



CRAY

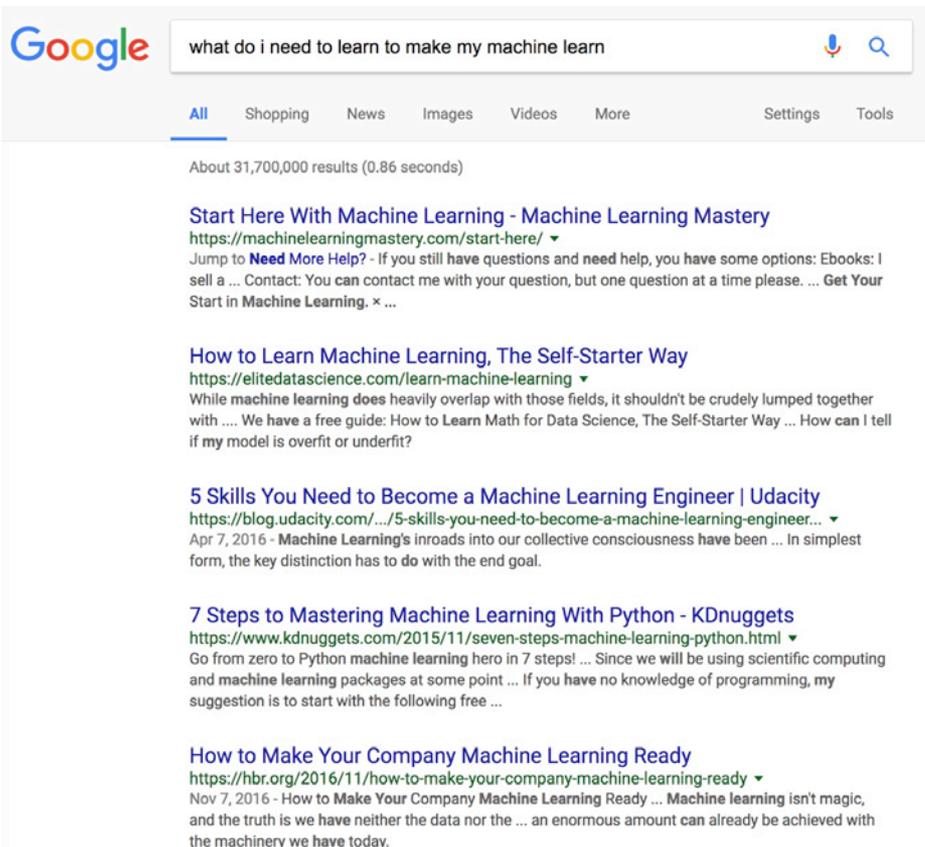
**What do I need to learn to make my
Machine learn?**

**Geert Wenes
Cray Inc.**

Thank you:

