Machine Learning Primer

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Lesson Objectives

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

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Machine Learning

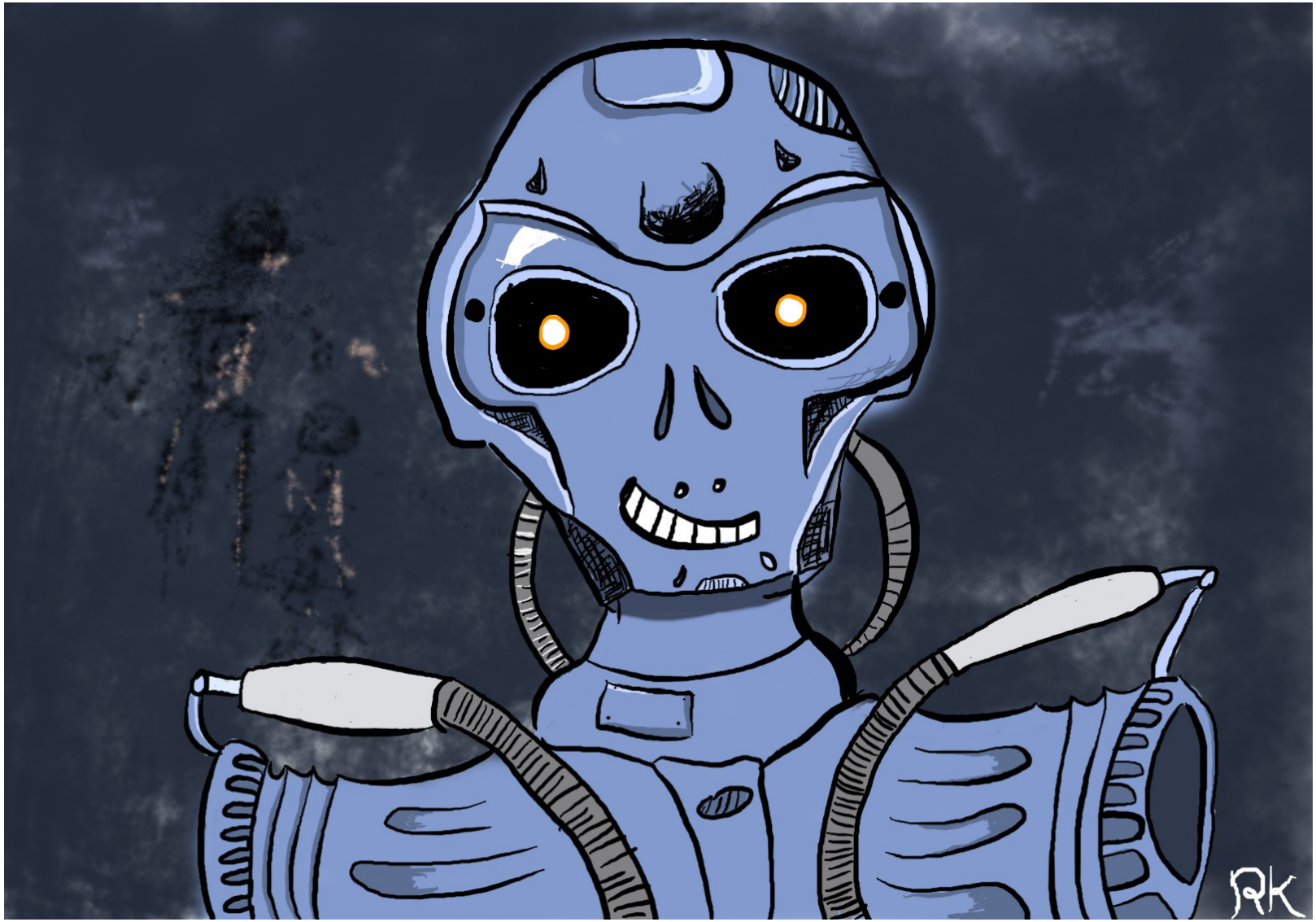


Image by Elephant Scale

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

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
  - Let it record the scenery and your reactions
  - Let it predict the next reaction

Scientific Definition of Machine Learning

- Tom Mitchell
  - "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

- Example: conducting dialogs
  - E = experience of many dialogs
  - T = task of talking like a human
  - P = probability that the program will convince you that it is a human

Formula of Machine Learning

- Usually expressed as an optimization problem

- P – performance (number)
- E – experience (data)
- M – model (formula)

- P(M/E) ➔ max

- That means, choose a model (M) that will give the optimal performance (P) over the data (E)

Linear Algebra

- What does linear algebra have to do with it?
- Consider house pricing prediction

- House features
  - Each house has square footage, number of bedrooms, etc.
  - So, each house is a “point”, H
  - Its features can be represented as H_1, H_2, H_3 , …, H_n
  - That is a vector, H_i
- Houses
  - There are many houses
  - They can be represented as H^1, H^2, H^3, …, H^m

