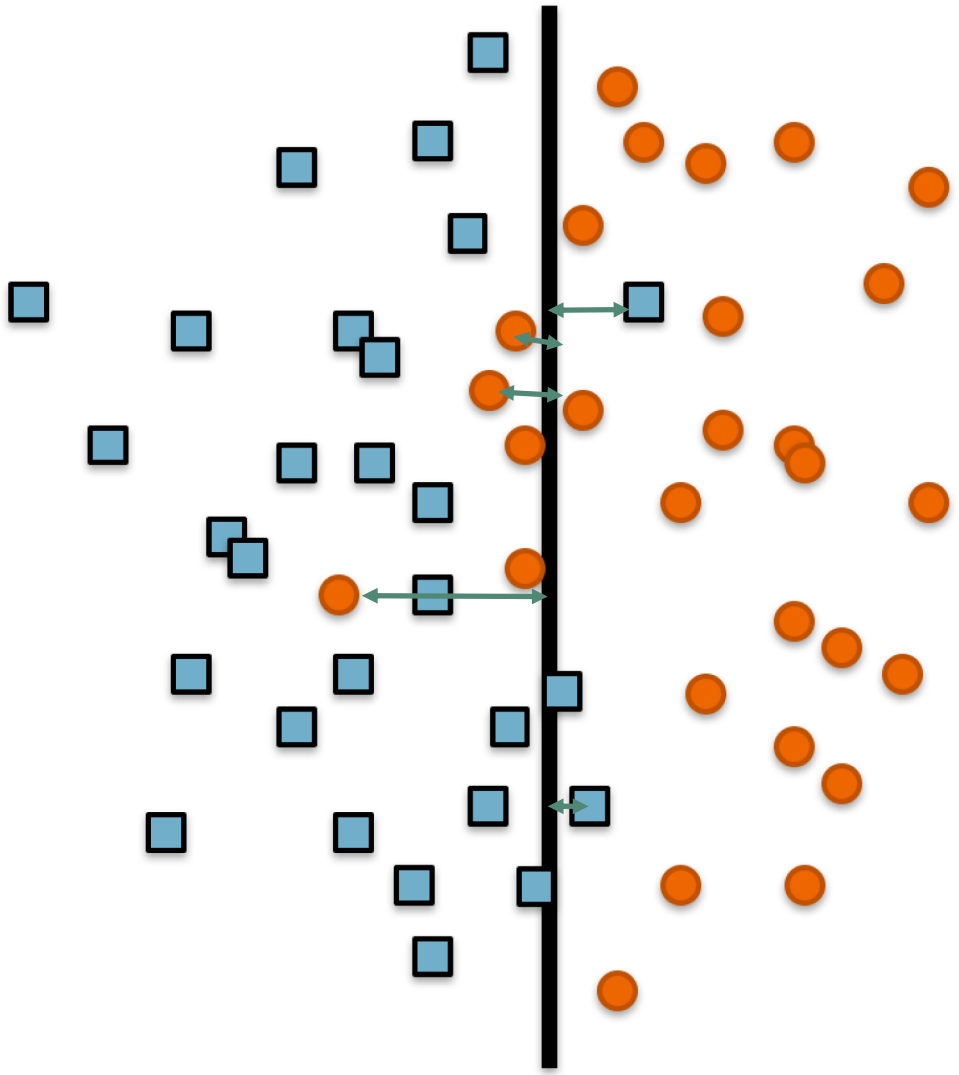
Source: H20.ai

Divide this



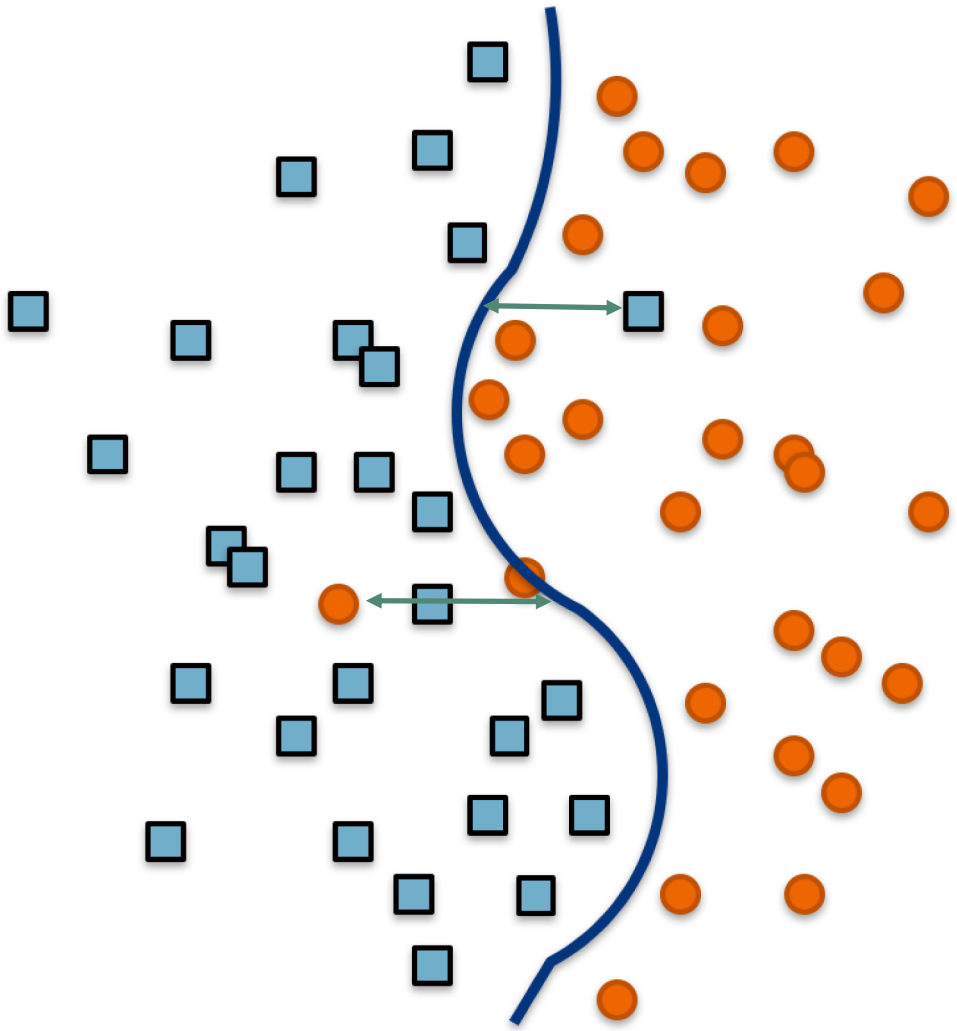
