Machine Learning Primer

Machine Learning Evolution

Machine Learning Use Cases

Machine Learning Ecosystem

Machine Learning Algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Lesson Objectives

- See the potential of machine learning
- Get the basic vocabulary
- Overview of major machine learning algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Machine Learning!

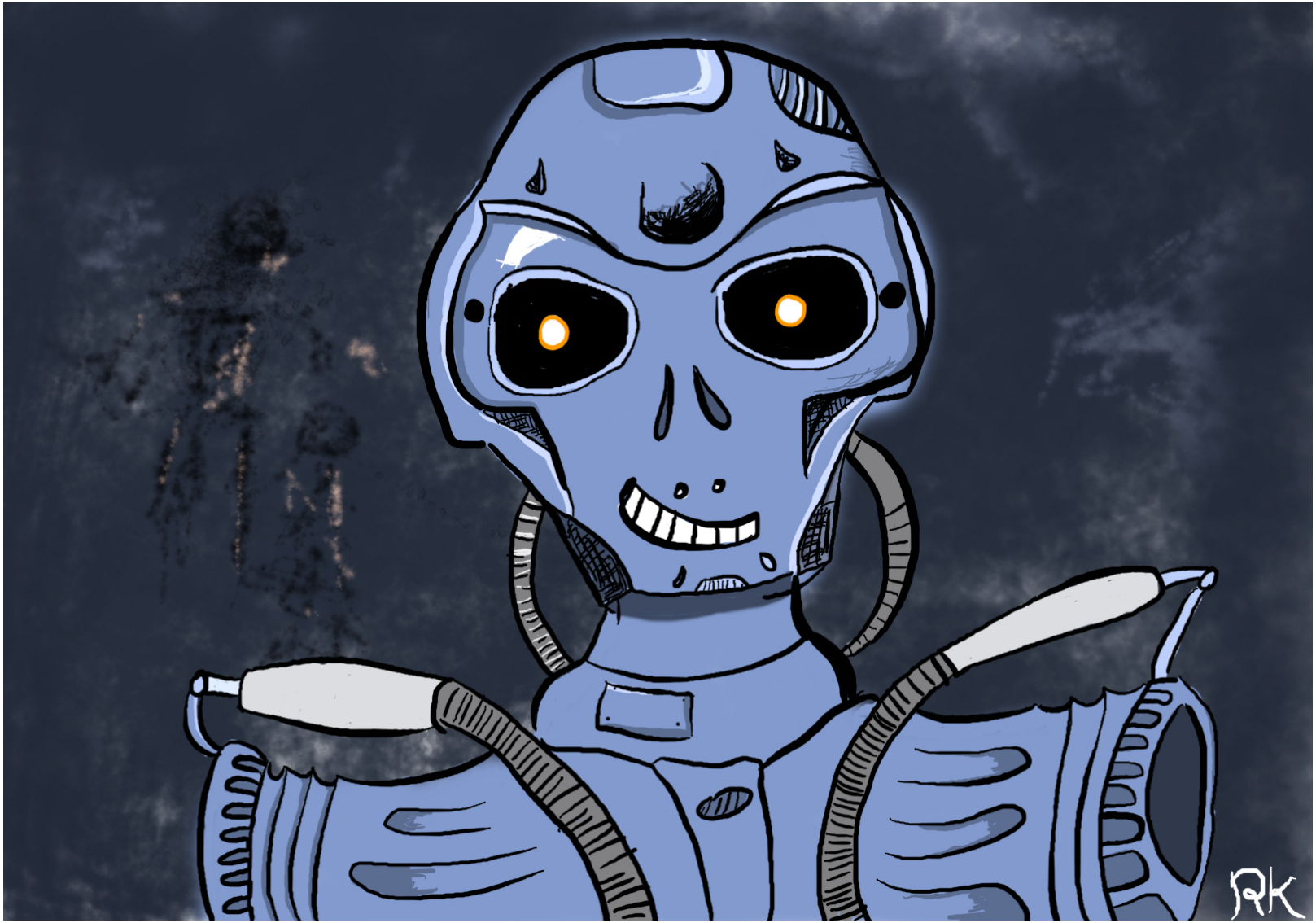


Image by Elephant Scale

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Machine Learning Evolution

➔ **Machine Learning Evolution**

Machine Learning Use Cases

Machine Learning Ecosystem

Machine Learning Algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Informal Definition of Machine Learning

- Arthur Samuel
  - **"The field of study that gives computers the ability to learn without being explicitly programmed."**

- Example: Self-driving cars
  - Tell the car the rules, or
  - Let it record the scenery and your reactions
  - Let it predict the next reaction

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

AI Evolution

- Initial AI thinking was TOP DOWN (or symbolic logic)
- Write a **big, comprehensive** program
  - Program **all the rules** (expert systems)
- Problem:
  - Too many rules
  - Works only for specific domain, e.g. math theorems or chess
- Success stories: playing chess at the grand master level
  - Domains with limited, clear rules
- Not so successful: image recognition

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Another AI Approach – Bottom Up

- Computers can learn from 'ground up' (data-driven)
- E.g. how babies learn to talk:
  - Learn from example
  - They don't know the 'whole dictionary' or 'grammatical rules'
- The focus shifts from **logic to data**
- More data → smarter systems
- Success stories
  - Image recognition
  - Language translation
  - Self-driving cars

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Spam Detection – Traditional (Rule Based) Approach

- Computers follow explicit instructions (code)
- Can be tens of millions code lines but still explicit instructions, something like this:

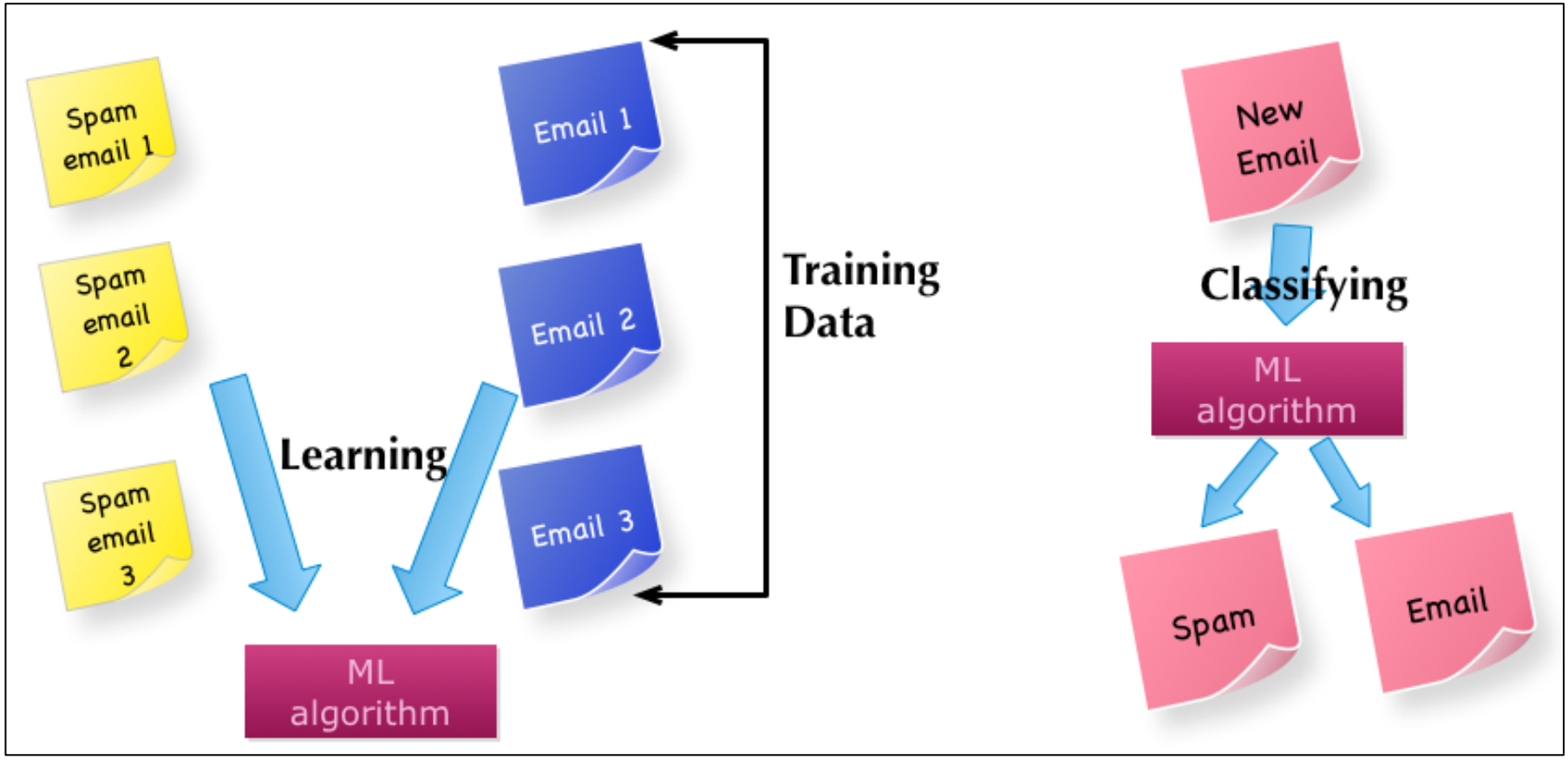
```
// explicitly coding the rules

if (email.from_ip.one_of("ip1", "ip2", "ip3")) {
  result = "no-spam"
}
else if ( email.text.contains ("free loans", "cheap degrees"))
{
  result = "spam"
}
```

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Spam Detection - Machine Learning Approach

- Show the algorithm with spam and non-spam emails
- Algorithm 'learns' which attributes are indicative of spam
- Then algorithm predicts spam/no-spam on new email



Translation - Early Approach

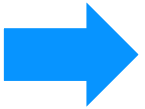
- Creating a translation system from English ← → Japanese
- Code in the following:
  - English dictionary + grammar rules
  - Japanese dictionary + grammar rules
  - Translation rules
- Now the system is ready to translate
- But this approach really doesn't work well:
  - Rules have too many exceptions
  - Context and subtle meanings are lost
- ”Minister of agriculture" → ”Priest of farming"

Translation - 'Bottom Up' Approach (Google Translate)

- Google Translate has been ported to 'Google Brain' on Sept 2016
- System learned from 'data'
- AI based system improved the accuracy many times over
- Link to case study

