EECS 349
Machine Learning

Instructor: Doug Downey

(some slides from Pedro Domingos, University of Washington)

Today

- Logistics
- ML Overview

Logistics

- **Instructor:** Doug Downey
  - Email: ddowney@eecs.northwestern.edu
  - Office hours: Mondays 4:00-5:00 (or by appt), Mudd 3111
- **TAs:** Dave Demeter, Zheng Yuan, 4 ugrad peer mentors
- **Web:** (linked from prof. homepage) http://www.cs.northwestern.edu/~downey/courses/349_Fall2018/
  - Also: Canvas, Piazza

Grading and Assignments (1 of 2)

Assignment	Points
Homework 1	15
Homework 2	5
Project Proposal	5+5 (peer mentoring)
Exam 1	10
Homework 3	10
Project Status Report	5+5 (peer mentoring)
Homework 4	10
Exam 2	10
Project Website	20
	100

A	A-	B+	B	B-	C+	C	C-	Etc…
93+	92-90	89-87	86-83	82-80	79-77	76-73	72-70	69…

Grading and Assignments (2 of 2)

- Four homeworks (40 pts)
  - Submitted via Canvas according to hmwk instructions
    - Late penalty 10% per day – must be within 1 week of original deadline
  - Significant programming, some exercises
    - Programming assignments in **groups of two (or one)**
- Exams (20 pts)
  - Monday of Week 4, Friday of Week 9
- Project (30 pts + 10 peer review)
  - Teams of ***k***
  - Define a task, create/acquire data for the task, train ML algorithm(s), evaluate & report

Expectations

- Grades
- Academic Integrity
  - You are expected to do your own work
    - More details in syllabus linked from course home page
  - Suspected violations of integrity policy will be referred to the administration
- Slides may not make sense if you don’t come to class

Prerequisites

- Significant Programming Experience
  - EECS 214, 325 or the equivalent
  - Example: implement decision trees (covered starting Monday)
  - **Python** is the language we'll use
    - You'll have skeleton code to help you
    - (also, I barely know Python)
- Basics of probability
  - E.g. independence

Source Materials

- Papers & Web pages
- Reading for next week:
  - Required:
    - Decision trees (see the Decision Tree notes when they're posted)
  - Optional:
    - Gartner 2016 Hype Cycle
      (machine learning at peak hype as of 2016)

Participation

- Think/Pair/Share (next)
- Peer Review

Today

- Logistics
- ML Overview

Think/Pair/Share

Why study Machine Learning?

Think

Start

End

Think/Pair/Share

Why study Machine Learning?

Think

Start

End

Think/Pair/Share

Why study Machine Learning?

Pair

Start

End

Think/Pair/Share

Why study Machine Learning?

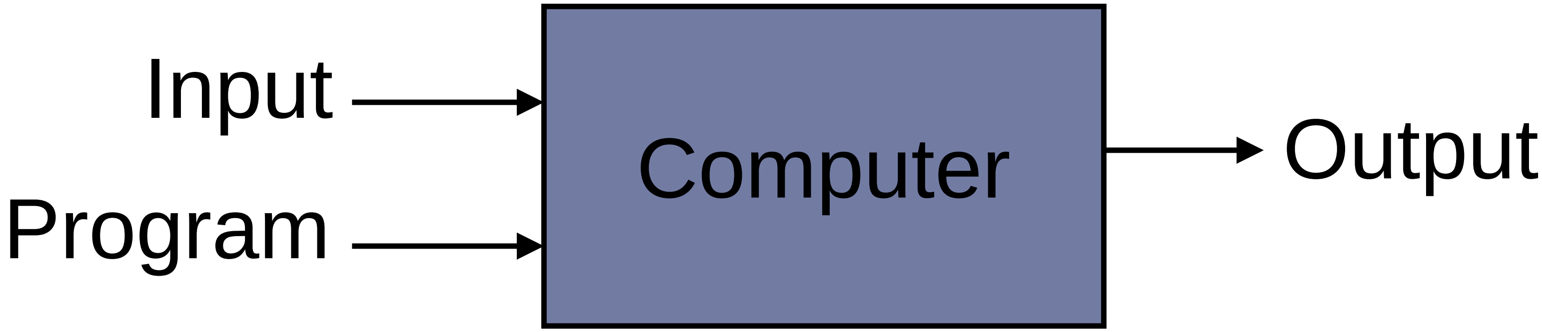
Share

What is Machine Learning?