Source: H2O.ai

Start training



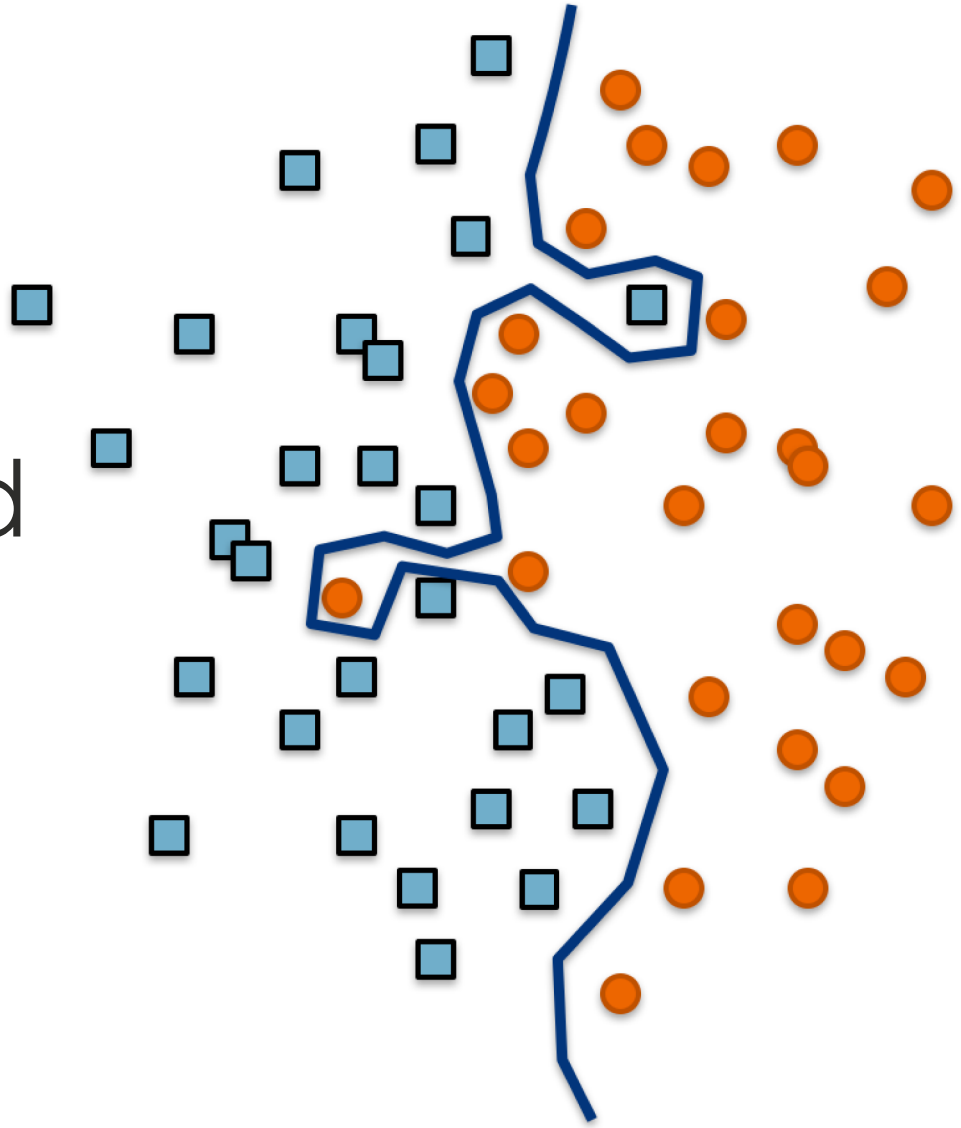
Source: H20.ai

