Introduction to ML with Tensorflow.rb

by @jimmynguyc

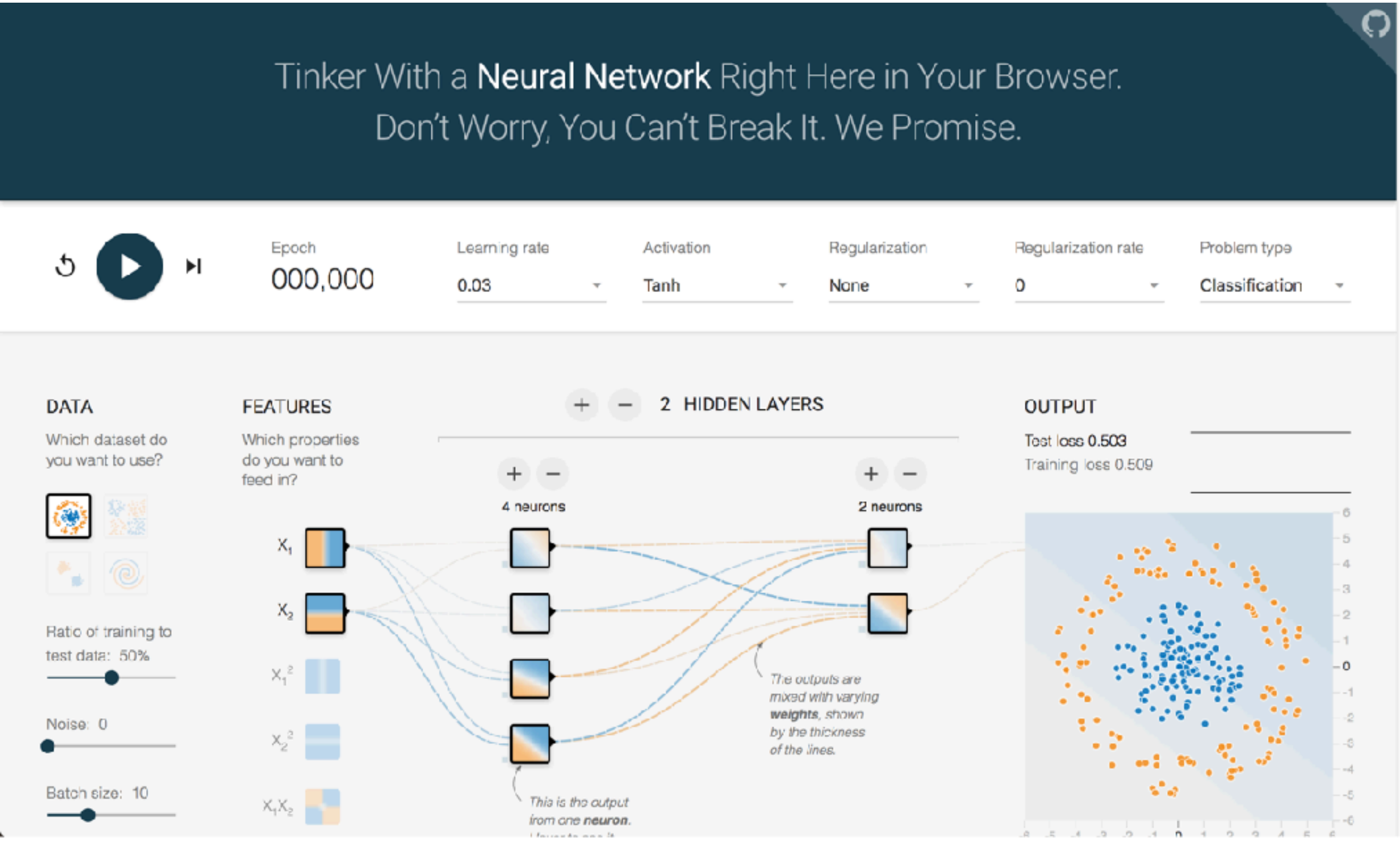
Hi

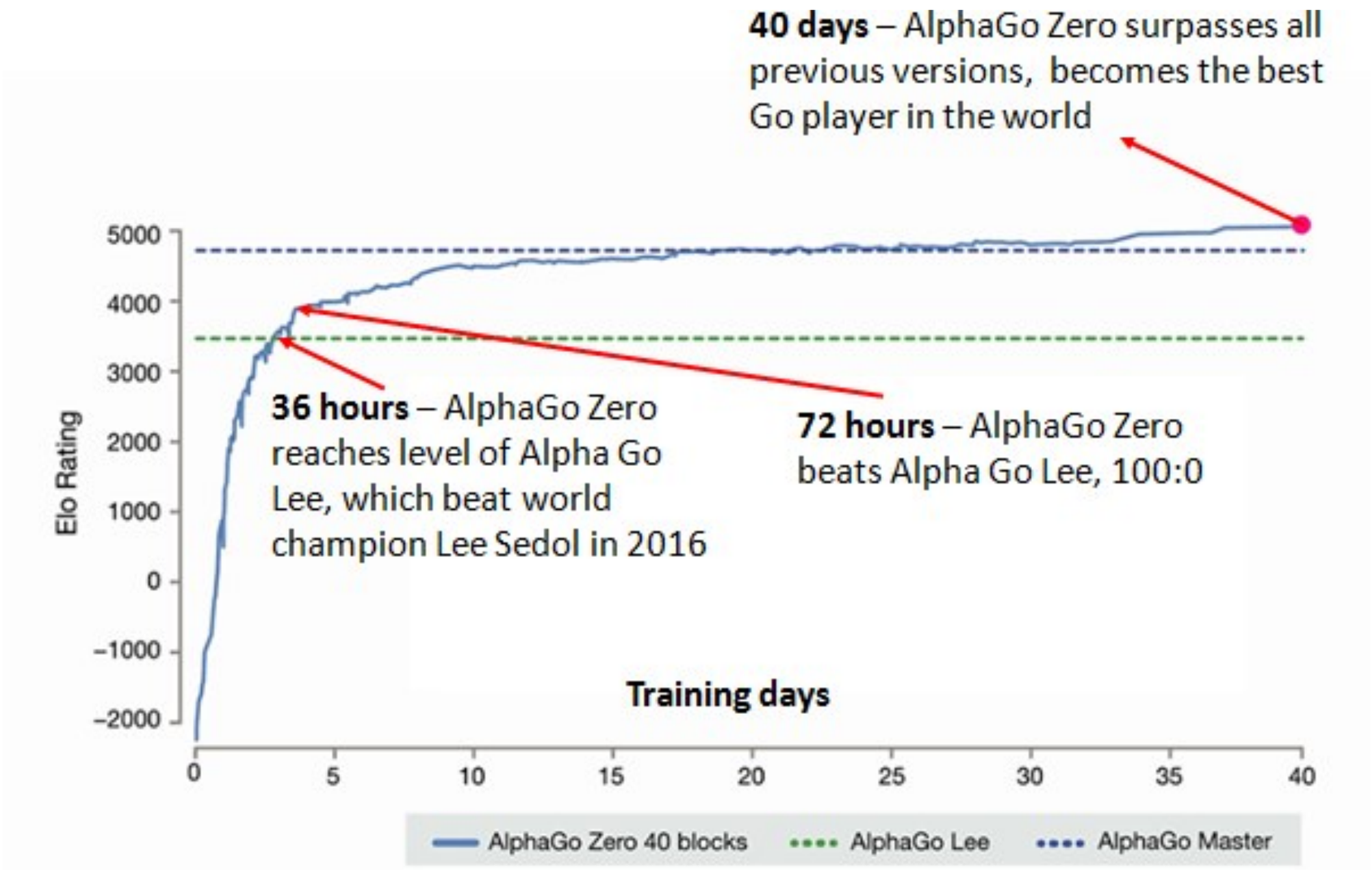
- Jimmy Ngu
- Engineering Team Lead @ RapidRiver Software
- KL Ruby Brigade, RubyConf MY
- I present a lot of useless talks :D

Motivation

- TensorFlow and Deep Learning Singapore
- A.I Day : Research - Prototype - Production
- Videos on Engineers.SG Youtube channel
- E.g. Machine Learning in Javascript, AlphaGo Zero : Under the Hood

http://playground.tensorflow.org/





Source: Mastering the Game of Go without Human Knowledge

What is ML?

- Using algorithms to parse data
- Learn from that data
- Make informed decisions based on what it has learned

Common Problems for ML

- Classification
- Regression
- Clustering

Classification Problems

- Given sets of medical parameters > Breast Cancer is benign / malignant?
- Given an image > Cat or Dog?
- Given email content > Spam or not?

Regression Problems

- Given size of house > Predict value

- Given sets of chess moves > Calculate chance of winning

Clustering Problems

- Given sets of unlabeled data > find clusters / groups

- E.g. Organizing unlabeled pictures based on characteristics (animals, faces, objects, etc)

Types of ML

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

- Given sets of input values corresponding outcomes / targets (i.e. good/correct vs bad/wrong outcomes)

- Train / generate a function that can predict outcome of undetermined input up to certain certain % of accuracy

