Artificial Intelligence, Machine Learning, Deep Learning, Cisco?

Tomáš Ondovčík
Systems Architect
29.10.2019

Part 1:
Introduction to
AI / Machine Learning, ...

An Introduction to ML/AI

Lip Reading

Deep Blue

Watson

Deep Mind

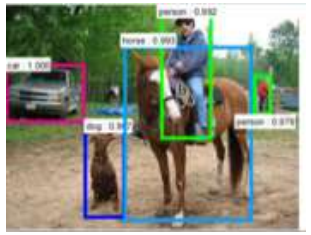
Playing Music

WaveNet (TTS)

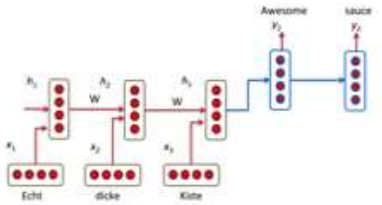
WaveNet

Counting people

Recognition



NLP Translation



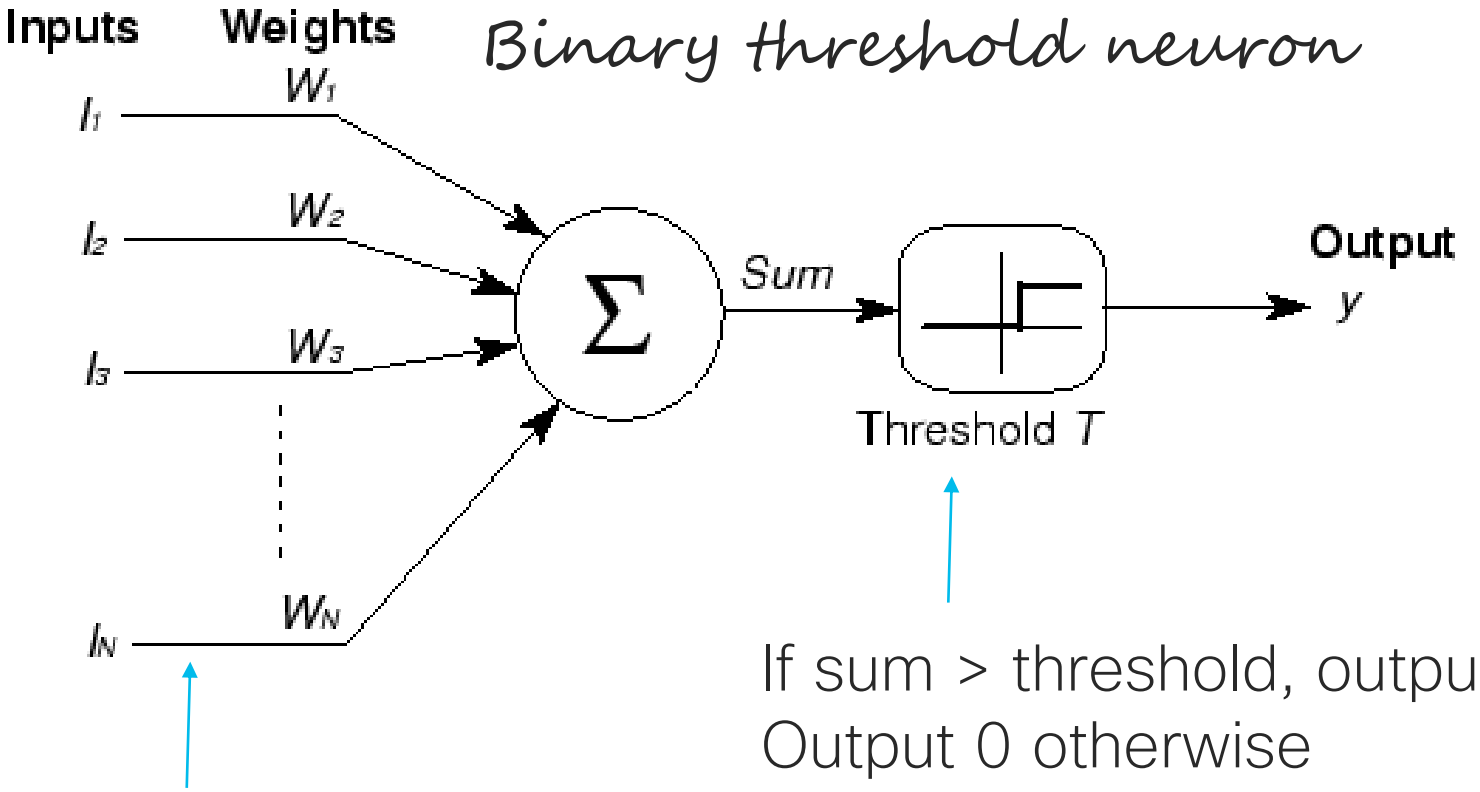
Self-driving

Networks

And Many more: CRM, Healthcare, Personal Assistants, ...

McCullogh & Pitts – 1943

Von Neumann used this logic when designing the "universal computer"



If sum > threshold, output 1
Output 0 otherwise

Weighted inputs

In the 1950s and 1960s, Principles of Neurodynamics were examined and Symbolic ML expanded

When I show these shapes to the camera

This IBM 704 computer can say “it’s a triangle”

Why Now?

Data "tsunami"

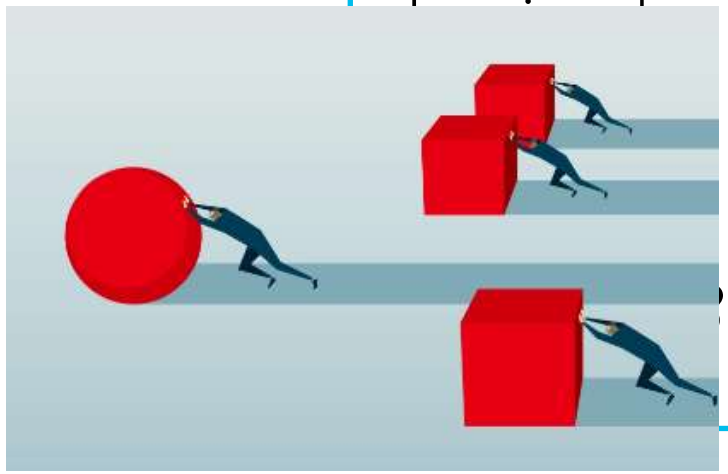
Computing power

$$Y = X\ \beta$$

$$\begin{bmatrix} \text{Price}_1 \\ \text{Price}_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & \text{Bedroom}_{11} & \text{Bathroom}_{12} & \cdots & \text{Age}_{1\ p-1} \\ 1 & \text{Bedroom}_{12} & \text{Bathroom}_{22} & \cdots & \text{Age}_{2\ p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{Bedroom}_{i2} & \text{Bathroom}_{i2} & \cdots & \text{Age}_{I\ p-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & \text{Bedroom}_{n1} & \text{Bathroom}_{n2} & \cdots & \text{Ag} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \end{bmatrix}$$

above equation for coefficients

$$= \left(X^T X\right)^{-1}\ X^T\ Y$$



Insatiable desire for efficiency / productivity

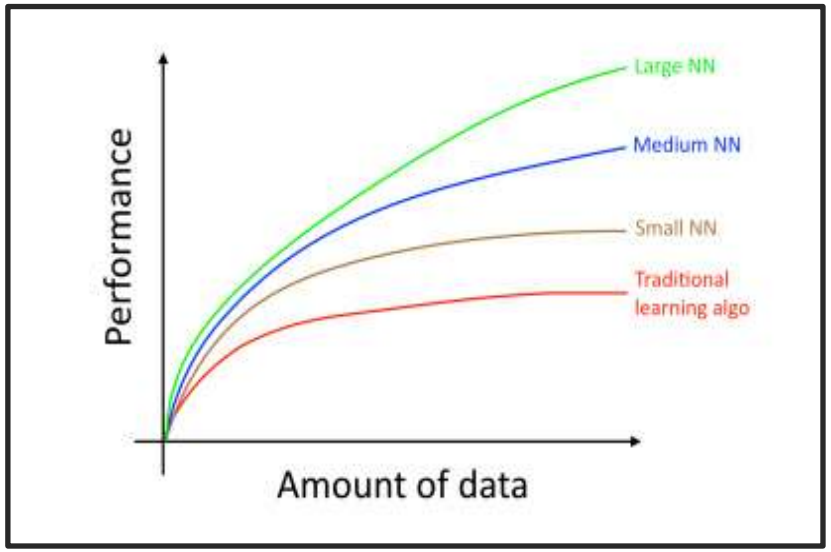
Availability of tools

The Emergence of Large Data Sets and ML

Data is the rocket fuel of Machine Learning – Andrew Ng

- Open data sets have been a crucial factor in the success of ML
  - http://yann.lecun.com/exdb/mnist/
  - http://image-net.org/
  - https://bis.lexisnexis.com
  - https://catalog.data.gov/dataset

- Allows for direct comparison of learning and inference algorithms
- The result has been improvement of error rates
  - Video analytics (facial recognition)
  - NLP/Voice recognition

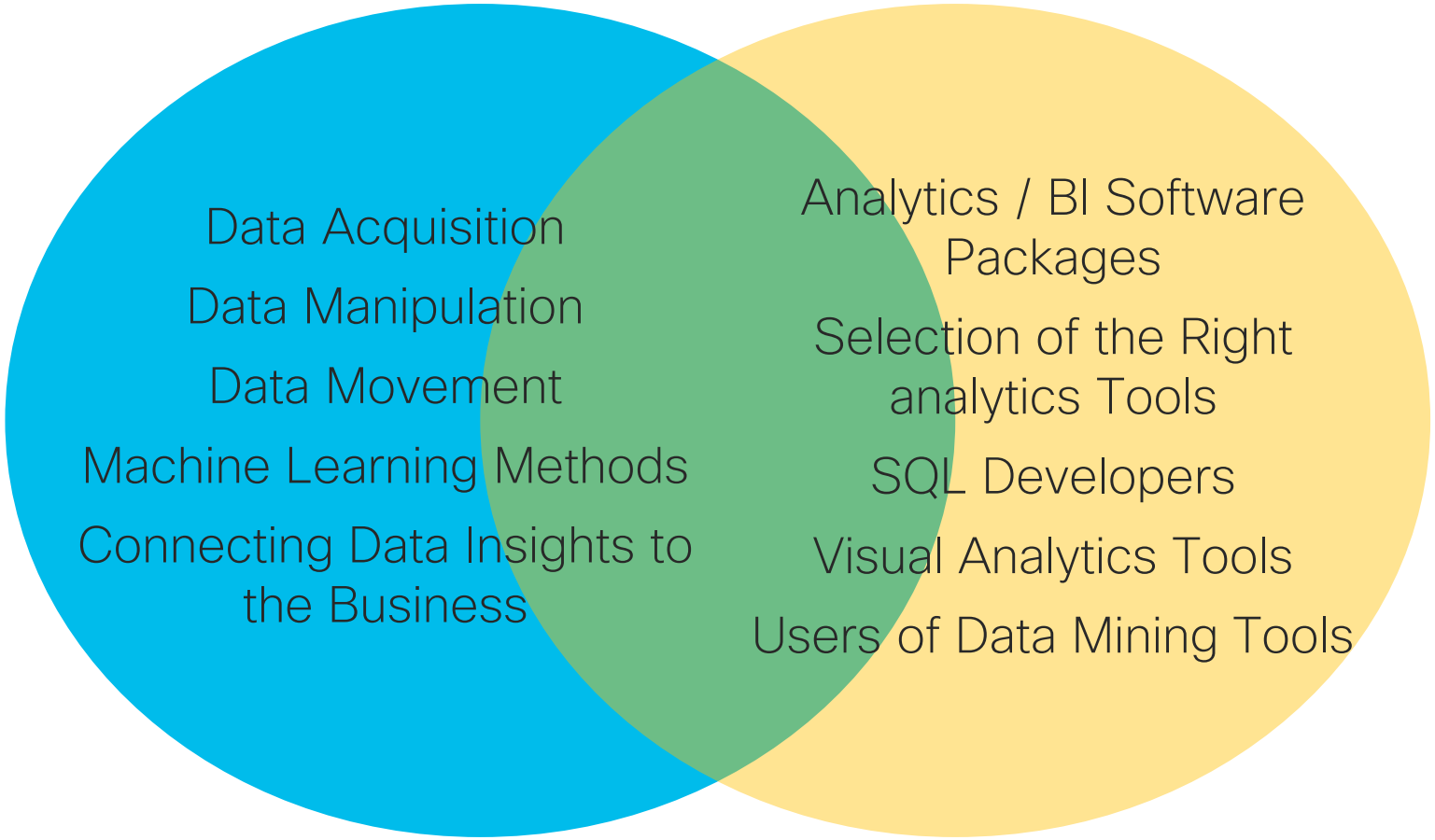


Graph courtesy Andrew Ng

MNIST – Handwritten Digit Data Set

- Training set of 60,000 examples (6000 per digit)
- Test set of 10,000 examples (1000 per digit)
- Each character is a 28x28 pixel box

Understanding the Roles



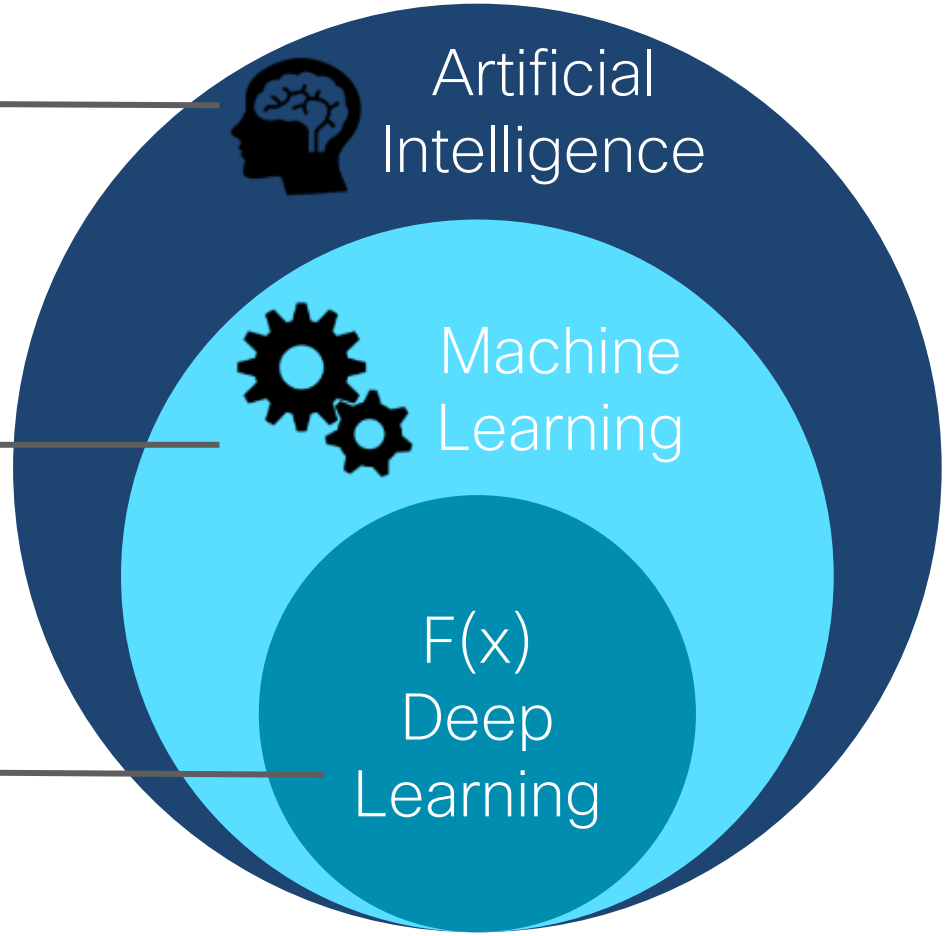
Network Engineers build the platform supporting the business

Artificial Intelligence, Machine Learning, and Deep Learning (AI/ML/DL)

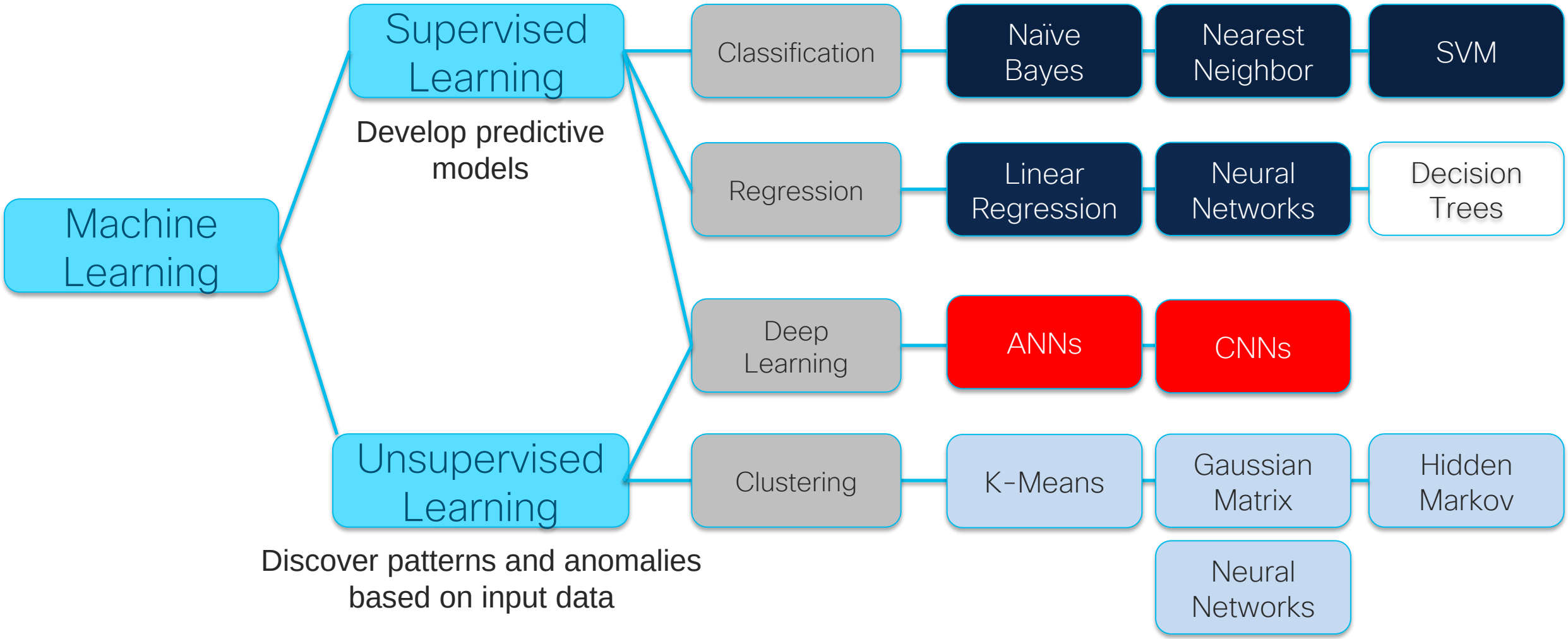
Artificial Intelligence
Technique where computer can mimic human behavior

Machine Learning
Subset of AI techniques which use algorithms to enable machines to learn from data

Deep Learning
Subset of ML techniques which uses multi-layer neural network to learn



It is a Complex Landscape



Machine Learning: The Main Methods

Supervised Learning

Learning with a labeled training set

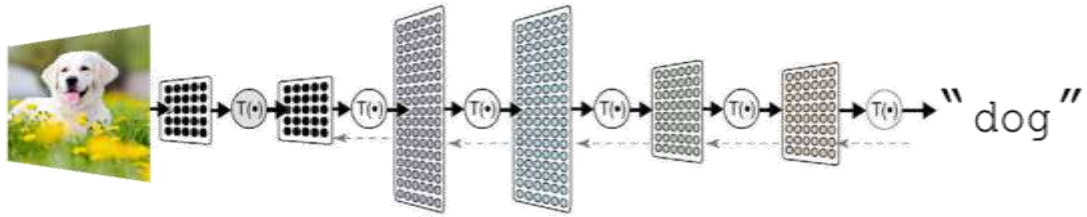
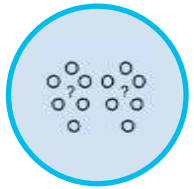
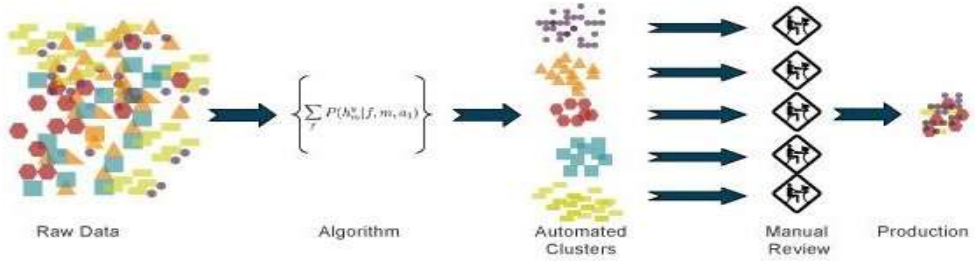


image by Jeff Dean



Unsupervised Learning

Discovering patterns in unlabeled data



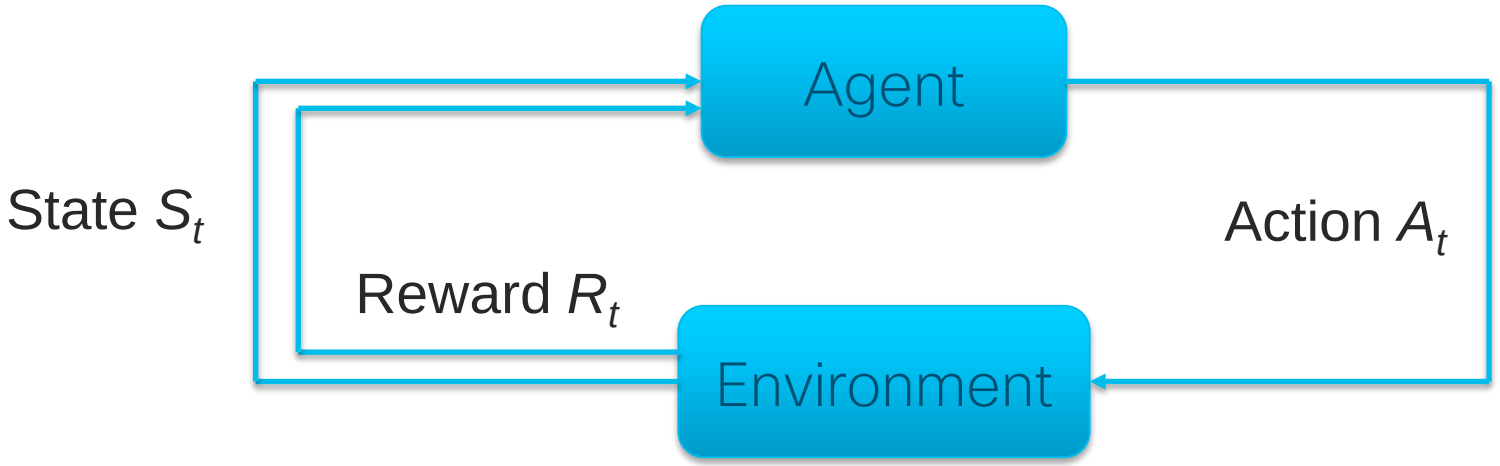
Inference (Statistical Learning)

How confident are you in the result?