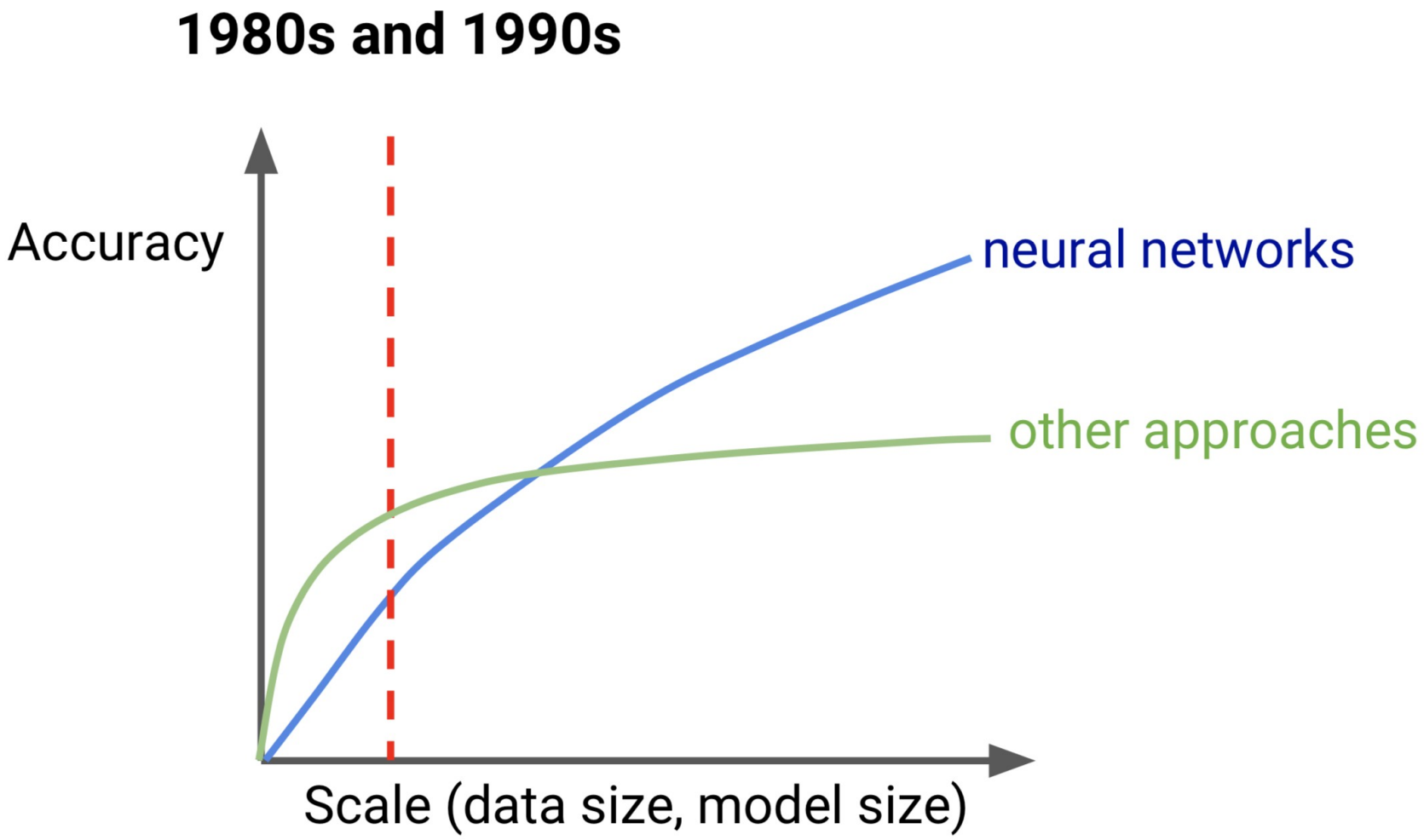
- Thus, our data is a matrix H, n*m

A Glimpse of AI History

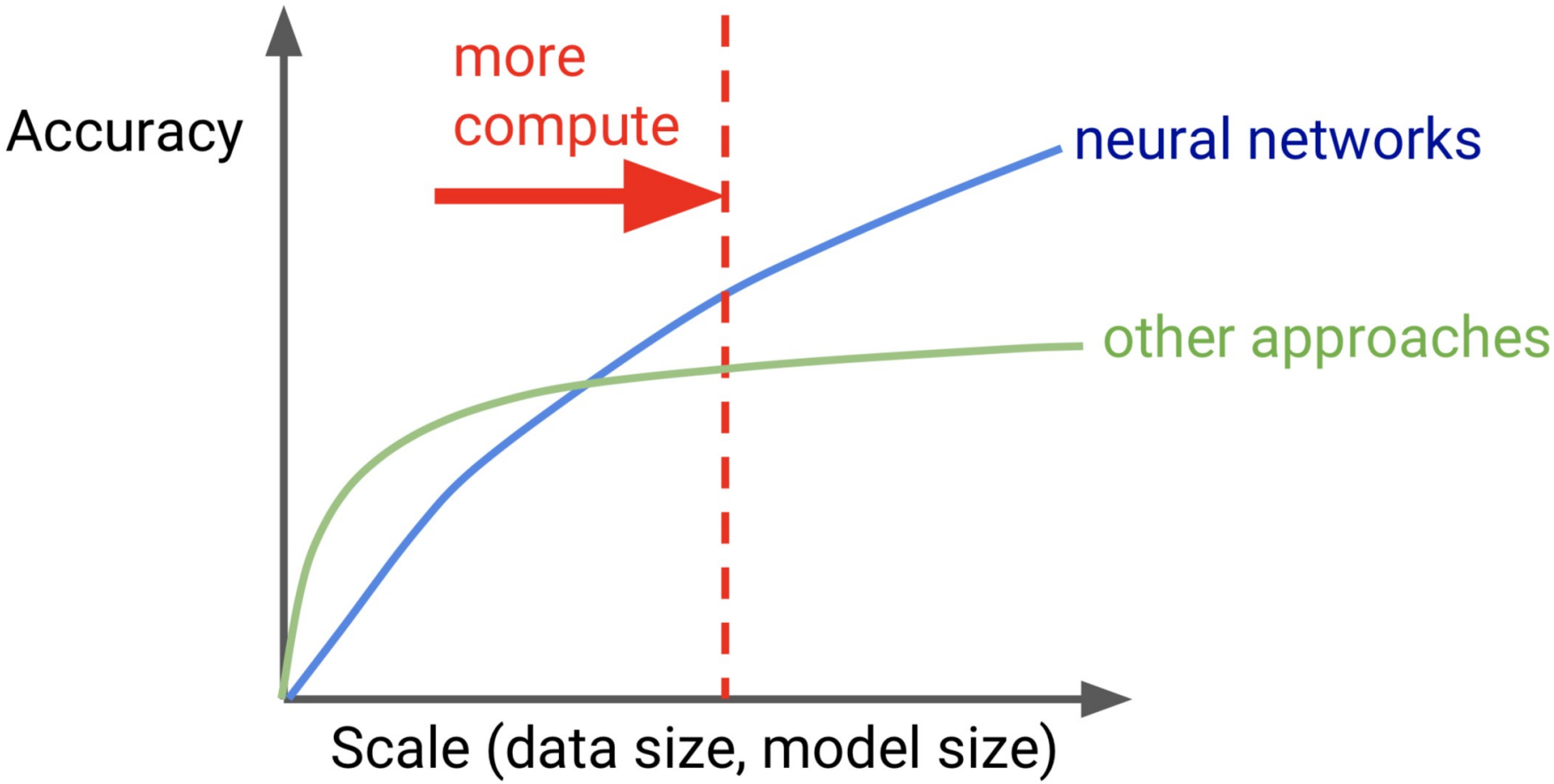
- Sixties
  - Commercial computers & mainframes
  - Computers play chess
- Eighties
  - Artificial intelligence (AI) gets 'oversold', doesn't live up to the promise and gets a bad rap
- 21st century
  - Big Data changes it all