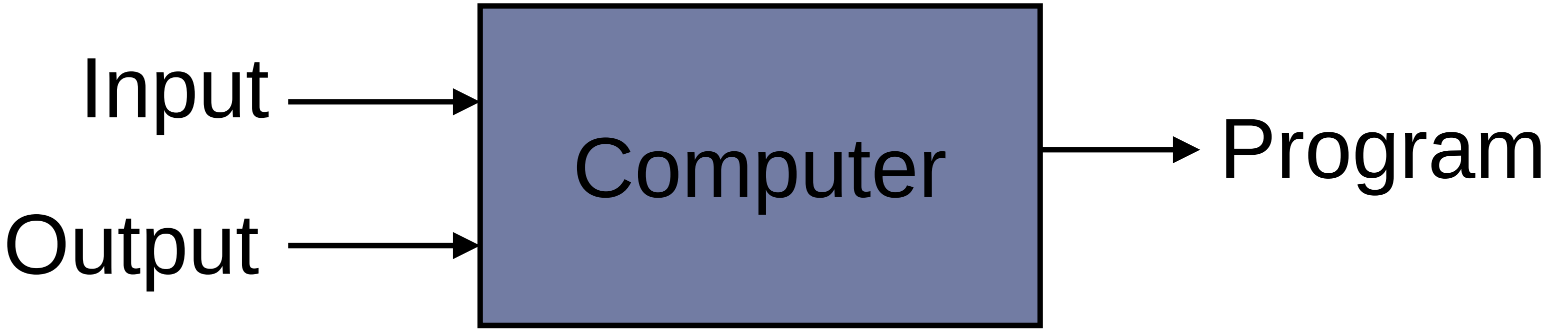
- “The study of computer programs that improve automatically with experience”
  T. Mitchell *Machine Learning*

