

Learning Machine Learning with Kaggle Challenges

(1) Introduction

Qiyang Hu
IDRE

About this series

- Not a complete course
 - No derivation of theories
 - Not covering every ML field
 - Not a guide for library (sklearn, tensorflow)
 - Not discussing an award-level Kaggle ranking
- Instead, we will
 - Give descriptive review (avoid math!)
 - Touch selected topics
 - Combine with slides and demos
- Expectations:
 - For beginners: get a general idea, flatten the steep learning curve
 - For experts: overview the knowledge structure, seek the collaborations

The series needs
your feedback!

Survey link:

<https://forms.gle/t3f8CztFQpeFFksy6>



Syllabus of the series

1. [Introduction to Machine Learning](#) (Oct 24)
2. [Classification](#) (Oct 31)
 - General machine learning and Scikit Learn
3. [Deep learning \(1\)](#) (Nov 7)
 - General Deep learning and Tensorflow 2.0
 - Convolutional Neural Networks
4. [Deep learning \(2\)](#) (Nov 14)
 - Data augmentation
 - Save/load models
 - Transfer learning
5. Deep learning (3) RNNs (*TBD*)
6. Reinforcement Learning — PPO (*TBD*)

The Series's
Github Repo

https://github.com/huqy/idre_learning_machine_learning



Learning Resources

- [Google Machine Learning Crash Course](#)
- Andrew Ng's Machine learning
 - [Coursera](#) or [Youtube](#)
- Aurélien Géron's Book:
 - ["Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems" 2nd Edition](#)
- Coding Tensorflow:
 - [Youtube](#)
 - [Udacity](#)
- Prakashan's Machine Learning/Deep Learning Session
 - [Notes on Google Sites](#)

ARTIFICIAL INTELLIGENCE

Any technique which enables computers to mimic human behavior



MACHINE LEARNING

AI techniques that give computers the ability to learn without being explicitly programmed to do so



DEEP LEARNING

A subset of ML which make the computation of multi-layer neural networks feasible



1950's

1960's

1970's

1980's

1990's

2000's

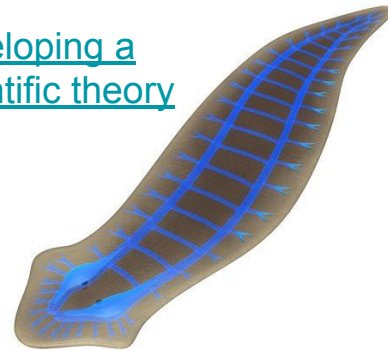
2010's

Some Amazing Machine Learning Achievements



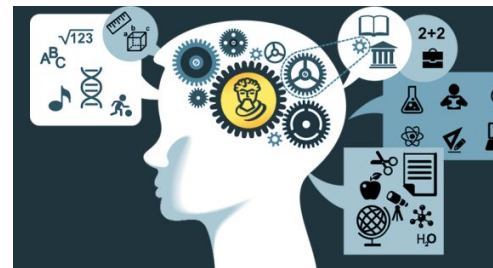
Lip Read

Developing a scientific theory



CycleGAN

Passing 8th grade Sci Exam

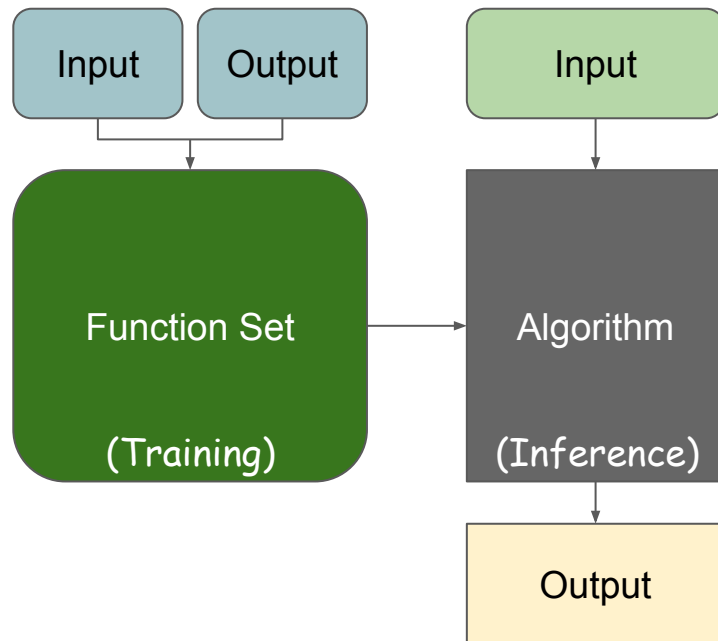


What is Machine Learning?