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

1980's and 1990's



Now



Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

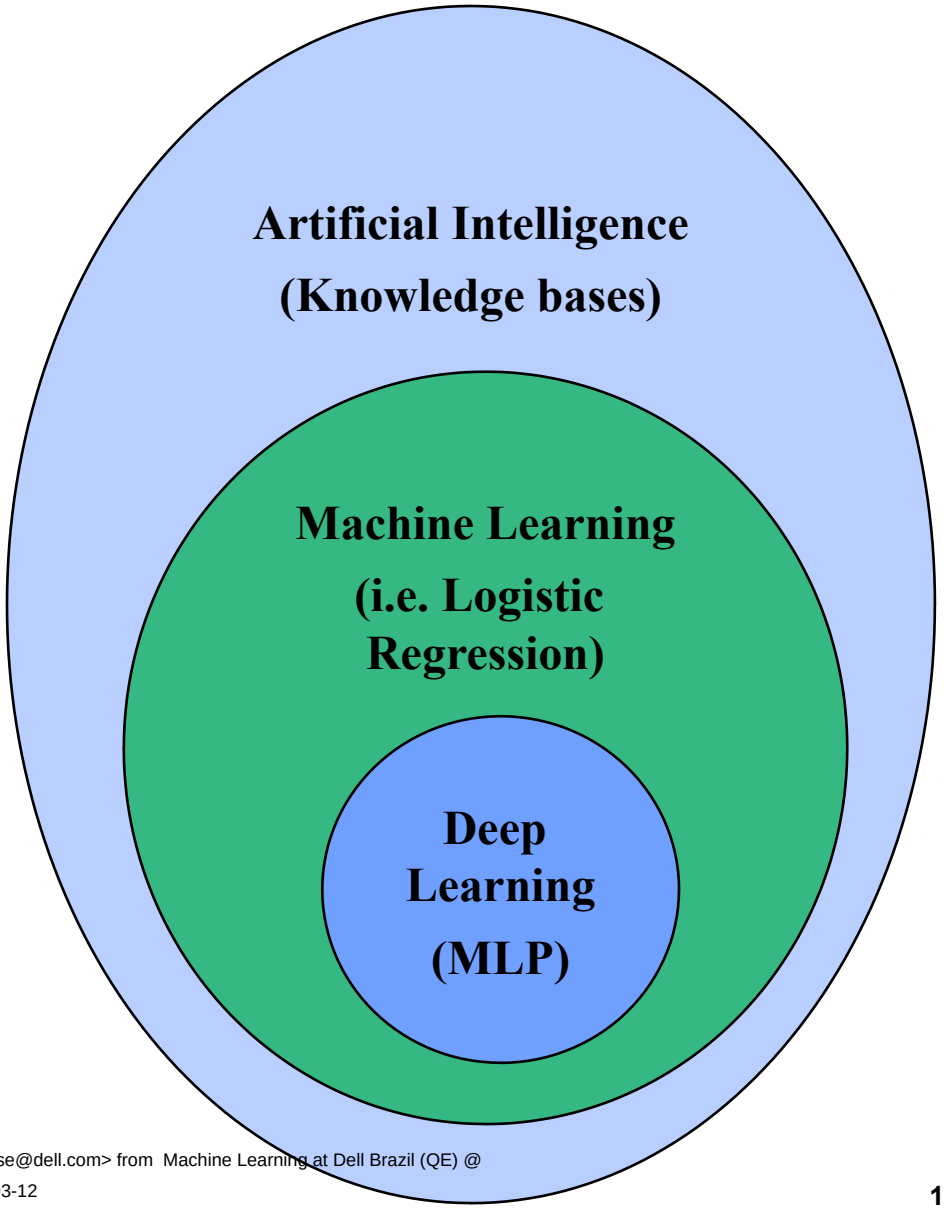
Machine Learning Applications

- 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

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

AI, Machine Learning, Deep Learning

- **Artificial Intelligence (AI)**
  - Combined learning technologies
- **Machine Learning**
  - Math and stats
- **Deep Learning**
  - Neural networks
  - Representation learning



