Defining Machine Learning

Dr. Alex Williams

August 21, 2020

THE UNIVERSITY OF TENNESSEE KNOXVILLE

COSC 425: Introduction to Machine Learning
Fall 2020 (CRN: 44874)

Syllabus Clarifications

#1: No textbook requirement. (See Daume in Canvas.)

#2: Added Office Hours link to Canvas.

#3: Alternative Course Website
http://web.eecs.utk.edu/~acw/teaching/cosc425/

COSC425: Intro to Machine Learning

Course Time: M/W/F @ 2:15-3:05
CRN: 44874, **Term:** Fall 2020

Instructor: Dr. Alex Williams.
Office Hours: Tues/Thurs 2:00-4:00 (**Reserve online!**)

Teaching Assistant: Zhuohang Li (zli96@vols.utk.edu)
Office Hours: By appointment (Book via e-mail.)

Teaching Assistant: Tuhin Das (tdas1@vols.utk.edu)
Office Hours: By appointment (Book via e-mail.)

Overview • Schedule • References • Office Hours

Note: The materials on this webpage are available on Canvas.

Overview

Machine learning is concerned with computer programs that automatically improve their performance through experience. This course covers the theory and practice of machine learning from a variety of perspectives. We cover topics such as clustering, decision trees, neural network learning, statistical learning methods, Bayesian learning methods, dimension reduction, kernel methods, and reinforcement learning. Programming assignments include implementation and hands-on experiments with various learning algorithms.

Schedule

Note: This schedule is subject to change.

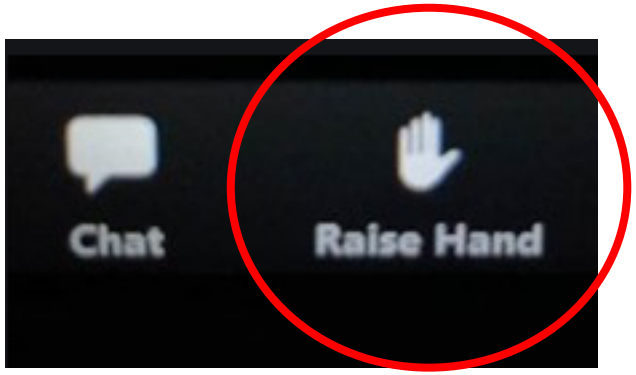
Week	Date	Topic	Notes
1	Aug 19	Introduction	
	Aug 21	Defining Machine Learning	**Reading:** Wagstaff.
2	Aug 24	Decision Trees I	
	Aug 26	Decision Trees II	
	Aug 28	The Limits of Learning	

Syllabus Clarifications

#4: Modern Machine Learning → Python

- LearnPython (http://learnpython.org)
- PythonTutor (http://pythontutor.com)
- Programming w/ Mosh (https://www.youtube.com/…)
  - YouTube Video → 6-hour Intro to Python.

Any Questions?



Use Zoom's "*Raise Hand*" feature, and I'll un-mute you.

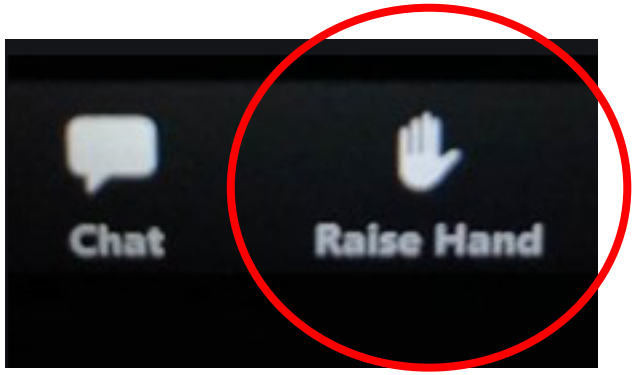
Today's Agenda

We will address:

1. What is "*Machine Learning*" (ML)?
2. How is ML operationalized?
3. What are the grand challenge of modern ML?

What is Machine Learning?

How would you define "machine learning"?



Use Zoom's "*Raise Hand*" feature, and I'll un-mute you.

"At a basic level, machine learning is about **predicting the future based on the past**."

- Hal Daumé III

"Machine learning is the field of study that gives computers the ability to learn **without being explicitly programmed**."