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

Traditional Programming



Machine Learning

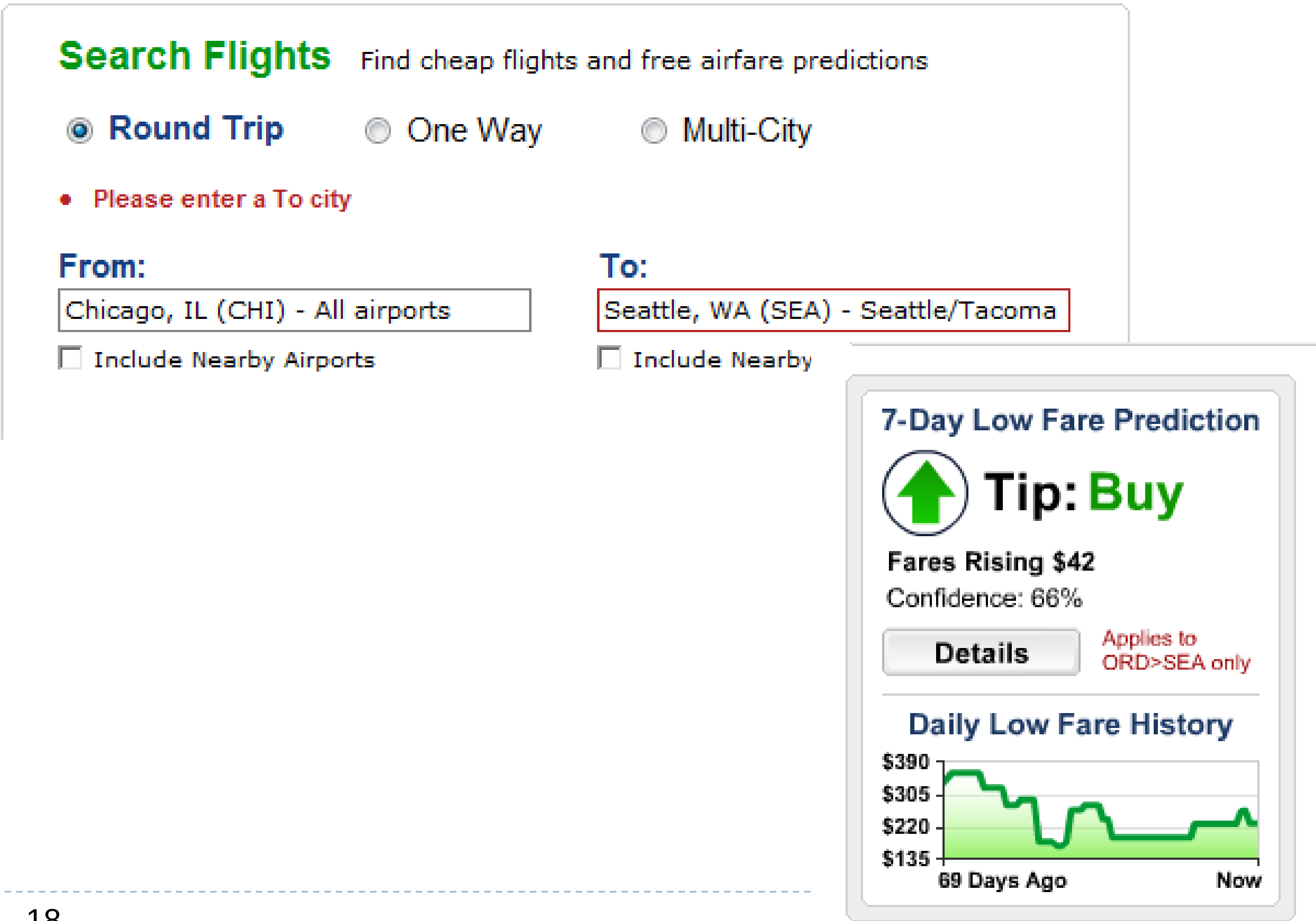


Magic?

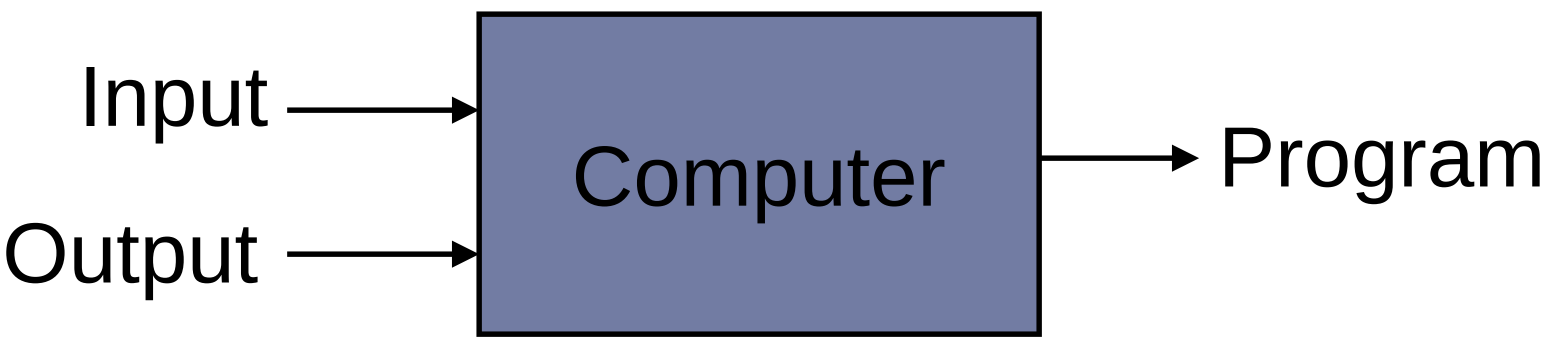
No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Example: Farecast