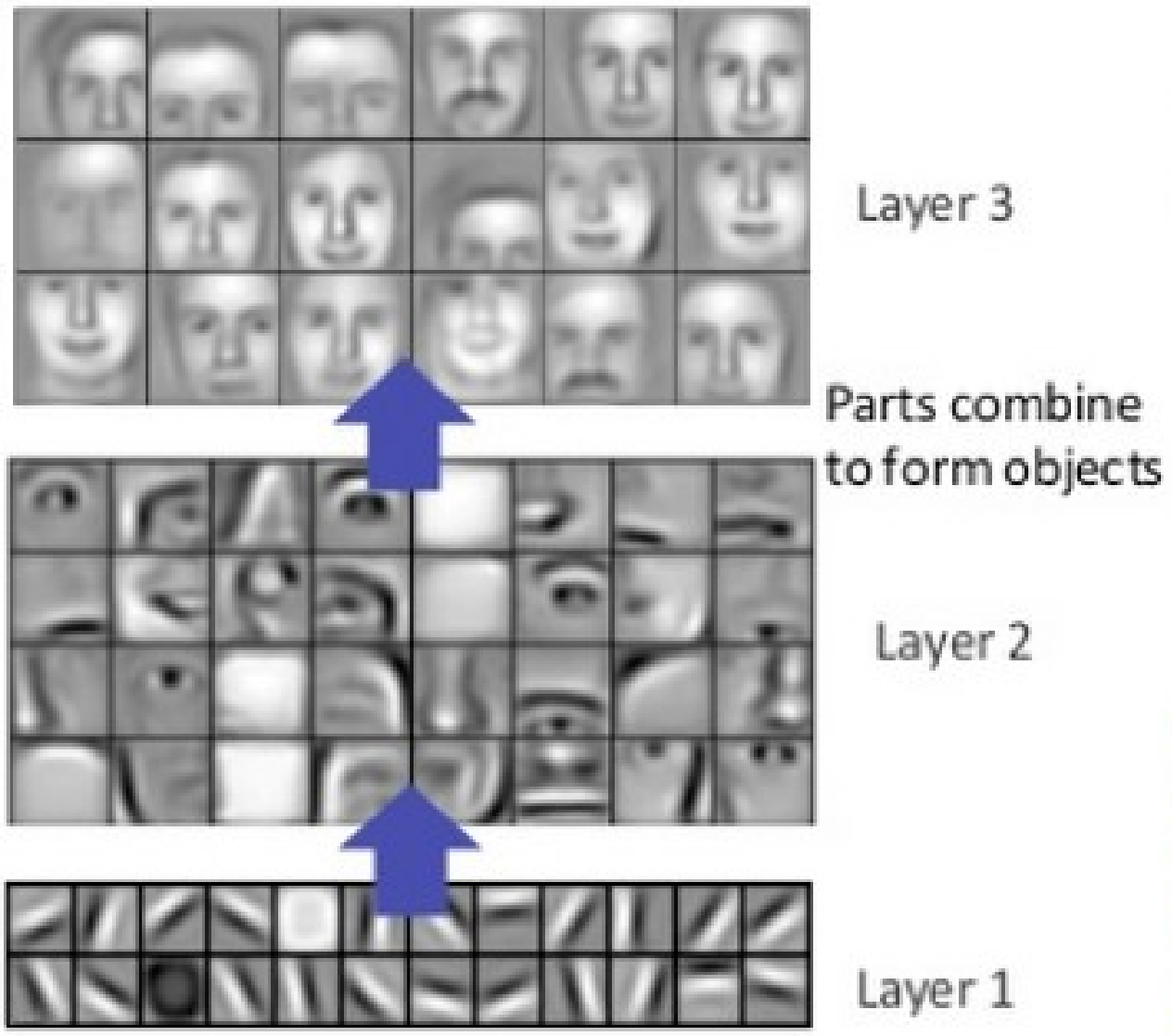
Neural Networks

- Modeled after human brain
- Recognized patterns
  - Numerical
  - Contained in vectors
  - Translated from real-world data
    Images, Sound, Text, Time series
- Invented in the 1960's
- "Re-invented" in 2012