- Arthur Samuel (1959)

"How can we build computer systems that **automatically improve with experience**, and what are the fundamental laws that govern all learning processes?"

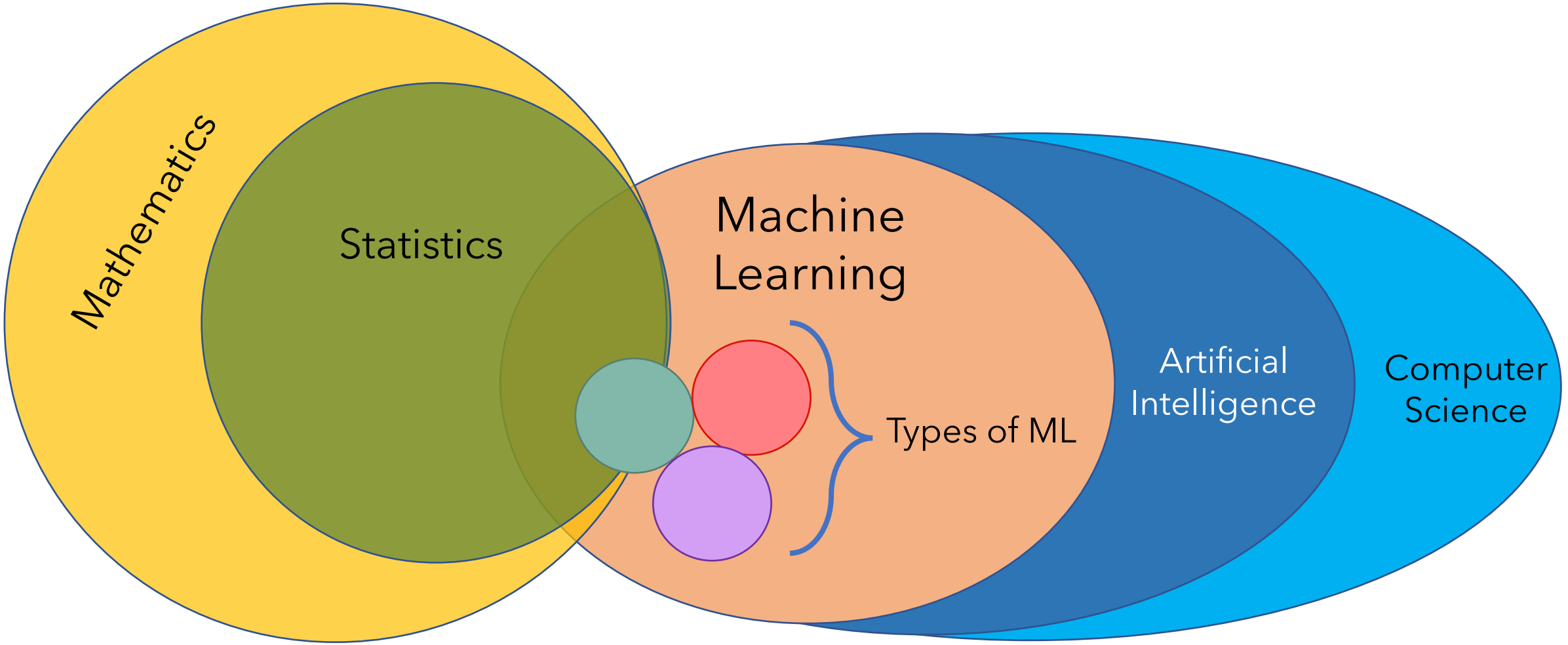
- Tom Mitchell (1998)

"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."

- Tom Mitchell (1998)

(Representation + Evaluation + Optimization) = Learning

- Pedro Domingos (2012)



- Ryan Urbanowicz (2018)

COSC 425 - Introduction to Machine Learning

COSC 425 - Introduction to Machine Learning

3 Credit Hours

Machine learning is concerned with computer programs that automatically improve their performance through experience. This course covers the theory and practice of machine learning from a variety of perspectives. We cover topics such as clustering, decision trees, neural network learning, statistical learning methods, Bayesian learning methods, dimension reduction, kernel methods, and reinforcement learning. Programming assignments include implementation and hands-on experiments with various learning algorithms.