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$

Reinforcement Learning

- A semi-supervised learning model
- No training data or correct/incorrect guidance needs to be given.
- Involves behavioural psychology
- In summary, a lot of trial and error

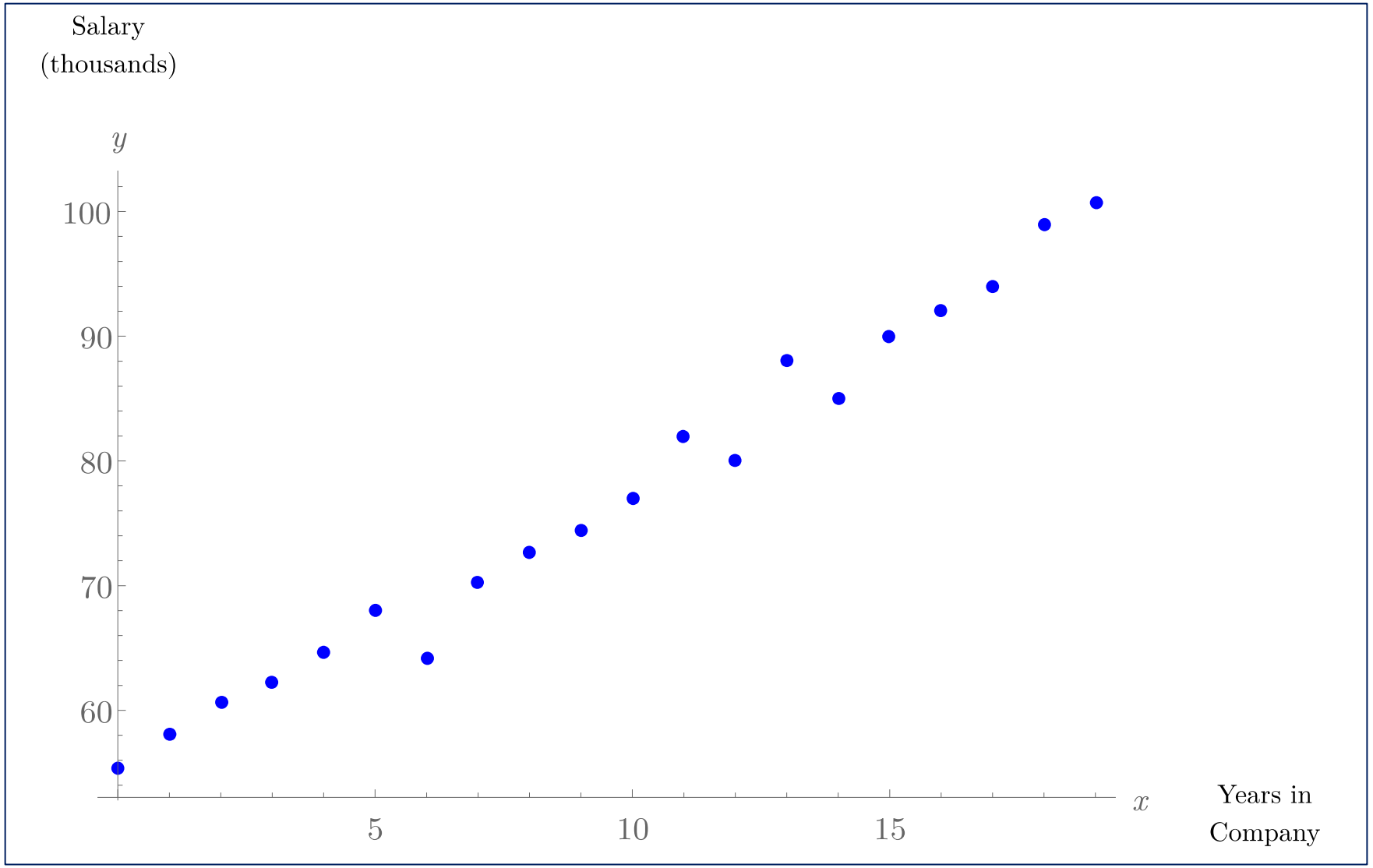


Supervised Learning Part 1
Linear Regression

Supervised Learning: Linear Regression (Single Variable)

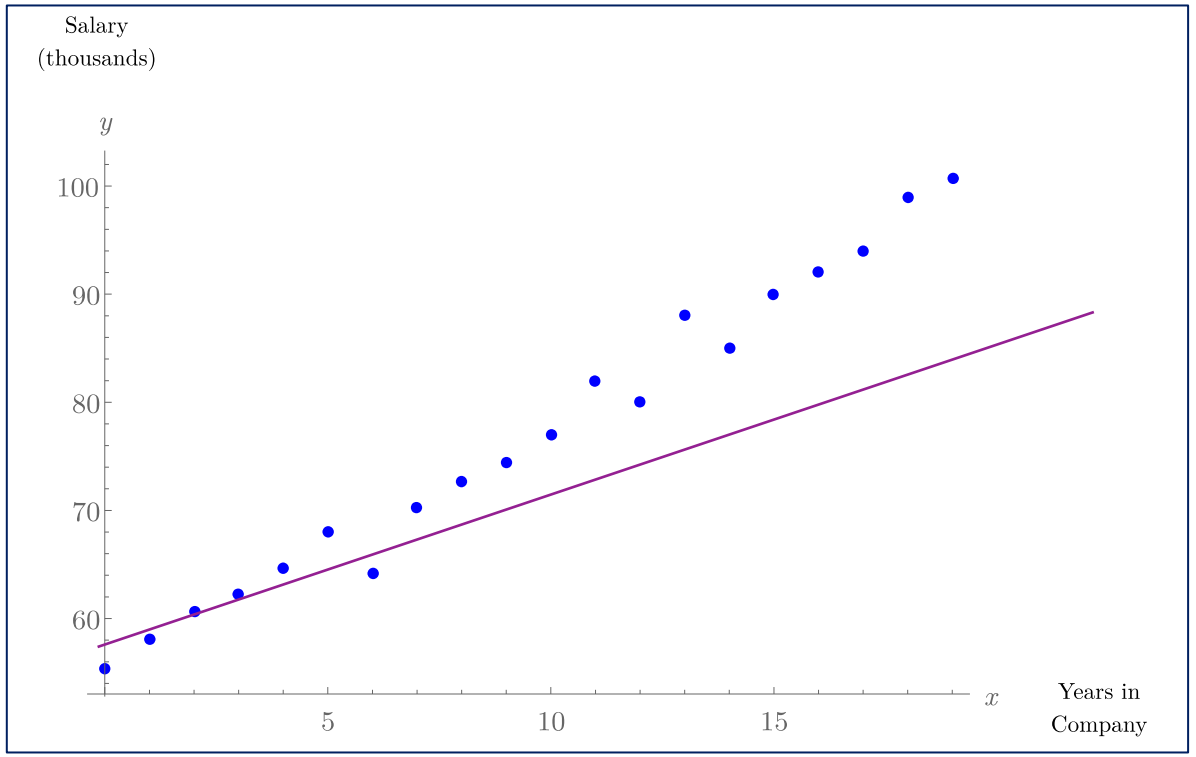
"How much is our headcount going to cost us in 5 years?"

Data sample from across company

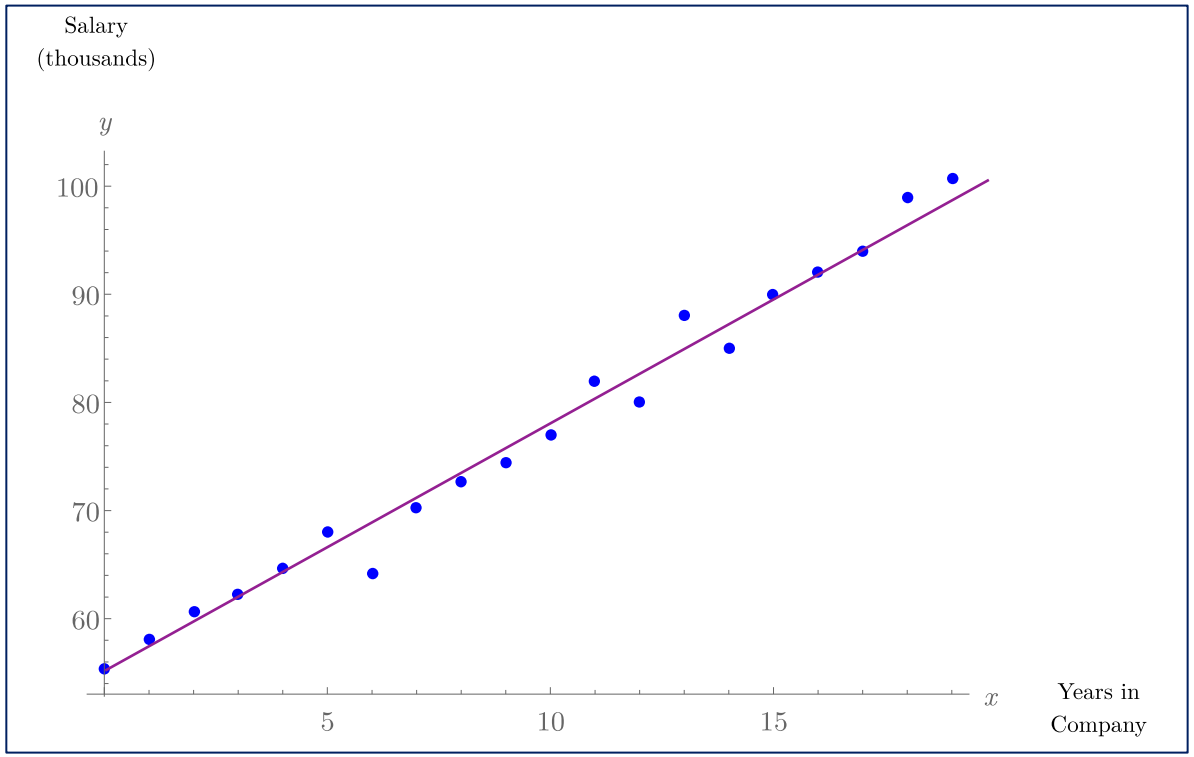


Supervised Learning: Linear Regression (Single Variable)

$$y = b + mX \longrightarrow Y = \beta_0 + \beta_1 X \longrightarrow \text{Salary} = \beta_0 + \beta_1 \text{Years}$$



Bad fit

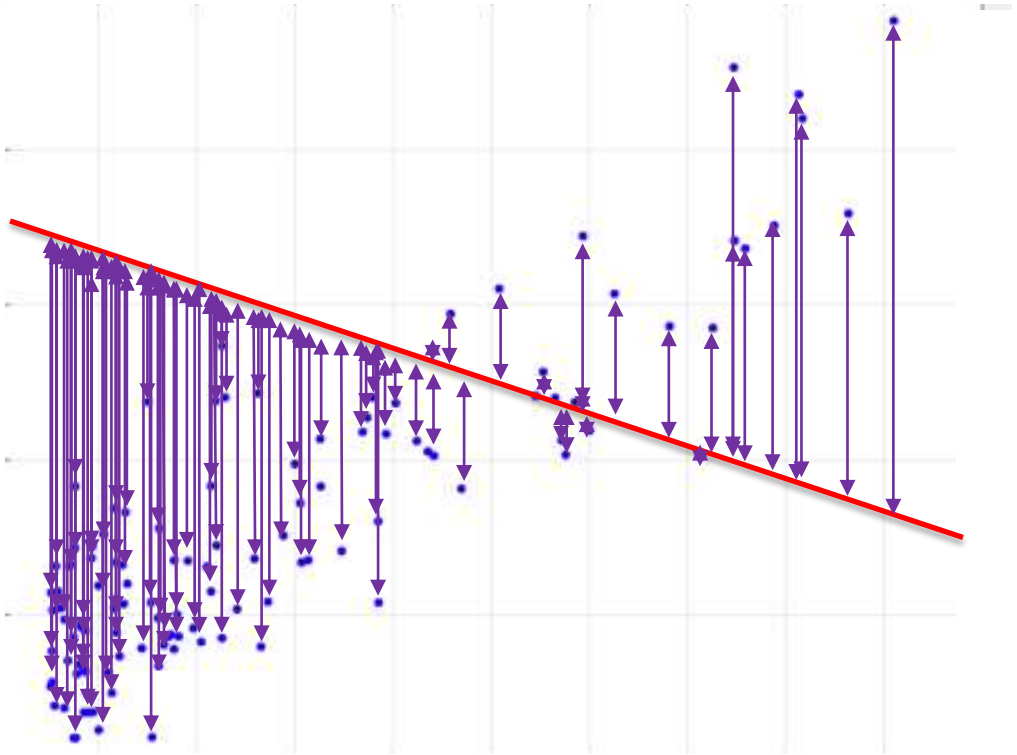


Good fit

"What makes one fit bad and another good?"

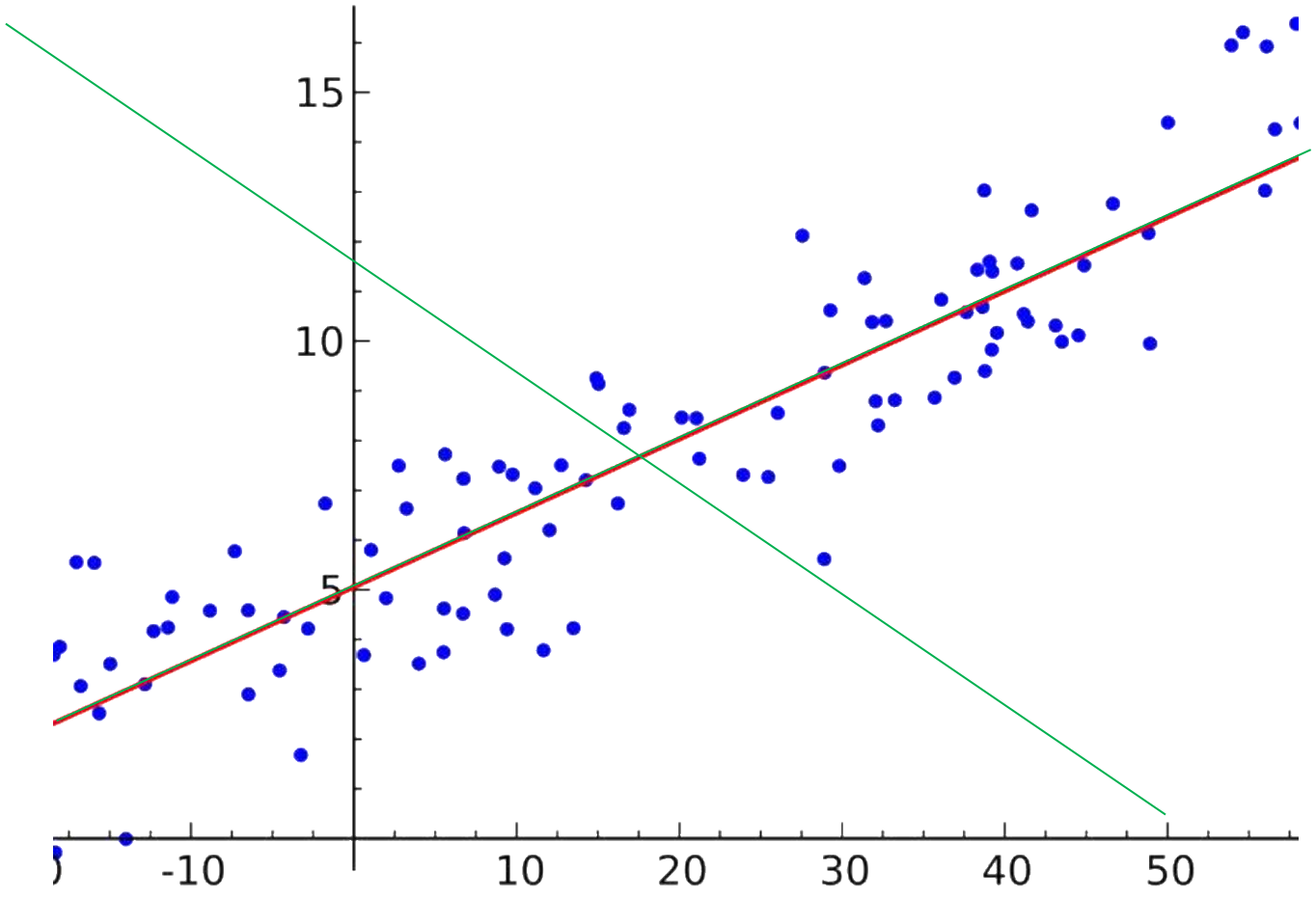
How Close Is Our Prediction to the Dataset?
The Cost Function

$$J\,(\theta_0, \theta_1) = \frac{1}{2} m \sum_{i=1}^{m} ((\theta_0 + \ \theta_1 x_i) - y_i)^2$$