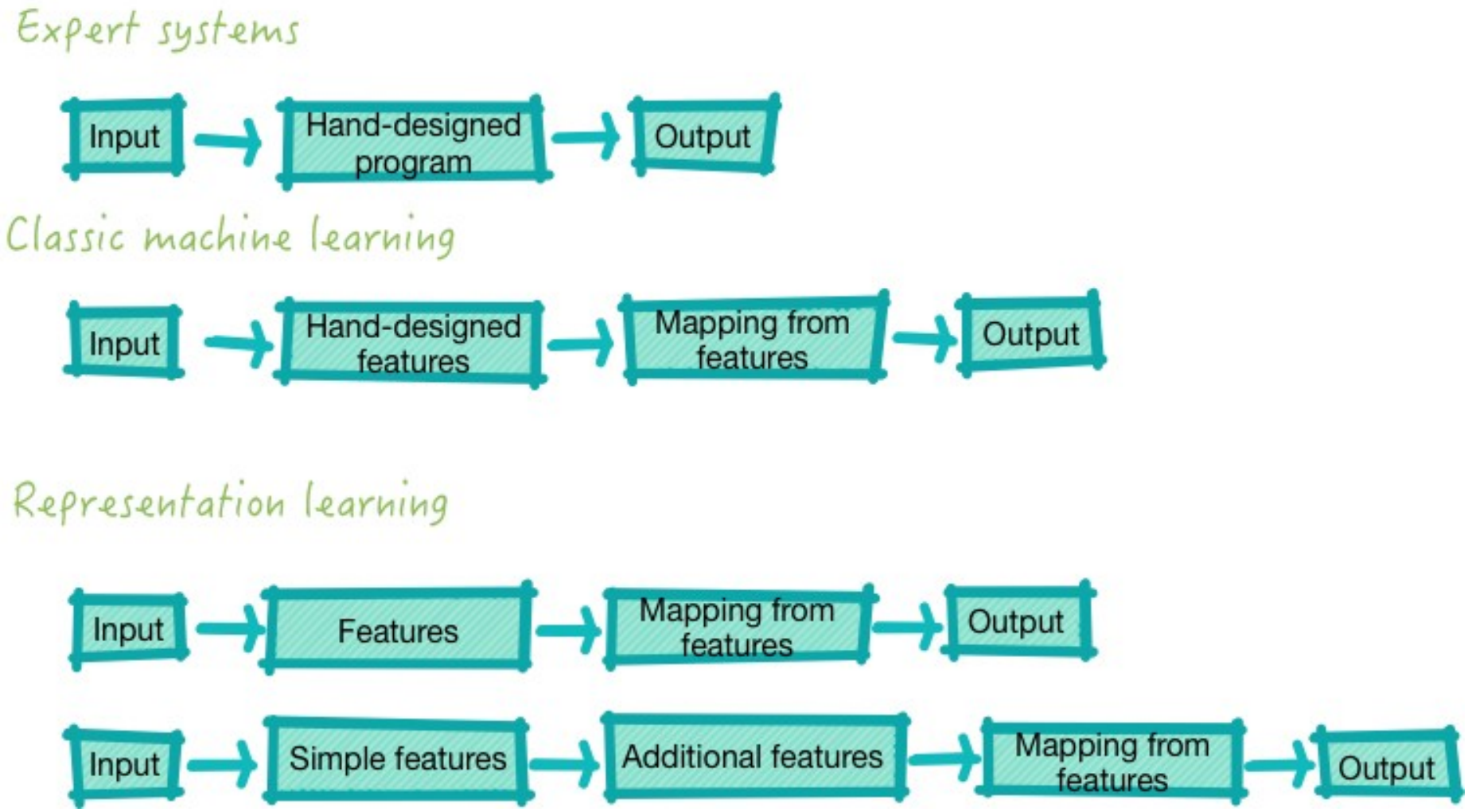
Image by Elephant Scale

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Deep Neural Network – Face Recognition



Machine Learning Levels



Types of Machine Learning

- **Supervised Machine Learning:**
  - A model is trained with labeled training data
  - Model is tested on test data to see performance
  - Model can be applied to unknown data
  - Regression and classification are 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 that contains unlabeled (usually a lot) and labeled (usually little) data

Machine Learning Types: Supervised

- Model learns from training data
- Then predicts on new data

Algorithms	Description	Applications
Regression	Continuous output	- House prices - Stock market
Classification	Categorical output	- Spam classification - Fraud detection

Supervised Learning

- Supervised learning
  - Answer is given
  - Examples:
    - House data ➔Price
    - Face ➔ Age

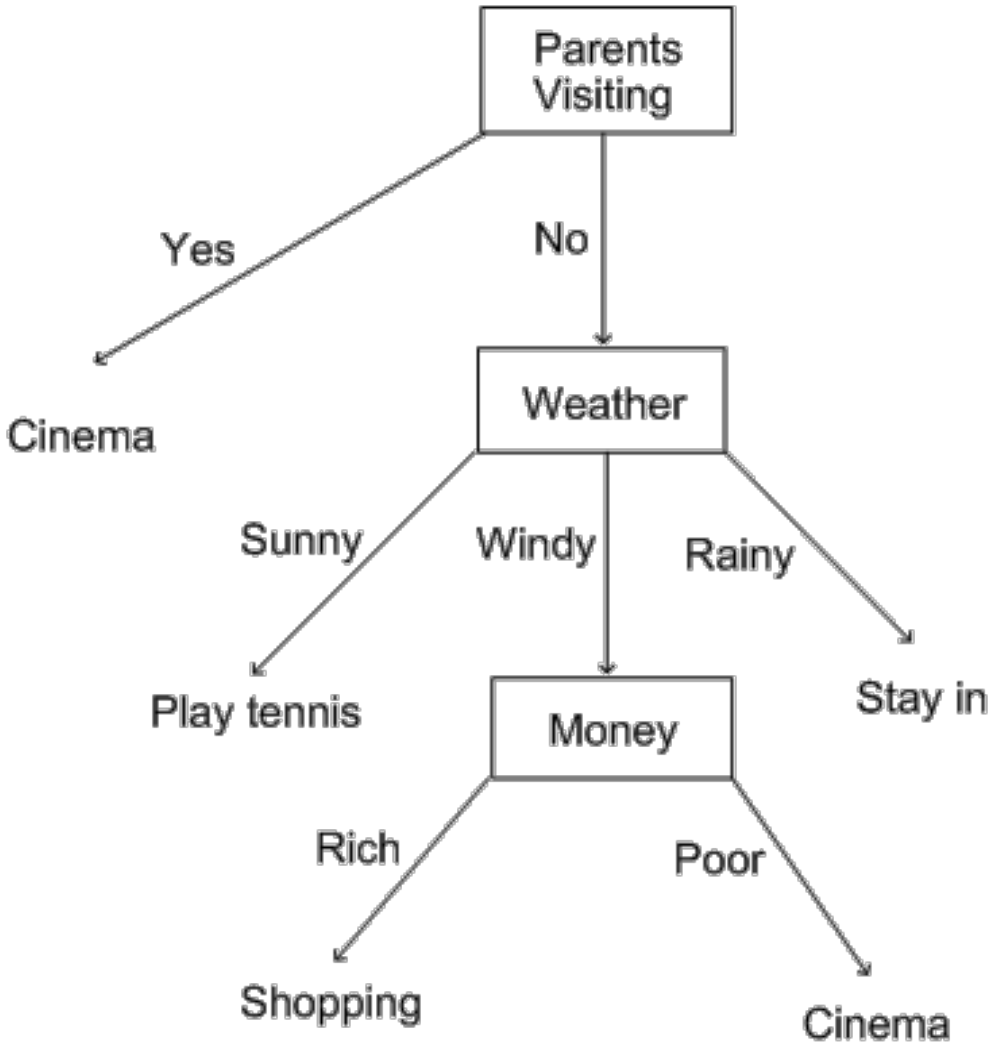
- Types of supervised learning
  - **Regression**
    - Continuous output
  - **Classification**
    - Animal pic ➔ Cat or Dog