Get better



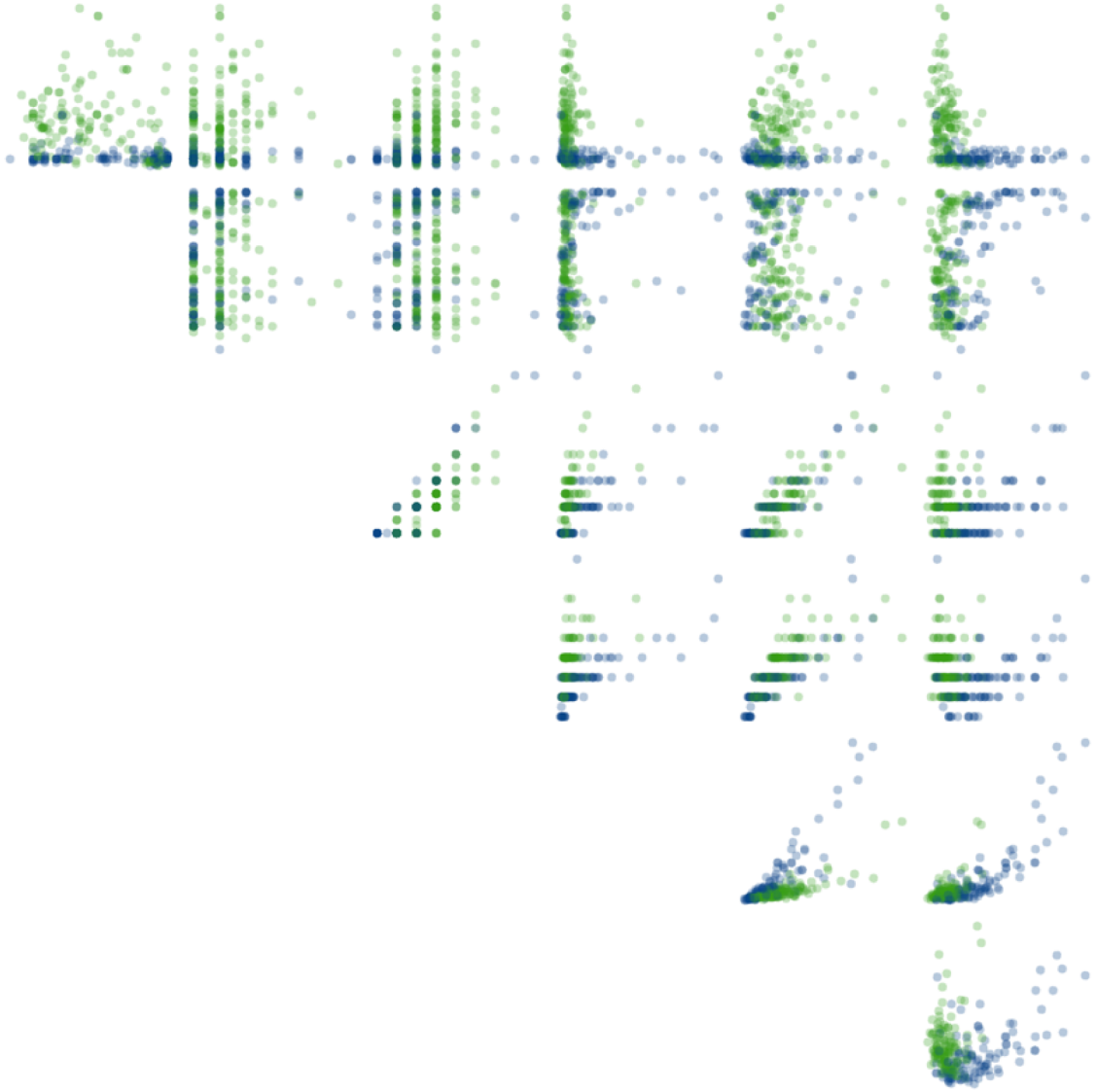
Source: H20.ai

Trying too hard (overfitting)