- E.g. Handwritten Digit OCR based on MNIST dataset

Unsupervised Learning

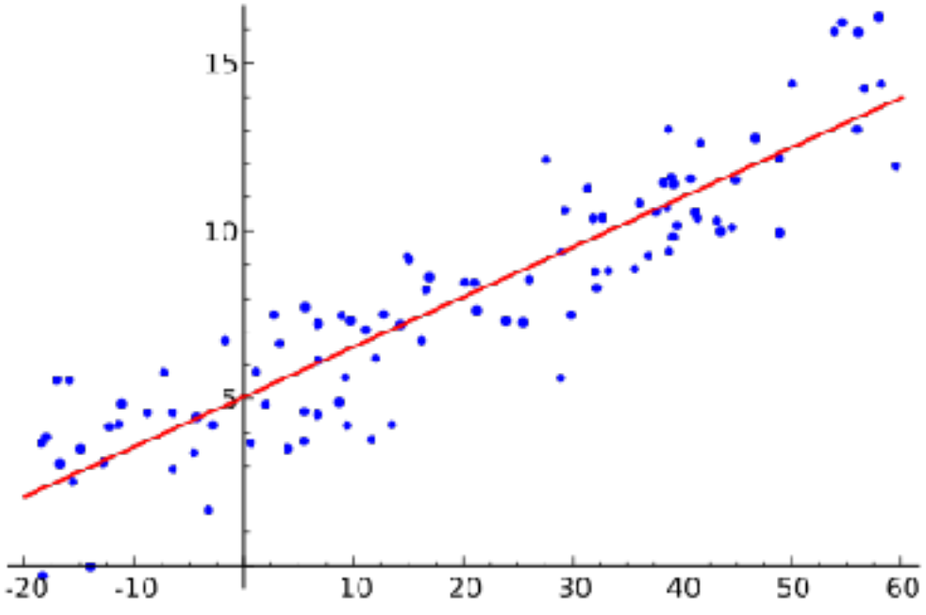
- Given sets of input values without specifying outcomes
- Find underlying structure / characteristics of datasets
- E.g. finding different groups of user behaviors from analytics

Reinforcement Learning

- Exposed to an environment
- Given ways to “interact” with the environment
- Environment return positive / negative rewards
- Learns from experience and tweaks decision making algorithm
- E.g. A program that plays Chrome’s Dino Game

Learning Algorithms

- Mathematical process of training a ML “model”
- Finds patterns in training data
- Function f(x) that takes inputs and gives expressive & useful outputs (for the problem)
- E.g. Linear regression