Uno no es lo que es por lo que escribe, sino por lo que ha leído

-- Jorge Luis Borges



Old translation

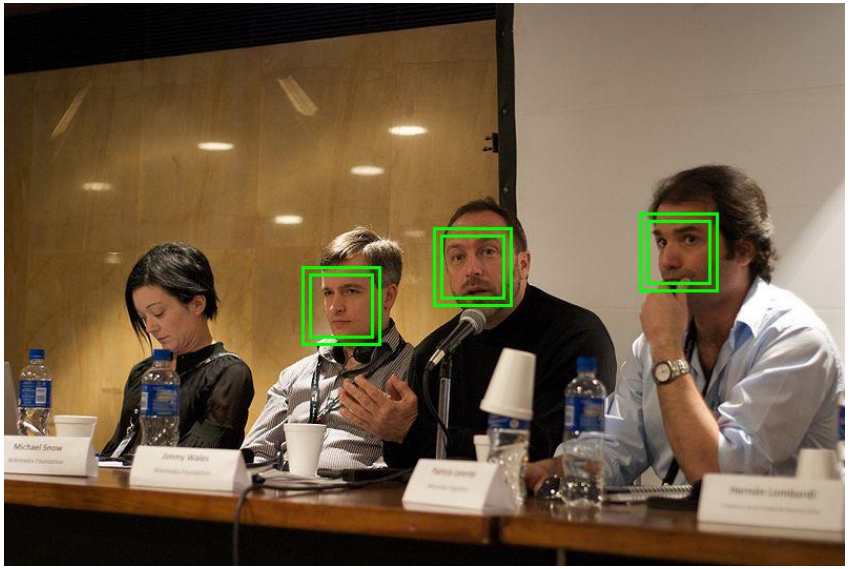
One is not what is for what he writes, but for what he has read

New AI powered translation

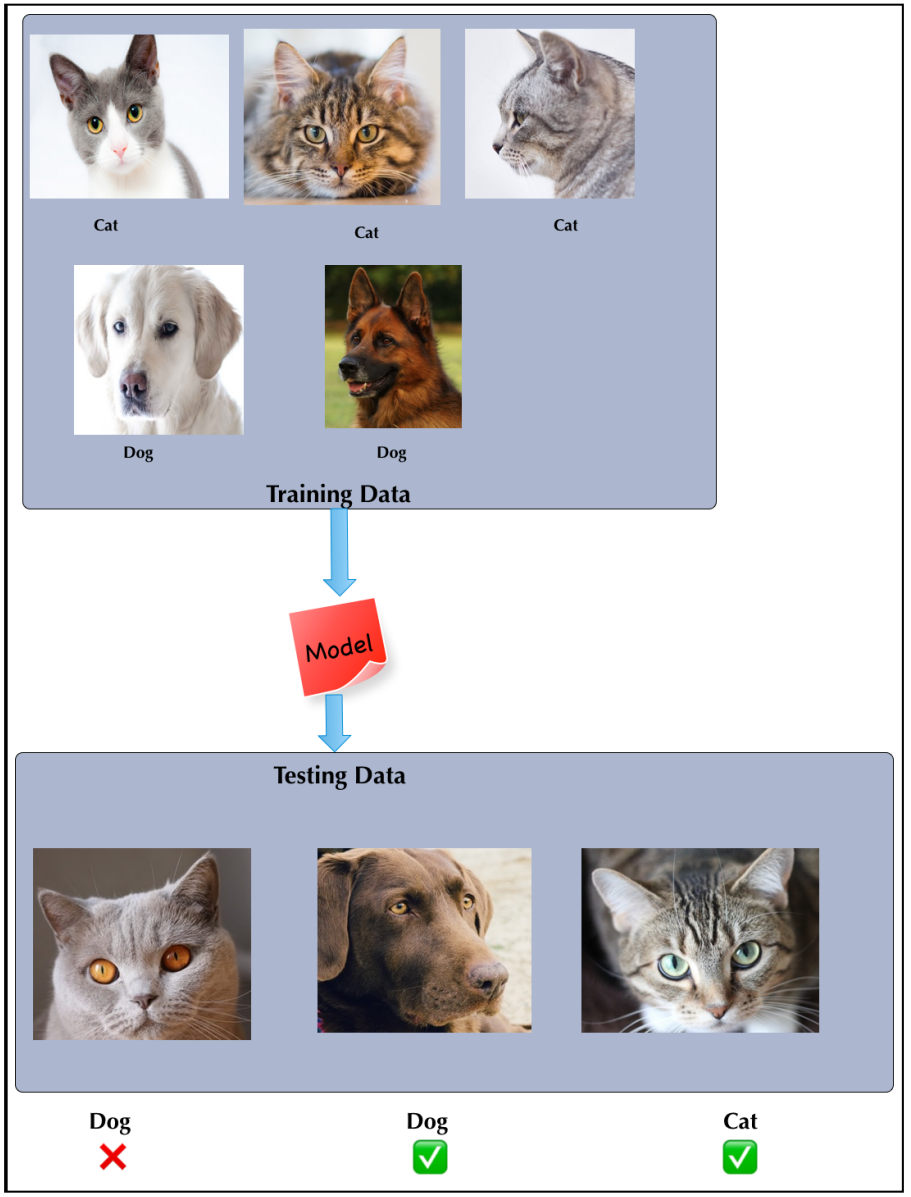
You are not what you write, but what you have read

Bottom Up AI Success Stories

- Image recognition
- Translation
- Self driving cars

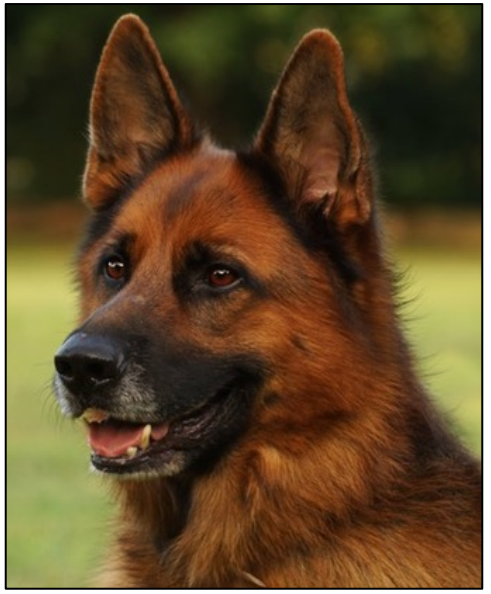


AI Success Story : Image Recognition: Cats & Dogs



Kaggle Competition

- Recognize dogs & cats
- Given 25,000 sample images to train
- Then tested on 15,000 test images
- Winning algorithm correctly classified 98.9% time !
- https://www.kaggle.com/c/dogs-vs-cats



A Glimpse of AI History

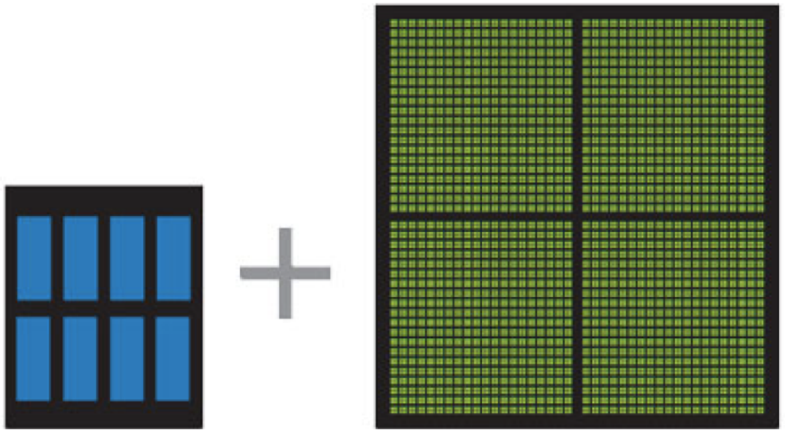
- Sixties
  - Commercial computers & mainframes
  - Computers play chess
- Eighties
  - Artificial intelligence (AI) get **'oversold'**, doesn't live up to the promise and gets a bad rap
- 21st century
  - Big Data changes it all

The Great AI Revival – 21st century (2010 on)

- AI is going through a resurgence now
- **'Big Data'** – now we have so much data to train our models
- **'Big Data ecosystem'** – excellent big data platforms (Hadoop, Spark, NoSQL) are available as open source
- **'Big Compute'** - **cloud** platforms significantly lowered the barrier to massive compute power
  - $1 buys you 16 core + 128 G + 10 Gigabit machine for 1 hr on AWS!
  - So running a 100 node cluster for 5 hrs → $500
- **Advances in hardware** – CPU / GPUs

Hardware - GPU

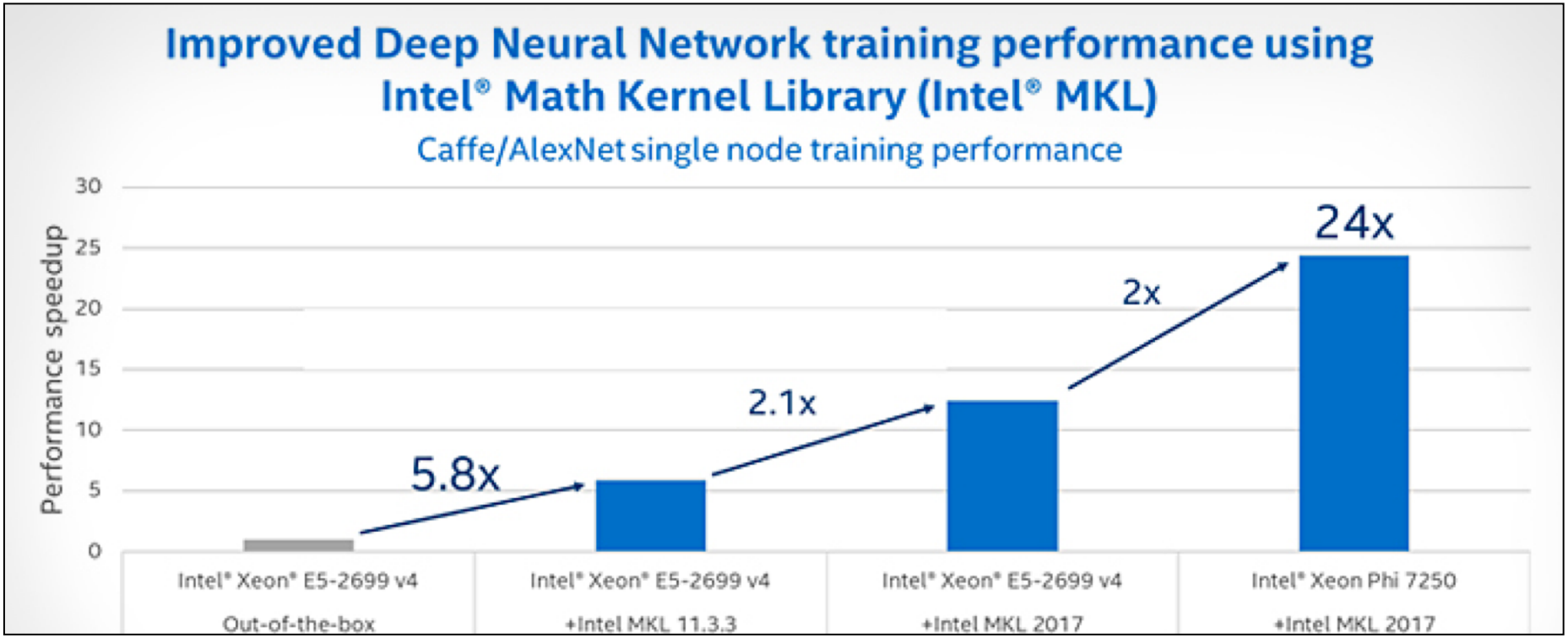
- Recently GPUs – Graphics Processing Units - have become popular (especially in Deep Learning)
- GPU cores are good at compute intensive calculations (math, matrix operations)
- Each GPU core is capable of executing small set instructions, but there are 1000s of core per GPU
  - Running in parallel



Hardware – Modern CPU

- Modern Intel Xeon CPUs (E5 or later) have vectorized linear algebra
  - Properly optimized, approaches speed of GPUs
  - And offers faster I/O performance for Big Data.