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

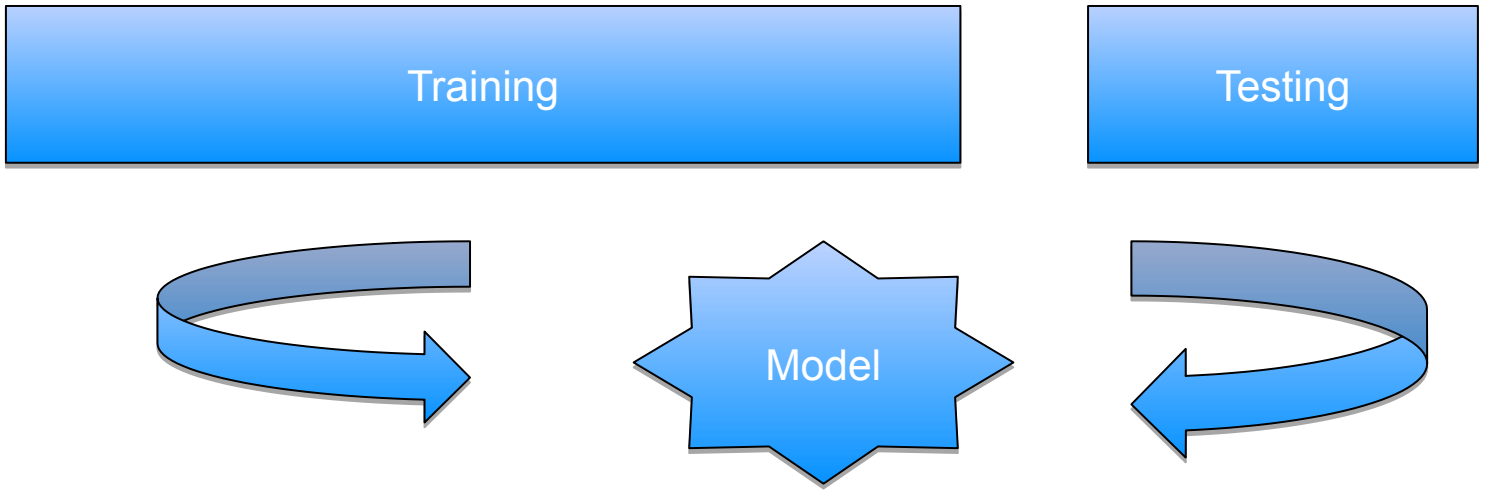
Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Decision Tree Example



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



Machine Learning Types: Unsupervised

- Example: Party Problem
- Algorithm tries to find patterns in data

Algorithms	Description	Applications
Clustering	Find groups	- Market segmentation - Document grouping
Anomaly detection	Find anomalies	- Engine assembly defects