THE SOCIETY OF
HPC
PROFESSIONALS

OK, Google – or Alexa, or Siri

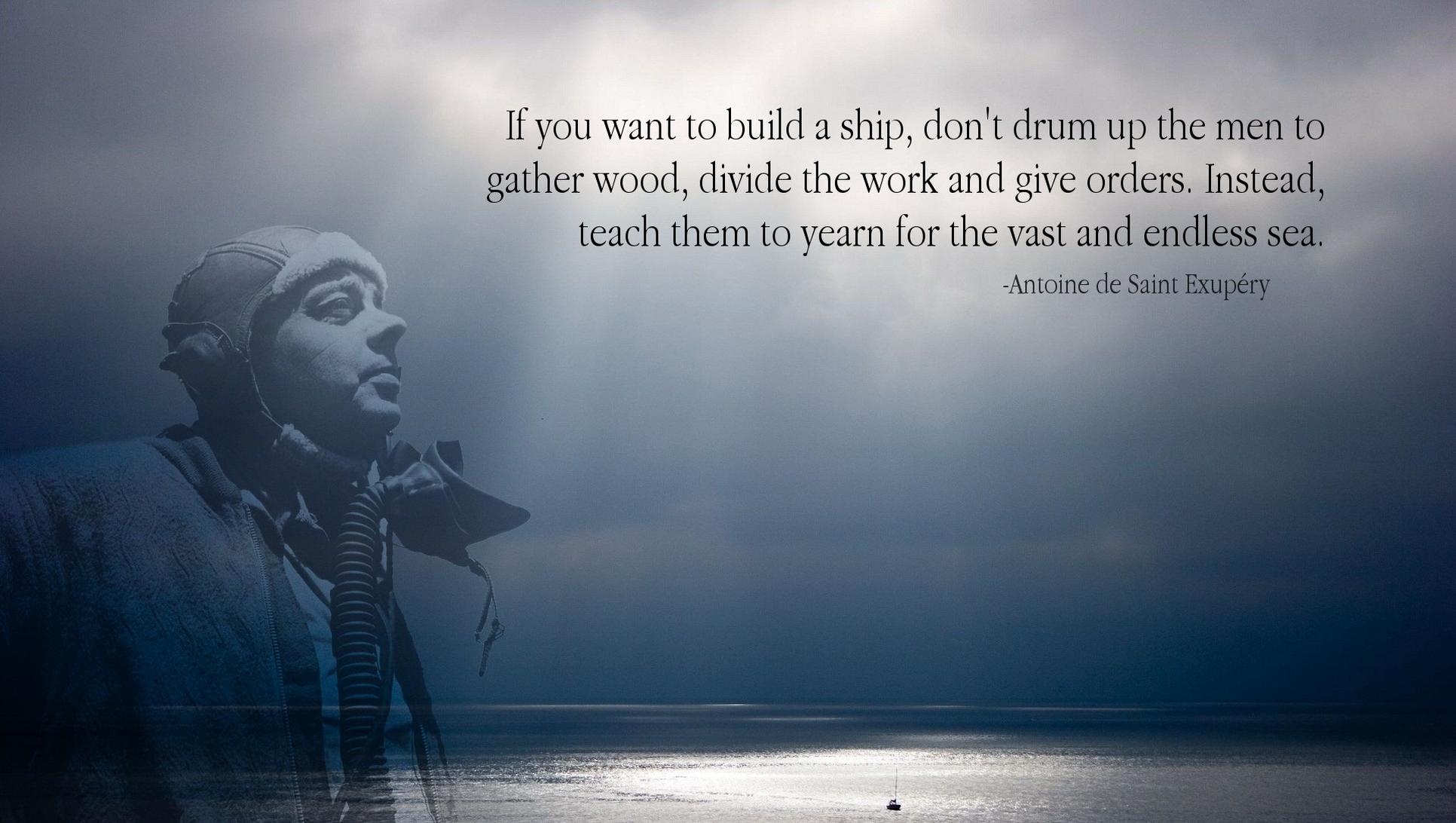


Google search results for "what do i need to learn to make my machine learn".

About 31,700,000 results (0.86 seconds)

- Start Here With Machine Learning - Machine Learning Mastery**
<https://machinelearningmastery.com/start-here/> ▼
Jump to **Need More Help?** - If you still have questions and need help, you have some options: Ebooks: I sell a ... Contact: You can contact me with your question, but one question at a time please. ... Get Your Start in Machine Learning. × ...
- How to Learn Machine Learning, The Self-Starter Way**
<https://elitedatascience.com/learn-machine-learning> ▼
While machine learning does heavily overlap with those fields, it shouldn't be crudely lumped together with ... We have a free guide: How to Learn Math for Data Science, The Self-Starter Way ... How can I tell if my model is overfit or underfit?
- 5 Skills You Need to Become a Machine Learning Engineer | Udacity**
<https://blog.udacity.com/.../5-skills-you-need-to-become-a-machine-learning-engineer...> ▼
Apr 7, 2016 - Machine Learning's inroads into our collective consciousness have been ... In simplest form, the key distinction has to do with the end goal.
- 7 Steps to Mastering Machine Learning With Python - KDnuggets**
<https://www.kdnuggets.com/2015/11/seven-steps-machine-learning-python.html> ▼
Go from zero to Python machine learning hero in 7 steps! ... Since we will be using scientific computing and machine learning packages at some point ... If you have no knowledge of programming, my suggestion is to start with the following free ...
- How to Make Your Company Machine Learning Ready**
<https://hbr.org/2016/11/how-to-make-your-company-machine-learning-ready> ▼
Nov 7, 2016 - How to Make Your Company Machine Learning Ready ... Machine learning isn't magic, and the truth is we have neither the data nor the ... an enormous amount can already be achieved with the machinery we have today.