- Intel Math Kernel Library - highly optimized, threaded, and vectorized math functions that maximize performance on each processor family



Hardware – TPU (Tensor Processing Unit)

- A Tensor processing unit (TPU) is an AI accelerator application-specific integrated circuit (ASIC) developed by Google specifically for neural network machine learning
- More capable the CPUs / GPUs in certain tasks
- Designed for Tensorflow
- Designed for high volume computes
  - A TPU can process 100 million photos a day
- Available in Google Cloud platform

Google TPU System in Data Center

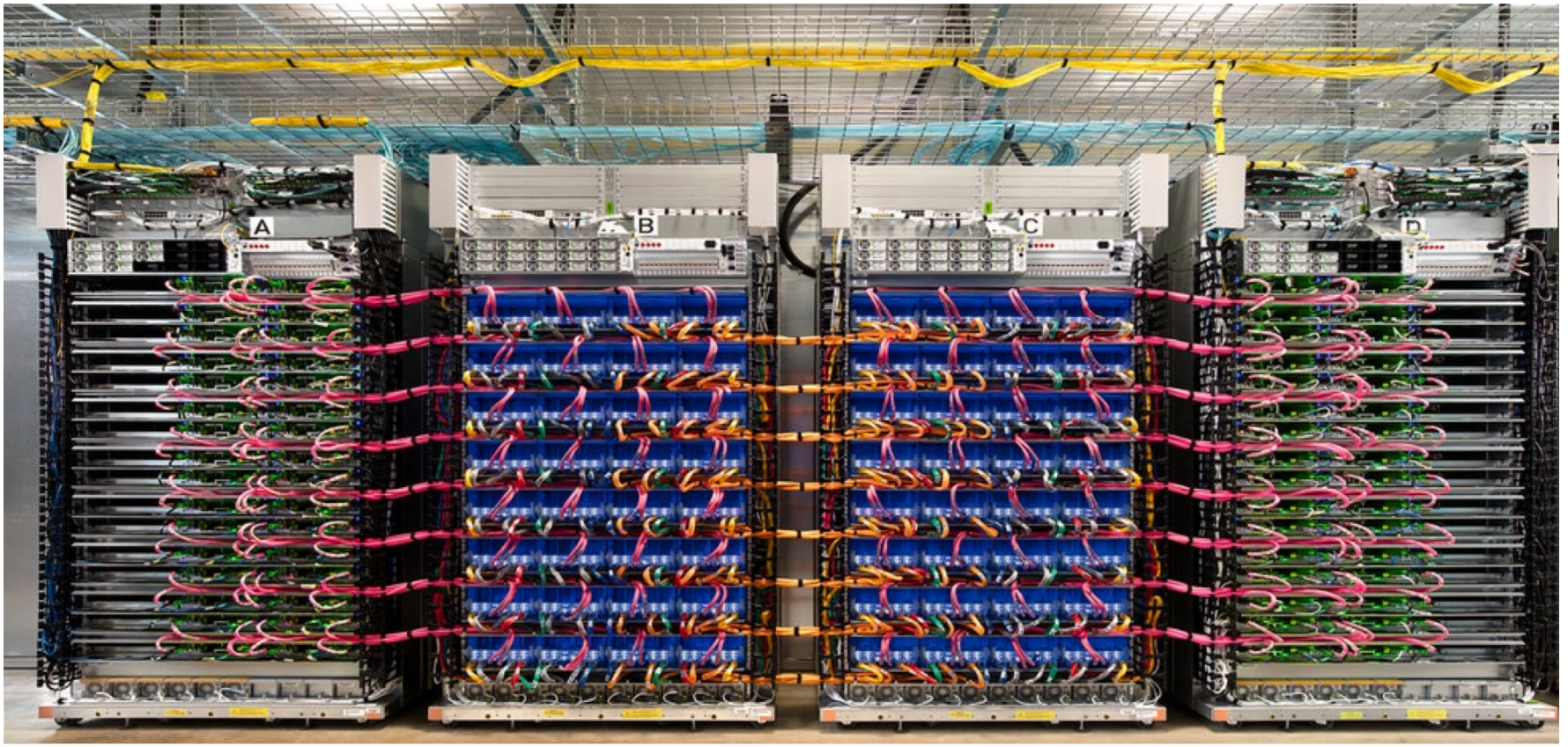


Image credit : NY times

Machine Learning Use Cases

Machine Learning Evolution

→ **Machine Learning Use Cases**

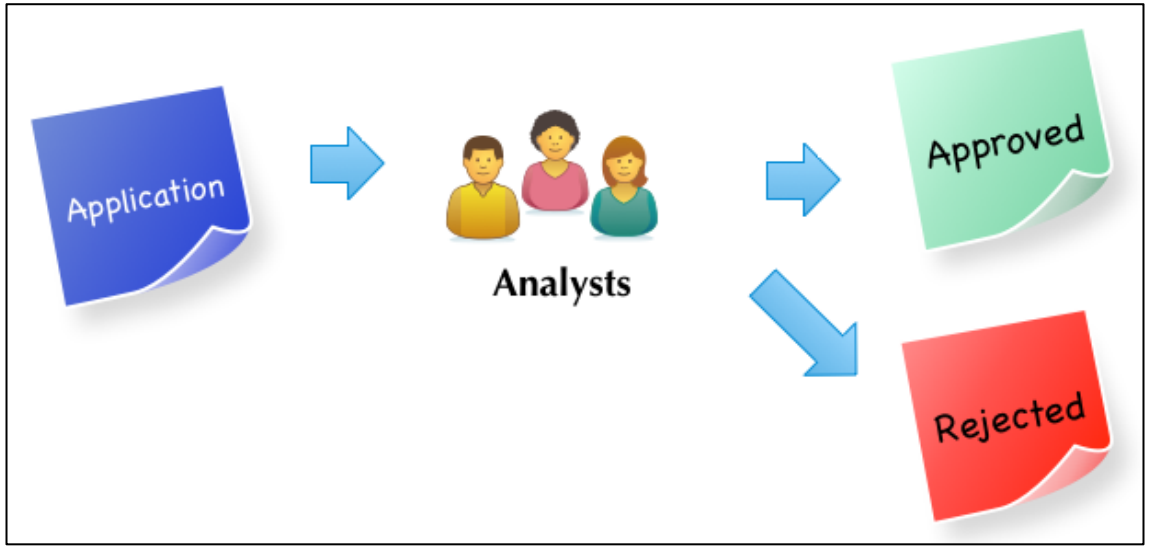
Machine Learning Ecosystem

Machine Learning Algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

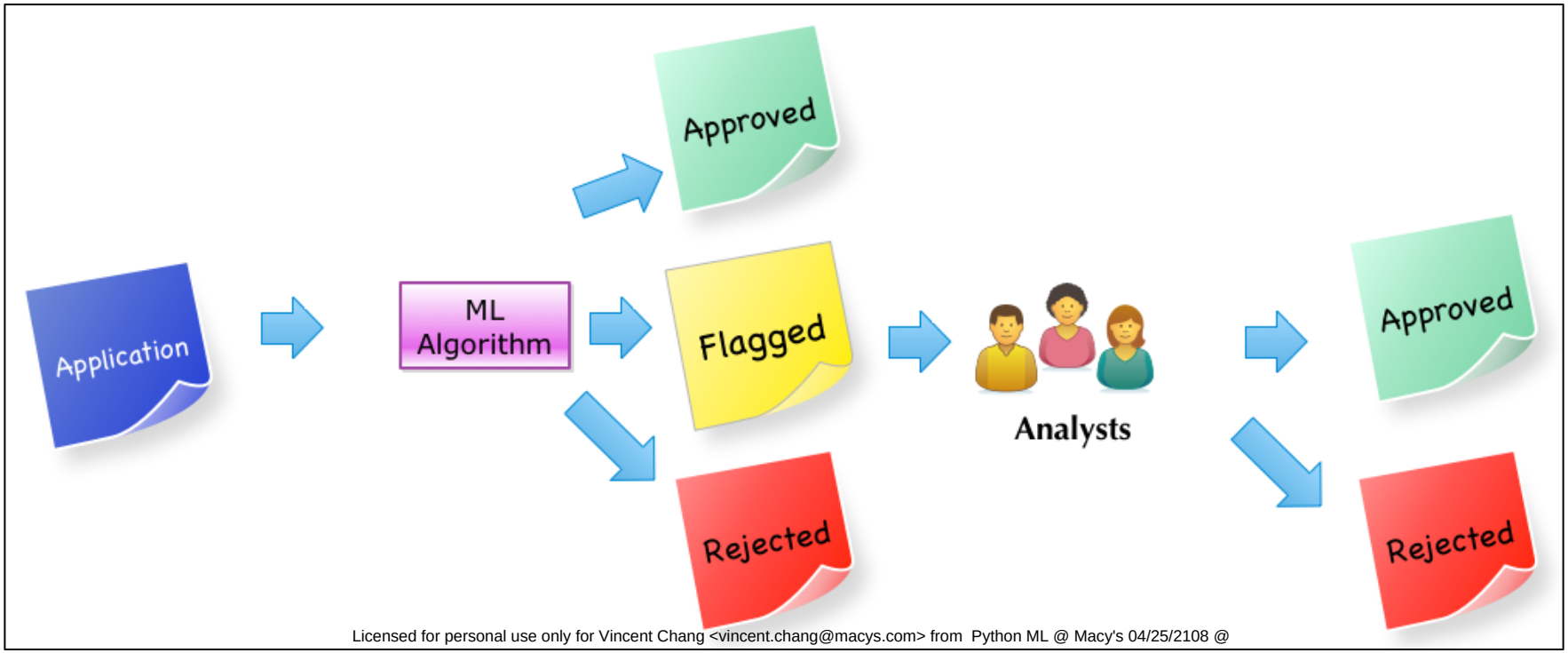
How ML Can Help a Business

- Credit Card Application use case
- In the beginning, all applications are reviewed manually by analysts
  - Approved or rejected based on criteria
- As the application volume goes up
  - Hire more analysts to keep up with volume
  - Human bias might lead to inconsistent or unfair approval process



How ML Can Help a Business

- Machine Learning algorithm can learn from past loan applications
  - E.g., if applicant already has a credit line and making minimum payments, he/she is likely to default on new credit
- ML can process applications very quickly and only send "flagged" applications for manual review