(RE) Prerequisite(s): ECE 313 or ECE 317 or MATH 323 with a grade of C or better; and MATH 251 or MATH 257 with a grade of C or better.
Comment(s): Prior knowledge may satisfy prerequisite with consent of instructor.

- Someone, at some point in time.

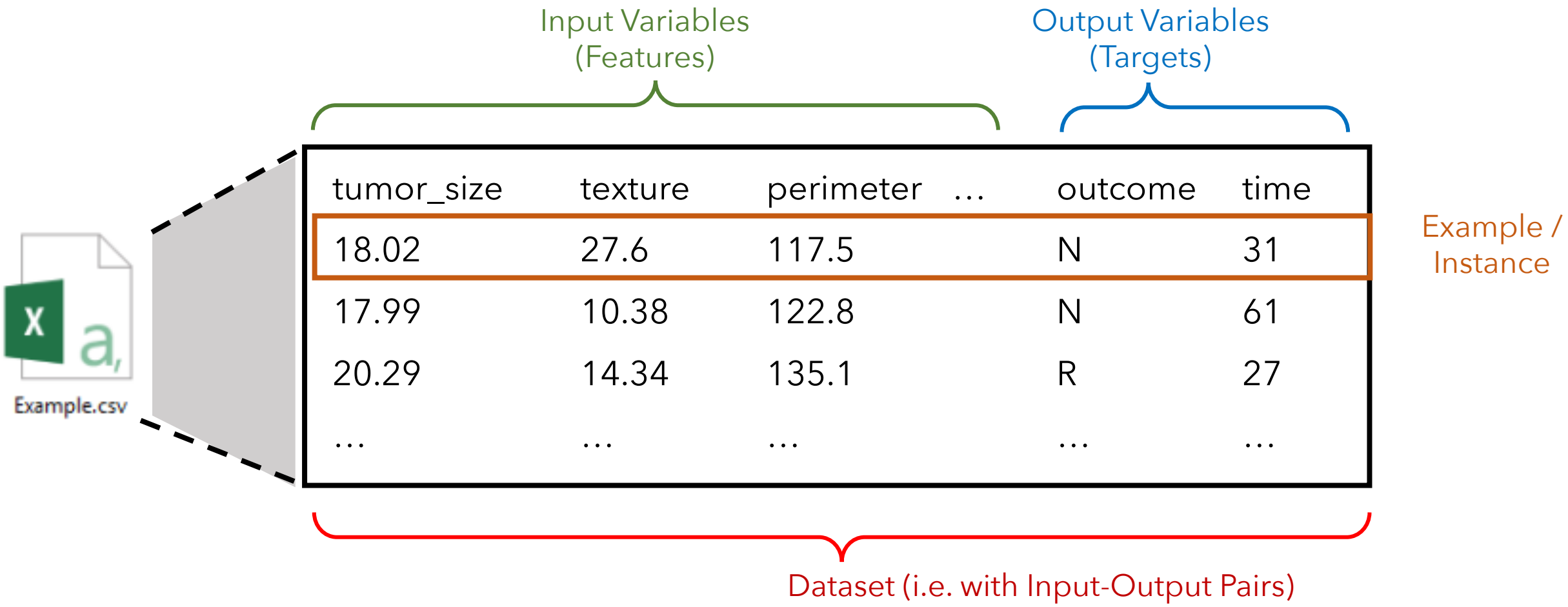
So, what's the right definition?

Technically: All of them.

The overarching goal of these methods is to learn a function from prior data.

Spoiler: Machine learning is (mostly) operationalized mathematics.

Terminology

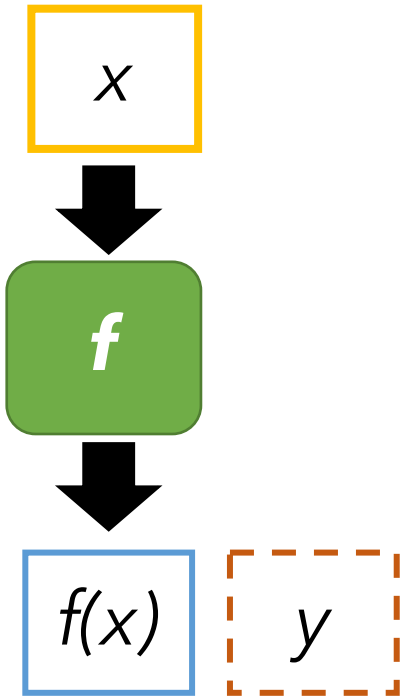


Goal: Maximize performance for any *x*.
Both in **Training and Test Data!**

Training Data

Input-output Pairs
(x_i, y_i)