- High-throughput in the inference stage
But takes quite an infrastructure to get there)
- Results' ordering
 - Not bad but literal
 - Training, training, training
 - Bias?
 - My personal search history
 - Somebody paid for it?
 - Search Engine Optimization (ESO)

A man wearing a flight helmet and a jacket is shown in profile, looking out over a vast ocean at sunset. The sun is low on the horizon, creating a bright glow and reflecting on the water. A small sailboat is visible in the distance on the right side of the frame.

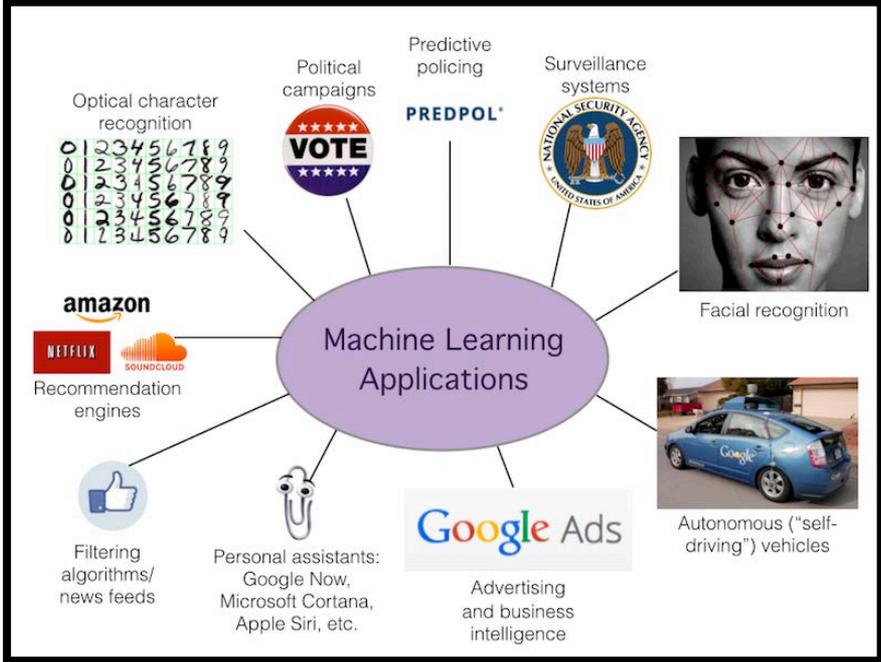
If you want to build a ship, don't drum up the men to
gather wood, divide the work and give orders. Instead,
teach them to yearn for the vast and endless sea.

-Antoine de Saint Exupéry

For a Company ... :