ML Advantages/Challenges

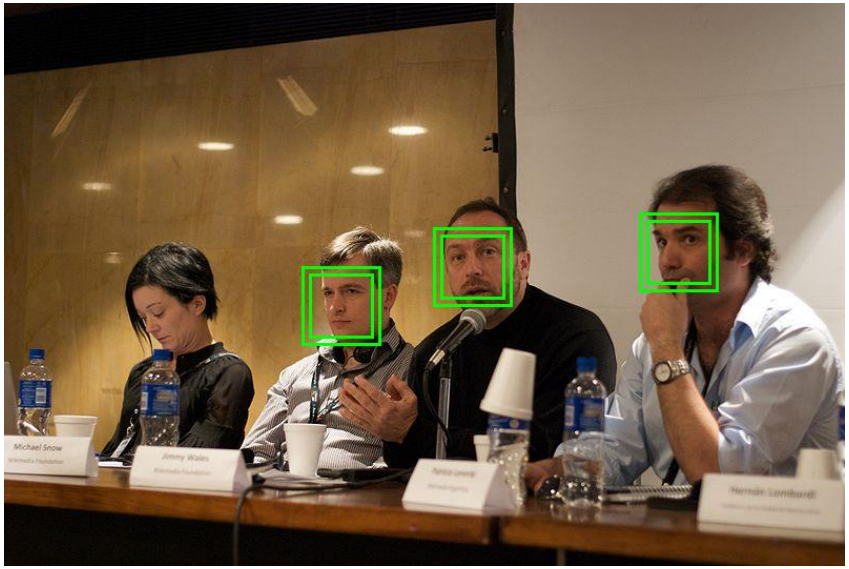
Advantages	Challenges
- **Accurate**: ML can learn from data, the more data it learns from the better it gets - **Automated**: Bulk of the decisions can be automated - **Fast**: ML can process data within milliseconds - **Customizable**: ML algorithms can be adopted for various scenarios - **Scalable**: ML algorithms can scale for large amount of data	- **Data prep:** Data may not be in ready-to-use form - **Accuracy**: Measuring accuracy can get complicated - **Algorithm Choice**: Different algorithms perform differently, choosing the best algorithm is very important

Machine Learning Applications

- Detect credit card fraud
  - Thousands of features
  - Billions of transactions
- Recommendations
  - Millions of products
  - To millions of users
- Genome data manipulation
  - Thousands of human genomes
  - Detect genetic associations with disease
- Language translation

Machine Learning Applications

- Self Driving Cars
  - ML system using image recognition
  - Where the edge of the road / road sign / car in front
- Face recognition
  - Facebook images
  - System learns from images manually tagged and then automatically detects faces in uploaded photos



Machine Learning Ecosystem

Machine Learning Evolution

Machine Learning Use Cases

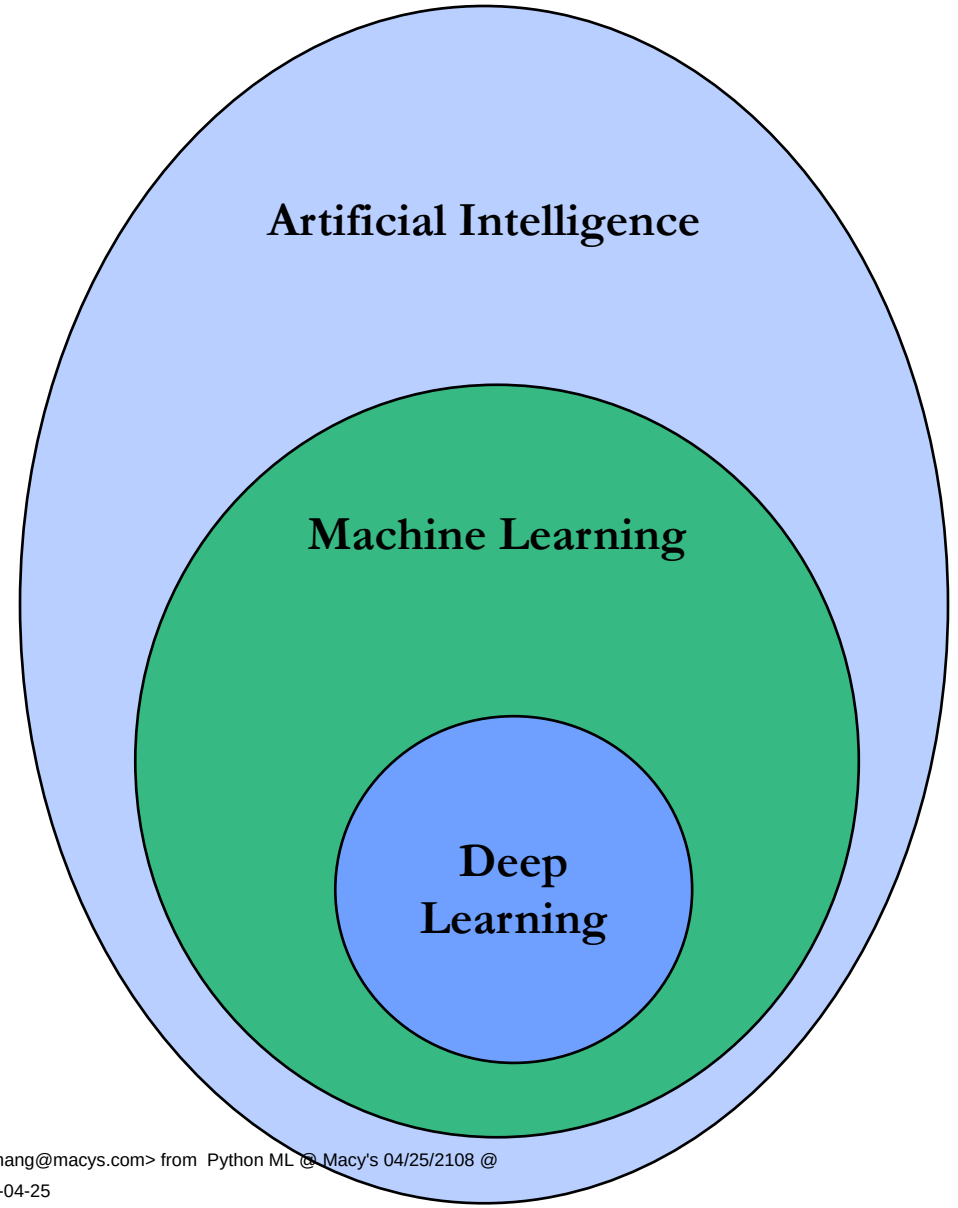
→ **Machine Learning Ecosystem**

Machine Learning Algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

AI / Machine Learning / Deep Learning

- **Artificial Intelligence (AI):** Broader concept of machines being able to carry out 'smart' tasks
- **Machine Learning:** Current application of AI that machines learn from data using mathematical, statistical models
- **Deep Learning: (Hot!)** Using Neural Networks to solve some hard problems

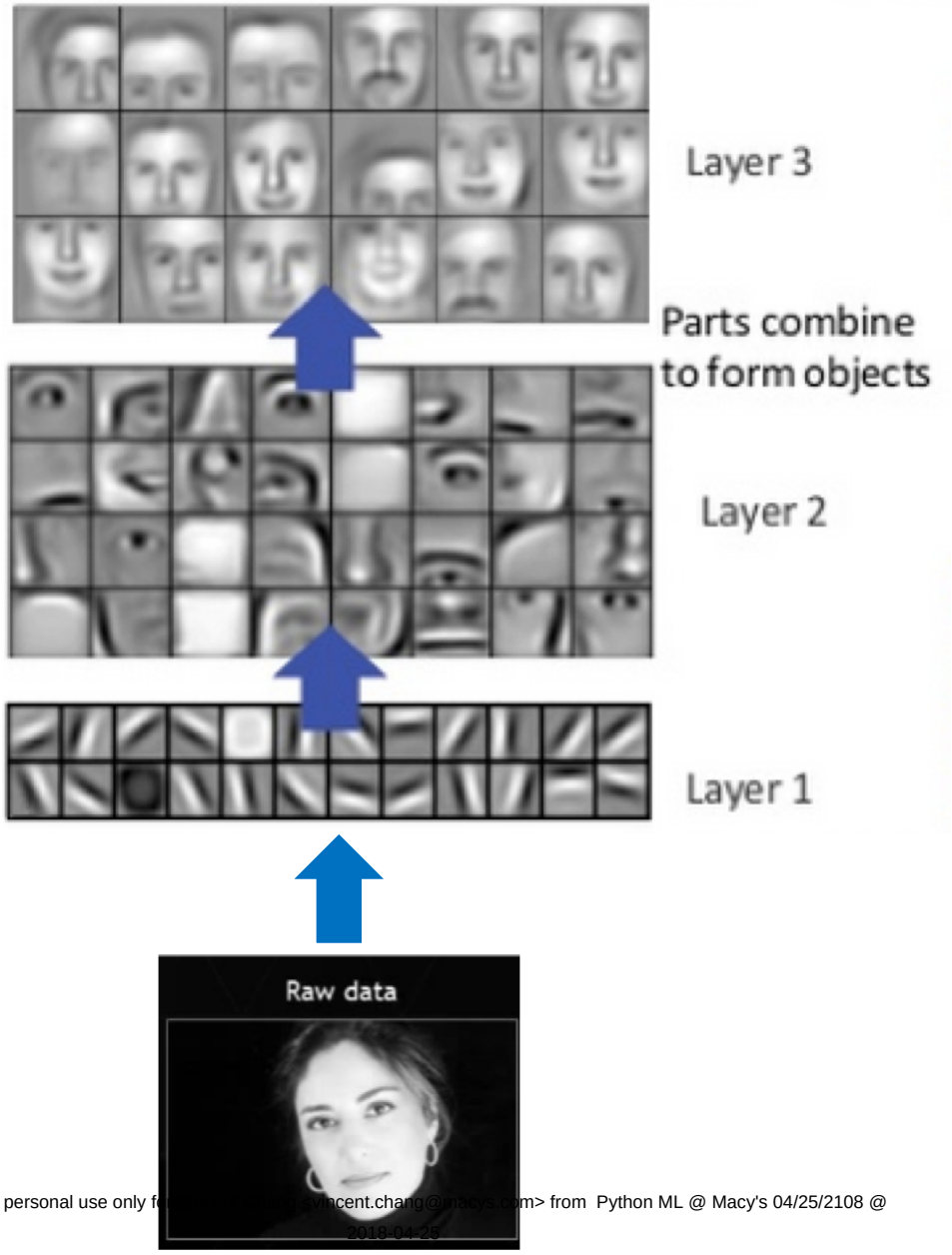


Deep Learning (DL)

- Deep Learning uses Neural networks techniques
- Neural Networks fell out of favor in the 90s as statistics-based methods yielded better results
- Now making a comeback due to Big Data & Big Compute ((cluster computing , GPU and TPU)
- Examples
  - Facebook Deep Face
  - Google Translate
  - Google DeepMind playing GO game
  - IBM Deep Blue winning Jeopardy