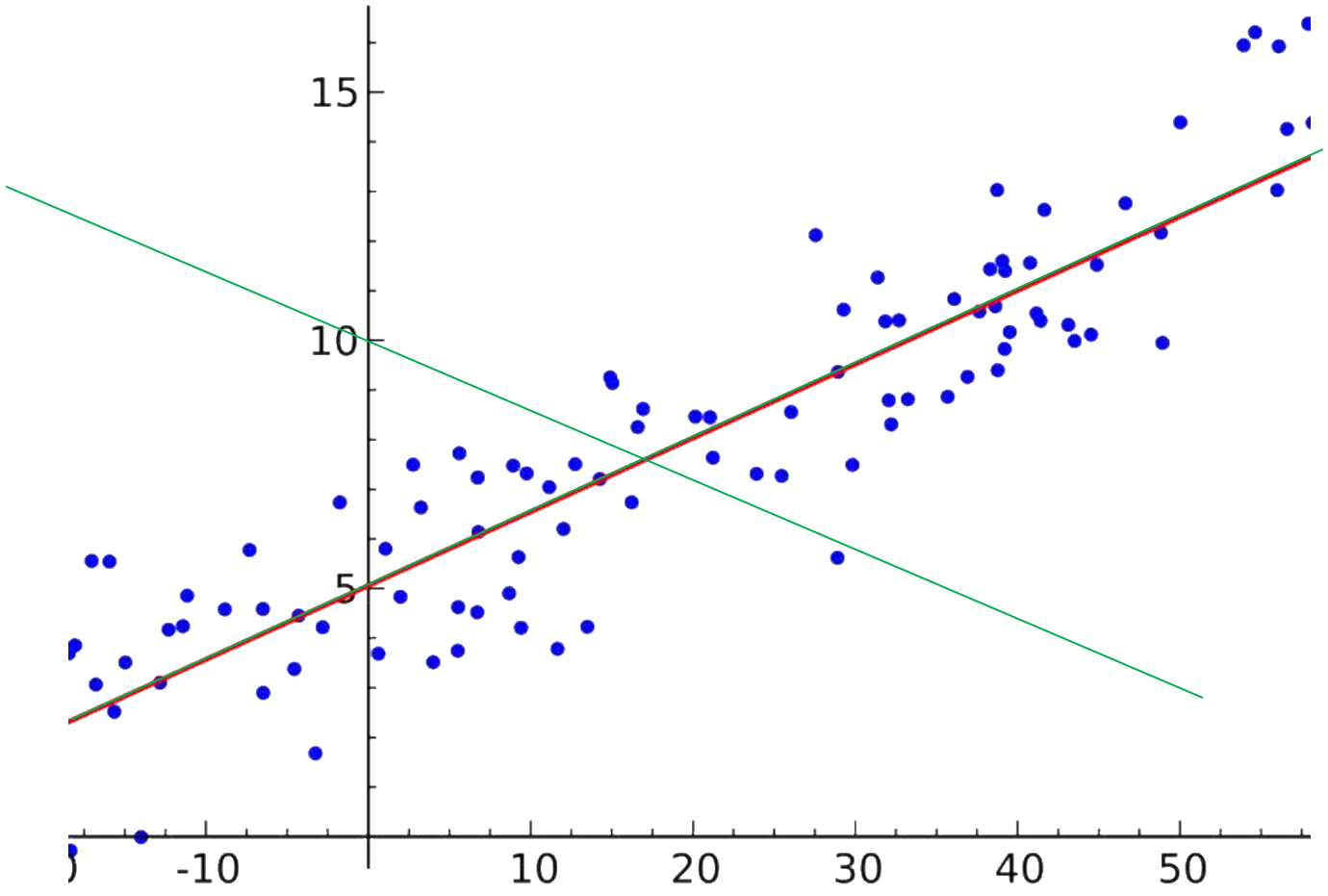


How far am I from the real y, if I use my random θ_0 and θ_1 and do θ_0 + θ_1x

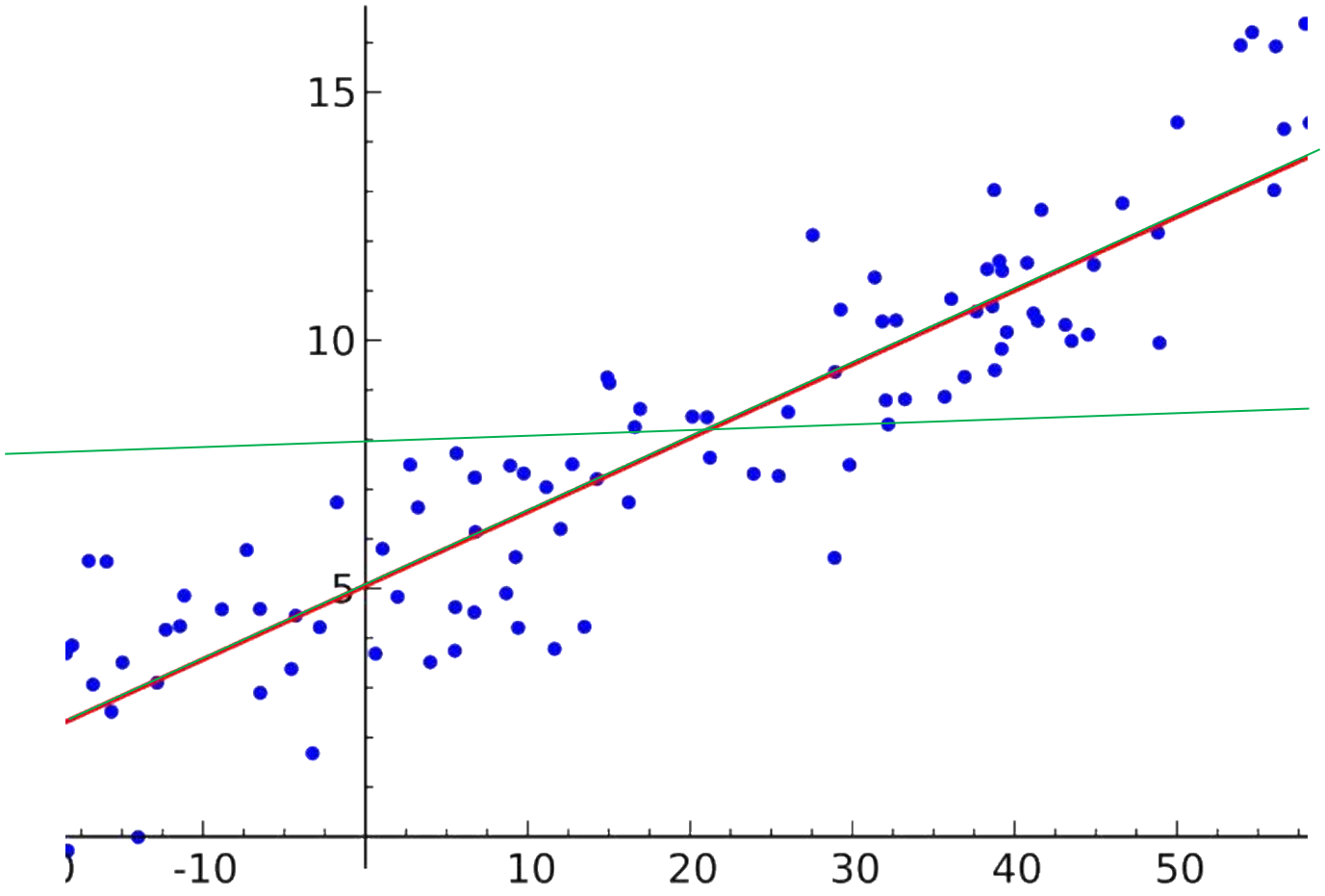
Gradient Descent
Getting Closer to the Answer



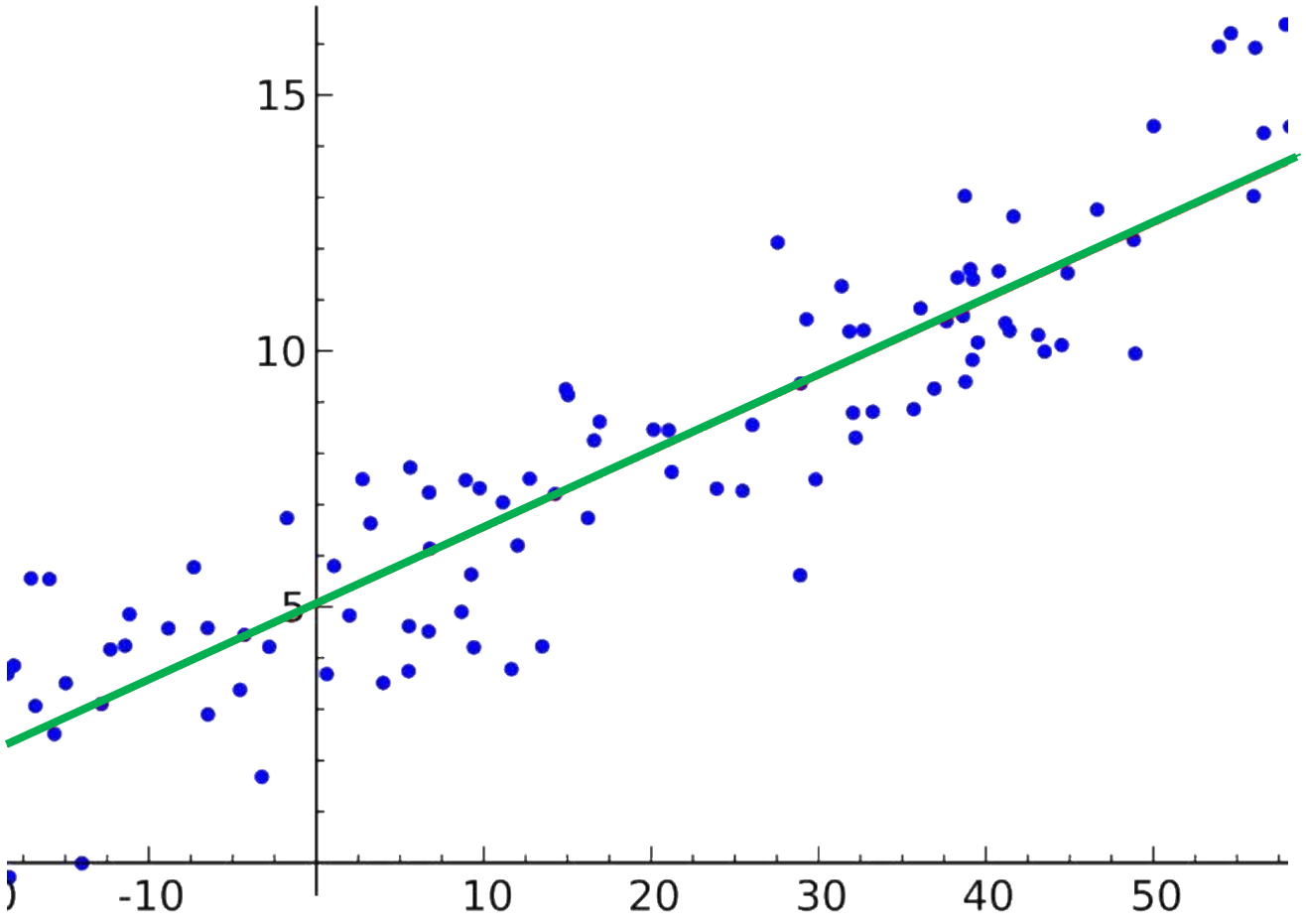
Gradient Descent
Getting Closer to the Answer



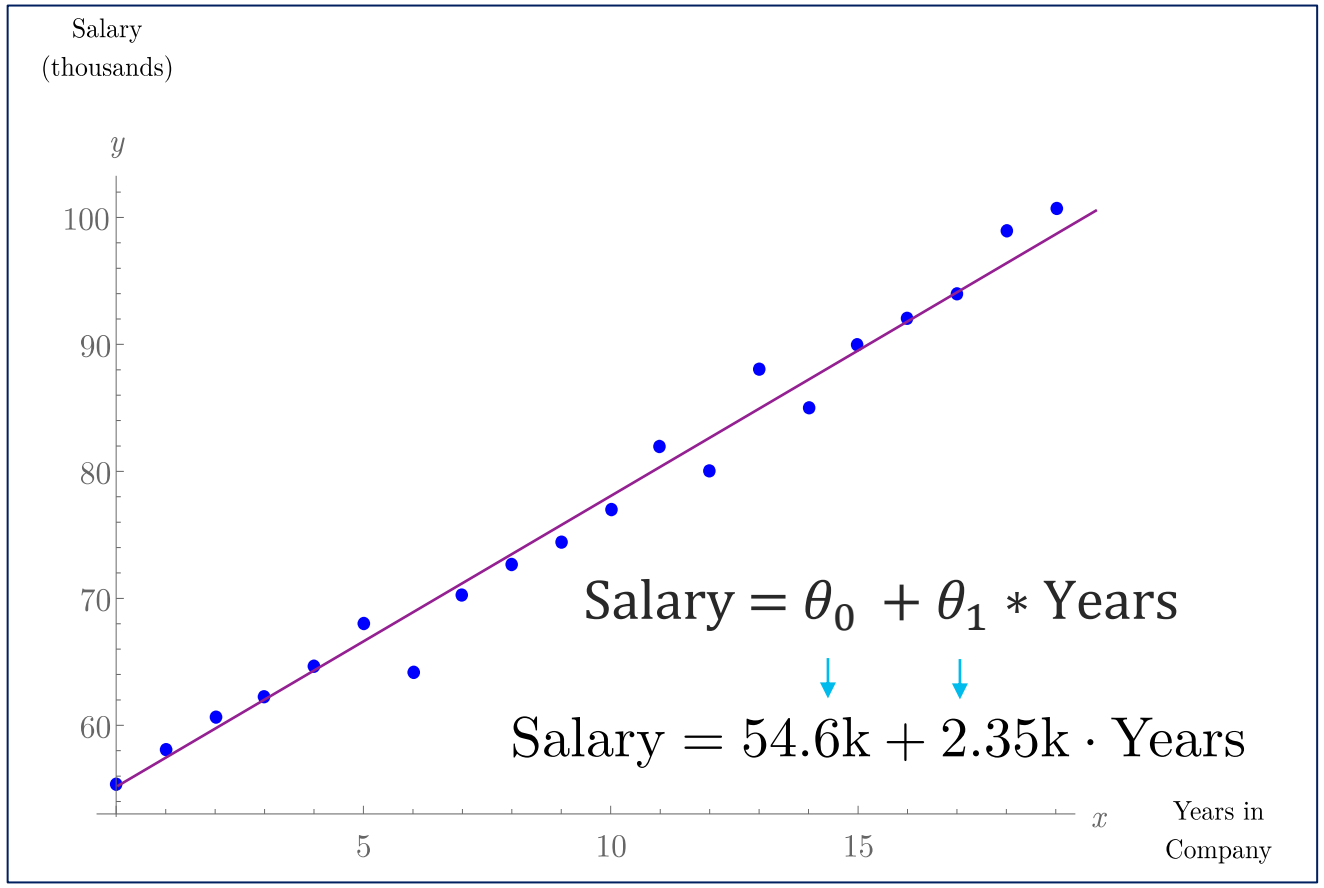
Gradient Descent
Getting Closer to the Answer



Gradient Descent
Getting Closer to the Answer



Now we can Predict a Future Event



$$\text{Salary} = f(\text{Years})$$

Now I can make some predictions:

"If a new person joined my team and they have been with the company 9 years, their salary will likely be:"

$54.6k + $2.35k x 9 = $75.75k

Expanding to Many More Dimensions

And instead of 2 dimensions (x,y),

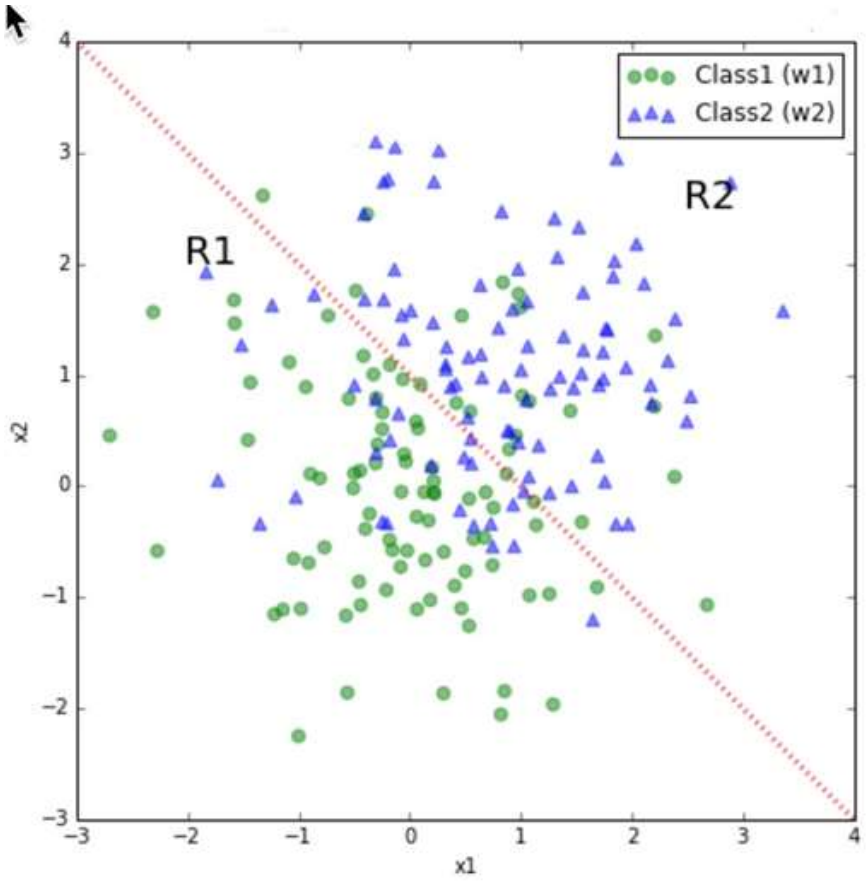
You might have 1 million dimensions!

(But the idea is the same)

Supervised Learning Part 2
Logistic Regression (Classification)

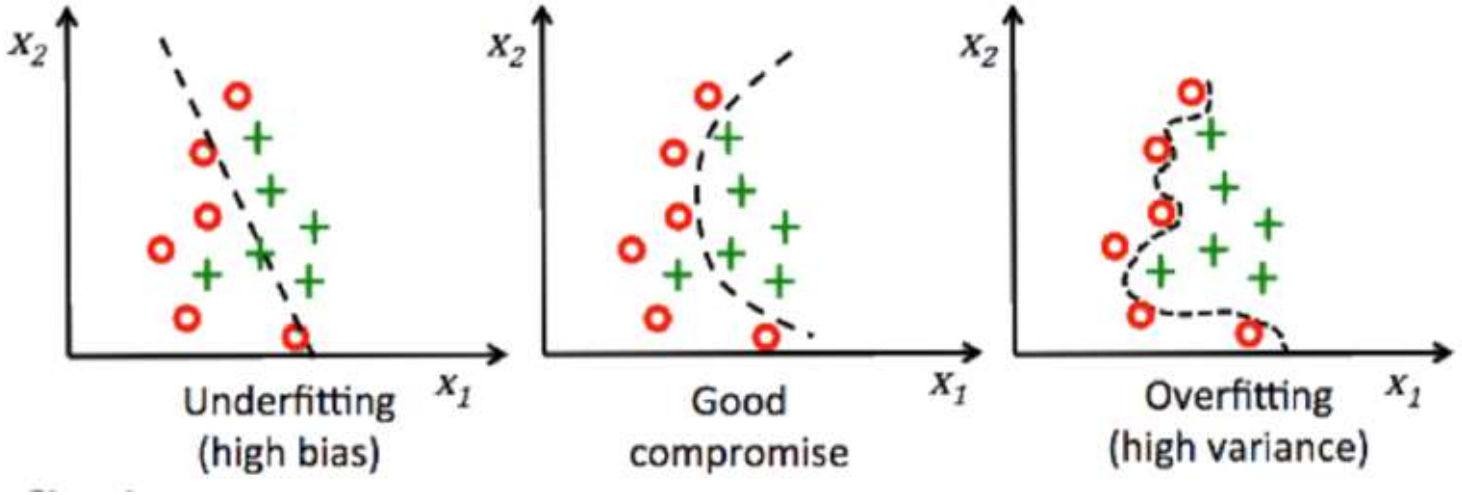
Supervised Learning: Classification

- With linear regression you try to find a number (y) that predicts an outcome.
- There is also classification, where your line separates two groups
- E.g. Is a credit card transaction fraudulent or is it safe?
- Also known as Logistic Regression