Sample Applications



Relationship of Machine Learning to…

- **Statistics**
- **Analytics / Data Science**
- **Artificial Intelligence**

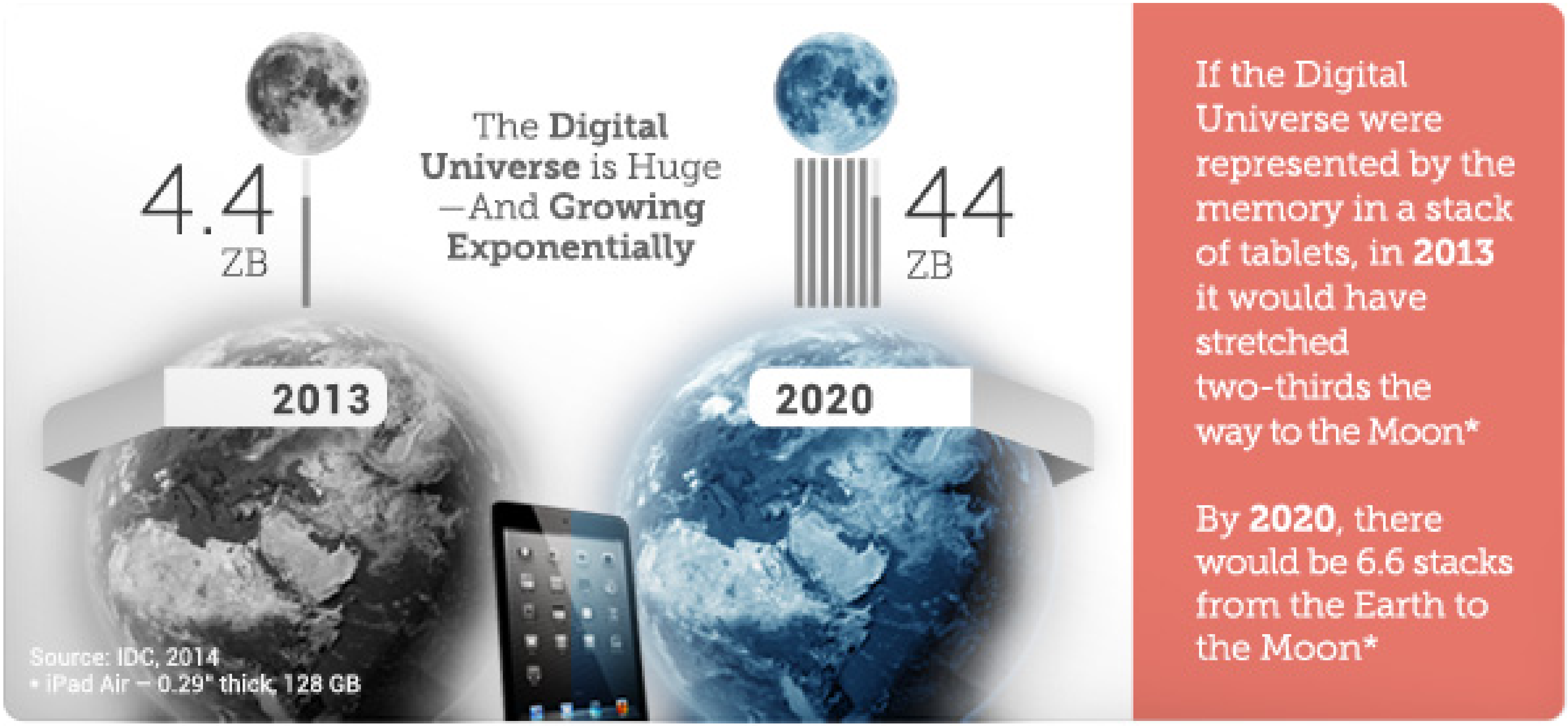
Why study Machine Learning? (1 of 4)

- "A breakthrough in machine learning would be worth ten Microsofts" (Bill Gates)
- "Machine learning is the next Internet" (Tony Tether, former Director, DARPA)

These quotes are ~10 years old
(e.g. Gates is from the NYT, 2004)