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

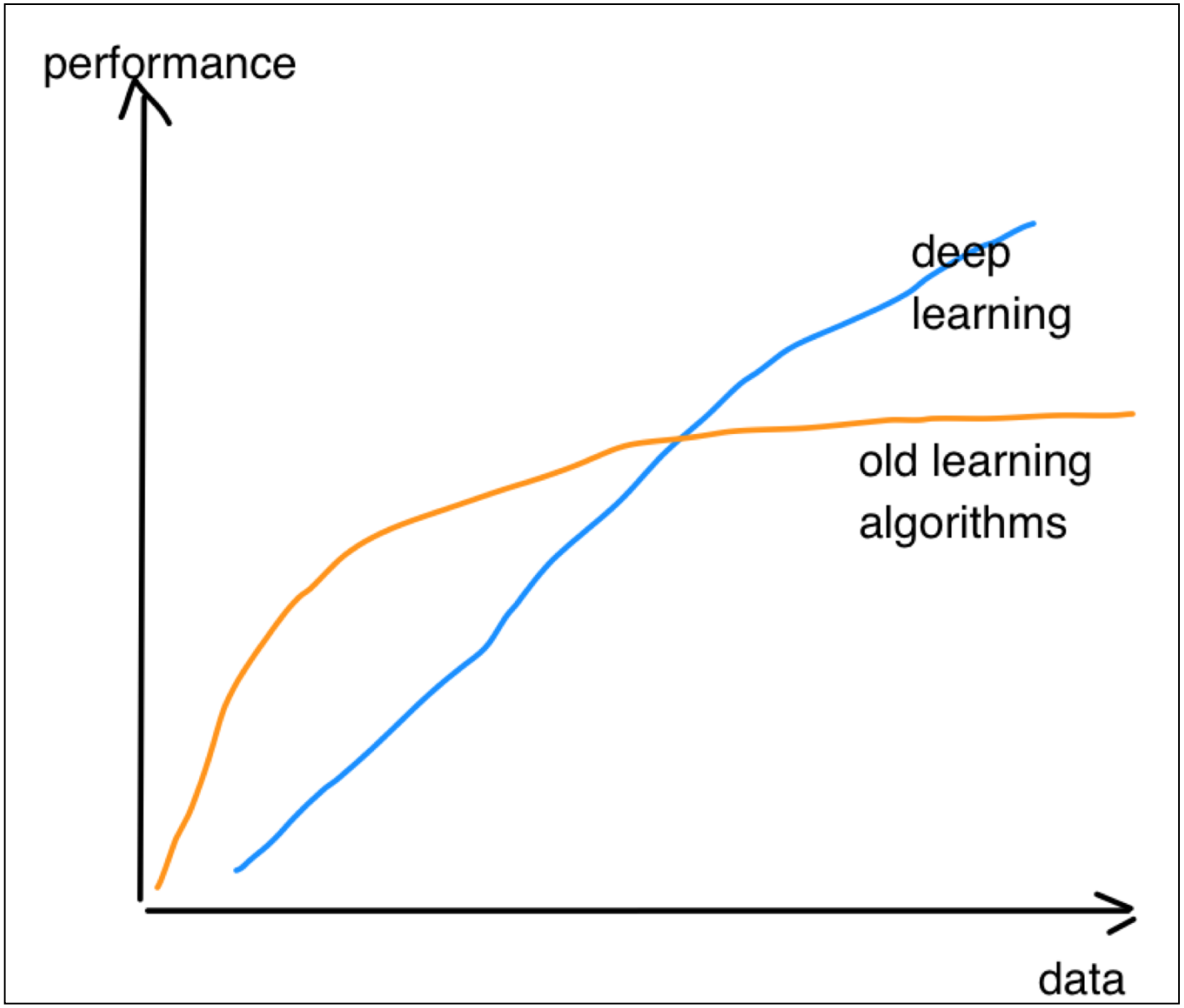
Deep Neural Network – Face Recognition



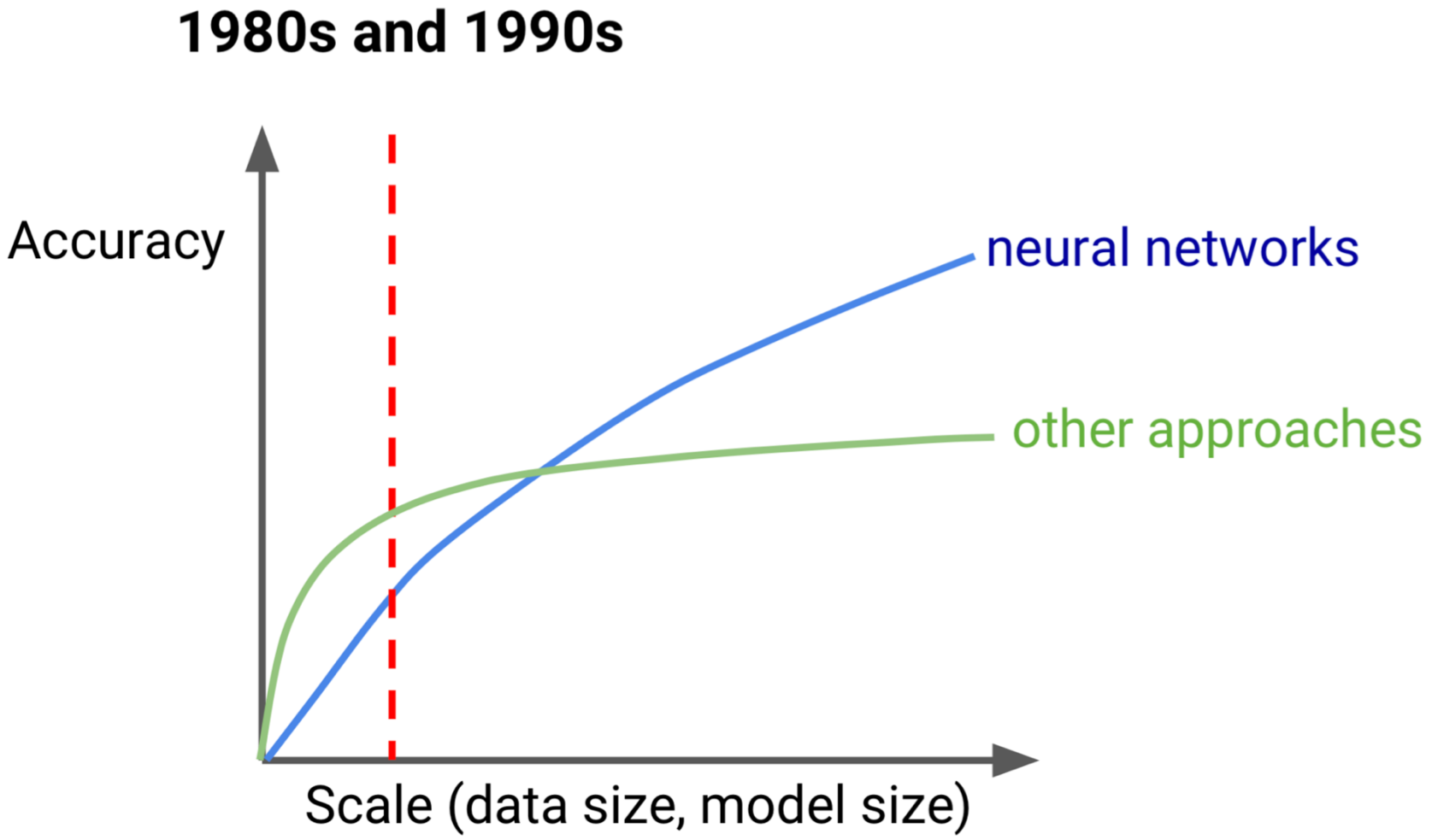
Machine Learning vs. Deep Learning

Features	Machine Learning	Deep Learning
Data size (see next slide for graph)	Performs reasonably well on small / medium data	Need large amount of data for reasonable performance
Scaling	Doesn't scale with large amount of data	Scales well with large amount of data
Compute power	Doesn't need a lot of compute (works well on single machines)	Needs a lot of compute power (usually runs on clusters)
CPU/GPU	Mostly CPU bound	Can utilize GPU for certain computes (massive matrix operations)
Feature Engineering	Features needs to specified manually (by experts)	DL can learn high level features from data automatically
Execution Time	Training usually takes seconds, minutes, hours	Training takes lot longer (days)
Interpretability	Easy to interpret	Hard to understand the final result