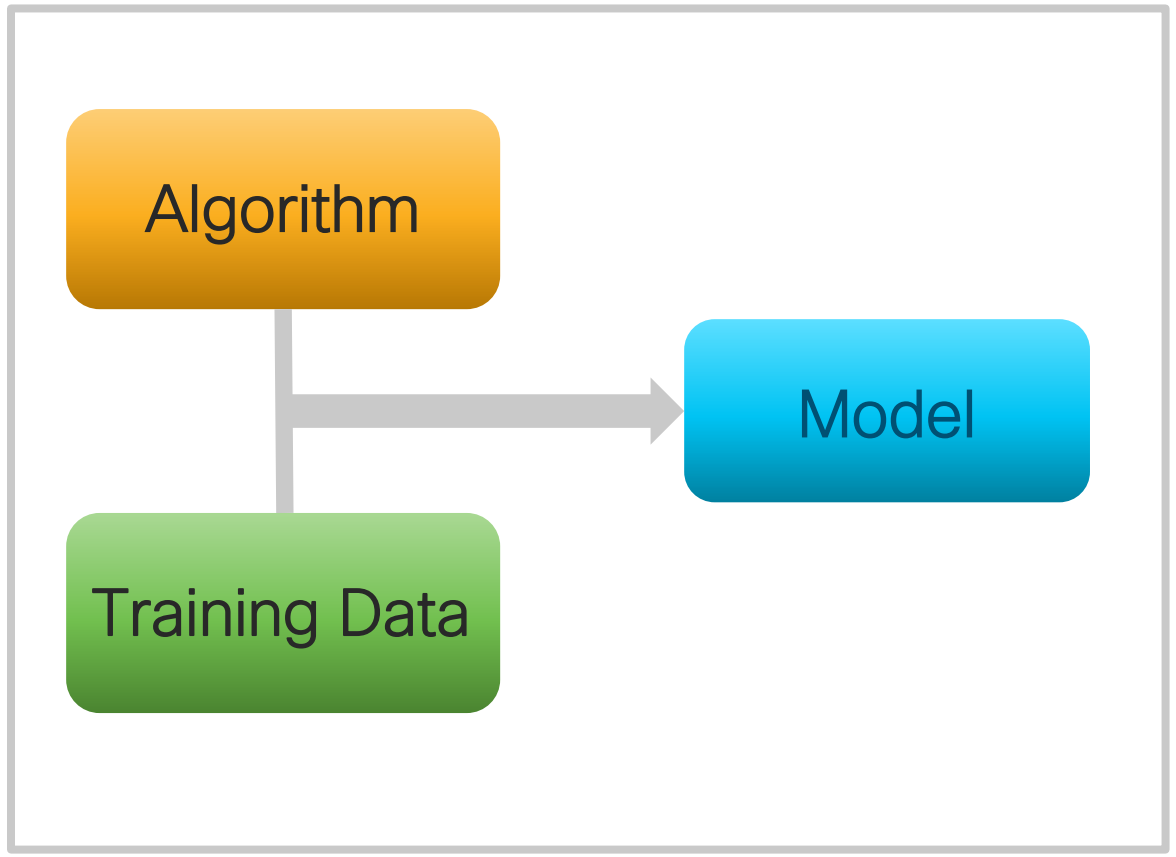


Fitting the Data

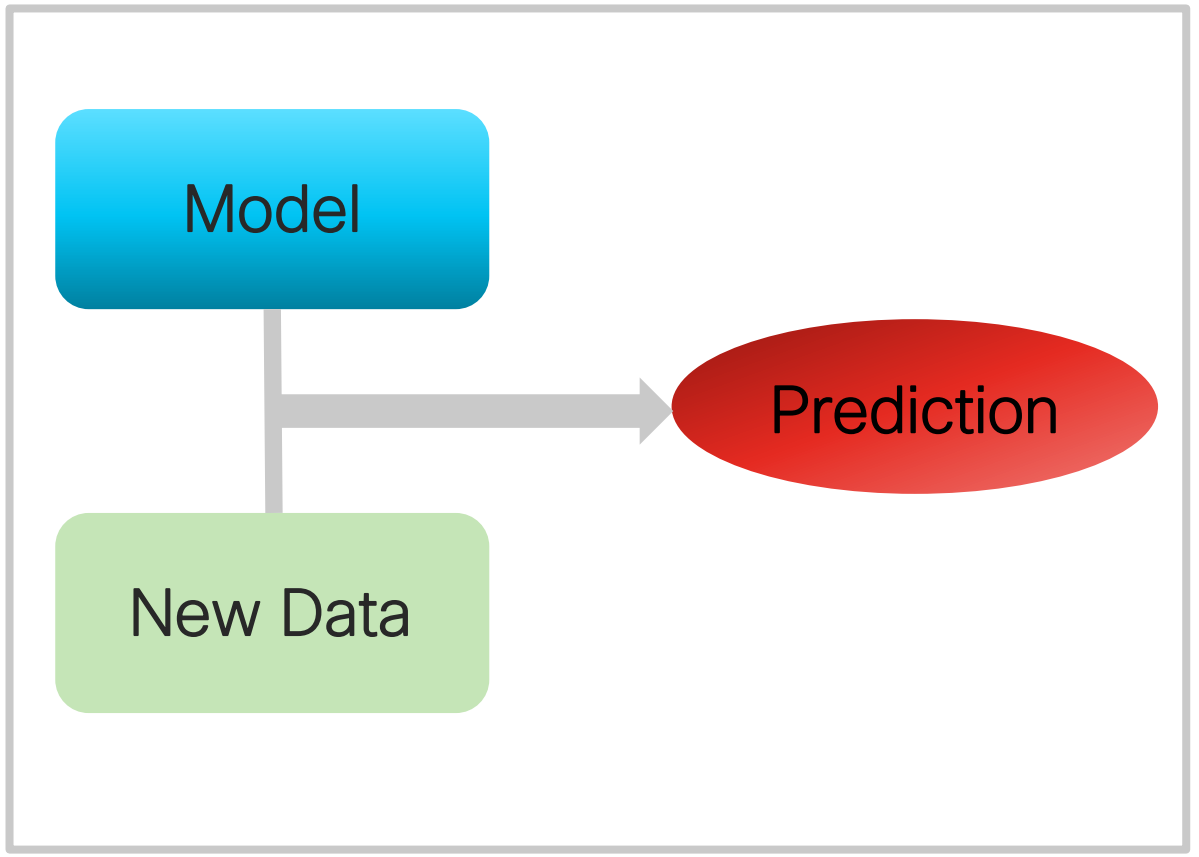
- The main challenge in Supervised Learning is to find the right equation… and figure out if the samples represent the full population



Regression Summary: Learn, then Predict

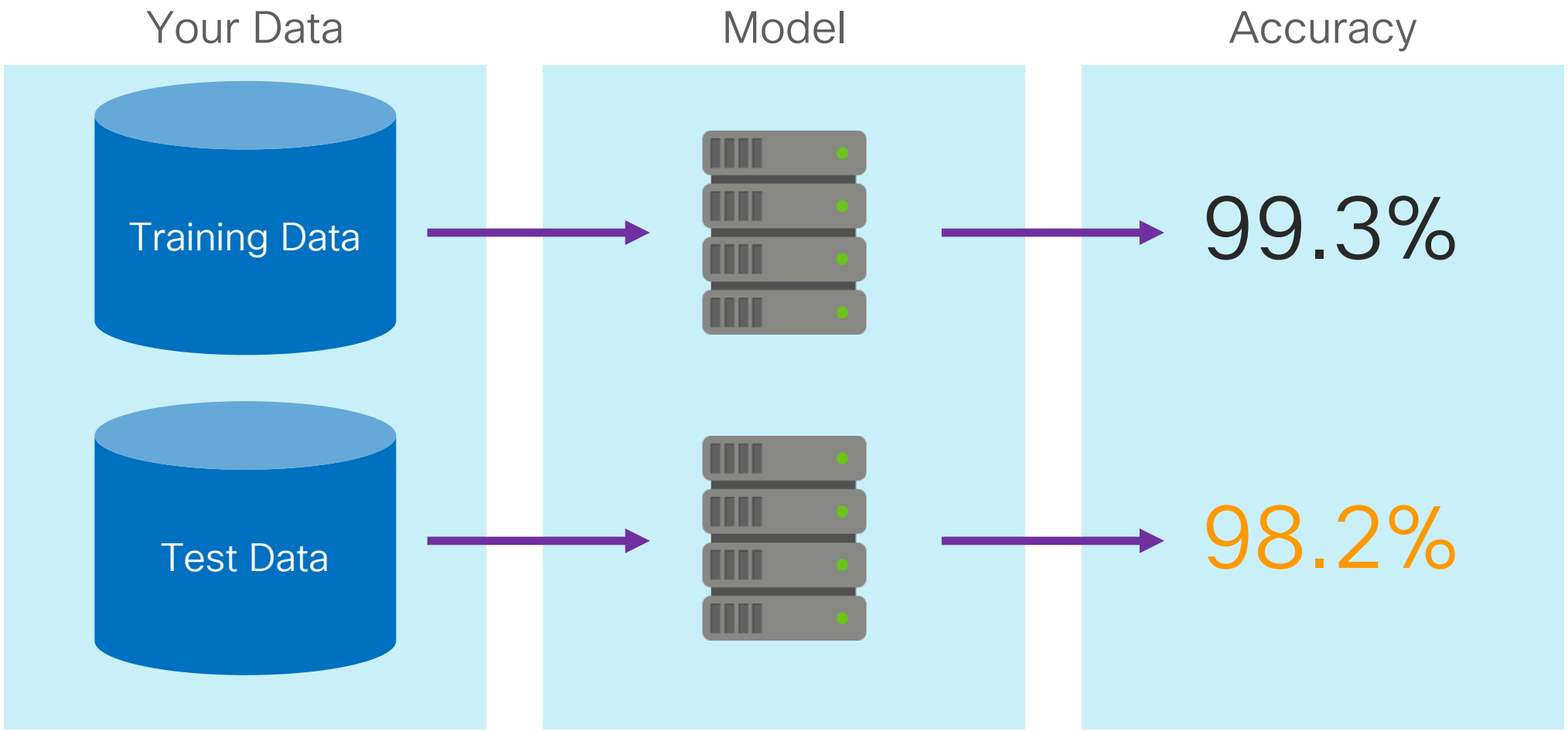


Learning (controlled experiments)

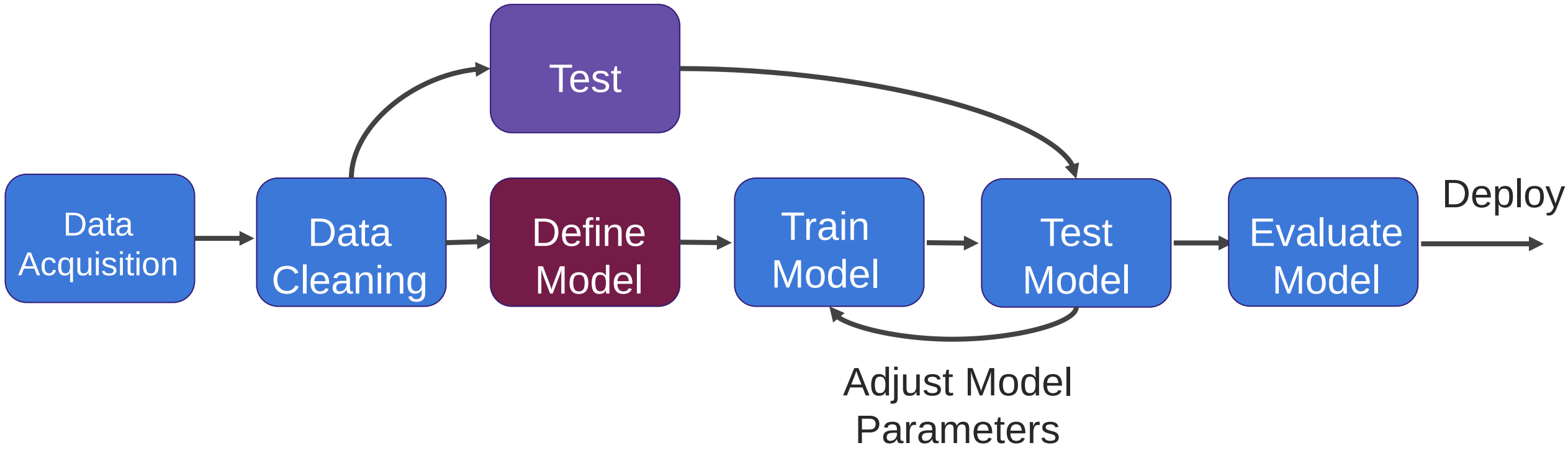


Predict (AI in the wild)

How Accurate is Your Model? Measure Your Accuracy!



A Typical Supervised Learning ML Work Flow



Unsupervised Learning

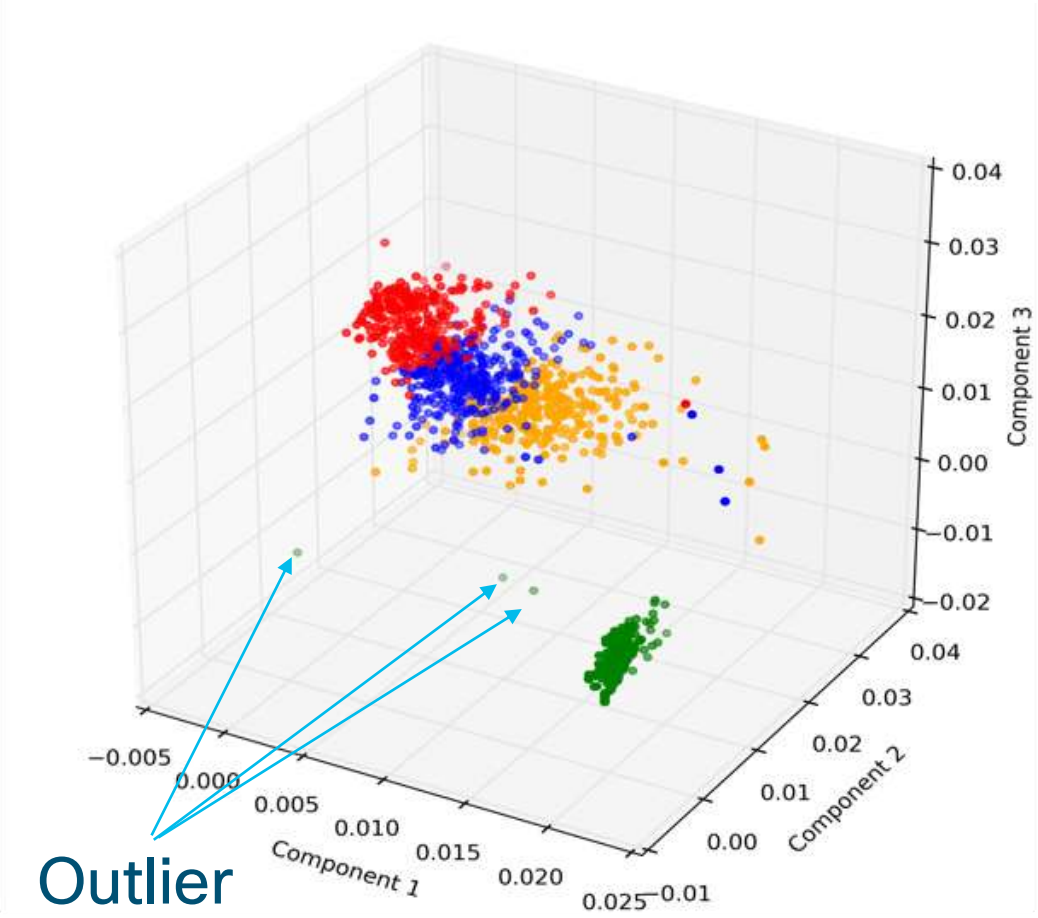
Unsupervised Learning

- You **do not** know the right answer, and there is too much data for you to guess
- Example: you manufacture small engines
- Some of them will fail
- You want to spot the failures before they get installed on mowers, chainsaws, etc.
- **How do you do that?**

Unsupervised Learning

- Your engine will group engines that have similar characteristics.
  - In math, this is simply grouping points that are close to one another
- And will spot the outliers
  - Those are the engines likely to fail (different from the others)

4 groups = 4 engine categories



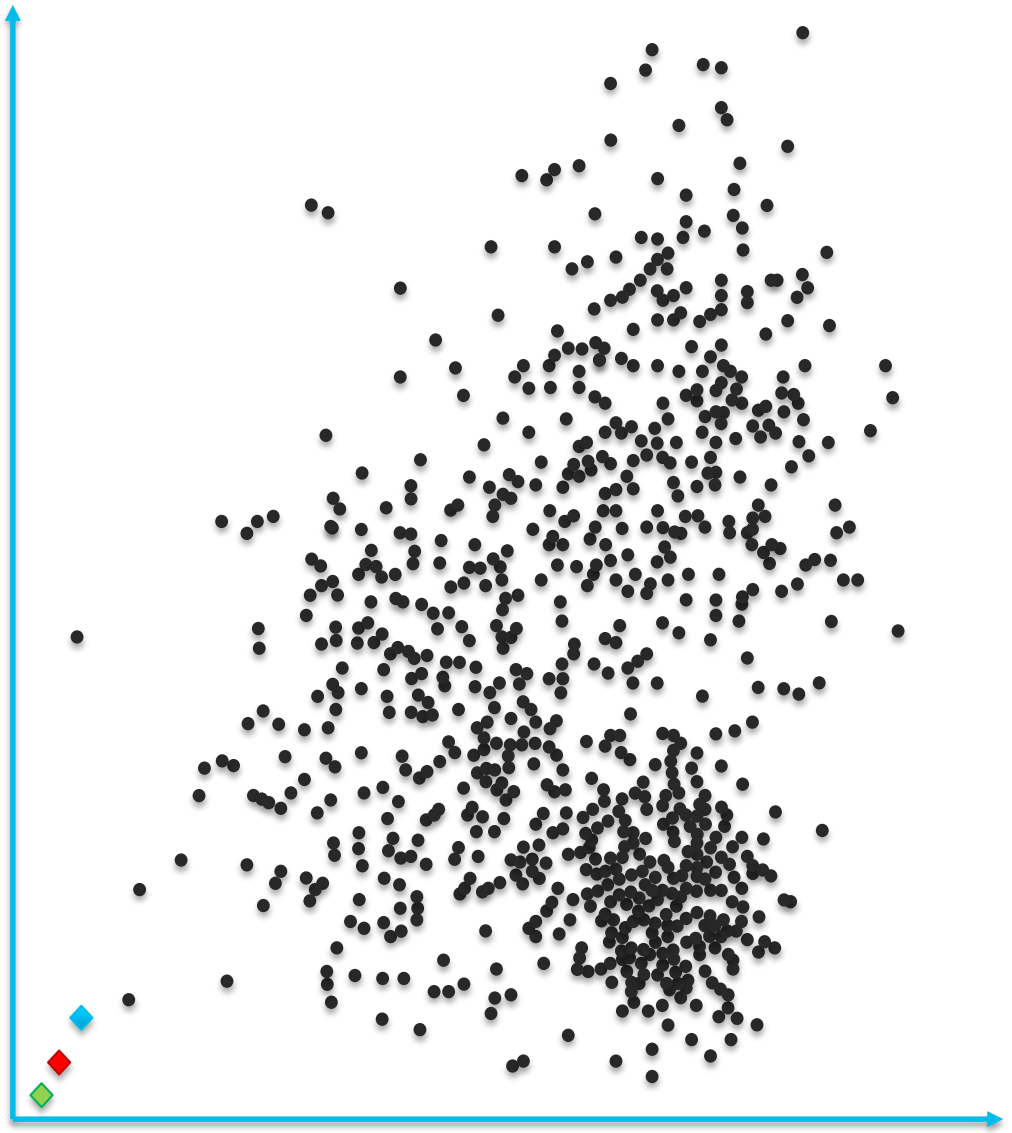
Unsupervised Learning

The math can take may forms, but a common form is K-means

Want 3 groups?

Take 3 random points (3 'centroids', $\mu_{c(i)}$)

Then take a known point, calculate which centroid is closest



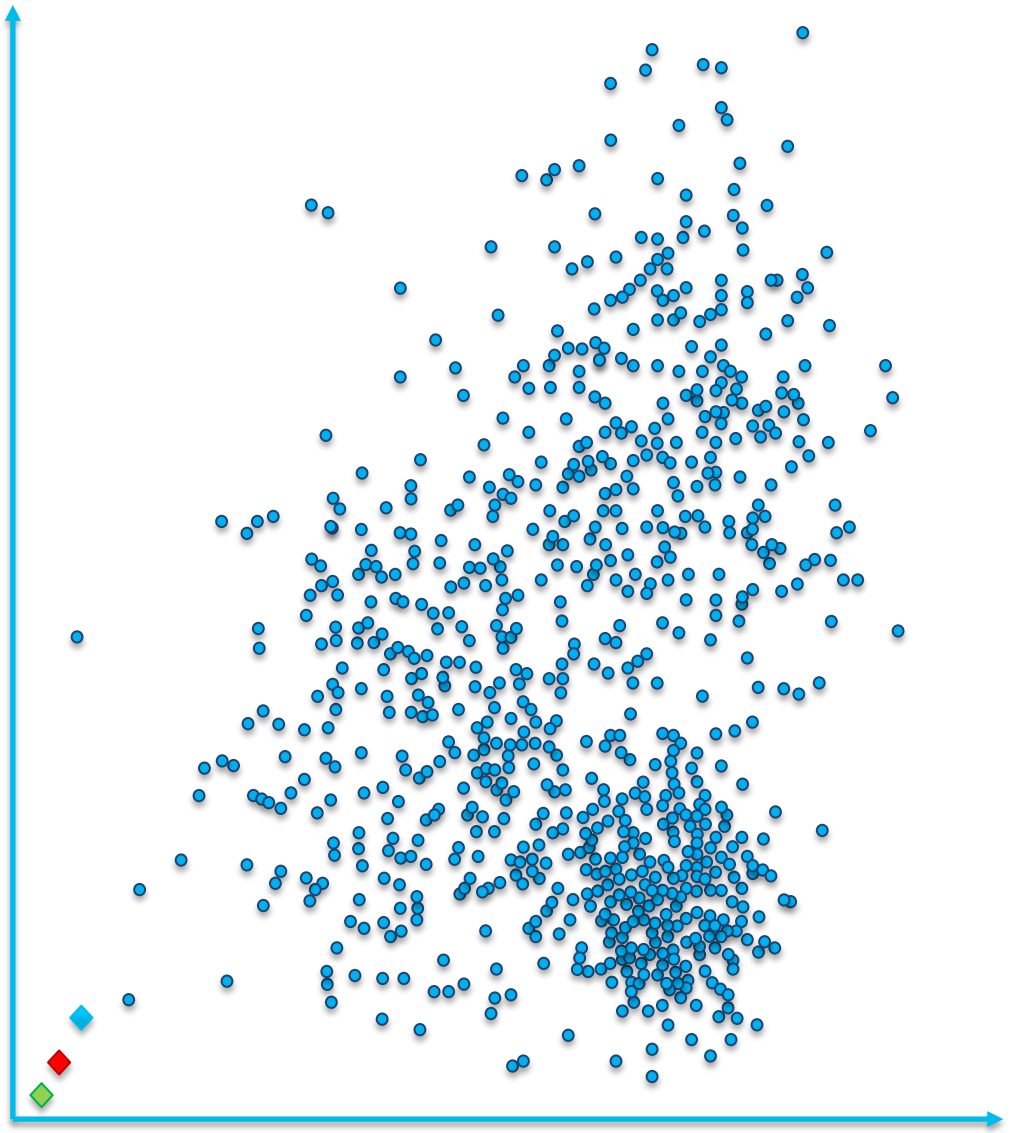
Unsupervised Learning

K-means

Want 3 groups?