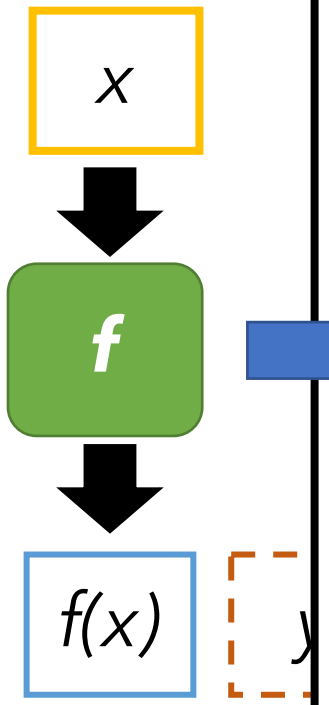
Testing Data



Major Assumption: You have access to y_i, (i.e. output variables).

What does "f" look like?

Testing Data



$$y = k + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Dependent Variable

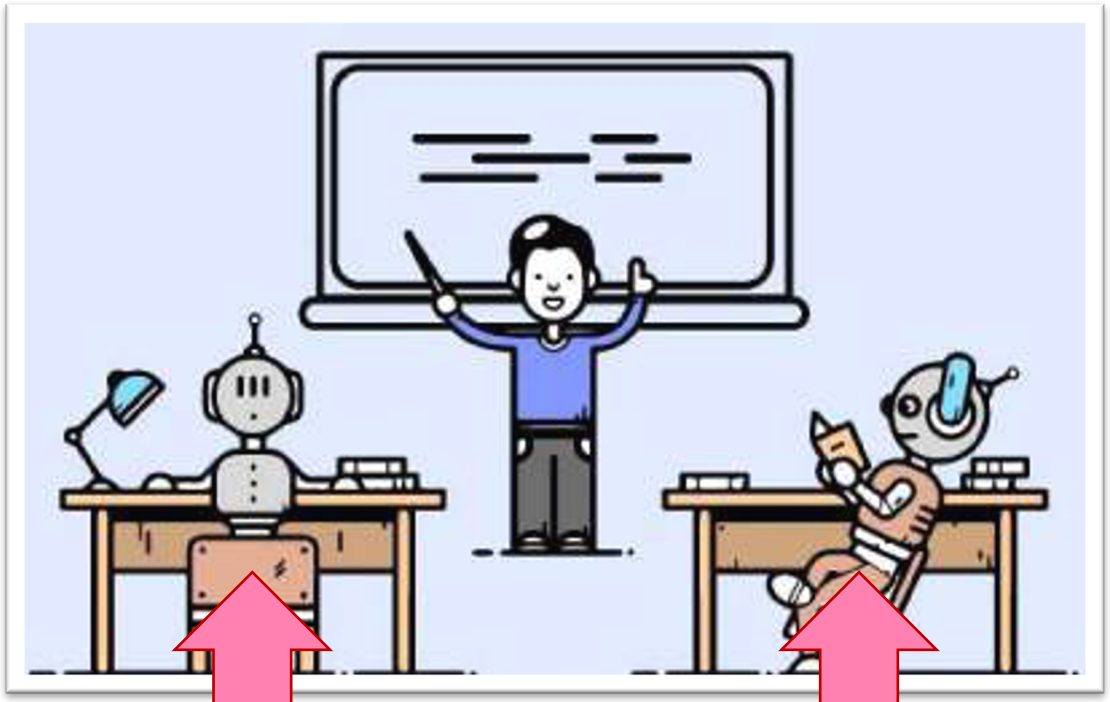
Coefficient

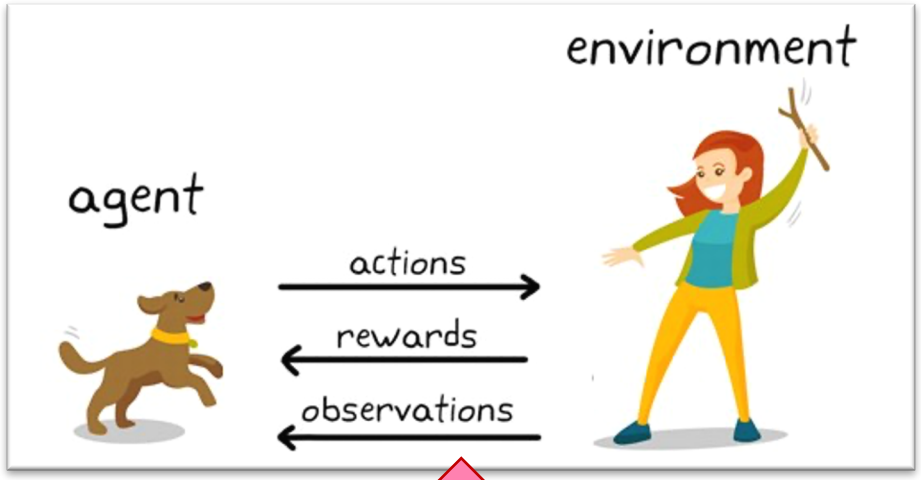
Predictors

Linear regression as an example.

Types of Machine Learning

Types of Machine Learning





Supervised Learning

Unsupervised Learning

Reinforcement Learning

Supervised Learning: Classification

Use-Case Criteria:

- You have output variables, i.e. y_i..
- Your OVs are **discrete / categorical**.

Example: Spam Filtering

- ***Goal***: Learn a function from categorical output.
- e.g. {spam, not spam}

	isUTKEmail	HeaderKeyword	Word 1	Word 2		isSpam
x1	Yes	CS425	Hi	Prof	…	No
x2	Yes	Orientation	Alex	You	…	No
x2	No	urgent	Dear	Sir	…	Yes
x4	No	cash	hello	I	…	Yes
x5	No	help	are	you	…	Yes
x6	Yes	Survey	Faculty	this	…	No
			…			

Supervised Learning: Regression

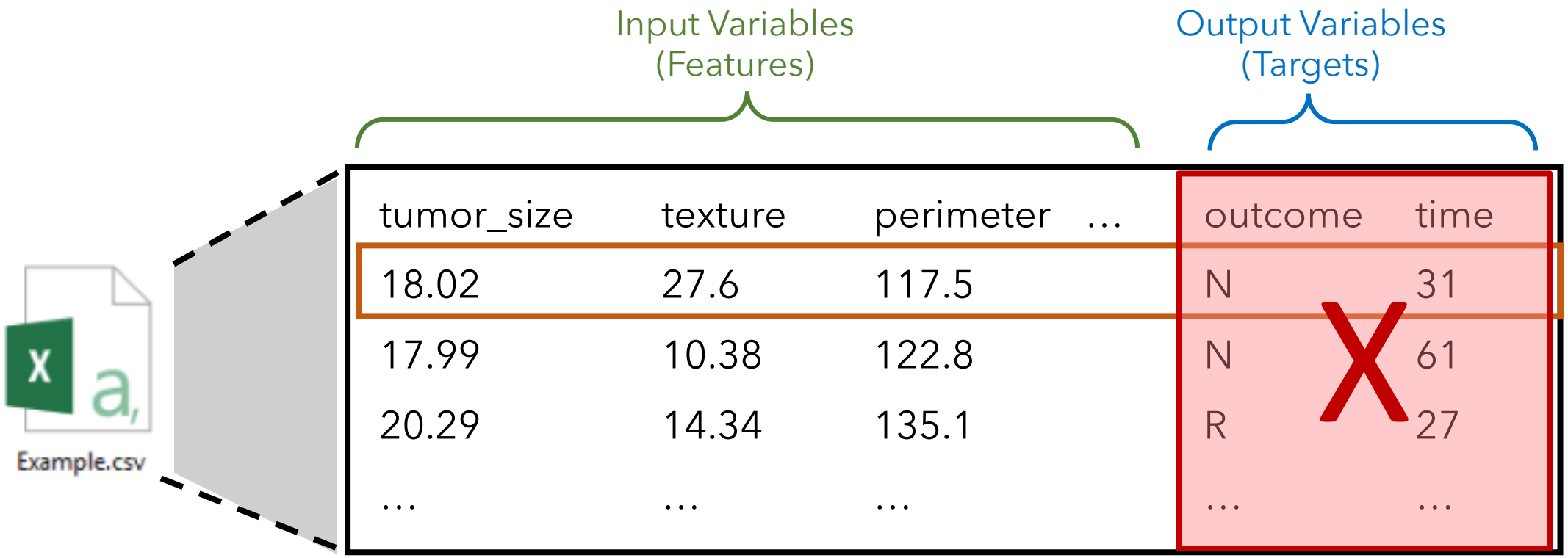
Use-Case Criteria:

- You have output variables, i.e. y_i.
- Your OVs are **continuous**.