Source: H20.ai

Over many dimensions



Source: H20.ai

Over many machines

Source: H20.ai

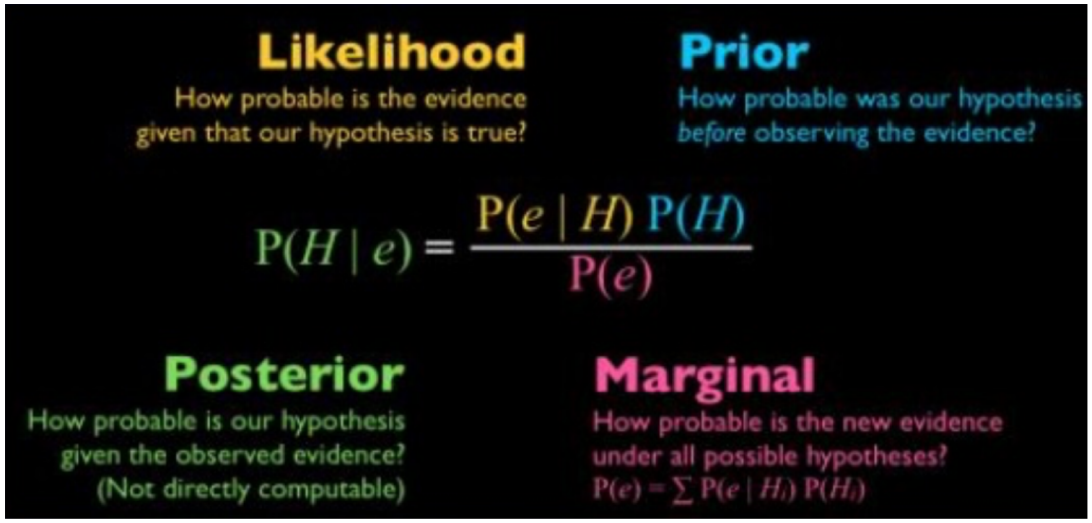
Five Tribes of Machine Learning

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

Source: Pedro Domingos

Machine Learning – Five Approaches

Bayesian Inference

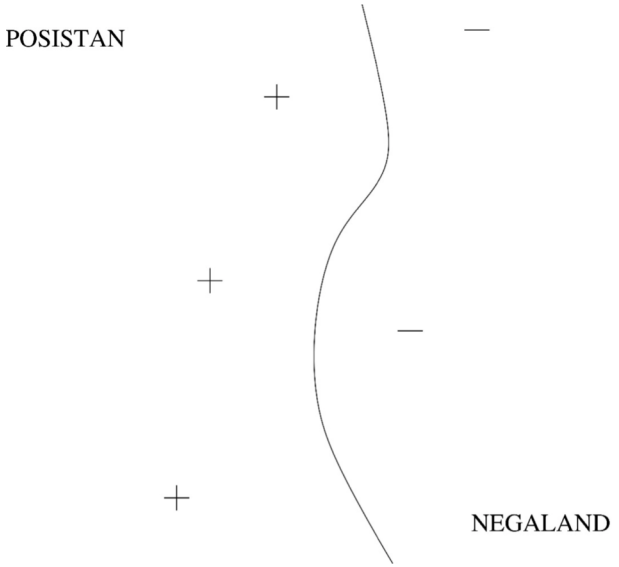


Fraud Detection
Spam Filtering

Source: Pedro Domingos

Machine Learning – Five Approaches

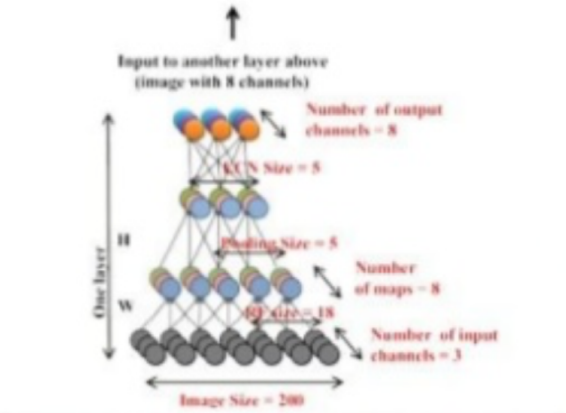
Kernel Machines



Recommender Systems – Collaborative Filtering
Clustering

Machine Learning – Five Approaches

Neural Networks
(aka Deep Learning)



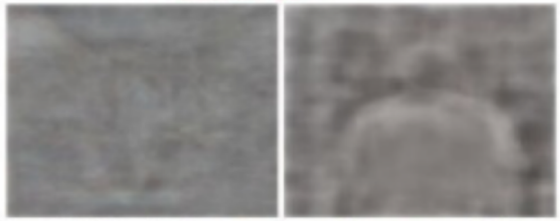
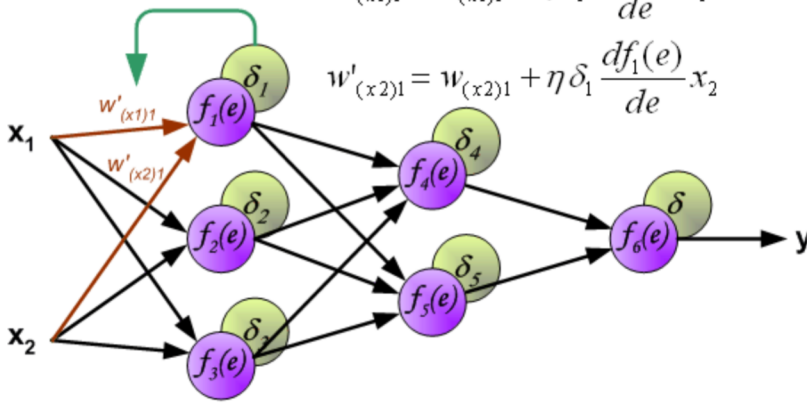


Figure 6. Visualization of the cat face neuron (left) and human body neuron (right).

www.kindcell.com

$$w'_{(x1)1} = w_{(x1)1} + \eta\delta_1 \frac{df_1(e)}{de} x_1$$

$$w'_{(x2)1} = w_{(x2)1} + \eta\delta_1 \frac{df_1(e)}{de} x_2$$



Face Recognition
Fraud Detection
Computer Vision
Speech Recognition
Natural Language Processing