Traditional Programming



Machine Learning



Hard-coded AI vs. Deep Learning AI



Key Terminology in Machine Learning

- Datasets:
 - Label: a desired output
 - Feature: a known input
- Model: relationship between input & output
 - Parameter: to be learned from data, e.g. weight, coefficients
 - Weight: a coefficient for a feature in linear model
 - Bias: an intercept or offset from an origin
 - Hyperparameter: often set by heuristics, e.g. learning rate, depth of trees, batch, epoch.
 - Batch: a subset from the division of training datasets
 - Epoch: all data in training sets has had an opportunity to update the internal model parameters

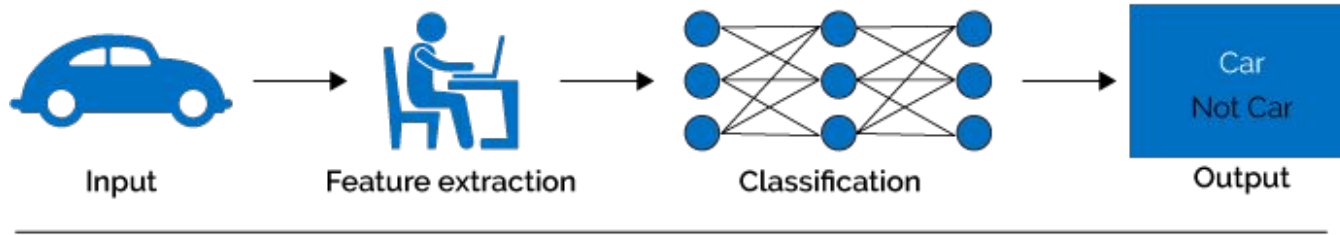
[Complete Glossary](#)

A lot of “Learning”s to learn

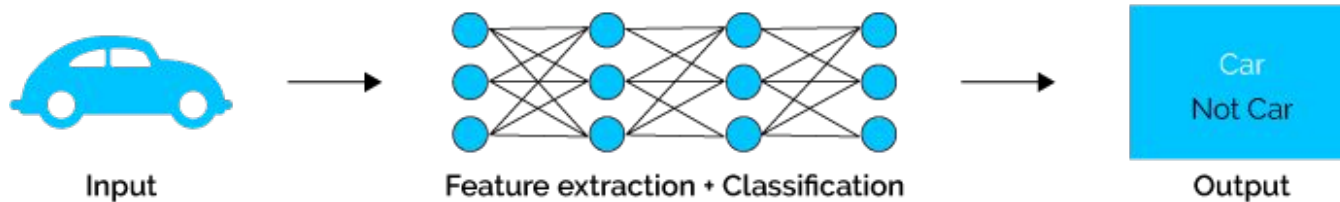
- Supervised Learning (data with labels)
 - Regression
 - Classification (SVM, Decision Tree, K-NN, **Deep Learning**)
- Unsupervised Learning (data without labels) (PCA, Clustering, Factor Analysis)
- Semi-supervised Learning (data with partial labels)
- Reinforcement Learning (reward rules to get data) (PPO, Deep Q-learning)
- Inverse reinforcement learning (no rules & no labels)
- Transfer Learning (data with unrelated labels)
 - (zero-shot learning, one-shot learning, few-shot learning, etc.)
 - ⇒ Continuous learning
 - ⇒ Meta Learning (MAML, LSTM)