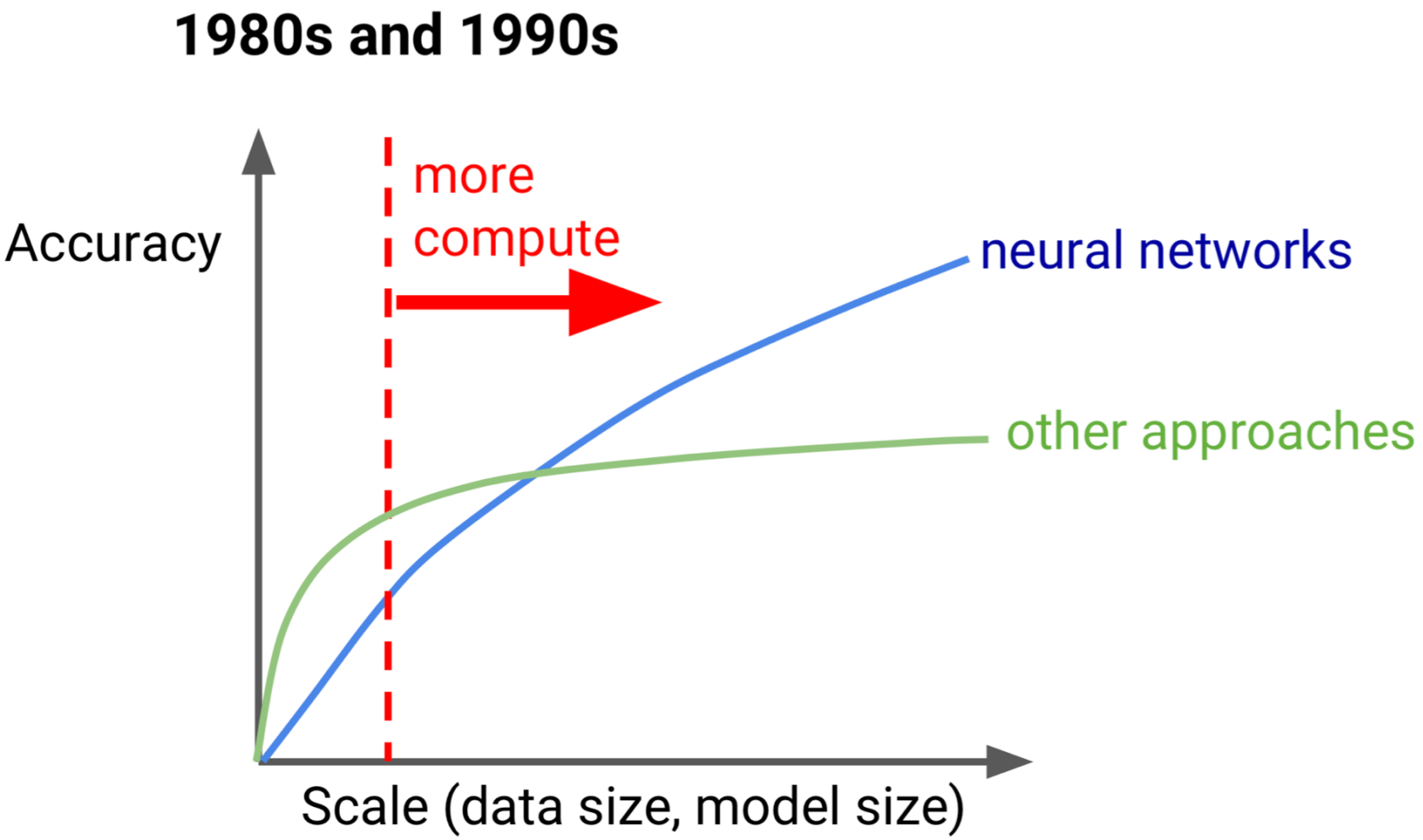
Machine Learning vs. Deep Learning



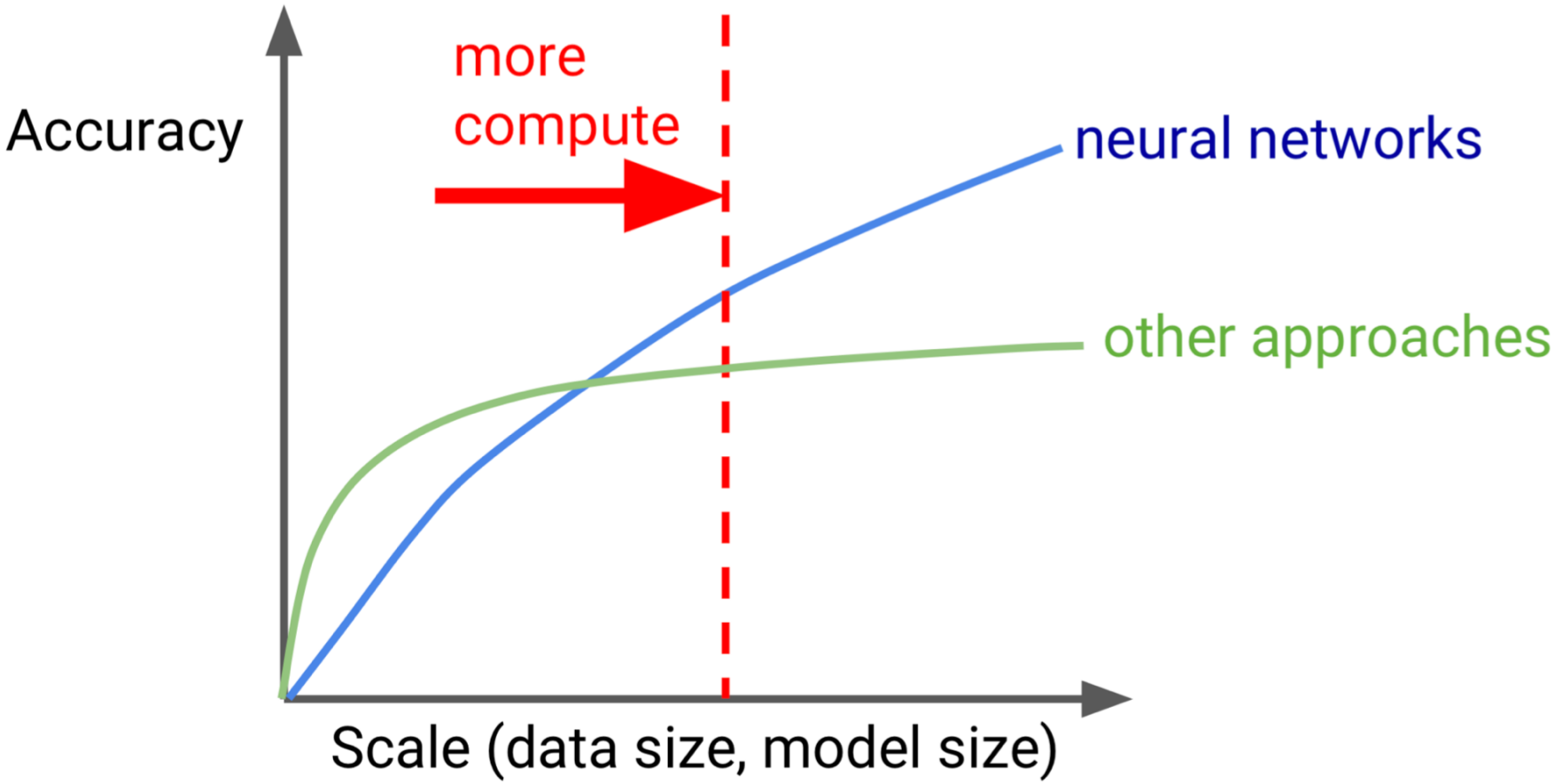
1980's and 1990's



Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25



Now



Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

AI Software Eco System

	Machine Learning	Deep Learning
Java	- Weka - Mahout	- DeepLearning4J
Python	- SciKit - (Numpy, Pandas)	- Tensorflow - Theano - Caffe
R	- Many libraries	- Deepnet - Darch
Distributed	- H20 - Spark	- H20 - Spark
Cloud	- AWS - Azure - Google Cloud	- AWS - Azure - Google Cloud

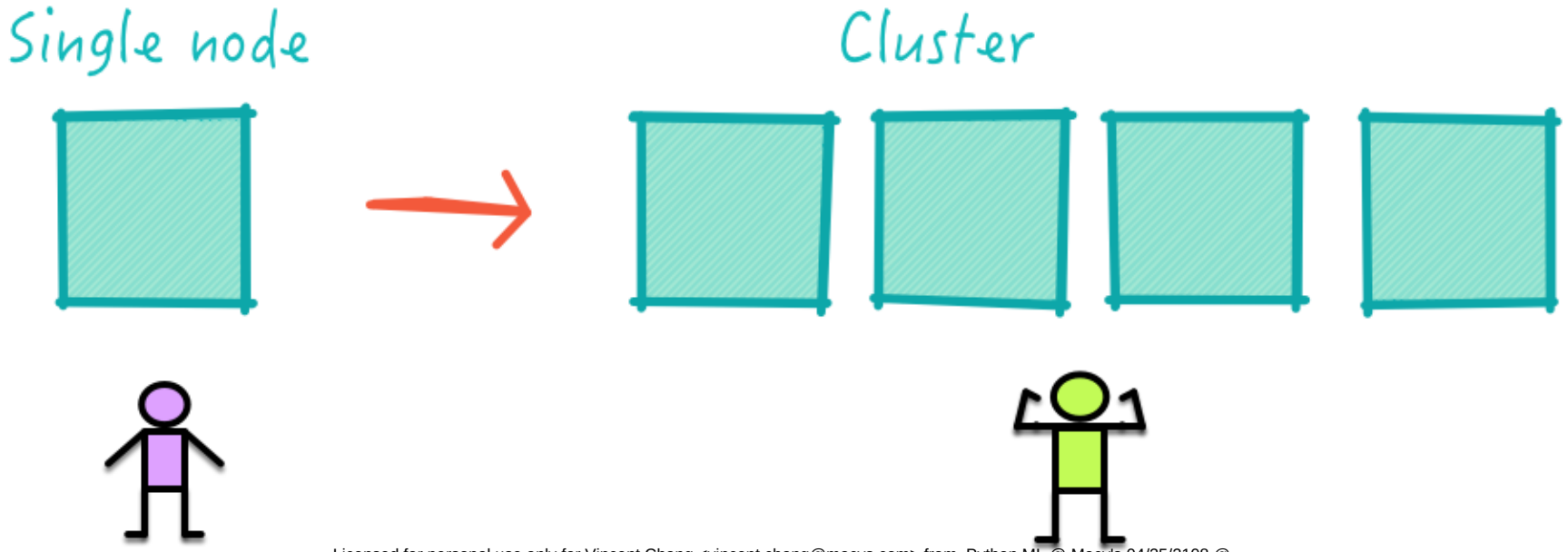
Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Technology Stack Comparison

Technology	Pros	Cons
R	- Rich environment - Thousands of libraries	- Rough on data cleanup - Not a general purpose language (may not be mainstream?) - **Data must fit on one machine**
Python	- General purpose programming language - Excellent libraries (Pandas / scikit-learn) - Gaining popularity in recent years	- **Data must fit on one machine**

Machine Learning and Big Data

- Until recently most of the machine learning is done on "single computer" (with lots of memory–100s of GBs)
- Most R/Python/Java libraries are "single node based"
- Now Big Data tools make it possible to run machine learning algorithms at massive scale–distributed across a cluster



Machine Learning vs. Big Data

Traditional ML	ML on Big Data
All (or most) data fits into single machine	Data is distributed across multiple machines
Data almost / always in memory	Memory is scarce
Optimized for heavy iterative computes	Optimized for single pass computes
Maintains state between stages	stateless
CPU bound	IO bound (disk / network).
GPU (Graphical Processing Unit) seldom engaged	GPUs are utilized increasingly

Tools for Scalable Machine Learning

- **Spark ML**
  - Runs on top of popular Spark framework
  - Massively scalable
  - Can use memory (caching) effectively for iterative algorithms
  - Language support: Scala, Java, Python, R

- **Amazon Machine Learning**
  - Ready to go algorithms
  - Visualization tools
  - Wizards to guide
  - Scalable on Amazon Cloud

Tools for Scalable Machine Learning

- **Azure ML Studio**
  - Built on Azure cloud (Microsoft)
  - Language support: Python, R

- **H2O**
  - Easy to use API
  - WebUI
  - Supports reading from multiple datasources (Excel/SQL/HDFS)
  - In memory compute
  - Works on top of Spark ("Sparkling Water")
  - Vendor: 0xData

Tools for Scalable Deep Learning

- **TensorFlow**
  - Based on "data flow graphs"
  - "Tensor" = batches of data
  - Language support: Python, C++
  - Run time: CPU, GPU

- **Intel BigDL**
  - Deep learning library
  - Built on Apache Spark
  - Language support: Python, Scala

Machine Learning Algorithms

Machine Learning Evolution

Machine Learning Use Cases

Machine Learning Ecosystem

➔ **Machine Learning Algorithms**

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

How to do Machine Learning

- **Collect data**
  More data we have, the better the algorithms become. This data can come from internal logs (clickstreams) or external sources (credit scores of customers)
- **Prepare Data**
  Raw data is hardly in a form to be used. It needs to be cleansed, tagged and curated before ready to use
- **Train a model**
  Feed the training data to model so it can learn
- **Evaluate the model**
  Test the model accuracy
- **Improve the model**
  Either by adding more training data, choosing a different algorithm ..etc