Example: Tesla Speed Control

- ***Goal***: Learn a function for a continuous output.
- e.g. {0-100 MPH}

Criticism:
Output Variables are Unknown.



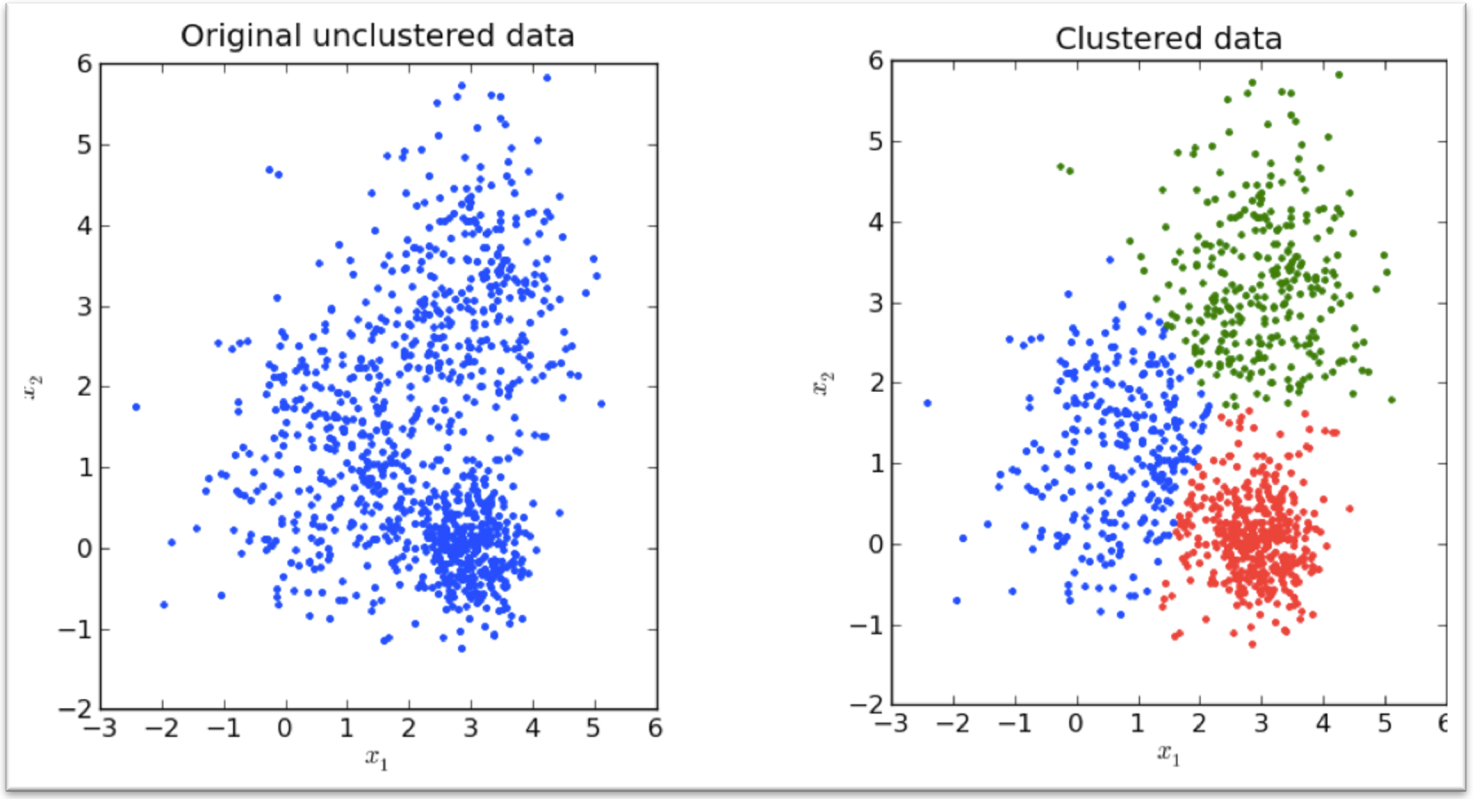
Unsupervised Learning: Clustering

Use-Case Criteria:

- You have no output variables.