Classical Machine Learning vs. Deep Learning

Classical Machine Learning



Deep Learning



Source: <https://www.xenonstack.com/blog/log-analytics-deep-machine-learning/>

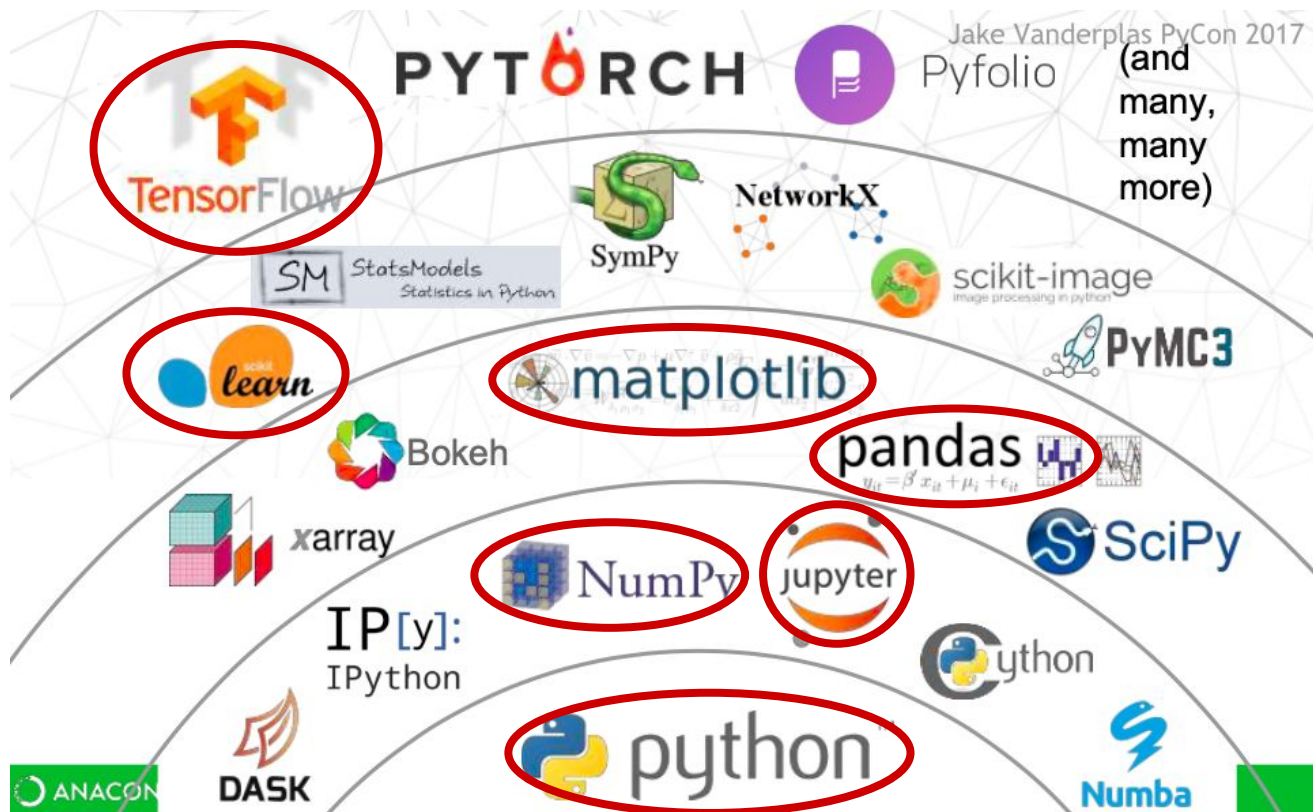
Machine Learning vs. Statistics

- Commons: same/interchangeable concepts & techniques:

Machine Learning	Statistics
Learning	Fitting
Supervised Learning	Regression/Classification
Unsupervised Learning	Clustering/Density Estimation

- Differences: [source](#)
 - Prediction vs. Explanation
 - Forward vs. Rearward Looking
 - Big vs. Small Data
 - Many vs. Few Variables