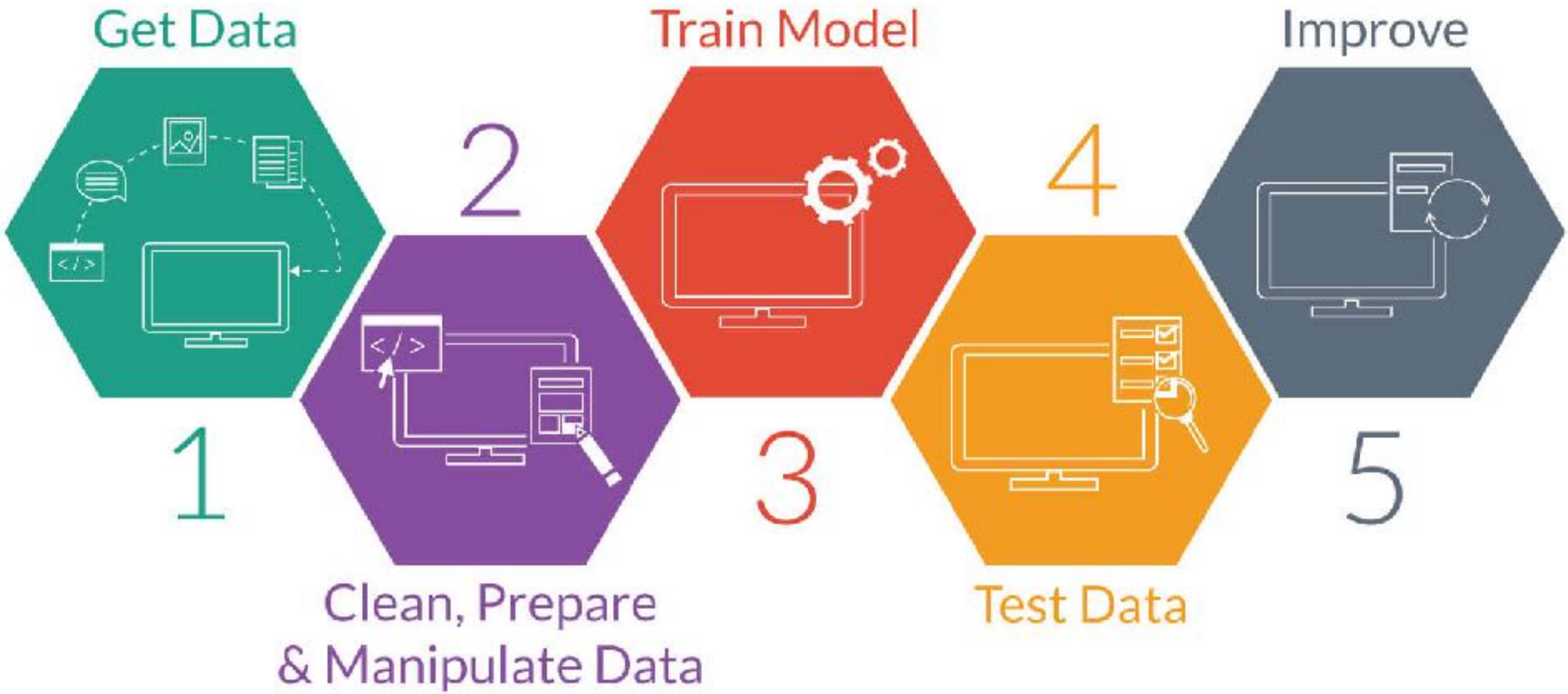
f(x) = (3x / 20) + 5

f(100) = (300 / 20) + 5 = 20 (predicted)

Machine Learning Algorithms

- Bayesian
  - Naive Bayes
  - Averaged One-Dependence Estimators (AODE)
  - Bayesian Belief Network (BBN)
  - Gaussian Naive Bayes
  - Multinomial Naive Bayes
  - Bayesian Network (BN)
- Decision Tree
  - Classification and Regression Tree (CART)
  - Iterative Dichotomiser 3 (ID3)
  - C4.5
  - C5.0
  - Chi-squared Automatic Interaction Detection (CHAID)
  - Decision Stump
  - Conditional Decision Trees
  - M5
- Dimensionality Reduction
  - Principal Component Analysis (PCA)
  - Partial Least Squares Regression (PLSR
  - Sammon Mapping
  - Multidimensional Scaling (MDS)
  - Projection Pursuit
  - Principal Component Regression (PCR)
  - Partial Least Squares Discriminant Analysis
  - Mixture Discriminant Analysis (MDA)
  - Quadratic Discriminant Analysis (QDA)
  - Regularized Discriminant Analysis (RDA)
  - Flexible Discriminant Analysis (FDA)
  - Linear Discriminant Analysis (LDA)
- Instance Based
  - k-Nearest Neighbour (kNN)
  - Learning Vector Quantization (LVQ)
  - Self-Organizing Map (SOM)
  - Locally Weighted Learning (LWL)
- Clustering
  - k-Means
  - k-Medians
  - Expectation Maximization
  - Hierarchical Clustering
- Deep Learning
  - Deep Boltzmann Machine (DBM)
  - Deep Belief Networks (DBN)
  - Convolutional Neural Network (CNN)
  - Stacked Auto-Encoders
- Ensemble
  - Random Forest
  - Gradient Boosting Machines (GBM)
  - Boosting
  - Bootstrapped Aggregation (Bagging)
  - AdaBoost
  - Stacked Generalization (Blending)
  - Gradient Boosted Regression Trees (GBRT)
- Neural Networks
  - Radial Basis Function Network (RBFN)
  - Perceptron
  - Back-Propagation
  - Hopfield Network
- Regularization
  - Ridge Regression
  - Least Absolute Shrinkage and Selection Operator (LASSO)
  - Elastic Net
  - Least Angle Regression (LARS)
- Rule System
  - Cubist
  - One Rule (OneR)
  - Zero Rule (ZeroR)
  - Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
- Regression
  - Linear Regression
  - Ordinary Least Squares Regression (OLSR)
  - Stepwise Regression
  - Multivariate Adaptive Regression Splines (MARS)
  - Locally Estimated Scatterplot Smoothing (LOESS)
  - Logistic Regression

Steps to Predictive Modelling



ML Models

- Training artifact of learning algorithms

- E.g. https://github.com/tensorflow/models

TensorFlow

- An open source machine learning framework
- Libraries for handling datasets, graphs, matrix / tensor operations
- GPU / TPU support
- Visualizations
- Saving / restoring models

tensorflow.rb

- TensorFlow for Ruby (duh …)
- https://github.com/somaticio/tensorflow.rb

- https://medium.com/@Arafat./image-recognition-in-ruby-tensorflow-df5d5c05389b

Demo

References

- http://playground.tensorflow.org/
- https://www.tensorflow.org/
- Engineers.SG Youtube Channel (https://www.youtube.com/channel/UCjRZr5HQKHVKP3SZdX8y8Qw)
- https://medium.com/applied-data-science/how-to-build-your-own-alphazero-ai-using-python-and-keras-7f664945c188
- https://medium.com/@Arafat./image-recognition-in-ruby-tensorflow-df5d5c05389b

Thank You ~