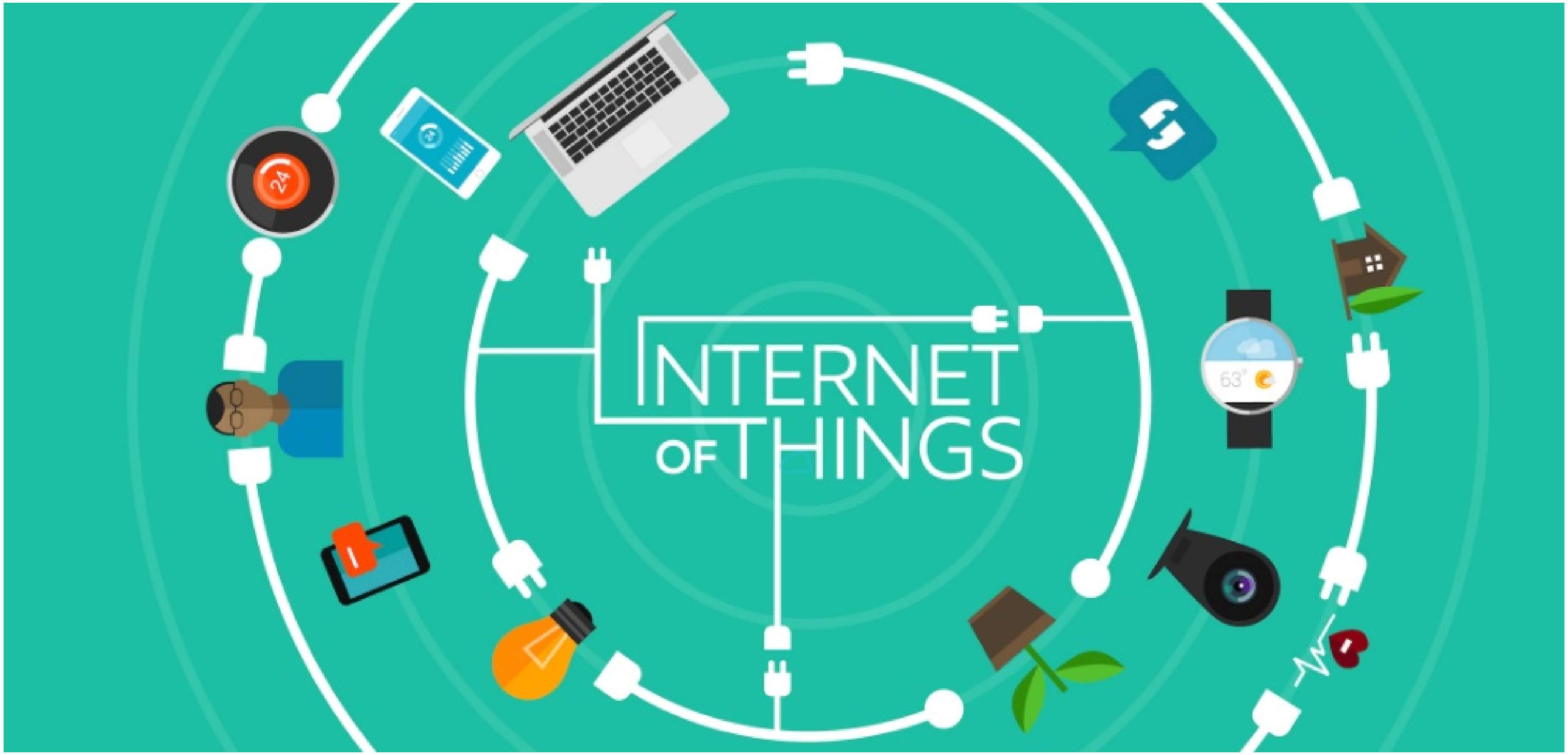
More recent:
"Artificial intelligence is one of the great opportunities for improving the world today," (Reid Hoffman, co-founder of $1B deep learning research center)

Why study Machine Learning? (2 of 4)