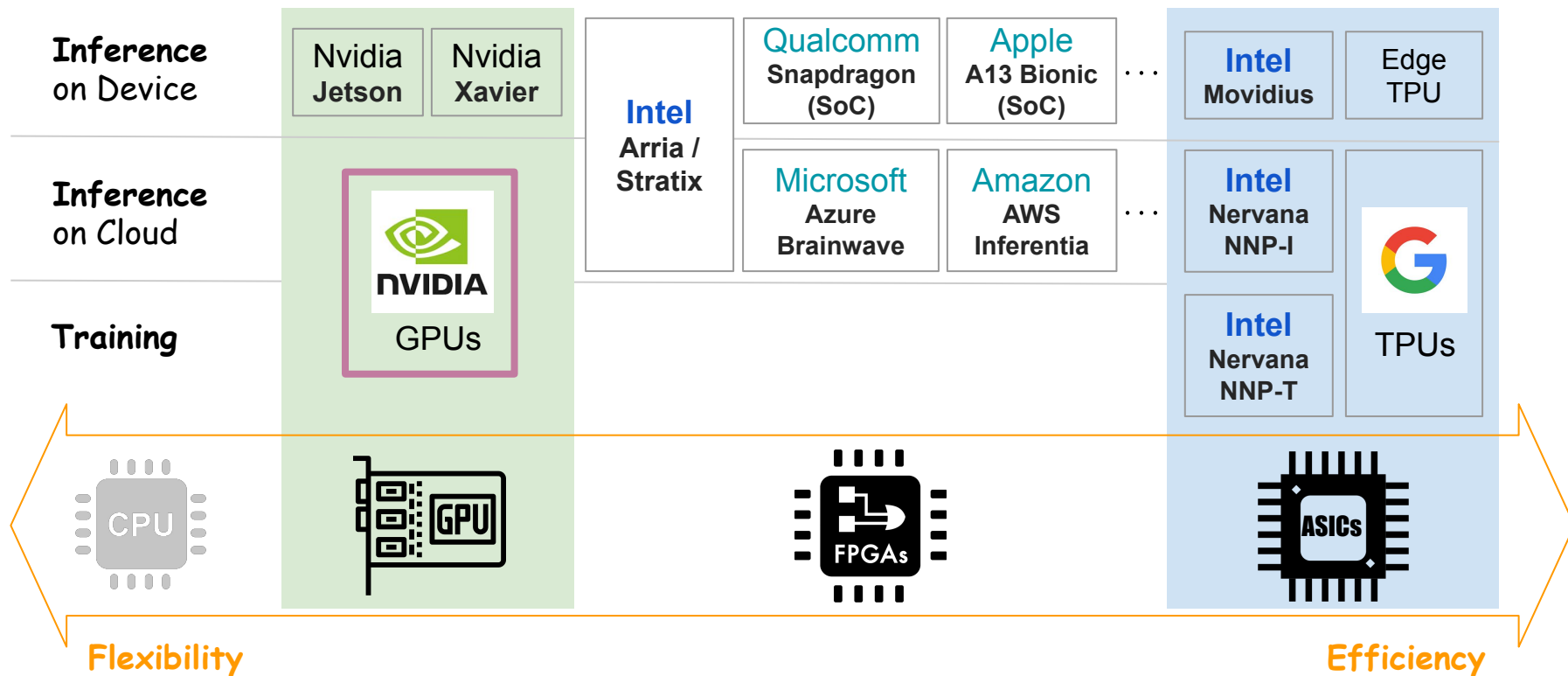
Python Scientific Ecosystem



What is **kaggle** ?

- An online platform and community for data scientists and machine learner
 - 1,000,000+ registered users in 194 countries in 2017
 - Founded in 2010, acquired by Google in 2017
 - Hosts 19K+ of datasets and 200K+ code snippets
 - Offers a cloud-based workbench with computational resources
 - Famous for the competitions with high rewards (accessible to anyone)
- Kaggle competitions ([active list](#))
 - **Featured**: full-scale, commercially-purposed, offering high prizes (e.g. from Lyft, Zillow...)
 - **Research**: experimental, usually no prizes (e.g. from Google, wikipedia, ...)
 - **Get-started**: tutorialized, easiest (e.g. [Titanic](#))
 - **Playground**: “for fun” (e.g. [Dogs-vs-Cats](#))
 - **Other types**: for recruitment, annual...