How much is 0.1% worth?
83,000,000 members + modest growth
\$10 * 12 months

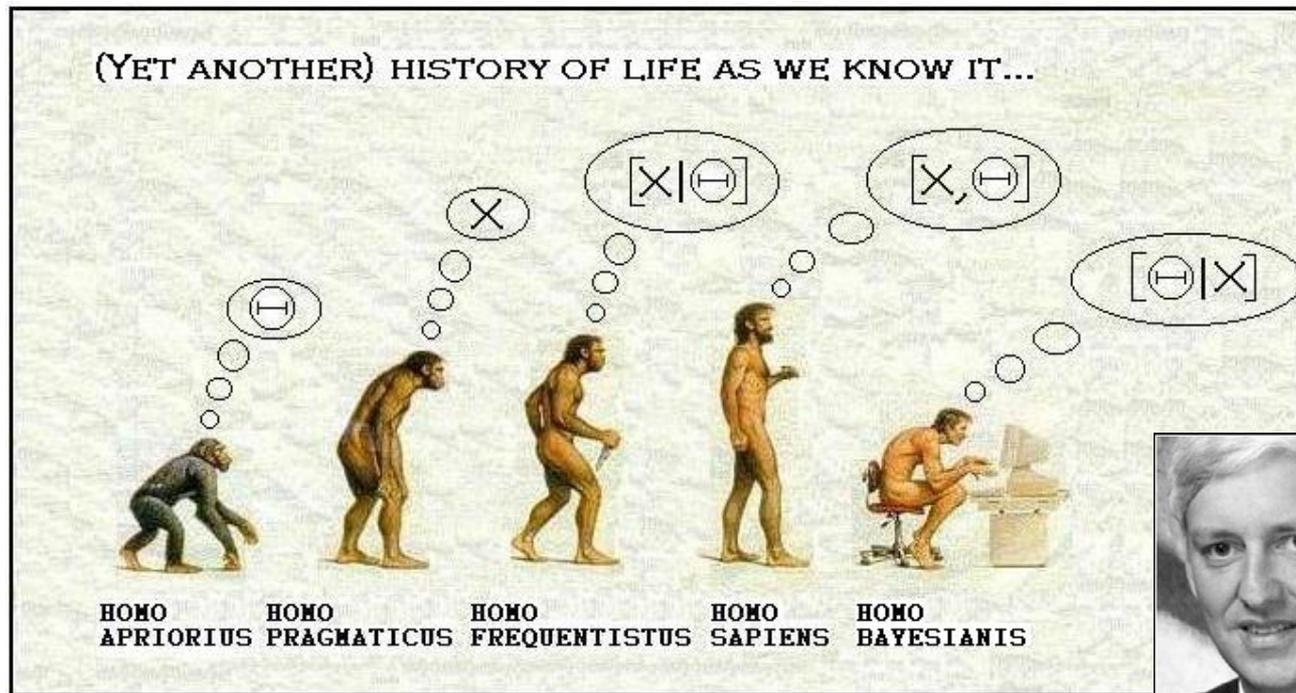
~5-100+ million a year



What is machine learning?



Have a machine learn the model from the data



All models are wrong, and increasingly you can succeed without them.

— Peter Norvig —

AZ QUOTES

COMPUTE | STORE | ANALYZE

What is Deep Learning (DL) ?



ARTIFICIAL INTELLIGENCE		Design of intelligent systems that augment human productivity		
Sense		Comprehend	Predict	Act and Adapt
ANALYTICS		MACHINE LEARNING		
Search datasets for insights		Learn patterns from the past to predict future		
Descriptive	What happened?	Unsupervised Group, cluster and organize content with domain-specific heuristic models.	Supervised Train mathematical predictive models with labelled data	
Diagnostic	Why did it happen?		DEEP LEARNING	
Predictive	What will happen?	Train and use neural networks as a predictive model		
Prescriptive	How to make it happen?	Vision	Speech	Language

COMPUTE | STORE | ANALYZE

It's a Zoo out there



Algorithms – of the non-linear type, for highly dimensional data

- New algorithms
- Old algorithms revisited



BUILD YOUR CAREER IN AI

Take our new Deep Learning courses, now open on Coursera

Enroll Now

This course provides a broad introduction to machine learning, datamining, and statistical pattern recognition. Topics include: (i) Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks). (ii) Unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning). (iii) Best practices in machine learning (bias/variance theory; innovation process in machine learning and AI). The course will also draw from numerous case studies and applications, so that you'll also learn how to apply learning algorithms to building smart robots (perception, control), text understanding (web search, anti-spam), computer vision, medical informatics, audio, database mining, and other areas.