Source: Pedro Domingos

Machine Learning – Five Approaches

Reverse Induction

2 + 2 = 4

2 + ? = 4

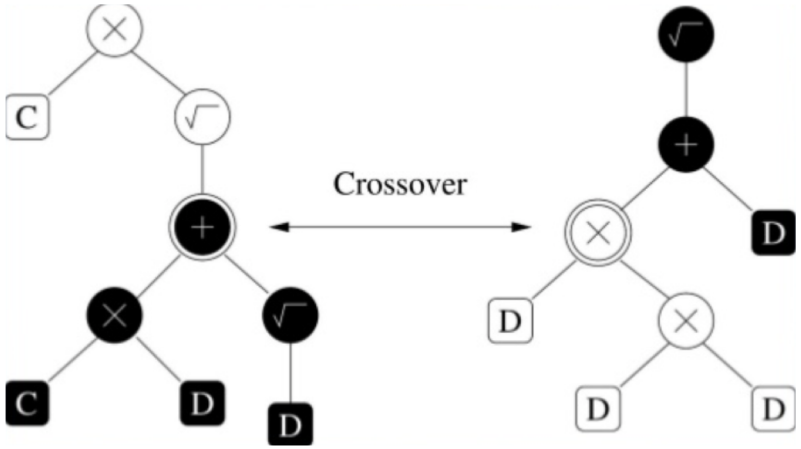
Queen
+
Woman
=
King
+
?