Machine Learning hardwares (AI Chips)



Available Free GPU Computation Resources

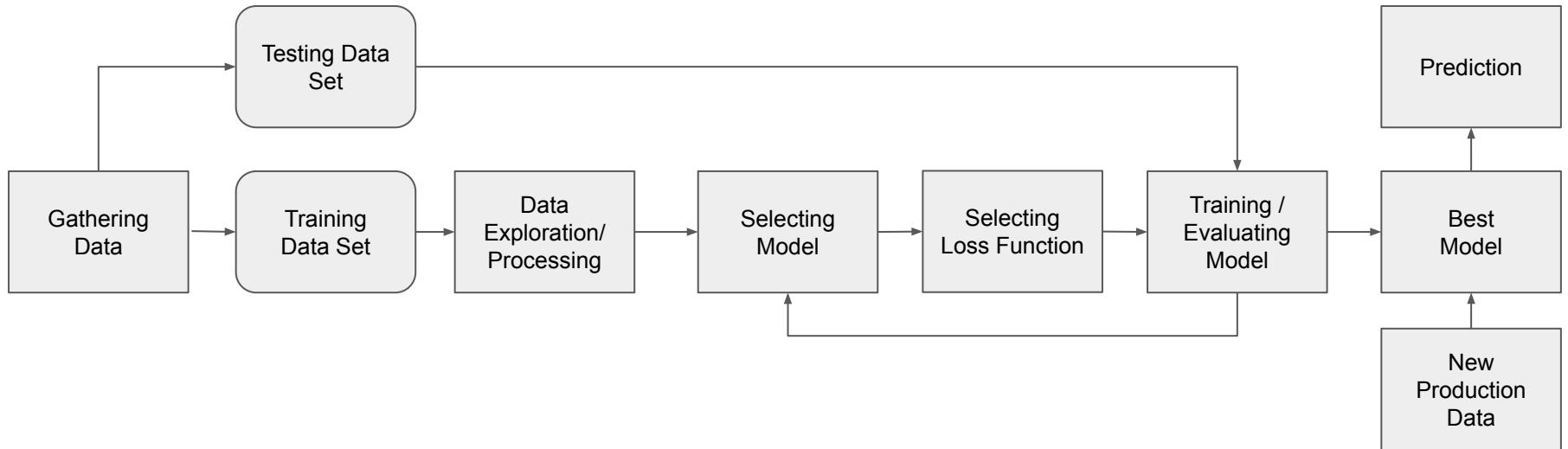
- Google Colaboratory
 - Google Drive -> New -> More -> Google Colaboratory
- Kaggle
 - Kaggle.com -> Log in -> Kernel -> New Kernel
- Hoffman2
 - Download [h2jupyterb](#)
 - `chmod +x h2jupyterb`
 - `./h2jupyterb -u [username] -t 8 -m 4 -s 8 -v anaconda3 -g yes`
 - [Info](#) about GPU resource on H2

	Colab	Kaggle	Hoffman2
CPU Type	Intel Xeon 2.30GHz	Intel Xeon 2.30GHz	Intel Xeon 2.80GHz
Slots/Threads available	1 core / 2 threads	1 core / 2 threads	8 cores w/o hyp-threads
RAM available	12 GB	18 GB	24 GB
Disk available	311 GB	626 GB	1 TB
GPU Type	Tesla T4 (2018)	Tesla P100 (2018)	Tesla P4 (2016)
GPU SP Floating-Point Perf	8.1 TFLOPs	10.6 TFLOPs	5.5 TFLOPs
GPU Memory	16 GB	16 GB	8 GB
Training Time Limit	8 hours	6 hours	24 hours

Before running the colab demos in this series

1. Register a Kaggle account
 - a. Kaggle.com → “Register”
2. Create Kaggle API token and download json file
 - a. Sign in → Your Profile → “My Account” → “Create New API Token”
3. Join the 2 competitions → “Join Competition”
 - a. [Titantic Challenge](#)
 - b. [Dogs-vs-Cats Challenge](#)
4. Get/run the colab files
 - a. Git clone the github [repo](#) and copy to google drive
 - b. Visit the github [repo](#) and open it directly (using chrome extension “[Open in Colab](#)”)

Workflow for a machine learning project



Don't forget

- Sign in your info to the class
 - To get the email notifications
- Fill out the survey to leave comments:
 - <https://forms.gle/t3f8CztFQpeFFksy6>



- Contact me if any questions or discussions
 - hugy@idre.ucla.edu
 - Office: Math Sci #3330
 - Phone: 310-825-2011