Types of Machine Learning

- **Supervised Machine Learning:**
  - A model is “trained” with human labeled training data
  - Model is tested on test data to see performance
  - Model can be applied to unknown data
  - Classification and regression are usually supervised
- **Unsupervised Machine Learning**
  - Model tries to find natural patterns in the data
  - No human input except parameters of the model
  - Example: Clustering news stories
- **Semi-Supervised Learning**
  - Model is trained with a training set which contains unlabeled (usually lot) and labeled (usually little) data
  - Example: Large images archive only a few of them are labeled (cat, dog, person) and majority are unlabelled

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Machine Learning Types: Supervised

- Model learns from (training) data
- Then predicts on 'unseen' data

Algorithms	Description	Applications
Classification	Categorize things into groups	- Spam classification - Fraud / no fraud
Regression	Dealing with numbers and calculate the probability something happening	- Predict house prices - Predict stock market

Machine Learning Types : Un Supervised

- No training needed
- Algorithm tries to find patterns in data

Algorithms	Description	Applications
Clustering	Find naturally present patterns in data	- Identify news stories (sports / business) - Gnome clustering
Association	Find similar patterns	- people who buy A also buy B
Dimensionality Reduction	Reduces number of features	- Reducing 1000s of variables into manageable size

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Supervised

Machine Learning Evolution

Machine Learning Use Cases

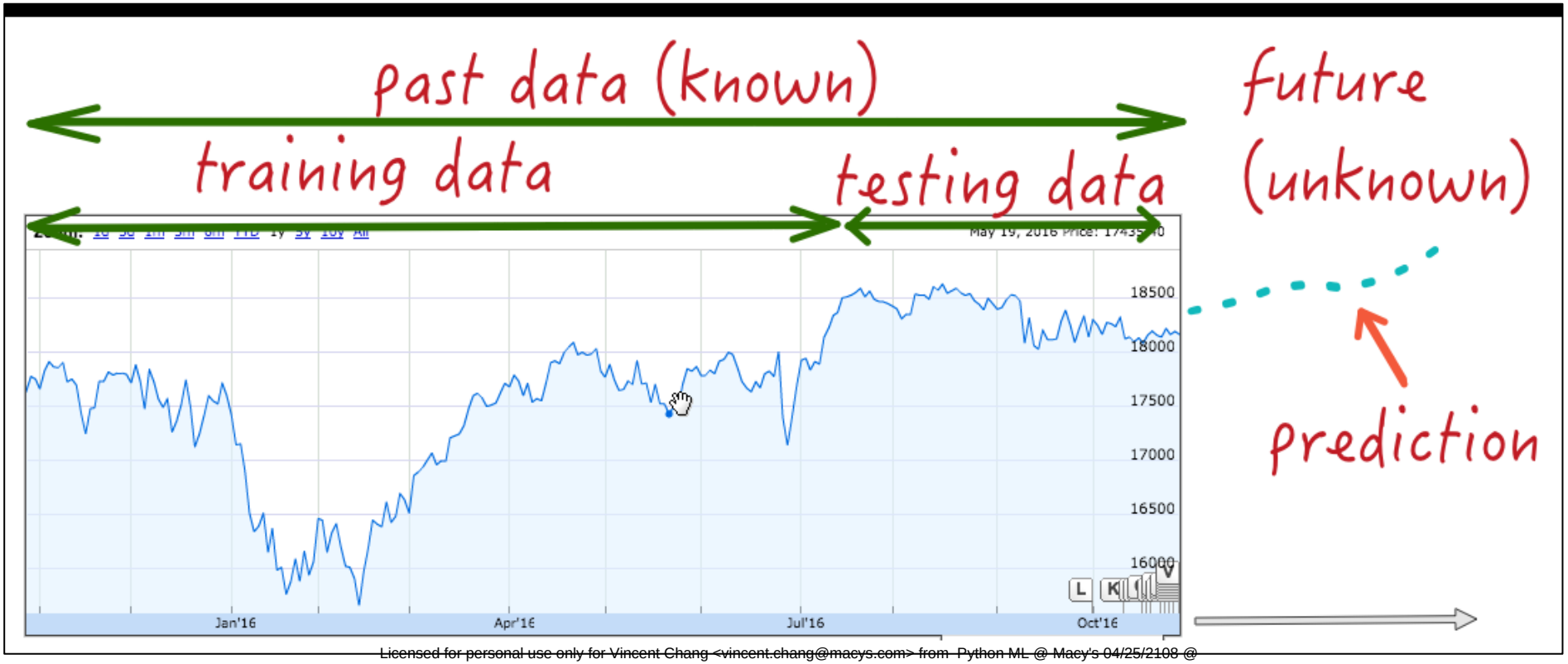
→ **Machine Learning Ecosystem**

Machine Learning Algorithms

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

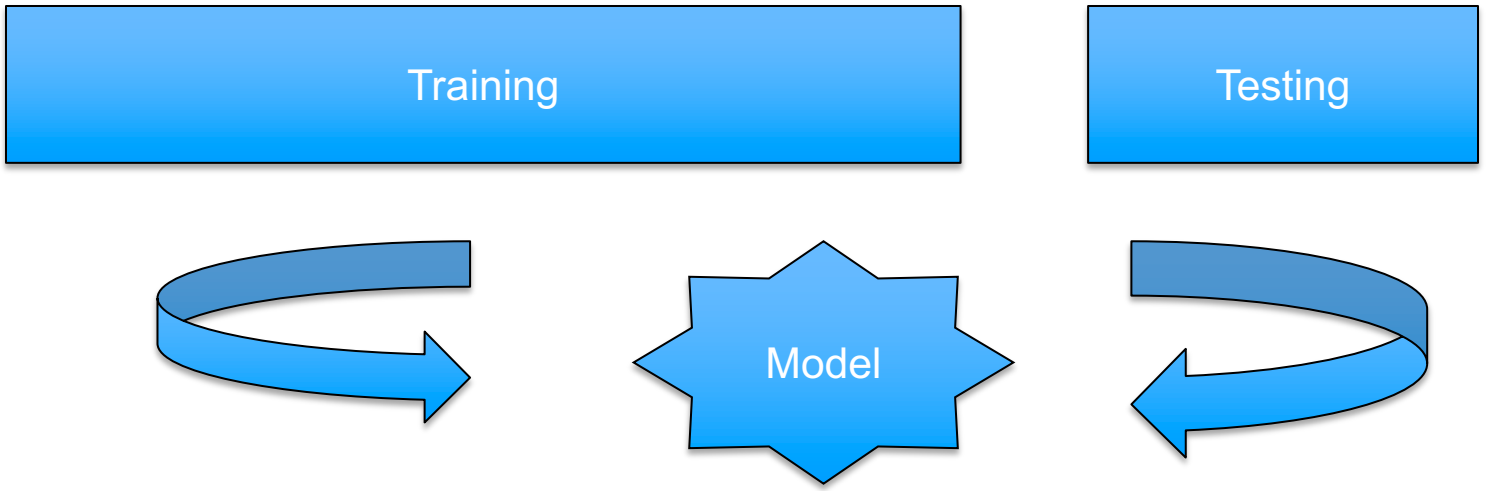
Supervised Learning Example – Regression

- Predicting stock market
- Train the model using training data (already known)
- Test performance using test data (already known)
- Predict no new data (unseen)



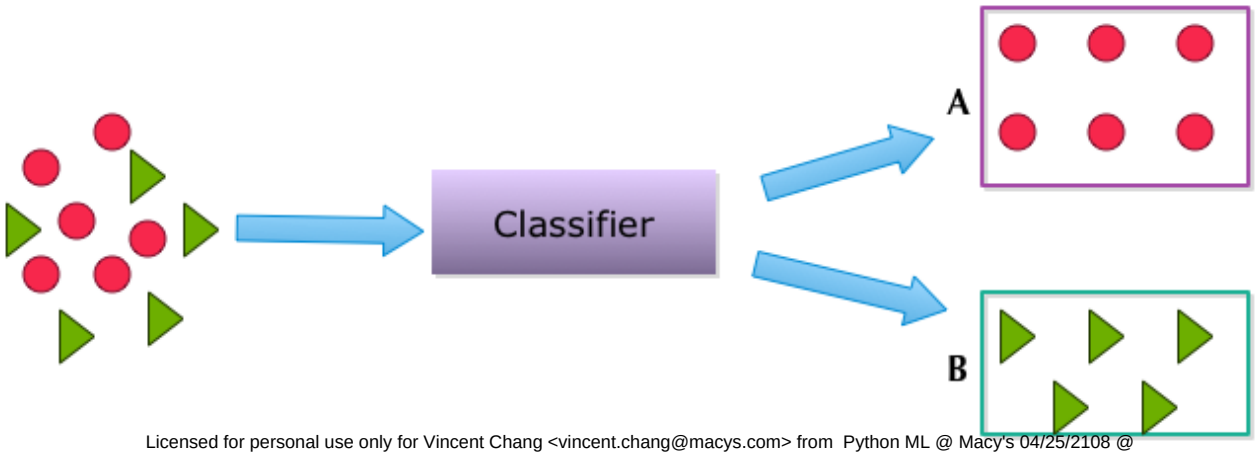
Supervised Learning Methodology

- Split the data set into
  - Training set: Train model
    (Training set should represent data well enough)
  - Test set: Validate the model
- Initially 70% training, 30% test
- Sometimes, 60% training, 20% cross-validation, 20% test
- Tweak the dials to increase or decrease the proportion



Supervised Learning – Classification

- Classification is a model that predicts data into "buckets"
- Examples:
  - Email is **SPAM** or **HAM** (not-SPAM)
  - A cell is **cancerous** or **healthy**
  - Hand-written numbers → any digits 0, 1, 2,…, 9
- Classification algorithm learns from training data
  - Supervised learning
- Also predicted classes are "**discrete**" or "**qualitative**"



Classification Applications

- Web
  - Email is spam or not
  - Website is authentic or fraudulent
- Medicine
  - Is this cell cancerous or not?
- Finance
  - Credit card transaction fraudulent or not
- OCR
  - Recognizing characters and symbols

Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Unsupervised

Machine Learning Evolution

Machine Learning Use Cases

→ **Machine Learning Ecosystem**

Machine Learning Algorithms

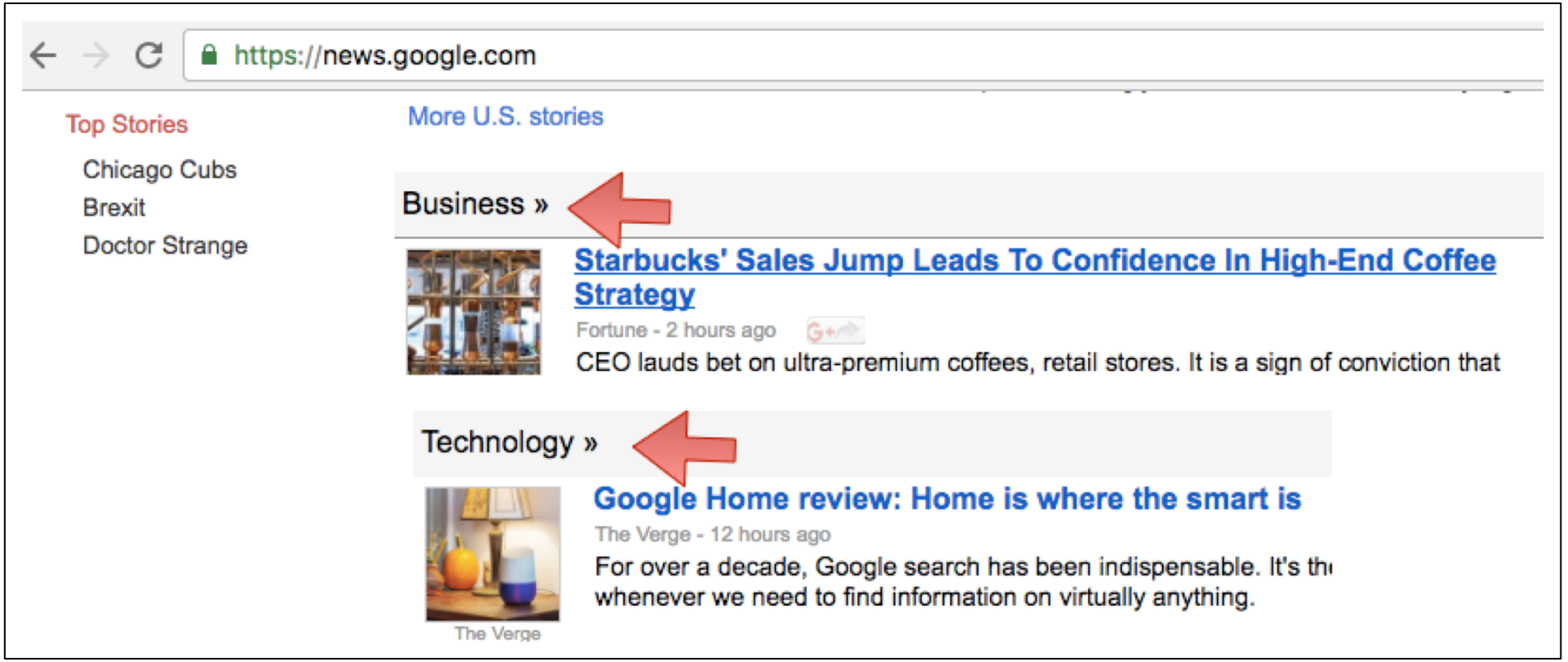
Licensed for personal use only for Vincent Chang <vincent.chang@macys.com> from Python ML @ Macy's 04/25/2108 @ 2018-04-25

Unsupervised Machine Learning

- Draw inference from input data without "labeled responses"
- Common clustering algorithms
  - K-means: Group data points into cluster
  - Hidden Markov Model: State transitions
- Example applications:
  - Find patterns in data
  - Gene expression analysis
  - Recover states from results of random transitions

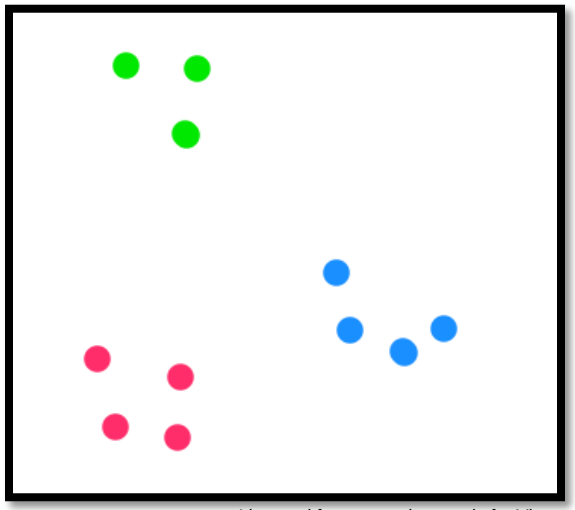
Unsupervised Example: Google News

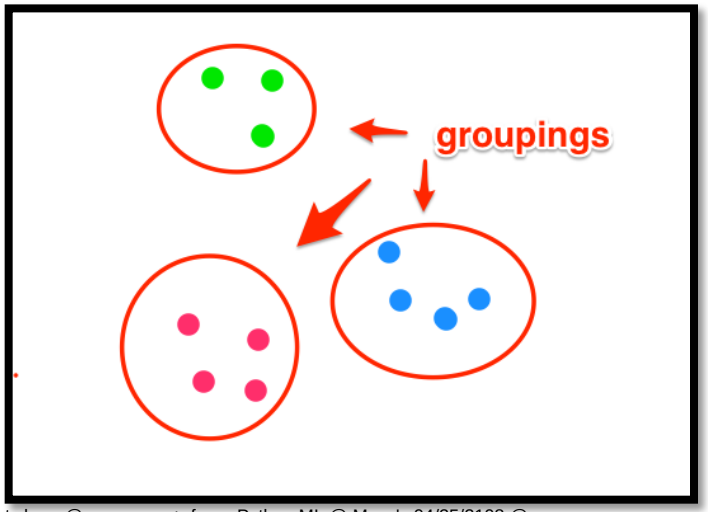
- Google News algorithm automatically groups **related news stories** into sections



Clustering

- Clustering finds natural groupings in data
- Humans naturally cluster data we encounter
  - Categorizing, organizing, etc.
  - Our brains seek patterns
- Why do we cluster?
  - To understand our data
  - To find “more like this”





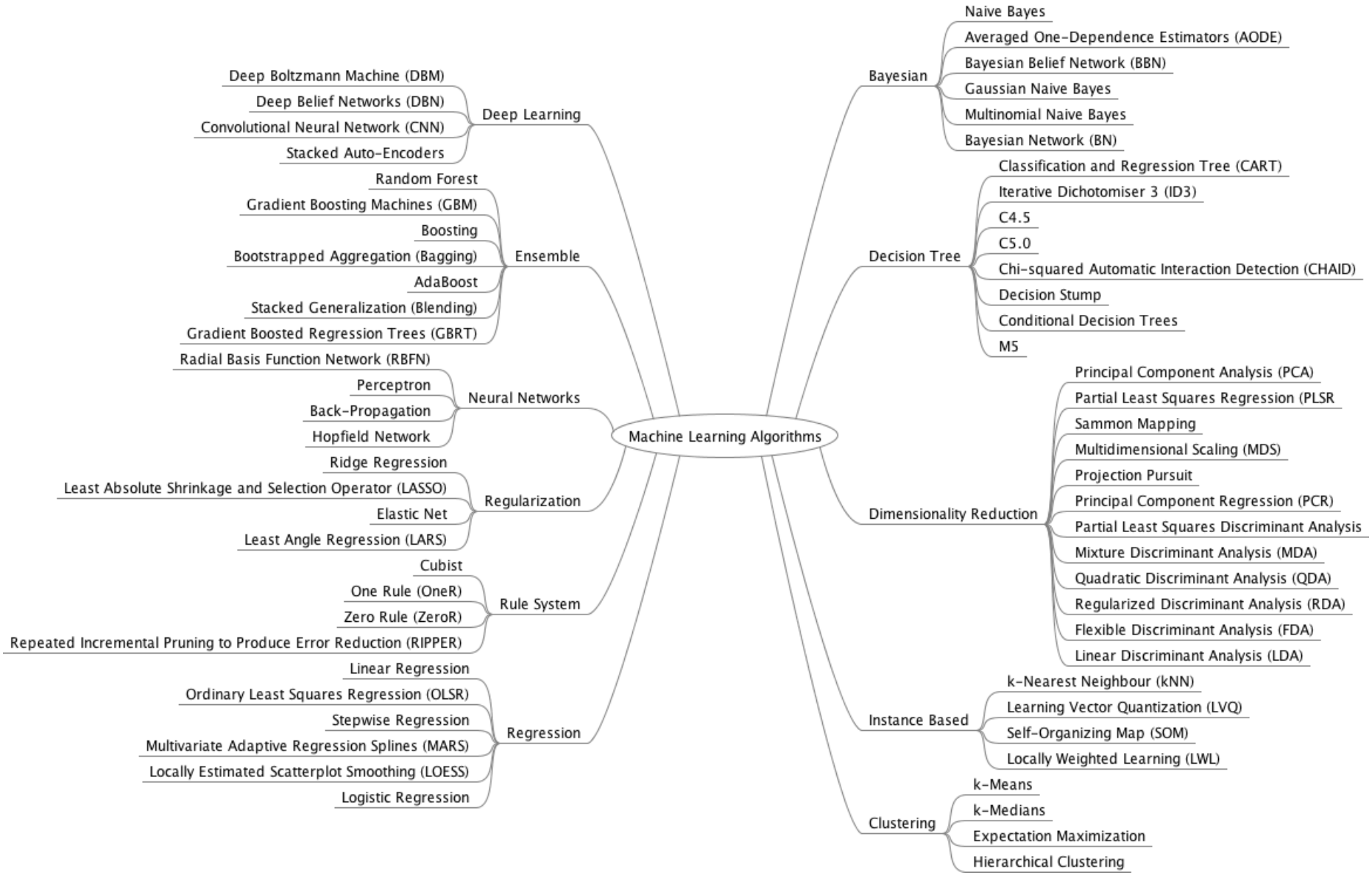
Clustering Applications

- Biology
  - Genomics grouping
- Medicine
  - Xray/CAT image analysis
- Marketing
  - Consumer grouping ("soccer mom"...etc.) and behavior analysis
- Web
  - Search result grouping
  - News article grouping (Google news)
- Computer Science
  - Image analysis
- Climatology
  - Weather pattern analysis (high pressure/warm regions)

Algorithm Summary

Category	Sub Category	Example	Algorithms
Supervised	Regressions	- Predict house prices - Predict stock price	- Linear Regression - Polynomial - Stepwise - Ridge, Lasso, ElasticNet
	Classifications	- Cancer or not - Spam or not	- Logistic Regression - SVM - Naïve Bayes - K Nearest Neighbor (KNN)
	Decision Trees	- Classification (credit card fraud detection) - Regression (predict stock prices)	- Decision Trees - Random Forests
Unsupervised	Clustering	- Group Uber trips - Cluster DNA data	- Kmeans - Hierarchical clustering
		- Dimensionality reduction	- PCA
		- Text mining	- Topic discovery
Recommendations		- Recommend movies	- Collaborative Filtering

ML Algorithm Cheat Sheet



Lab: Design ML Algorithm

- **Overview**:
  - A very practical design lab
  - Illustrates that even small entities can use ML and Big Data
  - It is a design and discussion lab, all on the slides
- **Problem**:
  Domestic tension
- **Solution**:
  Buy flowers
- **Questions**:
  - How much $$$ to spend
  - Which flowers to choose
  - (Our proposed solution is on the next slide)

Review Questions

- What is Machine Learning and how is it different from regular programming?
- Name a few of Machine Learning use cases
- How does Big Data help Machine Learning?
- What is supervised learning? Unsupervised learning?

Further Reading

- Great AI Awakening – New York Times profile of on Google Brain and the people behind it
- Gentle Intro to Machine Learning
- Machine Learning Basics