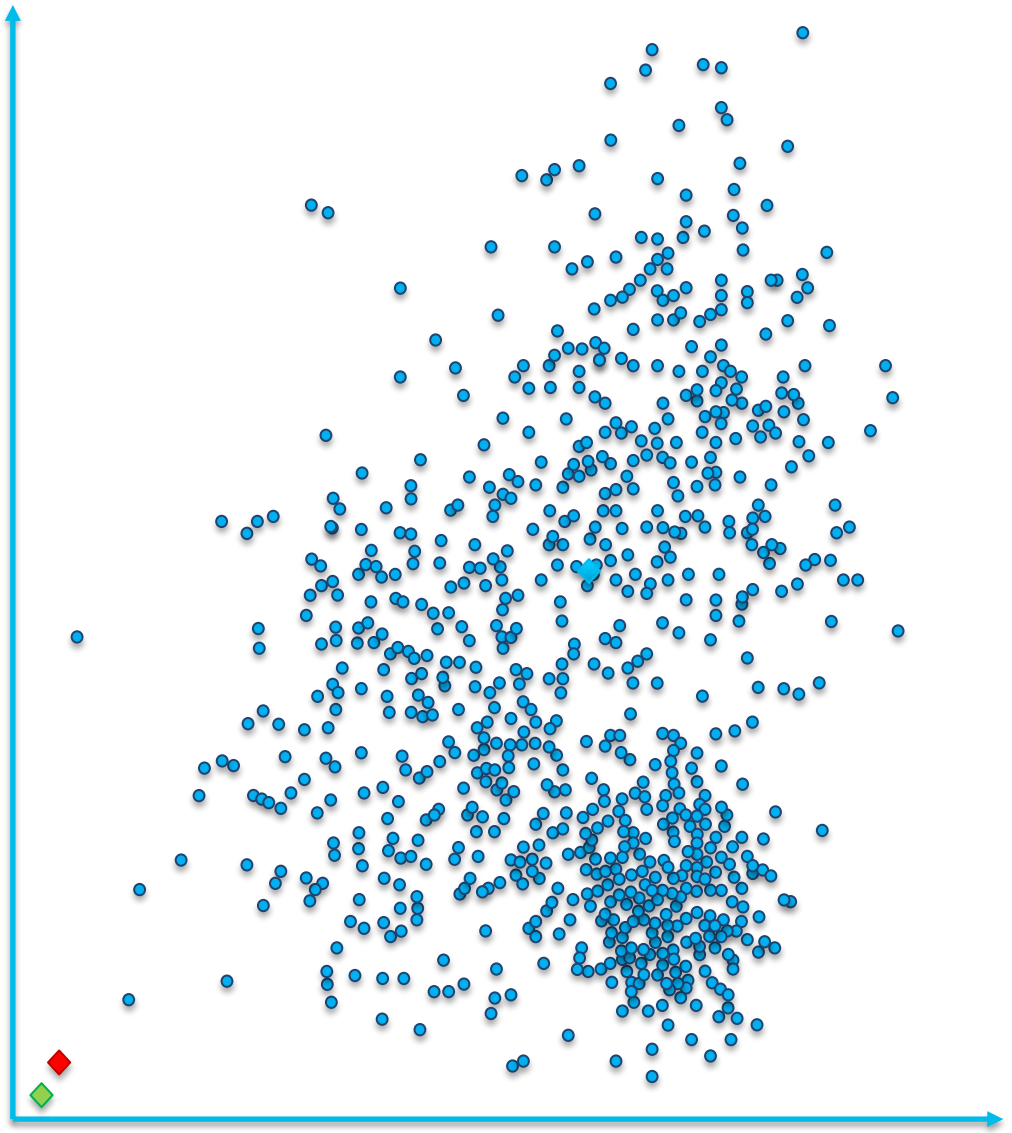
Repeat for all points. That's our usual distance equation:

$$\min_{c^{(i)}} \sum_{i=1}^{K} ||x^i - \mu_{c^{(i)}}||^2$$



Unsupervised Learning

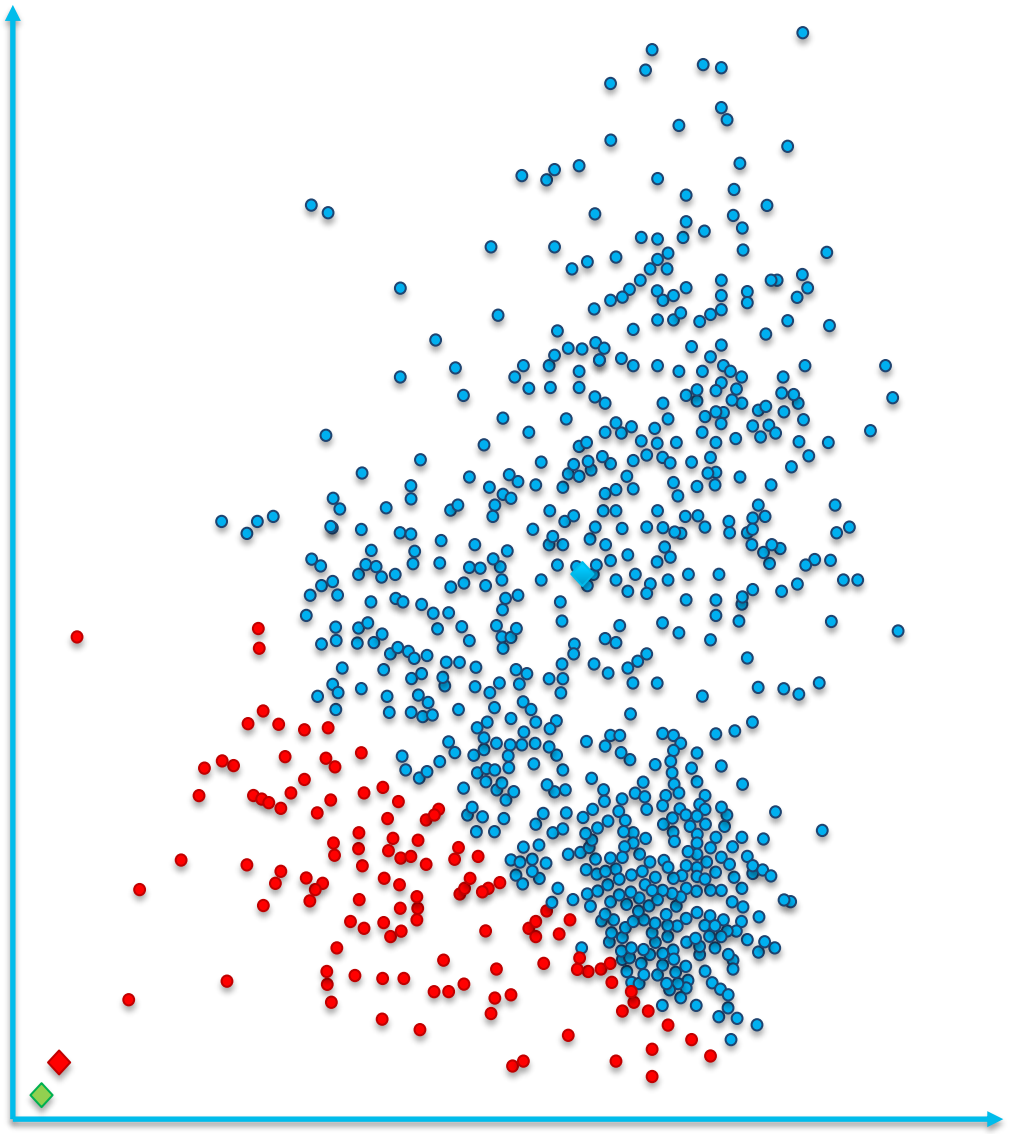
Then you move your ‘centroids’, $\mu_{c(i)}$ to the center (mean x,y) of each group you formed



Unsupervised Learning

And you repeat.

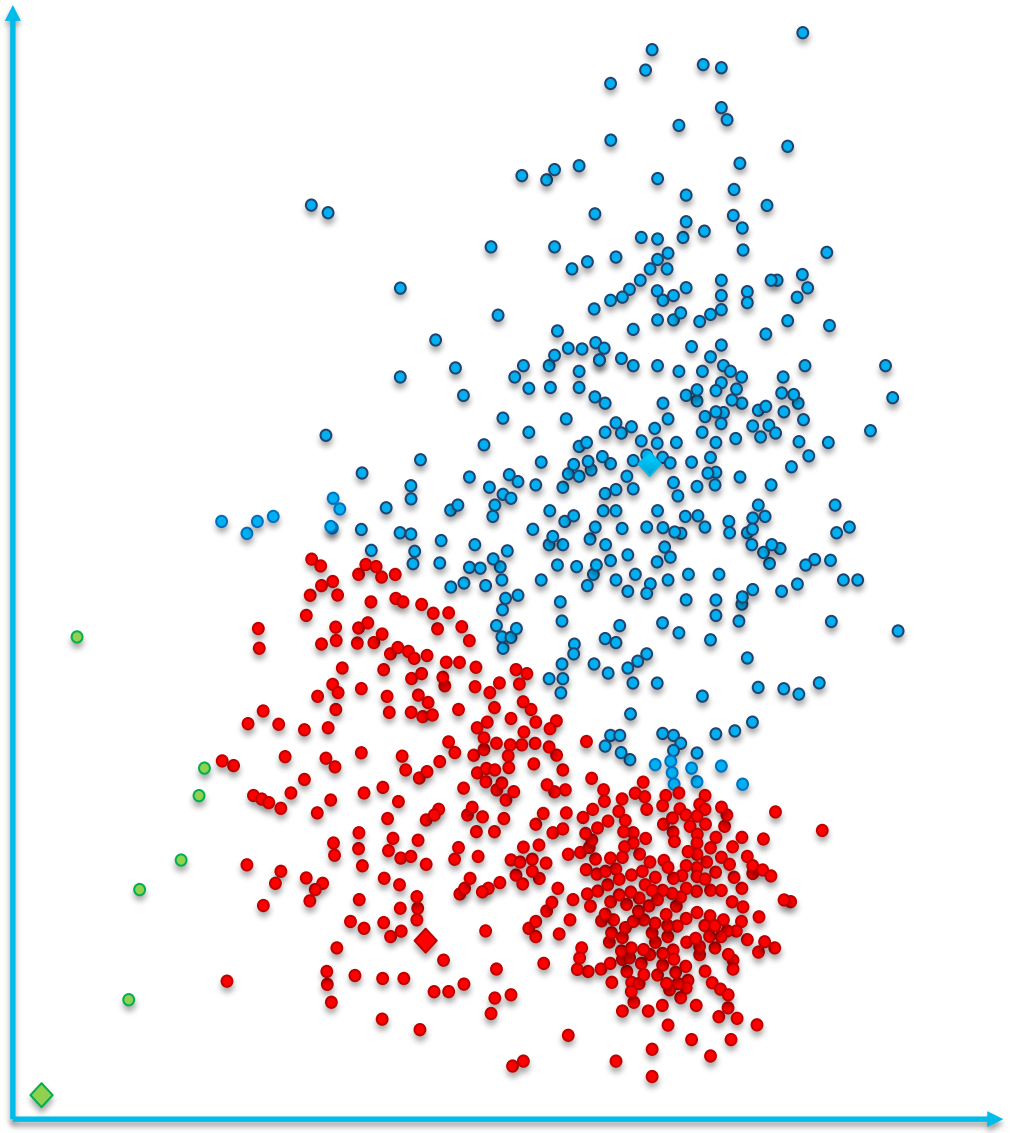
You are done when no point jumps to another group anymore



Unsupervised Learning

And you repeat.

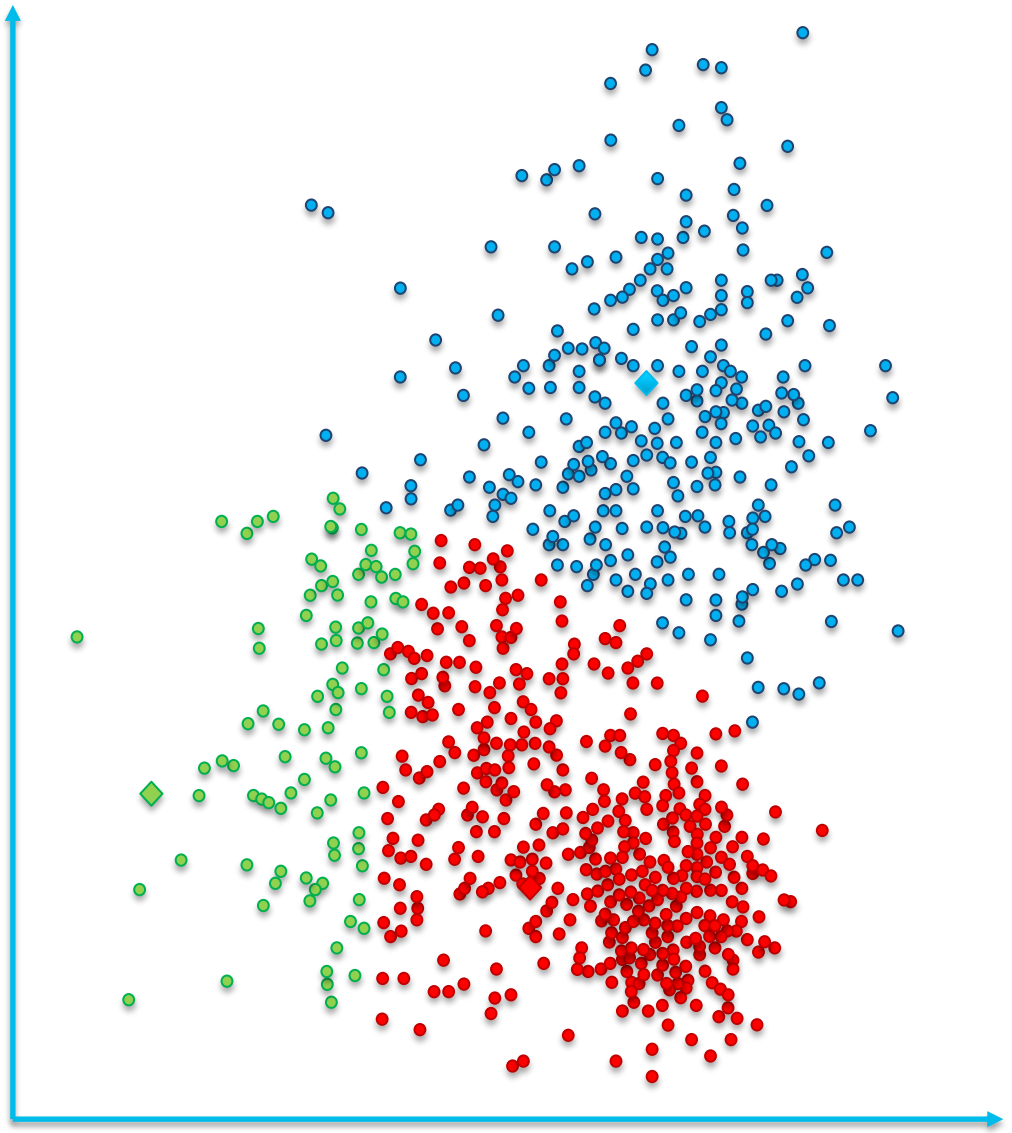
You are done when no point jumps to another group anymore



Unsupervised Learning

And you repeat.

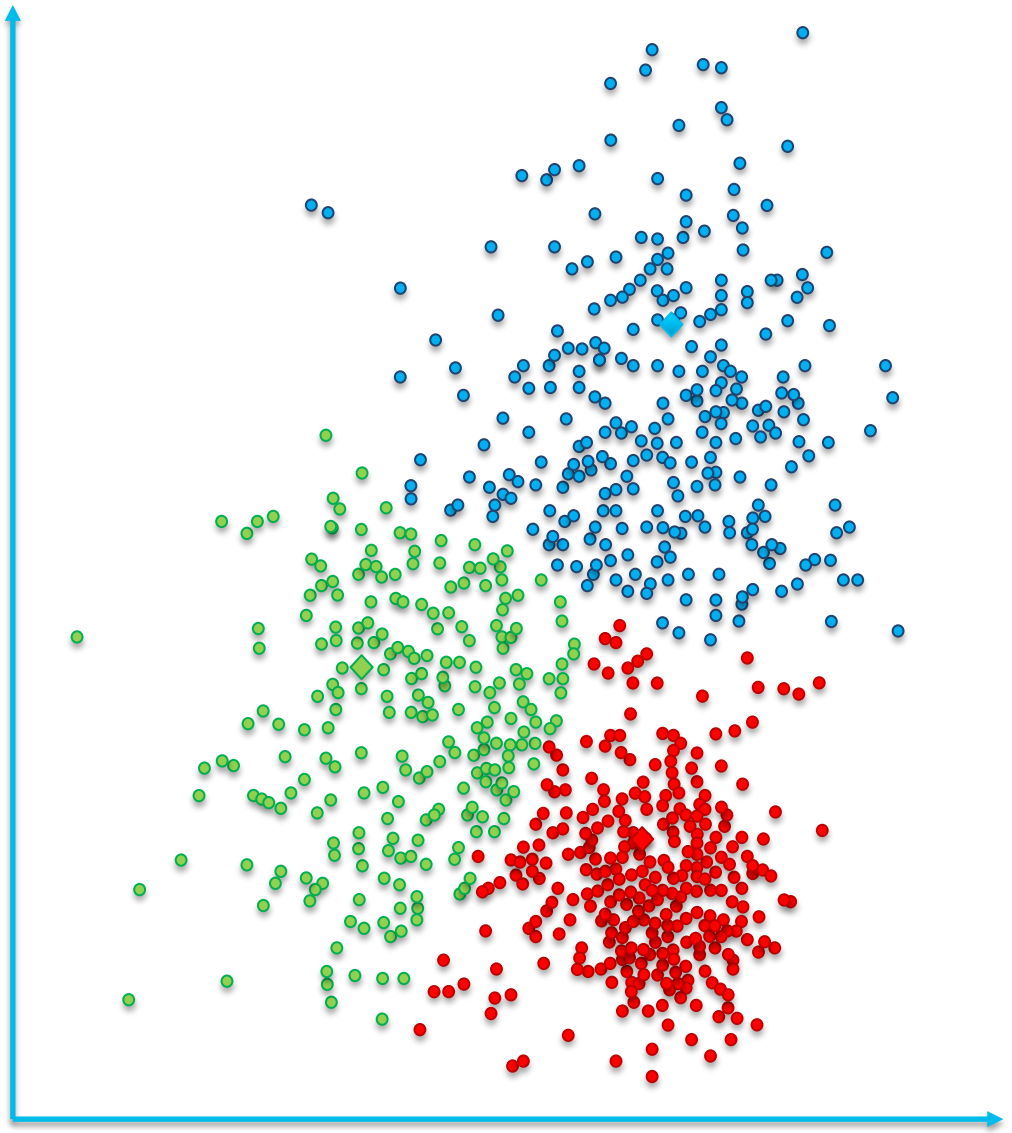
You are done when no point jumps to another group anymore



Unsupervised Learning

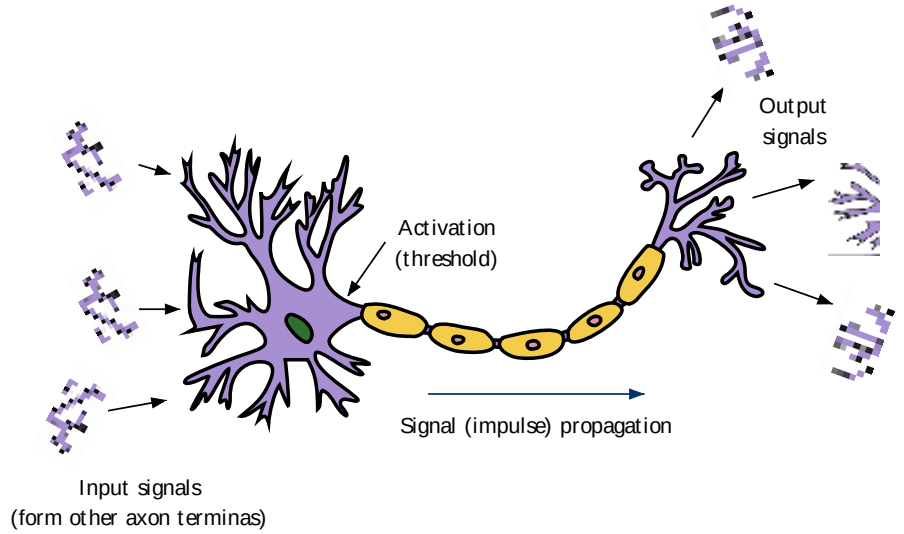
And you repeat.