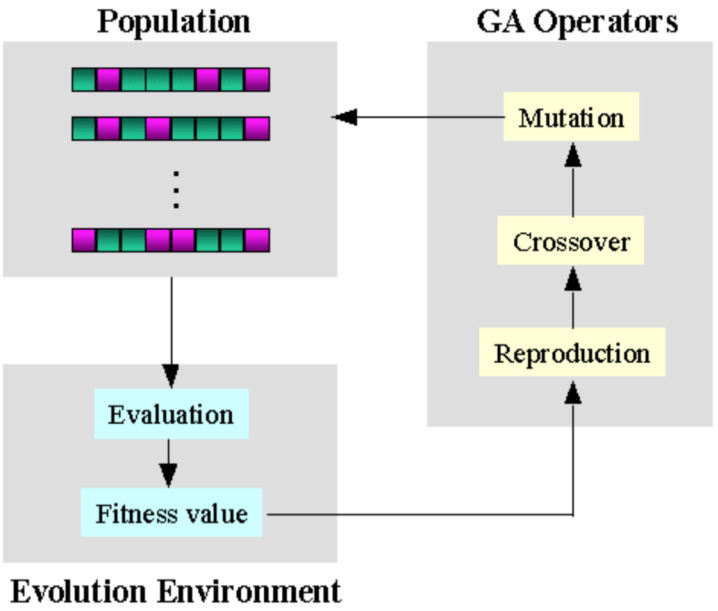
Gene Research
Dynamic Pricing

Source: Pedro Domingos

Machine Learning – Five Approaches

Genetic Programming





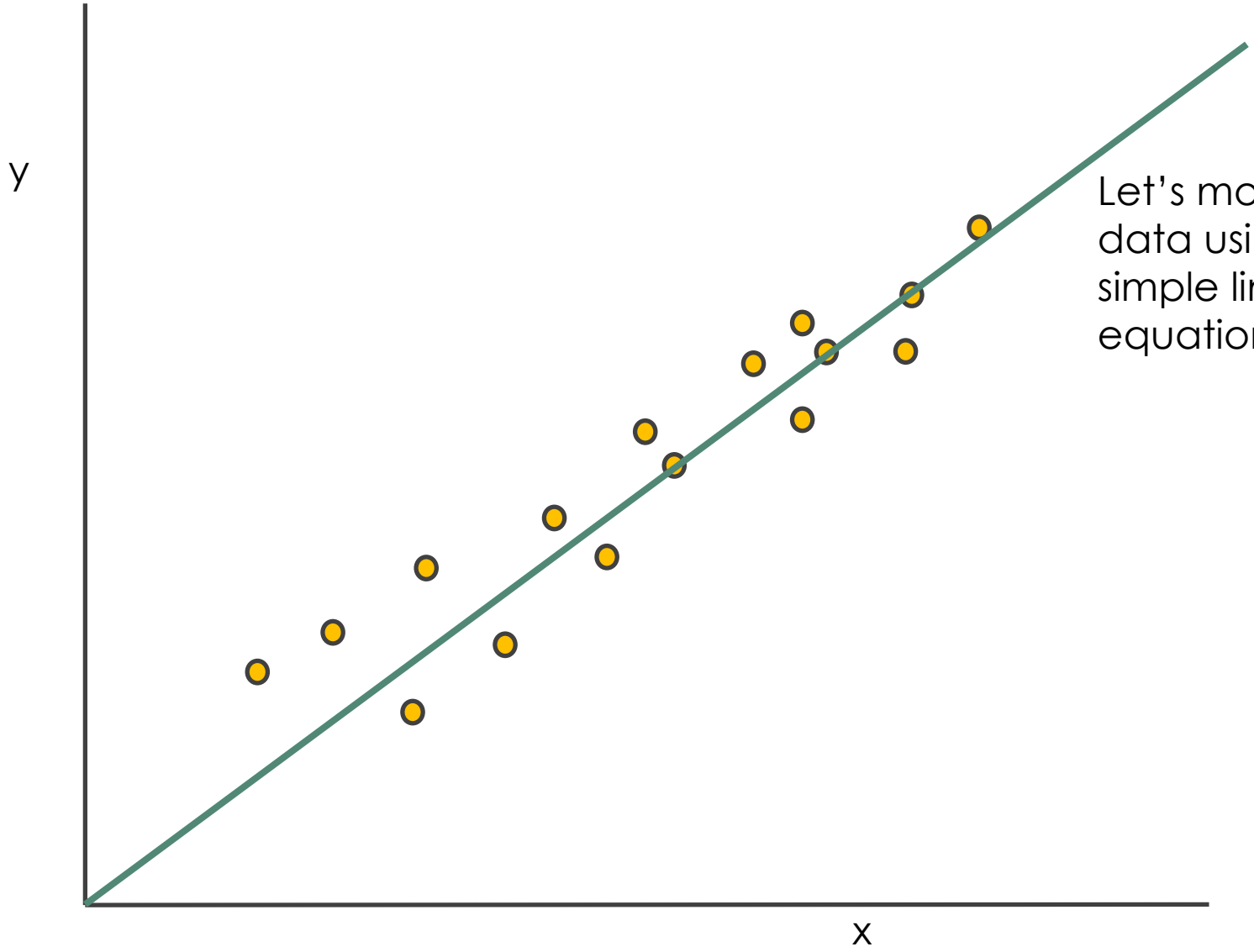
Robotics

Popular ML Algorithms

- Linear Regression
- Logistic Regression
- Random Forest
- Collaborative Filtering
- Support Vector Machines
- Deep Learning
  - Convolutional Neural Networks
  - Recurrent Neural Networks – LSTM, GRU
  - Generative Adversarial networks (GANs)
  - Deep Reinforcement Learning
- Ensemble Learning

LINEAR REGRESSION

A linear model



Let's model the data using a simple linear equation

Finding the model

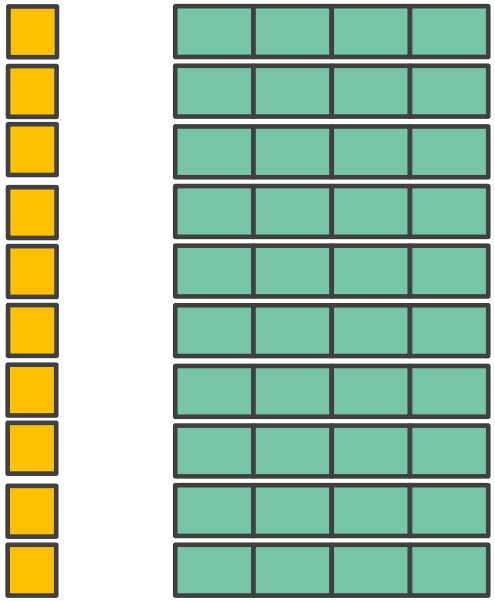
$$y = w_0 + w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + w_4 * x_4$$

Predictor

We could write this as,

$$y = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}$$

Our Model



Our Algorithm

- Our objective is to reduce the error.

```
Assume some random weights
repeat
  calculate our model y=f(x)
  come up with a way to measure error
  adjust weights in order to minimize error
until error (accuracy level) is acceptable
```

Objective function

- Our approach to measuring error and minimizing it, is called the objective function.

- Error is measured in many ways.
  This is referred to as **loss function** or **cost function**

- There some common ones,
  - Mean Square Error (MSE)
  - Mean Absolute Error
  - Logistic Error
  - Cross Entropy

Optimizing Training

- This general approach is called Gradient Descent
- Normally, this is a slow process. Some common optimizers include,
  - AdamOptimizer
  - RmsPropOptimizer
  - AdaGrad
- These optimizers find the optimal step
- Another approach that is taken is mini-batch.

LET'S TRY SOME LINEAR REGRESSION