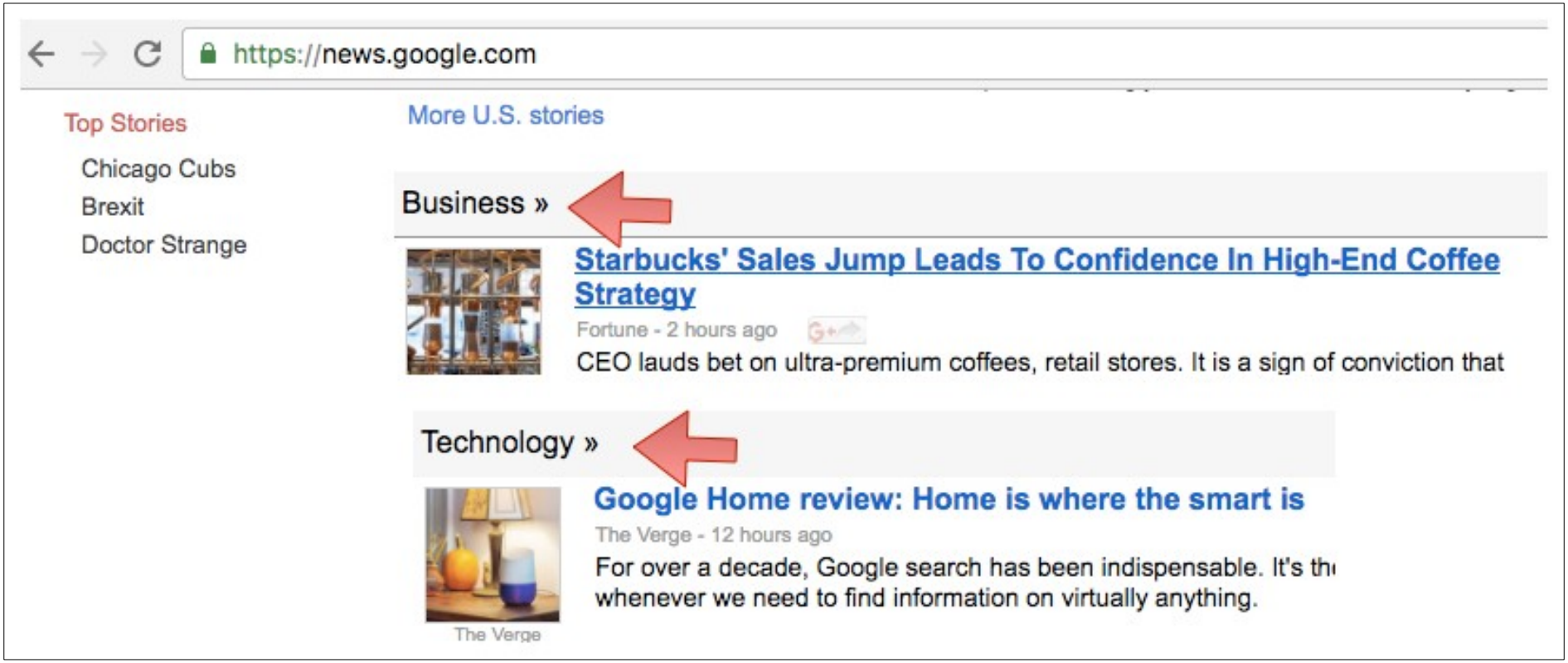
Unsupervised Machine Learning

- 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

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

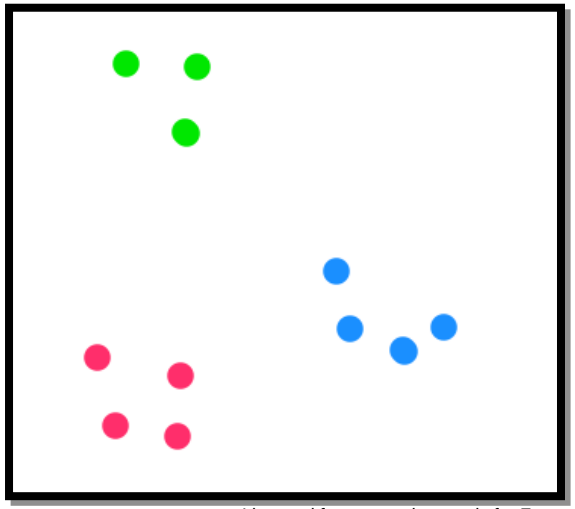
Unsupervised Example: Google News

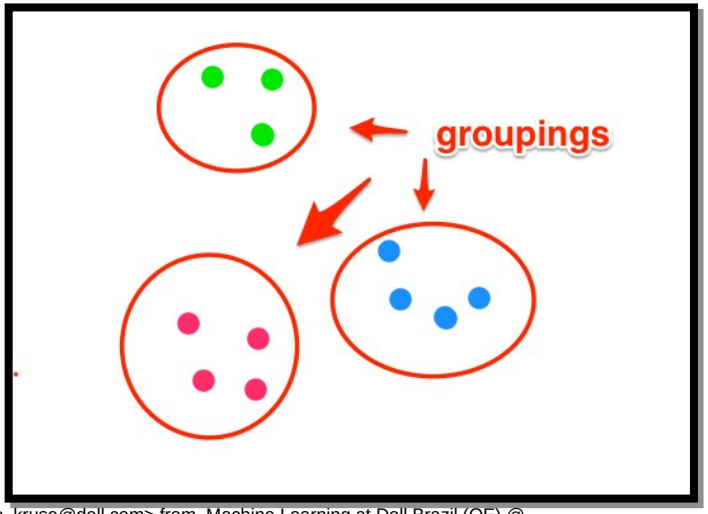
- Google News algorithm automatically groups **related news stories** into sections



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)

Technology Stack Comparison

Technology	Pros	Cons
R	- Rich environment - Thousands of libraries	- Rough on data cleanup - Not a general purpose language - **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**

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	
Cloud	- Google: GCP - Microsoft: ML on Azure - Amazon: SageMaker	

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 (SageMaker)
  - Ready to go algorithms
  - Wizards to guide
  - Scalable on Amazon Cloud
  - Integrated with AWS

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
  - http://www.h2o.ai/

Tools for Scalable Deep Learning

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

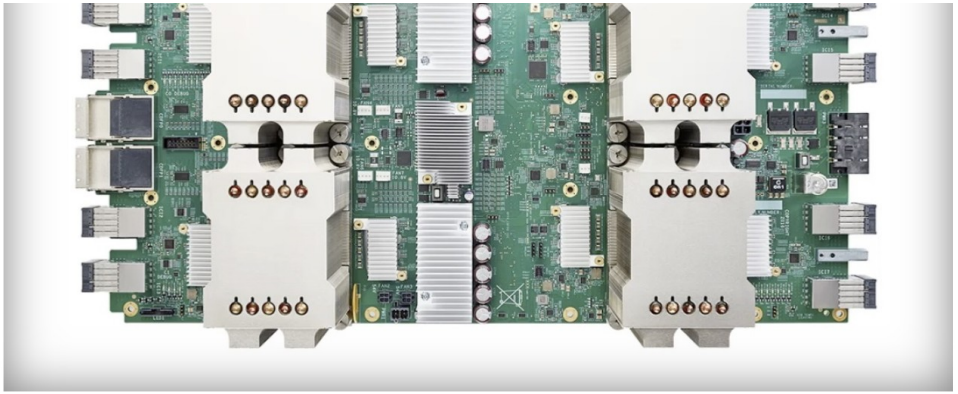
- **Intel BigDL**
  - Deep learning library
  - Built on Apache Spark
  - Language support: Python, Scala