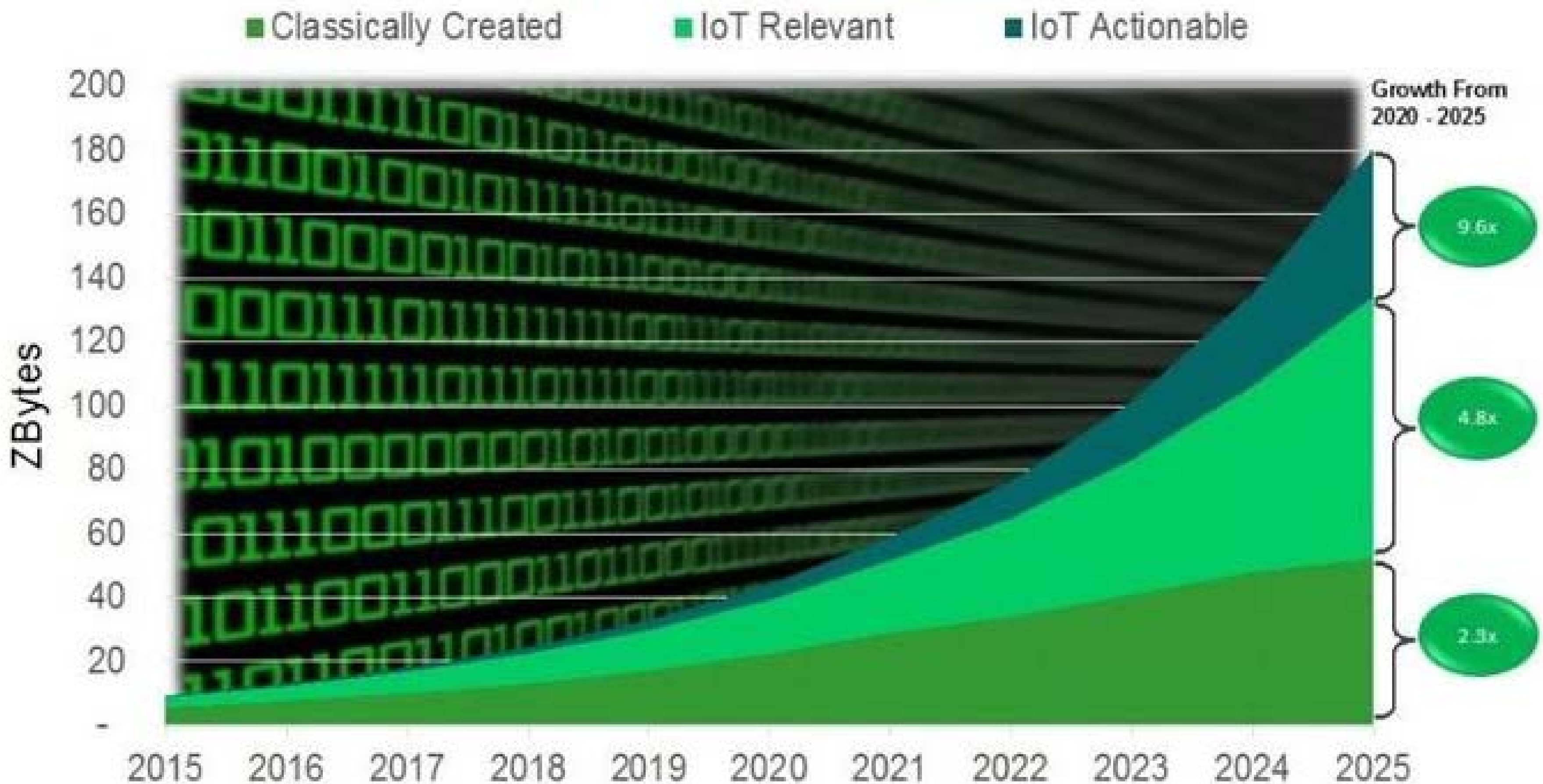
http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

Why study Machine Learning? (3 of 4)