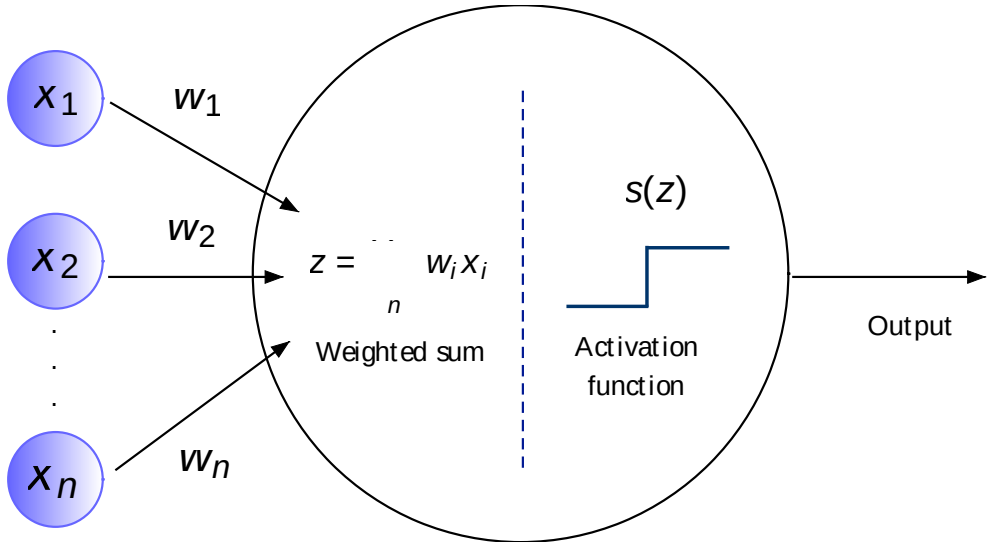
You are done when no point jumps to another group anymore



Neural Networks

Neural Networks: From Neurons to Perceptrons

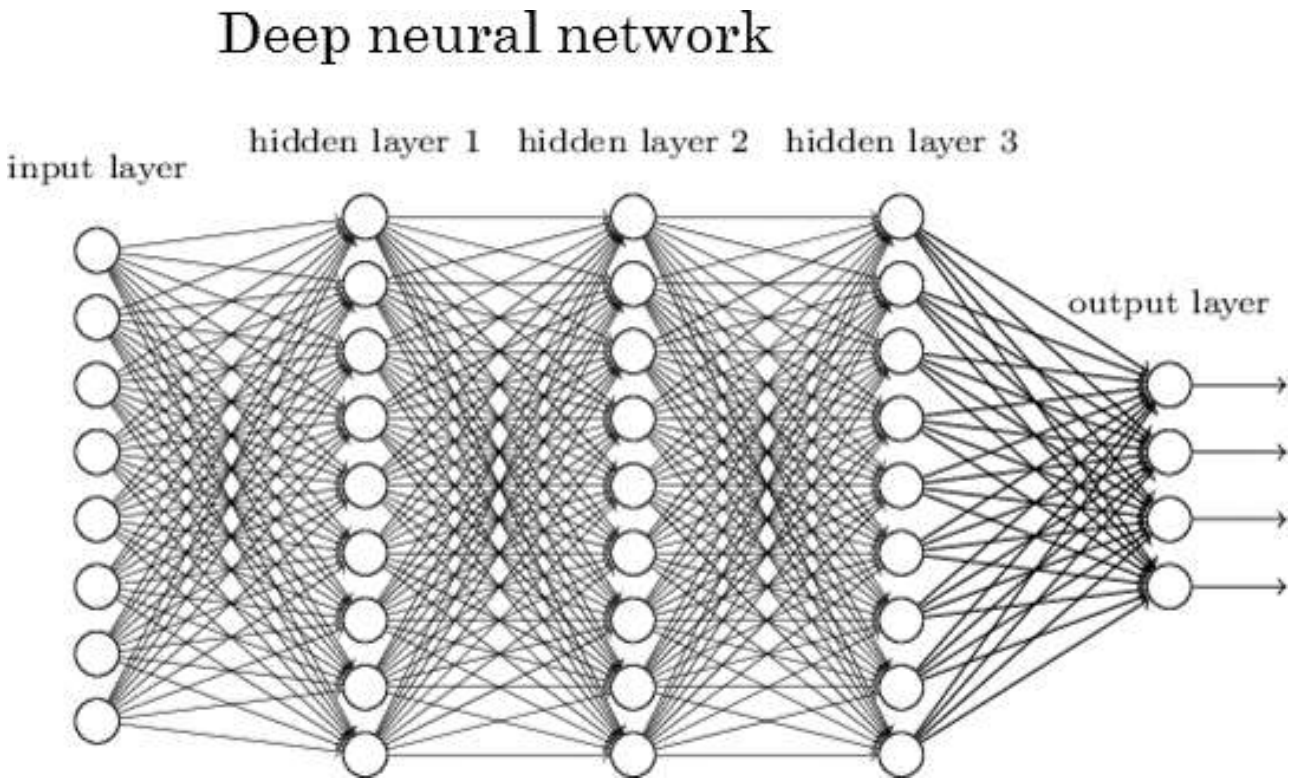




1. Inputs arrive from other neurons / nodes.
2. Receiving neuron / node processes the inputs → weighted sum.
3. Receiving neuron / node applies an activation function (sigmoid).
4. Output of activation function passed to subsequent neurons / nodes.

Neural Networks

- Then you put tons of units, possibly in multiple layers (in this case, it is called deep learning)
- Also called Artificial Neural Networks (ANNs)
- Convolutional Neural Networks (CNNs) are a common sub-class of ANNs

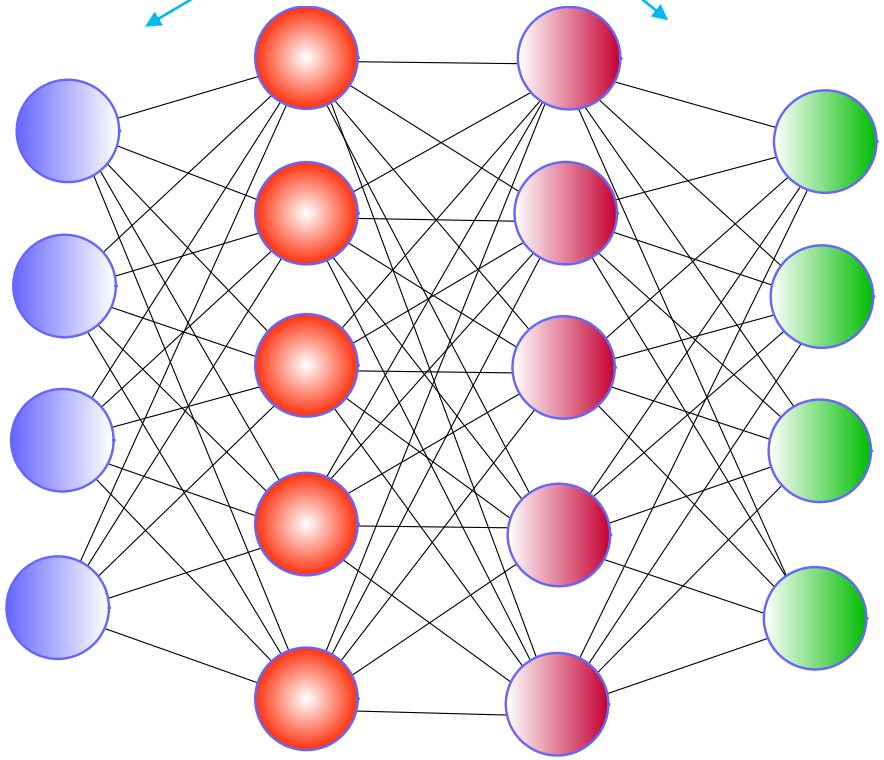


Neural Networks: Image Recognition

Backpropagation (just 10 years ago by Geoffry Hinton)

Adjust weights

Training Images



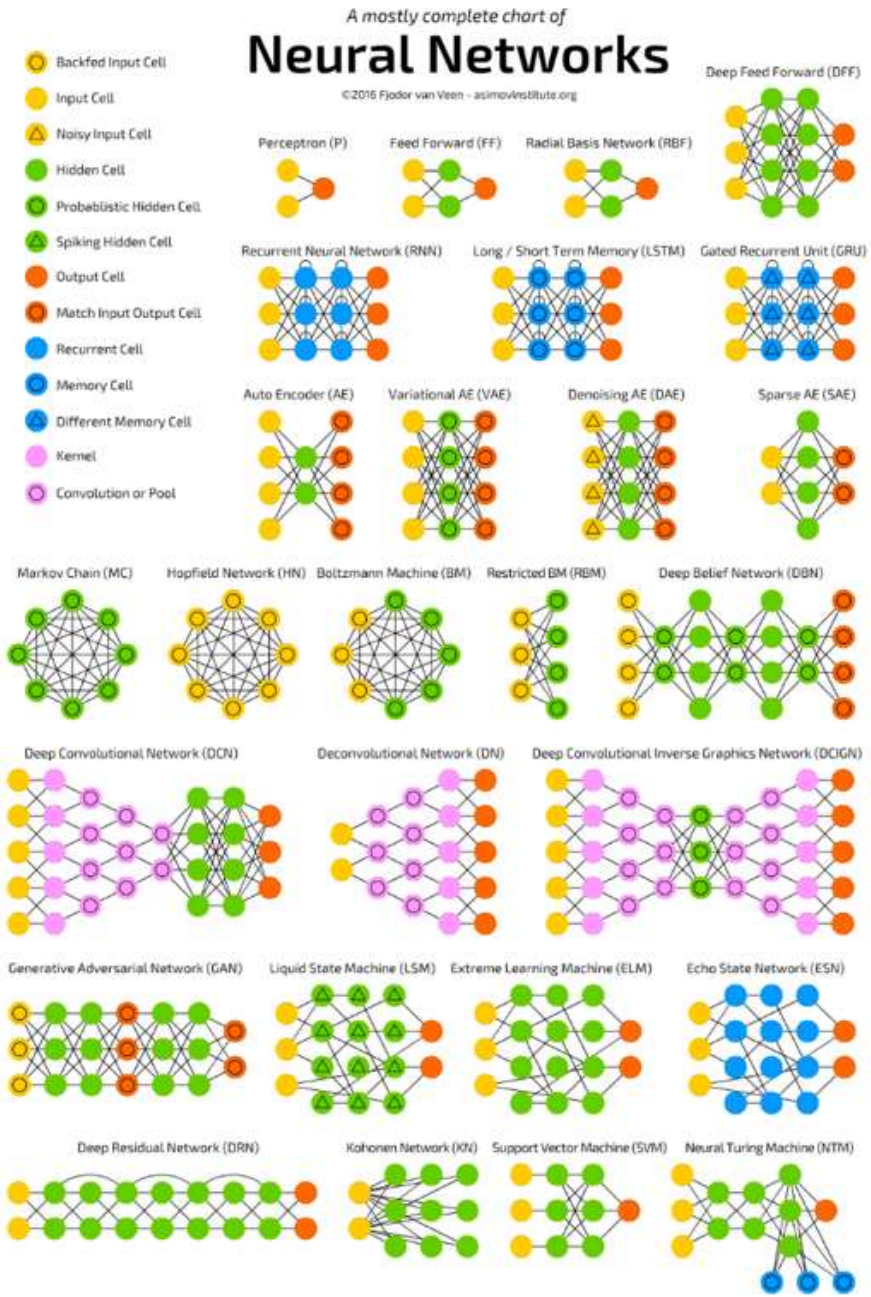
Input Layer Hidden Layer Hidden Layer Output Layer

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

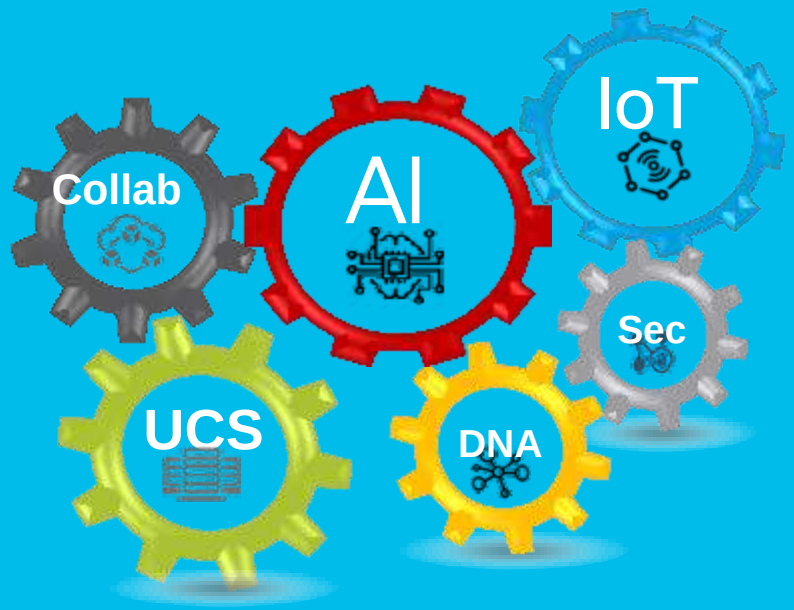
Eventually, you train model to recognize “car”

Types of Neural Networks

- The way you connect the units can vary immensely
- And this is what makes this family very rich
- Tons of possible applications depending on what data you are looking at, and what you try to find



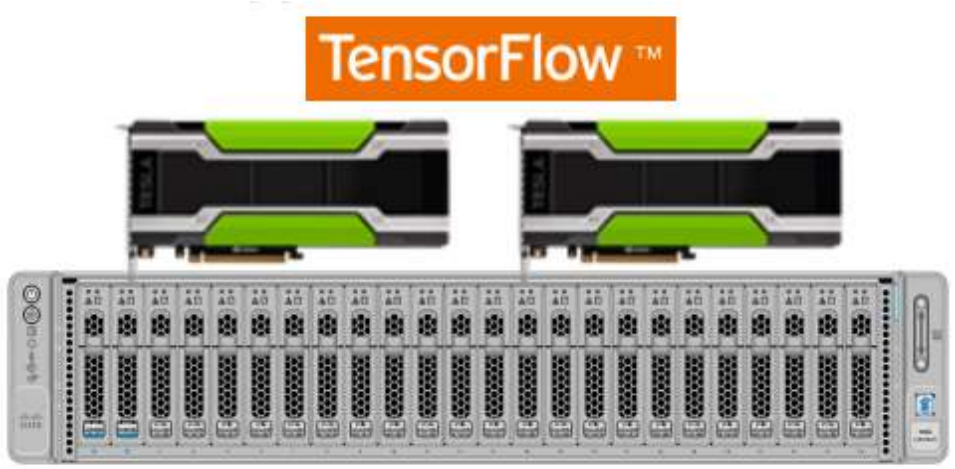
Part 2:
The AI/ML Landscape at Cisco



How Cisco Approaches AI/ML

Consumption

Products use AI/ML to do things better



Enablement

Infrastructure Supporting AI/ML workloads

AI/ML By Product Category

Reinvent the Network

- DNA
- ETA
- Network Early Warning
- SD-WAN

Security is Foundational

- AMP
- Cloudlock
- Cognitive Threat Analytics
- NGFW
- Stealthwatch
- Talos
- Umbrella

Power a Multicloud World

- Hyperflex
- Intersight
- UCS

Unlock the Power of Data

- AppDynamics
- Kinetic
- Tetration

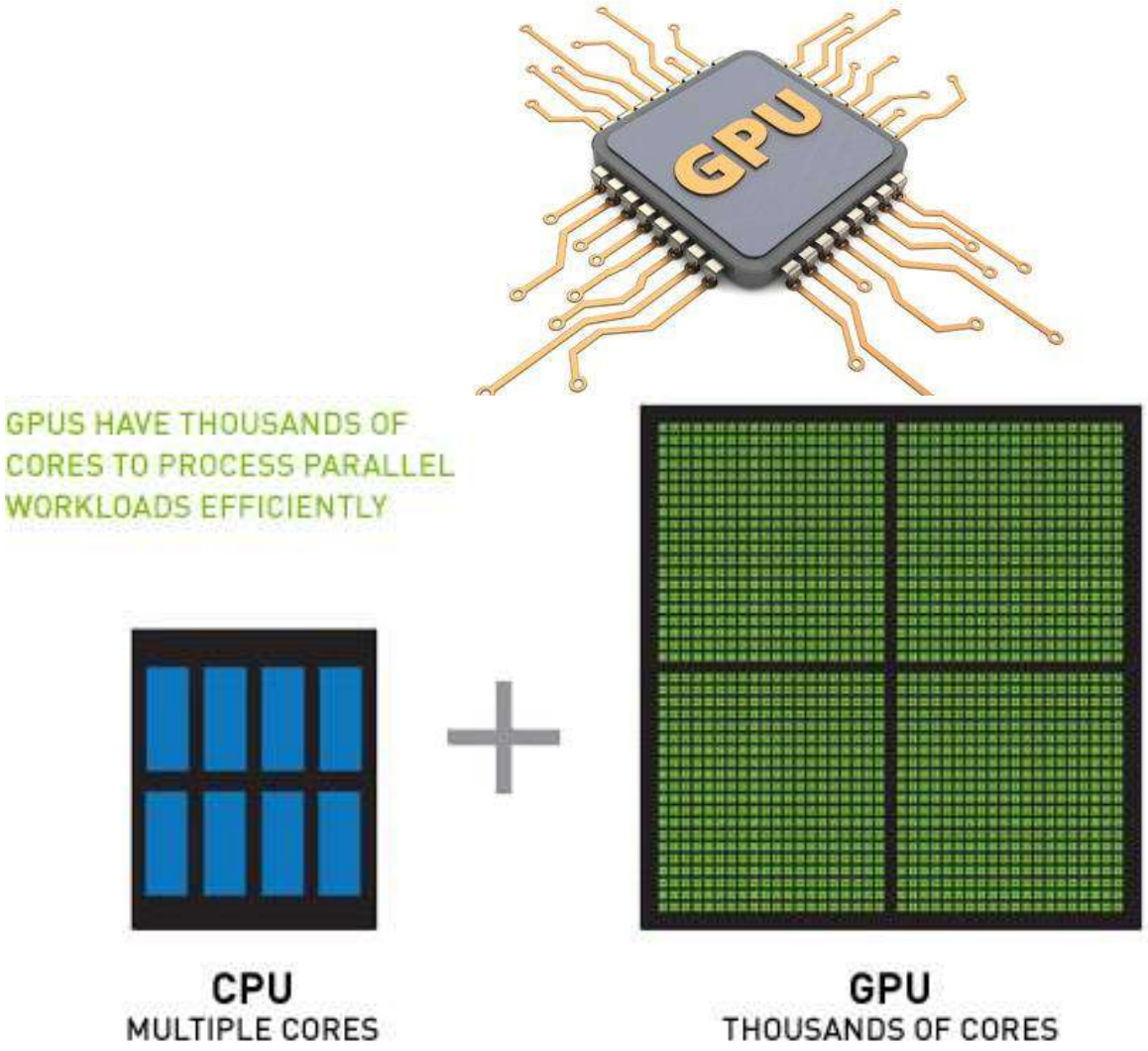
Create Meaningful Experiences

- Accompany
- CAM
- Meraki
- MindMeld
- Talent Trends
- Webex

AI/ML At Cisco:
The Data Center

The Power of GPUs for Deep Learning

- Graphical Processing Units are specialized types of electronic circuitry designed to rapidly manipulate memory for graphics
- GPUs support parallel processing, accelerating their ability to execute algorithms that require parallel processes
- GPUs are at the heart of deep learning and neural networks