Hardware Progression

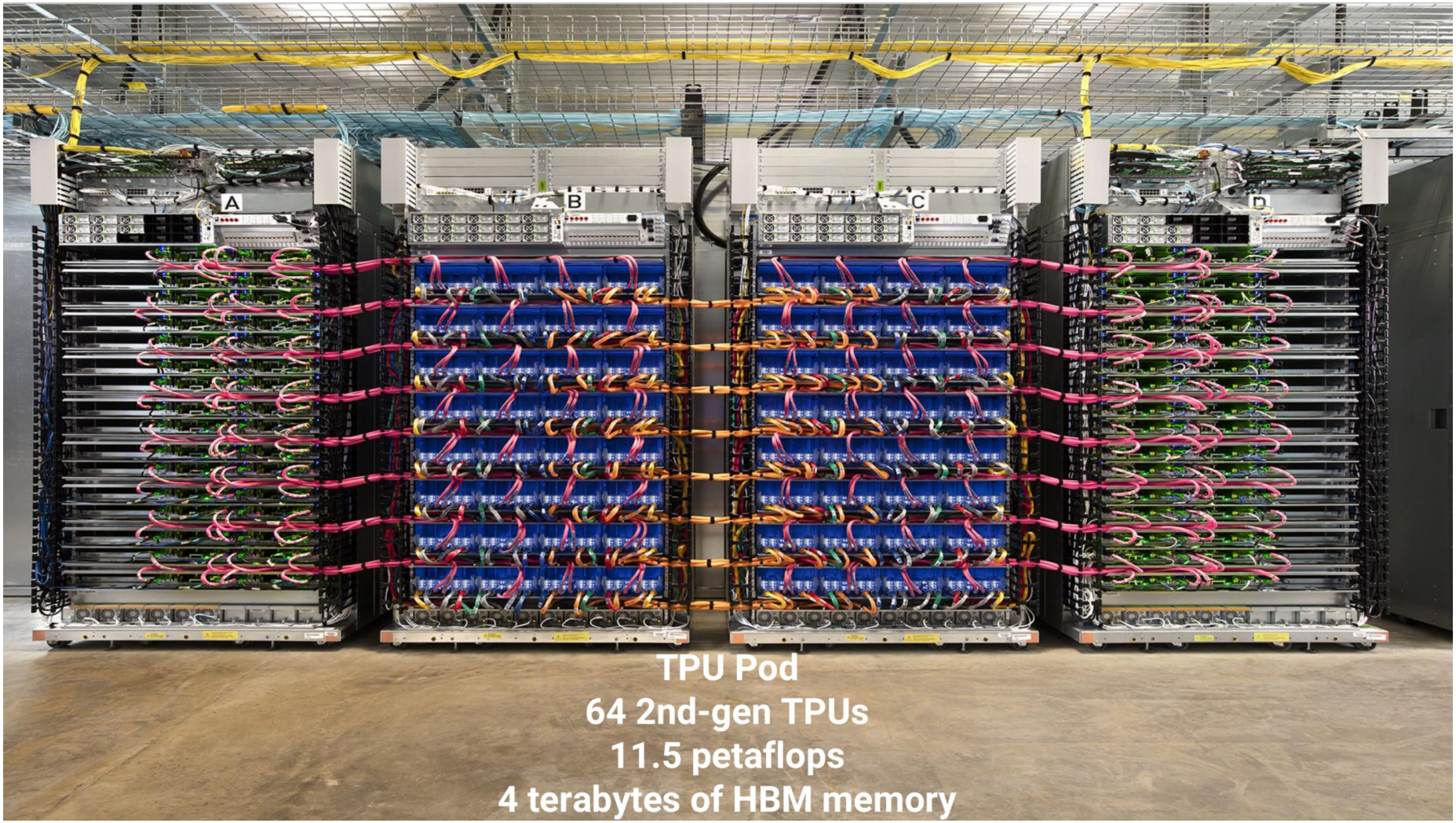
- CPU
  - Moore's law
  - Number of transistors x2 in 2 years
  - Till 2012
- GPU
  - Performance x1000
  - Scala, Go
- ASIC
  - Application-specific integrated circuit
- Computation-specific hardware

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 than CPUs or 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

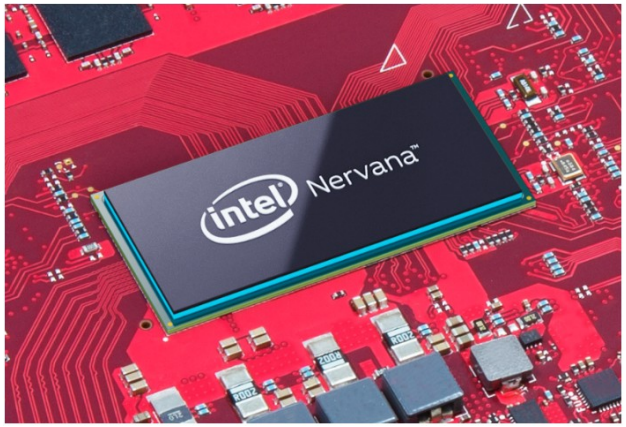


Google TPU Computer



AI Chips

- Intel+Facebook “Nervana”
  - NNP – Neural Network Processor
  - Pre-trained learning
- Nvidia is the current market leader
- Amazon (”Inferentia”)
- Alibaba
- Startups



Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12

Lab: Design ML Algorithm

Lab

- **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)

Lab: Design ML Algorithm

- **Data sources**:
  - Store registry for all buyers
    - Watch for patterns indicative of discord and harmony restoration
  - Studies (especially old, archived studies, family counseling, etc.)
  - Search history, blogs, tweets
- **Feature extraction**:
  - Collect features (what we know about people)
  - History, results, flower involvement, price, type
- **Model**:
  - Which model might be a fit: linear regression, logistic regression
  - Train and test
- **Apply results to yourself**:
  - Celebrate!
  - Or go back to the drawing board

Review Questions

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

Licensed for personal use only for Fernando K <fernando_kruse@dell.com> from Machine Learning at Dell Brazil (QE) @ 2019-03-12