http://www.gartner.com/newsroom/id/3598917

The Content Created World 2015 – 2025 *

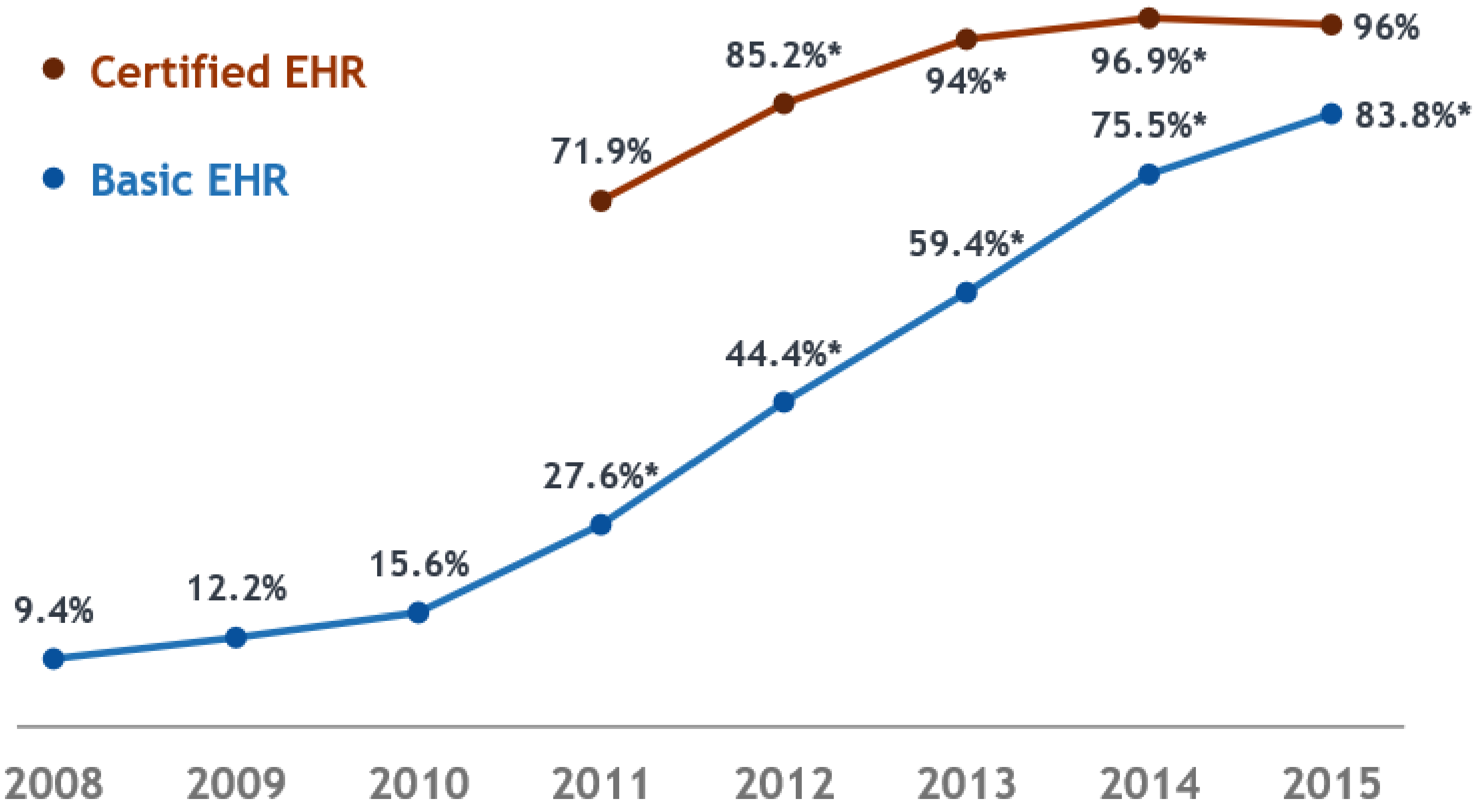


Source: The Internet of Things: Getting Ready to Embrace Its Impact on the Digital Economy (IDC #DR2016_GS4_VT)

@Vernonxt
@CMacIDC

Why study Machine Learning? (4 of 4)

Hospitals possessing (Certified EHR) or adopting (Basic EHR) electronic medical records. What will we be able to learn from these?



https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php

What You'll Learn in this Class

- How do ML algorithms work?
  - Learn by **implementing**, using
- When **should** I use ML?
- For a **real** problem, how do I:
  - Express my problem as an ML task
  - Choose the right ML algorithm
  - Evaluate the results

ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - **Representation**
  - **Evaluation**
  - **Optimization**

Representation

- *How do we represent the function from input to output?*
  - Decision trees
  - Sets of rules / Logic programs
  - Instances
  - Graphical models (Bayes/Markov nets)
  - Neural networks
  - Support vector machines
  - Model ensembles
  - Etc.

Evaluation

- *Given some data, how can we tell if a function is "good"?*
  - Accuracy
  - Precision and recall
  - Squared error
  - Likelihood
  - Posterior probability
  - Cost / Utility
  - Margin
  - Entropy
  - K-L divergence
  - Etc.

Optimization

- *Given some data, how do we* ***find*** *the "best" function?*
  - Combinatorial optimization
    - E.g.: Greedy search
  - Convex optimization
    - E.g.: Gradient descent
  - Constrained optimization
    - E.g.: Linear programming

Inductive Learning

- **Given** examples of a function $(\mathbf{x}, f(\mathbf{x}))$
- **Predict** function $f(\mathbf{x})$ for new instances $\mathbf{x}$
  - Discrete $f(\mathbf{x})$: Classification
  - Continuous $f(\mathbf{x})$: Regression
  - $f(\mathbf{x})$ = Probability($\mathbf{x}$): Probability estimation

- Example:
  - $\mathbf{x}$ = <Flight=United 102, FlightDate=May 26, Today=May 7>
  - $f(\mathbf{x})$ = +1 if flight price will increase in the next week, or
-1 otherwise

What We'll Cover

- **Inductive learning**
  - Decision tree induction
  - Instance-based learning
  - Neural networks
  - Bayesian Learning
  - Logistic Regression
  - Support vector machines
  - Learning theory
  - Reinforcement Learning
- **Unsupervised learning**
  - Clustering
  - Dimensionality reduction

Today

- Logistics
  - 4 homeworks, 2 exams, course project. No final.
  - Take a look at the course Web page for more.
- ML Overview
  - Like gardening
  - data = rich source of fuel for ML
  - More soon…