Comparing CPUs and GPUs

- CPUs are capable of almost any task – but at a price

CPUs are like a swiss army knife

- GPUs are highly-specialized processors used to solve complex math problems

GPUs are like specialized surgical instruments

Cisco AI/ML – Compute Portfolio

Testing and development, and model training

Cisco UCS C240

2 x P100/ V100

Cisco HyperFlex 240

2 x P100/V100 per node

Option of GPU-only nodes

Deep learning/training

Cisco UCS C480

6 x PCIe P100/ V100

Cisco UCS C480 ML

8x V100 with NVLink

Inferencing

C/HX 220
C/HX 240

2 x P4
6 x P4

Unified management

CISCO INTERSIGHT

CISCO UCS Manager

Cisco IMC

XML API

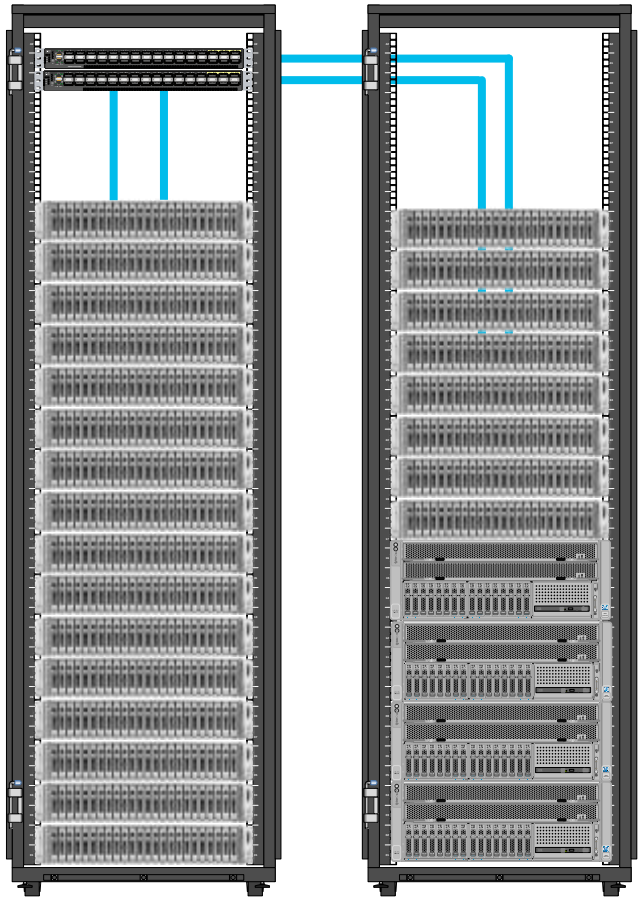
python SDK

Simplified management, customer choice, Cisco Validated Design

Tying it Together: Big Data with Machine Learning

Cisco Validated Hadoop Design with Cloudera on GPU-Powered AI/ML Workloads

2 x UCS 6332
Fabric Interconnect



16 x UCS C240 M5
datanode

C240 M5
datanodes

Cisco UCS® C480 ML M5 each with
Compute:
8x V100 Nvidia Tesla GPU
2x6132 CPU

Nvidia CUDA containers
orchestrated with YARN for Deep
Learning

Popular ML/DL frameworks

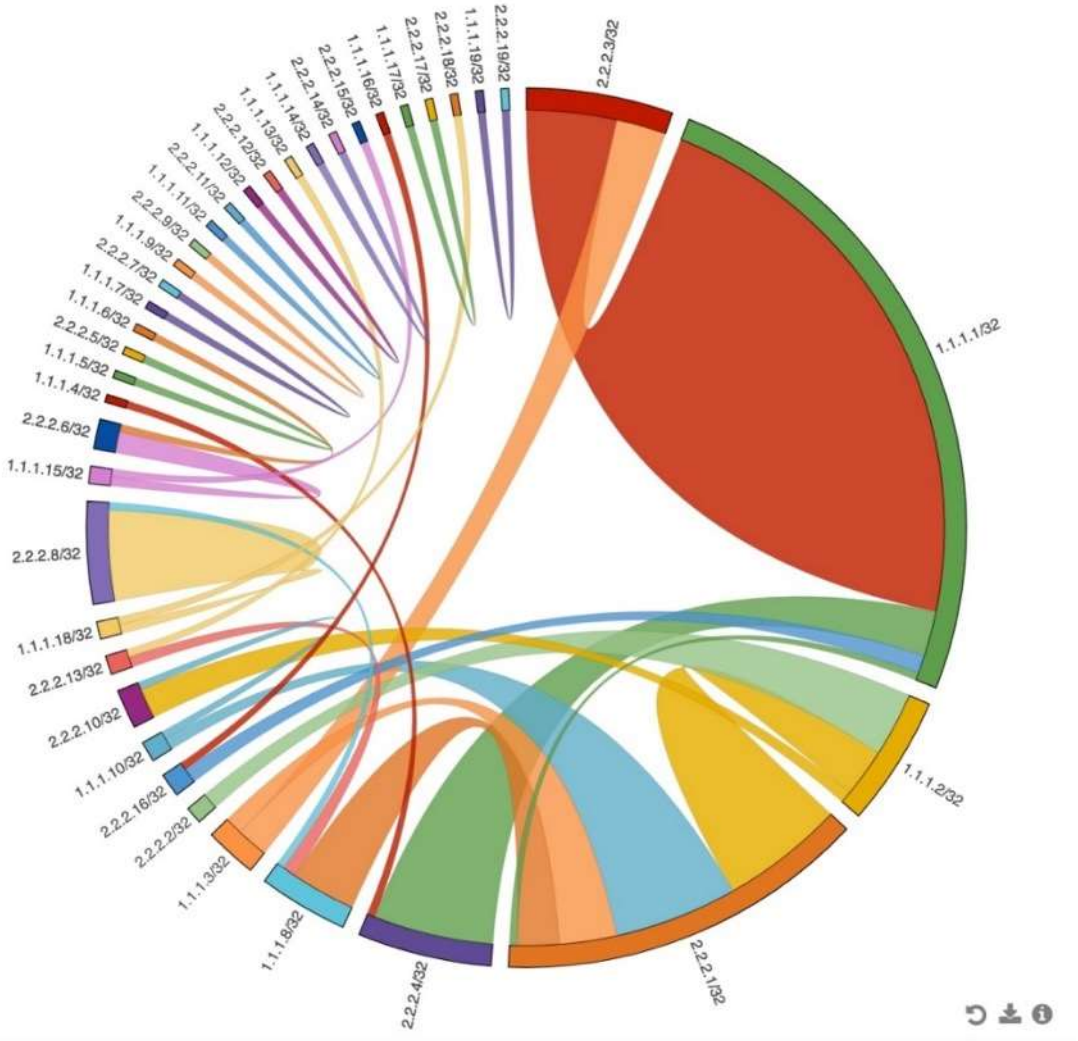
TensorFlow · Keras · PyTorch · theano · mxnet

cloudera

Cisco Cloudera CVD:
https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_Integrated_Infrastructure_for_Big_Data_with_Cloudera_28node.html

iCAM

Intelligent Comprehensive Analytics and Machine Learning on Nexus switches



iCAM Overview:

- Analytics & Telemetry, natively on the switch/router
- Security access control analytics
- Internal hardware tables usage analytics
- Top/bottom heavy hitters
- Anomaly visualization
- Build apps on top of iCAM
- Historical Analytics
- Predictive Analysis
- Streaming telemetry

Benefits:

- **Order of magnitude OPEX savings** : reduction in configuration, and ease of deployment.
- **Order of magnitude CAPEX savings** : Natively on the switch/router: Wiring, Power, Rackspace and Cost savings
- **Scalability** : Multi-Terabits/s
- Compute & Storage for Analytics, Historical Data

AI/ML At Cisco:
Cognitive Collaboration

Webex Endpoints Built on Powerful AI

Webex Board 70

Webex Board 55

Room 70D

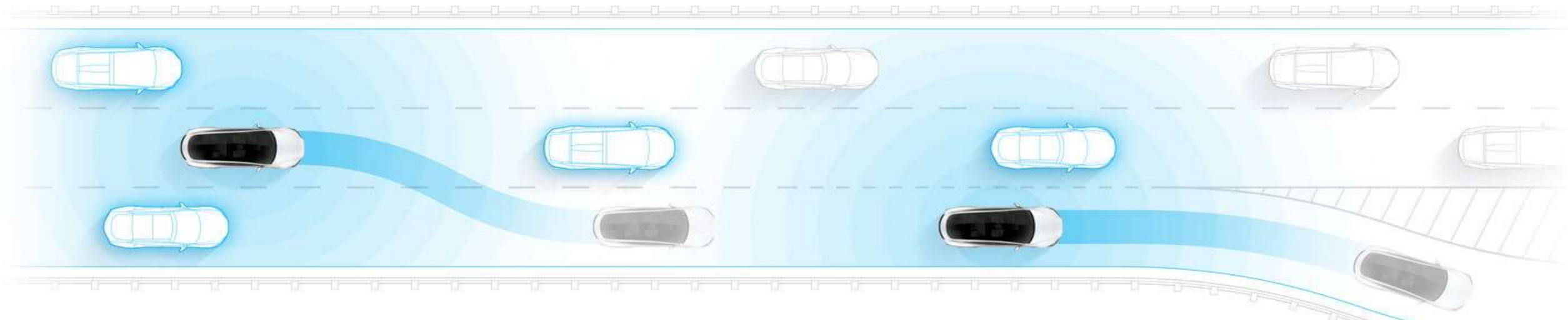
Room 70S

Room 55

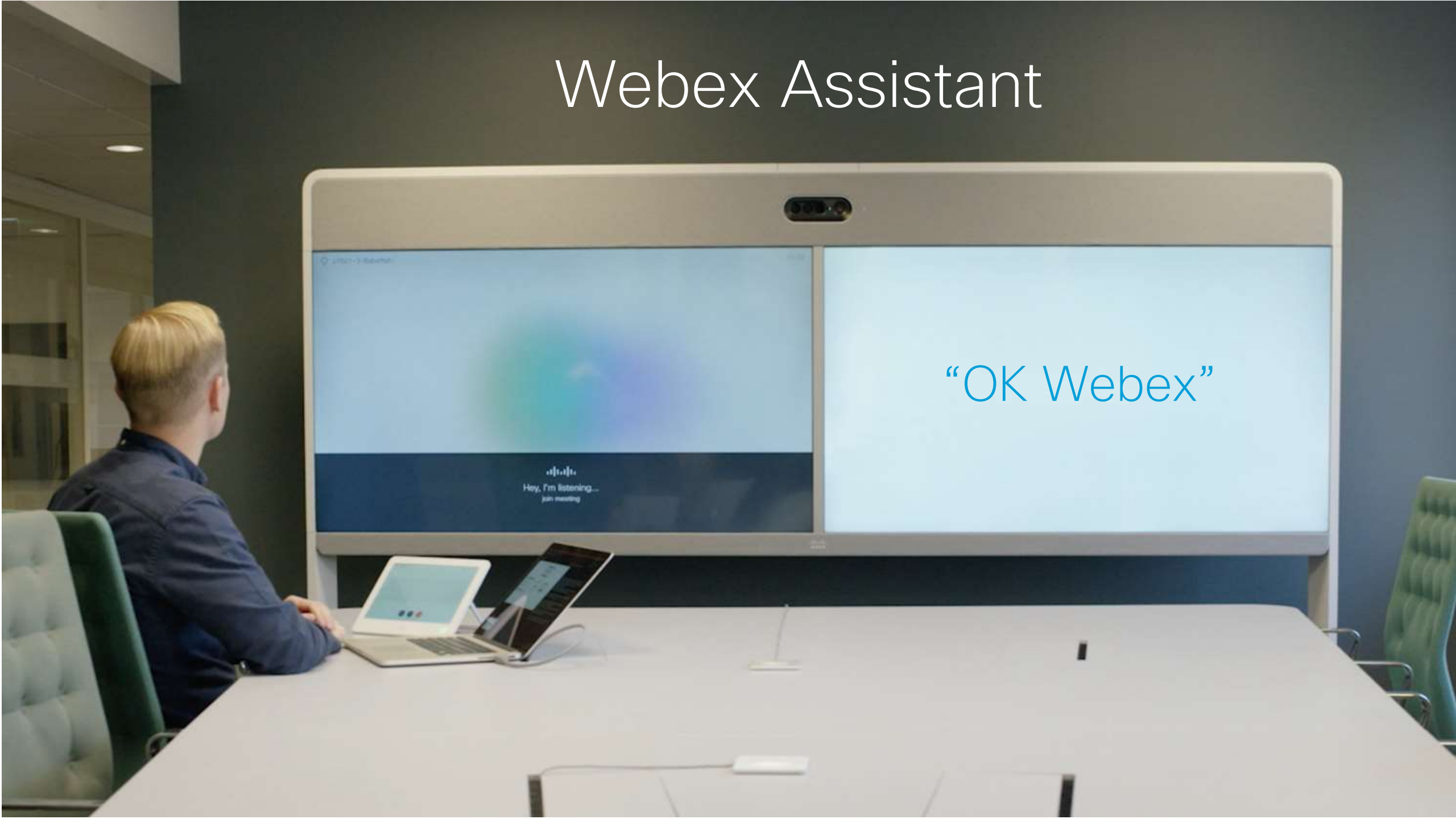
Room Kit Plus

Room Kit

NVIDIA Jetson Platform – The same electronics engine powering self-driving cars

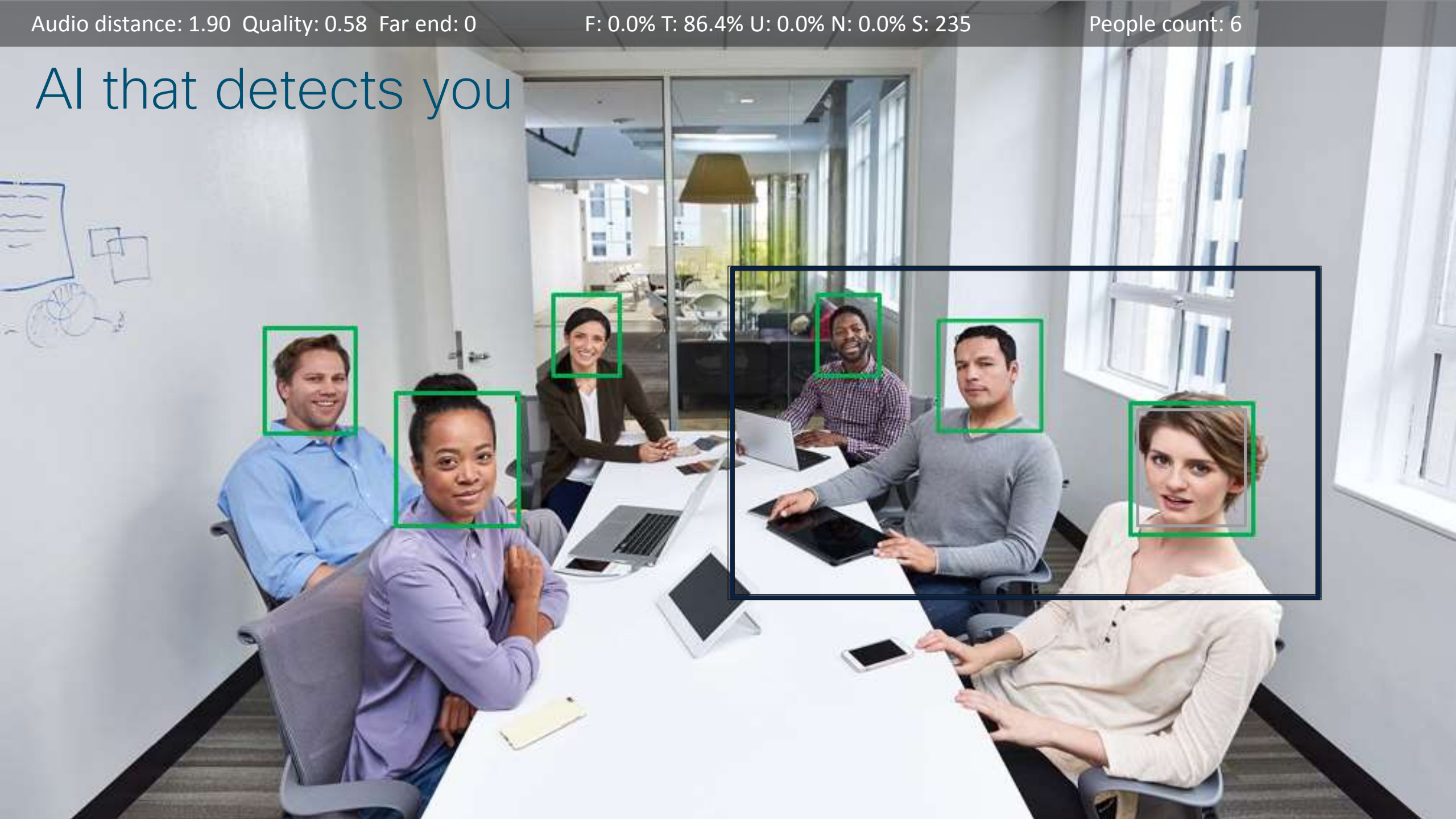


Webex Assistant



Audio distance: 1.90 Quality: 0.58 Far end: 0 F: 0.0% T: 86.4% U: 0.0% N: 0.0% S: 235 People count: 6

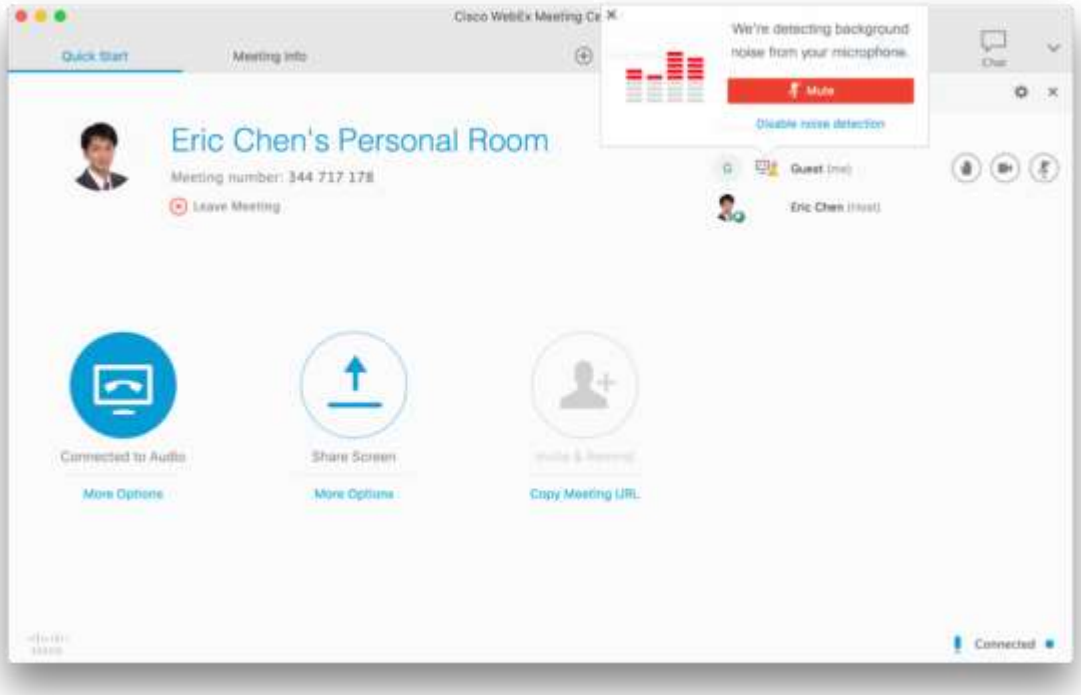
AI that detects you



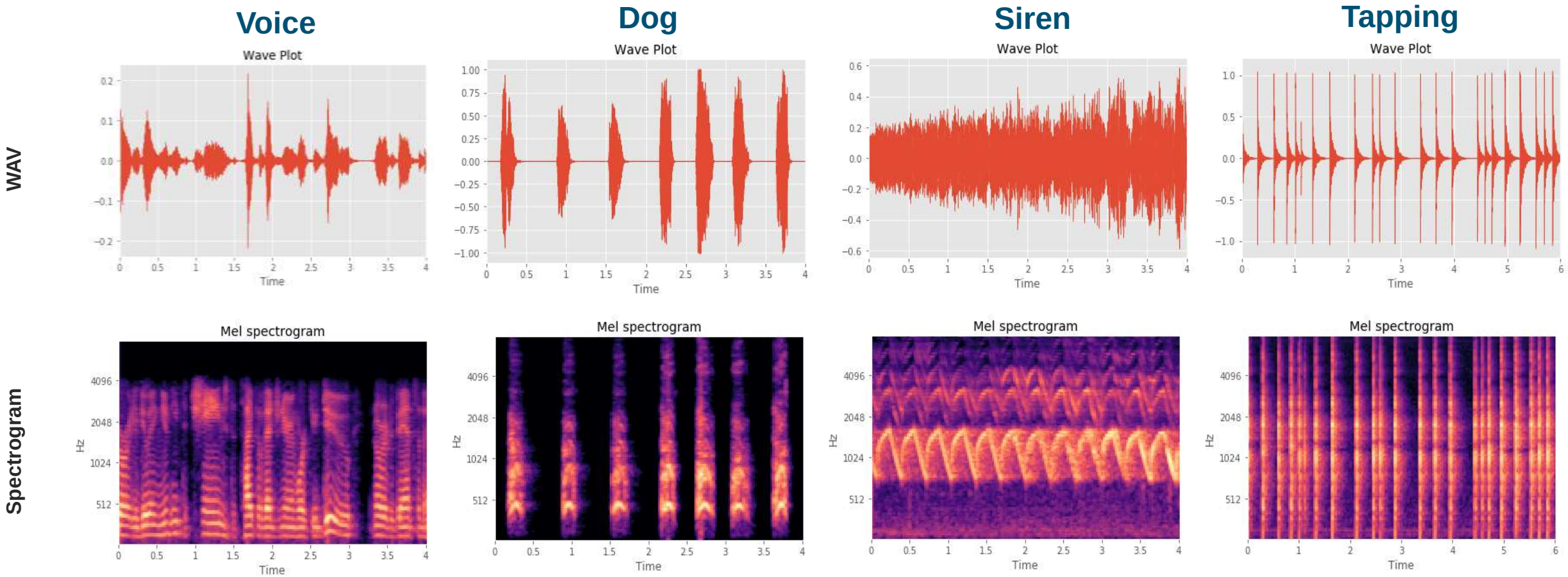
AI that Recognizes you

Noise Detection and Attenuation

- Noise detection on desktop clients generally available since FY18-Q1



Classification: From Signals to Images



Voices and "noise" have a distinct "image" that can be detected and filtered.