ANDREW NG

In five courses, you will learn the foundations of Deep Learning, understand how to build neural networks, and learn how to lead successful machine learning projects. You will learn about Convolutional networks, RNNs, LSTM, Adam, Dropout, BatchNorm, Xavier/He initialization, and more. You will work on case studies from healthcare, autonomous driving, sign language reading, music generation, and natural language processing. You will master not only the theory, but also see how it is applied in industry. You will practice all these ideas in Python and in TensorFlow, which we will teach.

ANALYZE

Deep Learning := Highly Evolved Type of Machine Learning

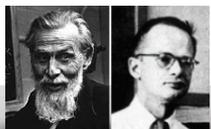


Electronic brain Perceptron ADALINE XOR Backpropagation SVM Deep Learning

Golden Age AI Winter

1943 1957 1960 1969 1986 1995 2006

1940 1950 1960 1970 1980 1990 2000 2010



S. McCulloch – W. Pitts



F. Rosenblatt



B. Widrow – M. Hoff



M. Minsky – S. Papert



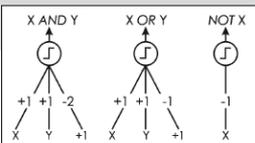
D. Rumelhart – G. Hinton – R. Williams



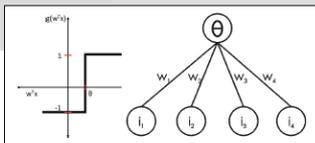
V. Vapnik – C. Cortes



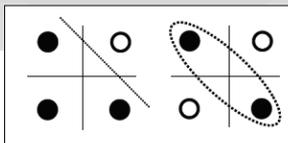
G. Hinton – S. Ruslan



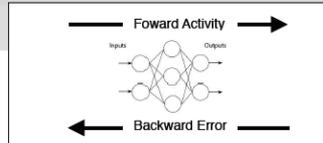
Adjustable weights
Weights are not learned



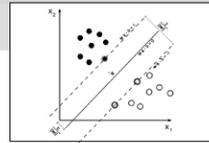
Learnable weights and
threshold



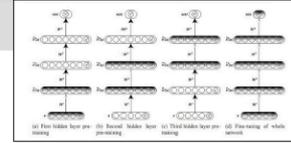
XOR Problem



Solution to nonlinearly separable
problems
Big computation, local
optima/overfitting



Limitations of learning
prior
Kernel function: Human
intervention



Hierarchical feature learning

COMPUTE | STORE | ANALYZE

HPC Professionals' Environment

Who uses supercomputers upstream?

- **The Wizards**

- First principles, direct physics
- Modeling and simulation

- **The Statisticians**

- Model builders & - refinement
- Uncertainty quantification

- **The Forecasters**

- Ad hoc analysis
- Trends and predictions



Competitive advantage
~ Integration level

Stochastic Optimization
(including variability)

Optimization
(cost criteria → best outcome)

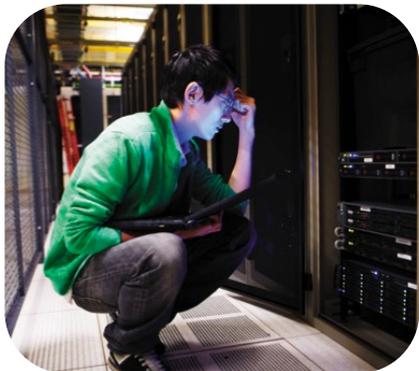
Predictive Modeling
(forward/backward/iterative)

Statistical analysis
(trends, features)

COMPUTE | STORE | ANALYZE

Complexity

Challenges



- “AI systems still demand considered design, knowledge engineering and model building”, Forrester AI TechRadar Q1 2017
- A lot to learn for practitioners and end-users:
 - Large, complex workflows
 - Different ML Toolkits + Data Movement + Network
 - Defining the value returned to the business
- Real data sets and large scale workloads are challenging libraries, implementations and HW:
 - Fake Data / Small Data have negative influence on performance optimization targets
- Machine Learning is changing how people think about HPC:
 - Data Movement, Workload Resiliency, etc.
 - Performance Optimizations