Deep Learning at Work in Cisco Collaboration Systems

AI/ML At Cisco: Enterprise Networking

The Power of AI/ML in the Network

- Anomaly detection
  - Dynamic network performance at different times and on different network conditions
  - Different expected performance on different SSIDs and/or locations for the same customer
  - Different expected performance for different customers
  - Static thresholds (even if configurable) would likely raise many false positives or miss relevant events
- Root cause analysis
  - Automatic selection of relevant KPIs explaining an issue
  - Cross-correlation across multiple devices
- Long-term trending
  - Automatically identifying trends and behavior changes on network entities/locations

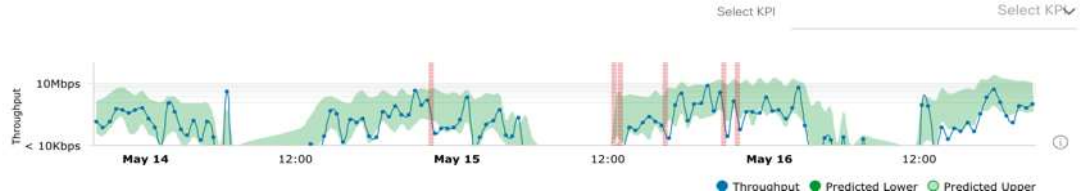
Cisco DNA Analytics

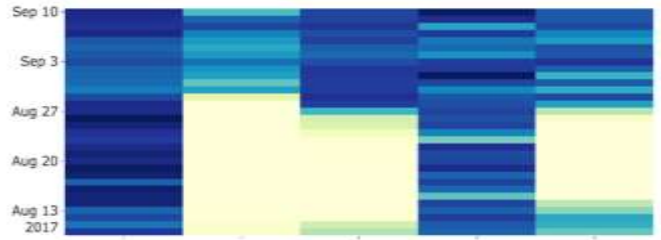
For Wireless, Wired Networks and IoT

Cognitive Analytics

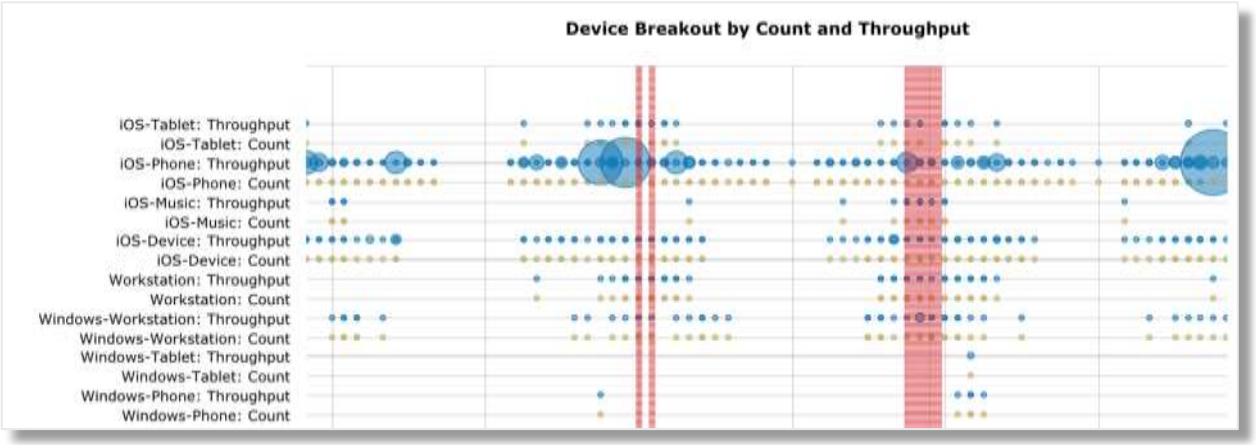
Anomaly detection
Root-cause Analysis
Long-term trending

Anomaly detection across hundreds of thousands of devices and thousands of networks





APs



App Throughput – High Packet Retries

Description

APs in network are experiencing a drop in Media Applications throughput. These radios are in the 5GHz band.

Impact of Last Occurrence

Aug 28, 2018 9:30 pm to Aug 28, 2018 10:30 pm

Location:

1 Building

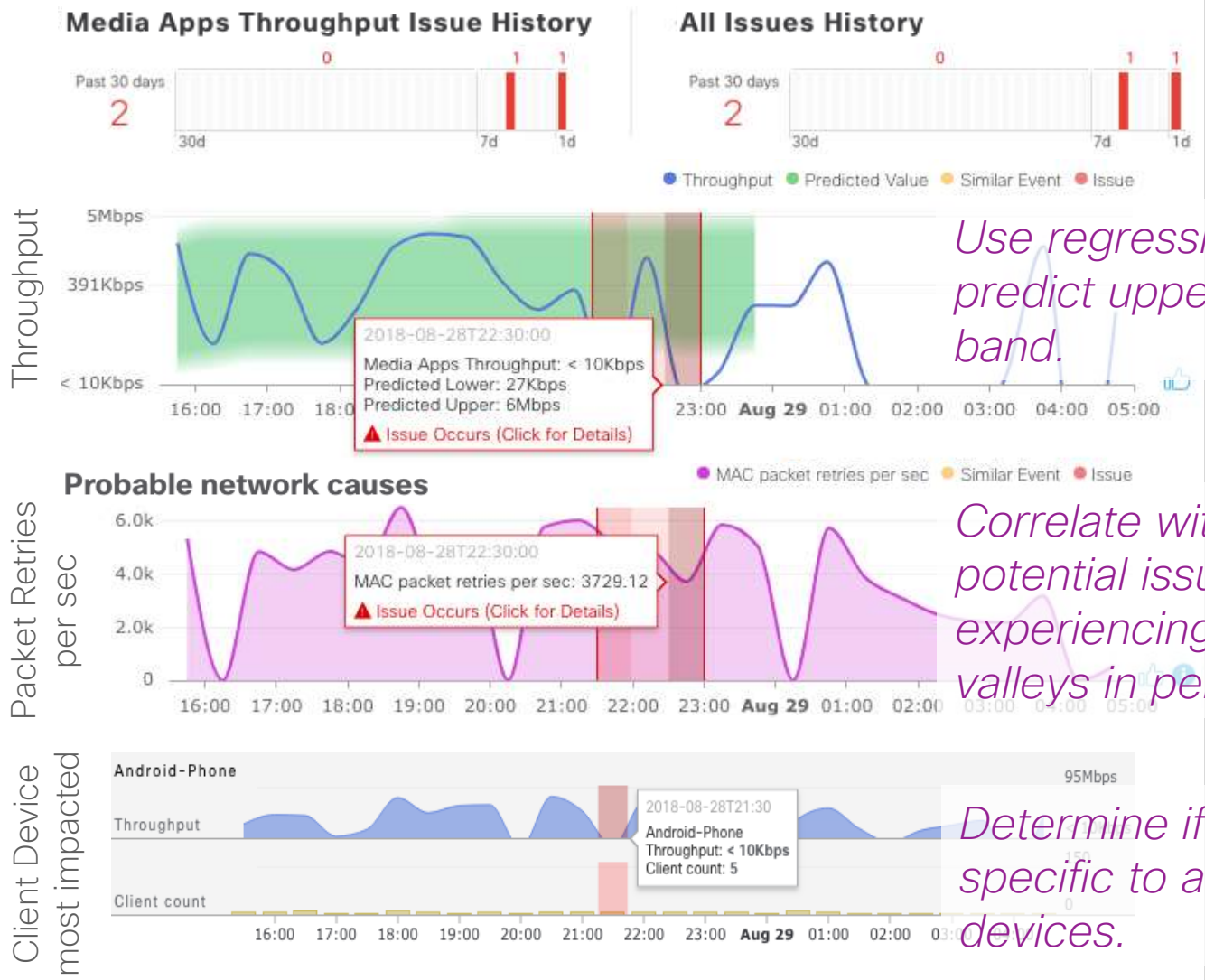
Clients

175 Wireless Clients

Additional Insight

Media Apps Throughput Issues Heatmap

Media Apps Throughput Peer Comparisons



Use regression to predict upper and lower band.

Correlate with other potential issues that are experiencing peaks / valleys in performance.

Determine if issue is specific to any specific devices.

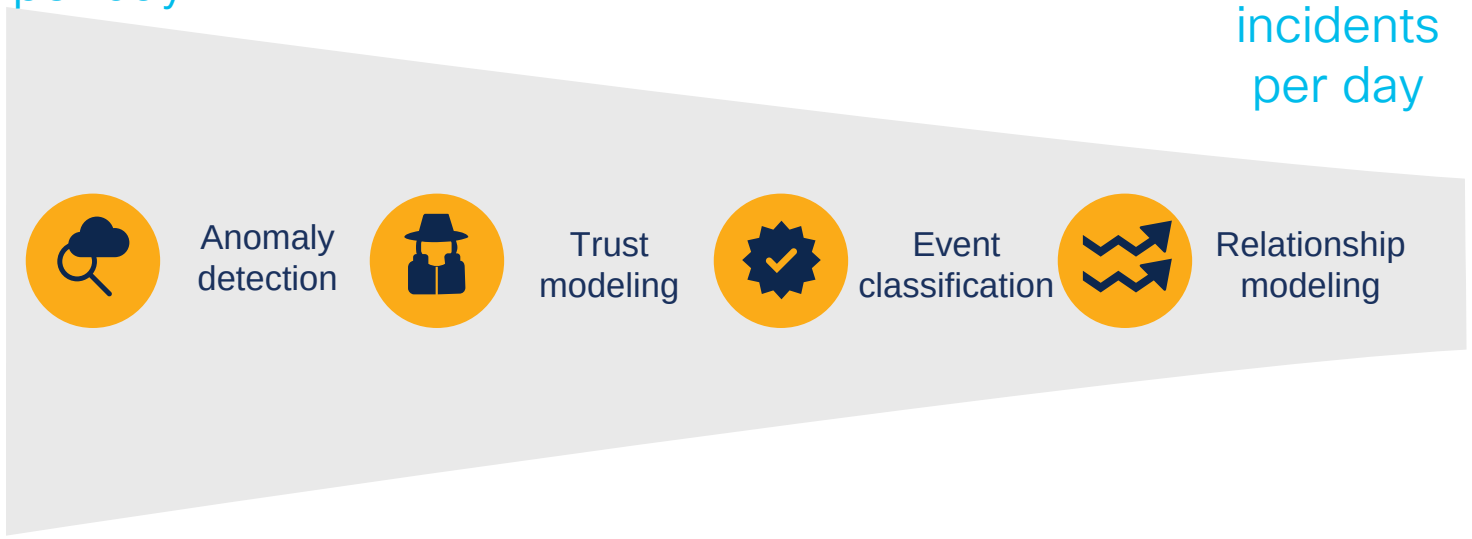
AI/ML At Cisco: Security

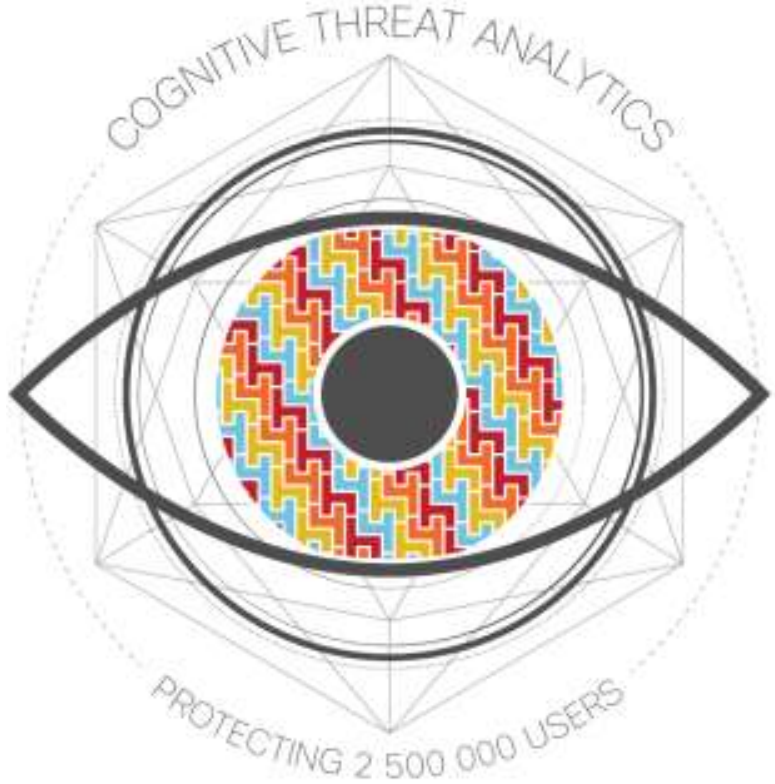
Cisco Cognitive Intelligence

Early Detection & Response with Artificial Intelligence

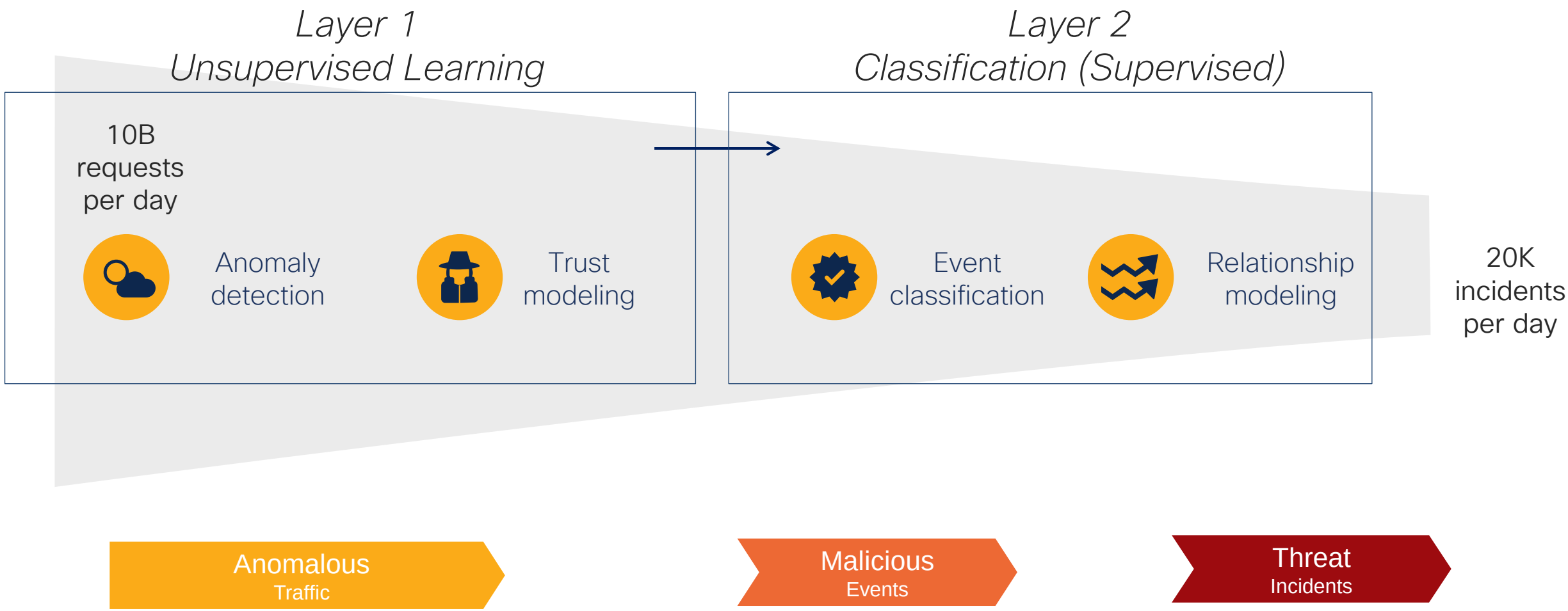
10B requests per day

20K incidents per day





Cisco Cognitive Intelligence

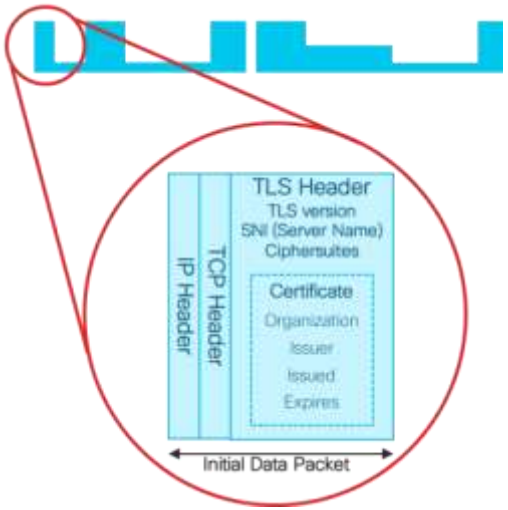


Stealthwatch for Security

Detecting Malware Embedded in Encrypted Traffic

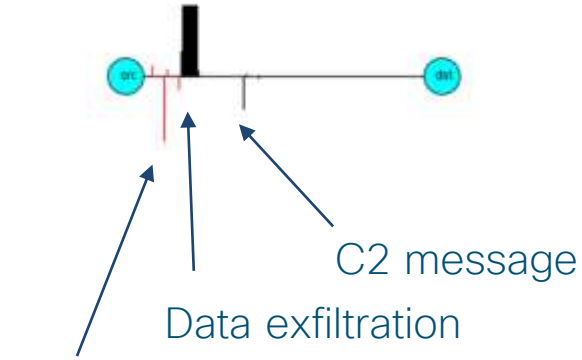
Initial Data Packet

Make the most of the unencrypted fields



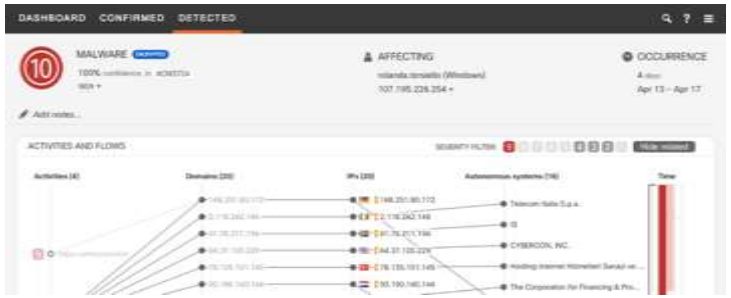
Sequence of Packet Lengths and Times

Identify the content type through the size and timing of packets



Global Risk Map

Who's who of the Internet's dark side



Broad behavioral information about the servers on the Internet.

Limitations, Society, Ethics, Future of AI/ML

Limitation of AI

- Performance limitation
- Not easy to explain
  - how AI make the decision
  - visualization for people
- Biased AI
  - through biased data - people doesn't want discrimination, but often they do such and if AI gets biased data, they could make a biased decision
  - AI unhealthy stereotypes - if man is king or father, then women is queen or mother, but when man is programmer, women is homemaker
  - facial recognition better works for light-skinned than dark-skinned - this is unfair and needs to be biased
- Combating bias:
  - technical solution:
    - "zero out" the bias in words ~ some words manually put to zero
    - use less biased and/or more inclusive data (e.g. for facial recognition system include data for multi-ethnics)
  - transparency and/or auditing processes - e.g. systematic check, if the system is right
  - diverse workforce

Ethical Questions are Emerging

- Adverse use of AI
- Adversarial attacks on AI
- How do we guard against mistakes made by machines? Who is liable?
- What about self-driving cars?
- Can we allow machines to judge other humans based on a learning mechanism?

AI Starting to Replace Humans for Certain Tasks

Bots starting to replace humans in customer service:

- The “Gootsman bot” fooled more than 100 raters into thinking they were talking to a human
- Will apply to vast array of customer service scenarios

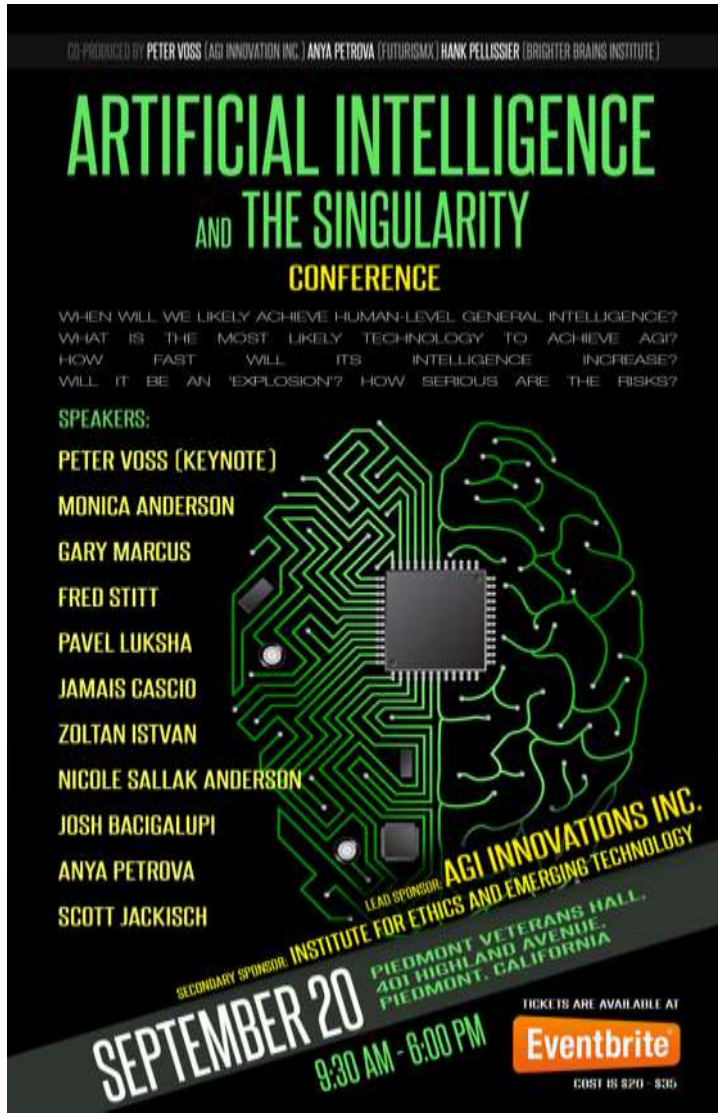
What About the “Singularity”?

- Will machines ever become self-aware?

Wikipedia Definition:

*The **technological singularity** (also, simply, the **singularity**) is the hypothesis that the invention of artificial superintelligence (ASI) will abruptly trigger runaway technological growth, resulting in unfathomable changes to human civilization.*

- How do we control these machines if they become self-aware?
- Today – still in the realm of science fiction



Many Mysteries Remain . . .

- Deep Learning Mysteries Remain
  - Back-propagation: Why is such a simple algorithm so powerful?
  - Many adversarial examples exist (when algorithms misclassify)
- How to scale up ML Algorithms
  - How can we scale to millions of training examples, thousands of features, hundreds of classes?
- One-Shot Learning
  - How can we learn from very few training examples?

Handy Resources

AI@ Cisco website	(www.cisco.com/go/intelligence) Website that describes which Cisco products have ML/AI in them, highlights some of our Cortex members (Cortex = Cisco's ML/AI Virtual Center of Excellence) and points to other resources
DevNet AI page	(https://developer.cisco.com/site/ai/) Shows developers how to get started with AI quickly using DevNet tools
	https://www.coursera.org


CISCO