Best Practices

ML@Scale

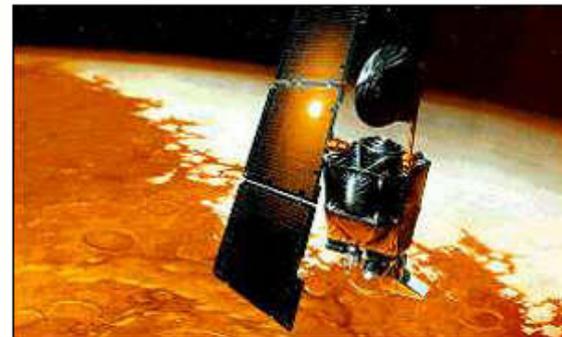
BBC NEWS

News in Audio News in Video Newyddion Hobb

Thursday, September 30, 1999 Published at 18:53 GMT 19:53 UK

Sci/Tech

Confusion leads to Mars failure



The Mars Climate Orbiter: Now in pieces on the planet's surface

The Mars Climate Orbiter Spacecraft was lost because one Nasa team used imperial units while another used metric units for a key spacecraft operation.

$$\min_m \left\| F(m) - d \right\|^2 + \lambda \left(\left\| \Theta \nabla(m - m_r) \right\|_{L^1} + \int_{\Omega} P \cdot \nabla(m - m_r) \right)$$

data misfit

weighted TV

Steerable Variation



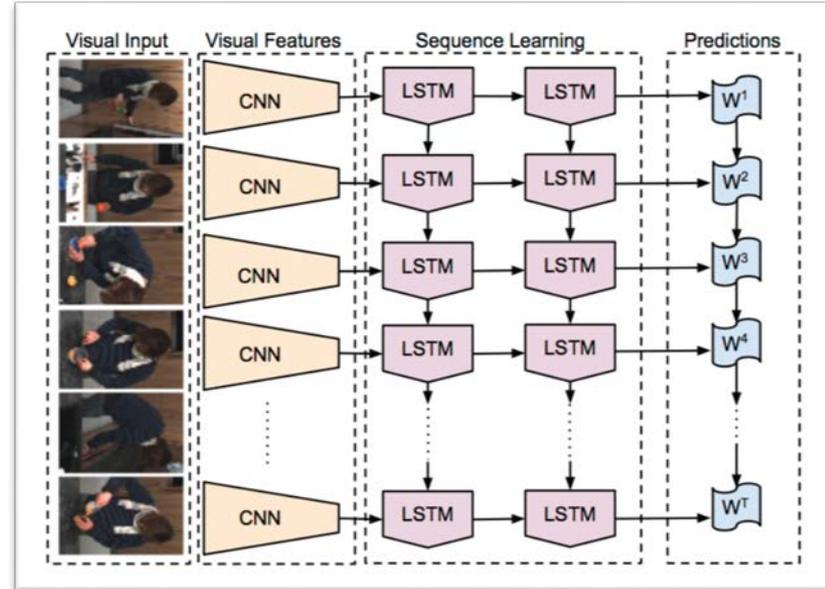
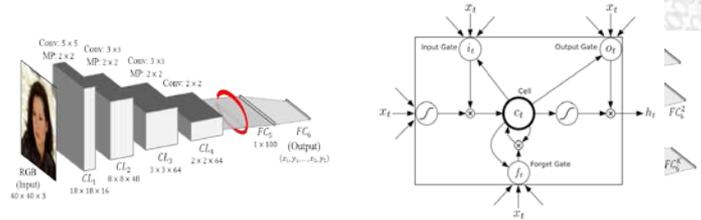
Algorithm Selection

Focus on the Data/Example

- Scientific Images (Radar Data, other)
 - Convolutions/Spatial
- At different heights
 - Tensor
- Regularly refreshed
 - Time Series/Temporal

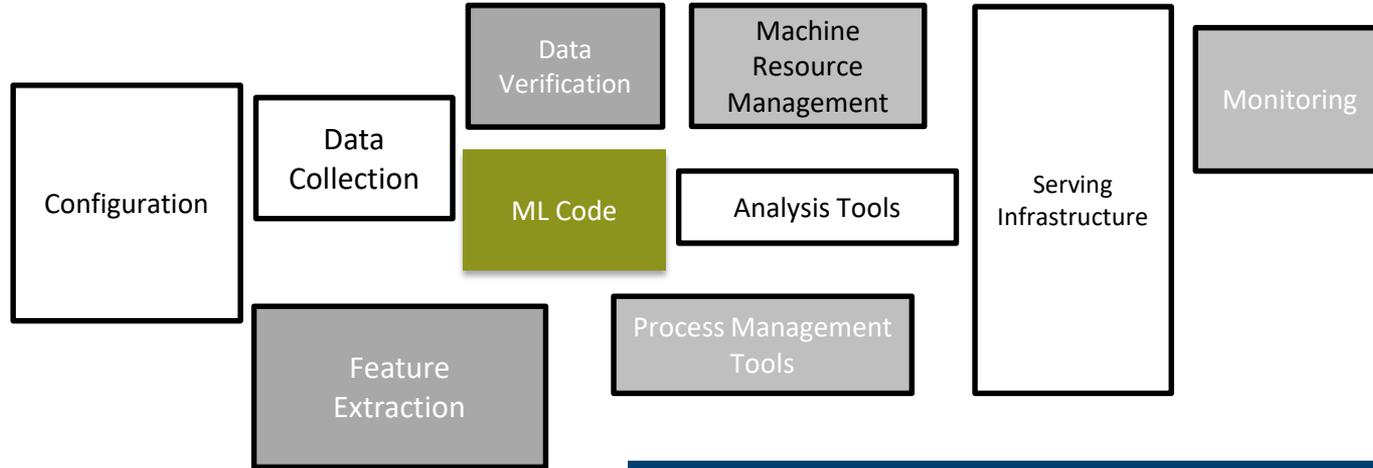
Algorithm: ConvLSTM

Almost all DL Nets (and model DBs) are part of almost all available packages, e.g. DIGITS, ...





Infrastructure and Integration



“Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.”

-Adapted from *Hidden Technical Debt in Machine Learning Systems*,
Sculley et. al., NIPS '15

Future Proofing



COMPUTE | STORE | ANALYZE

But ...



It takes all the running you can do, to keep in the same place.

If you want to get somewhere else, you must run at least twice as fast as that.

The Red Queen
Lewis Carroll
Through the Looking Glass



Figures-of-merit	State-of-practice	In 2-5 years (projected/expected)
Training-time to best accuracy	<4 days	2+ hours
Model Cost / TB (AWS GPUs)	~\$25K (ResNet training on 80 GPUs for 5 days)	~10K
Hardware Efficiency	O(~25 Gflops) Network Depth: Flops::20x: 16x (based on AlexNet-2012 and ResNet-2015)	O(Teraflops)
Statistical Efficiency	O(~25 Gflops) Depth: Accuracy:: 20x:13+ (based on AlexNet-2012 and ResNet-2015)	O(Teraflops)
Need for compute as data grows	O(~465 Gflops) Data: Flops: Accuracy:: 2x: 5x: 3+ (based on DeepSpeech1 and DeepSpeech2)	O(Petaflops)
Model creativity	Trial and error (e.g. Resnet, Inception, etc.)	Reconfigurable, Self-tuning (e.g. Ensemble, Model-of-models, etc.)
Training Cadence	~ Monthly	~ Daily
# of models per organization	1x	10-100x

COMPUTE | STORE | ANALYZE

The punchline: Deep Learning is a High Performance Computing problem

- **Delivers benefits similar to HPC in other disciplines**
 - The value is in the decisions that are enabled
- **Characterized by the same underlying factors**
 - Large amount of computation
 - Large amount of data motion (I/O and network)
- **The same methods work**
 - HPC Technology and HPC Best Practice apply directly to DL



Andrew Ng ✓
@AndrewYNg

Follow

Why's HPC/High Performance Computing speeding up deep learning research?
youtu.be/c_55gZfUK1E