Example: Unlabeled Data

- ***Goal***: Learn a function from input.
- e.g. Organize the data!



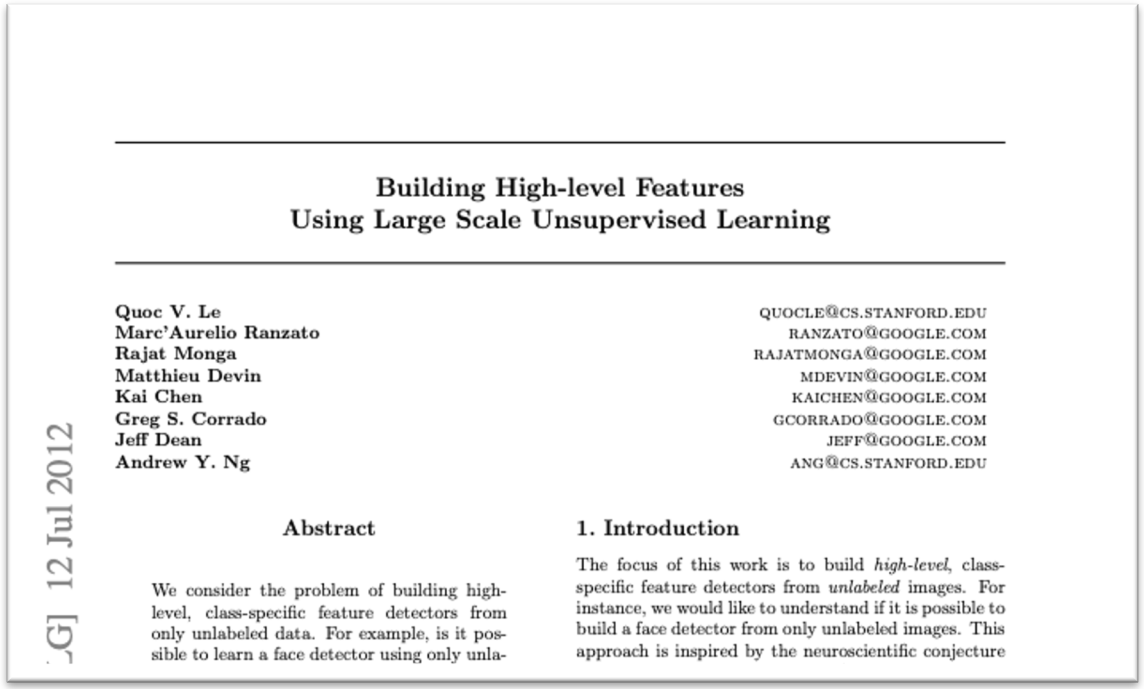
Unsupervised Learning: Feature Selection

Long-Term Goal.

- Figure out which inputs matter.

Feasible, but Challenging.

- Data, data, and more data.



Building High-level Features
Using Large Scale Unsupervised Learning

Quoc V. Le QUOCLE@CS.STANFORD.EDU
Marc'Aurelio Ranzato RANZATO@GOOGLE.COM
Rajat Monga RAJATMONGA@GOOGLE.COM
Matthieu Devin MDEVIN@GOOGLE.COM
Kai Chen KAICHEN@GOOGLE.COM
Greg S. Corrado GCORRADO@GOOGLE.COM
Jeff Dean JEFF@GOOGLE.COM
Andrew Y. Ng ANG@CS.STANFORD.EDU

[cs.LG] 12 Jul 2012

Abstract

We consider the problem of building high-level, class-specific feature detectors from only unlabeled data. For example, is it possible to learn a face detector using only unla-

1. Introduction

The focus of this work is to build *high-level*, class-specific feature detectors from *unlabeled* images. For instance, we would like to understand if it is possible to build a face detector from only unlabeled images. This approach is inspired by the neuroscientific conjecture

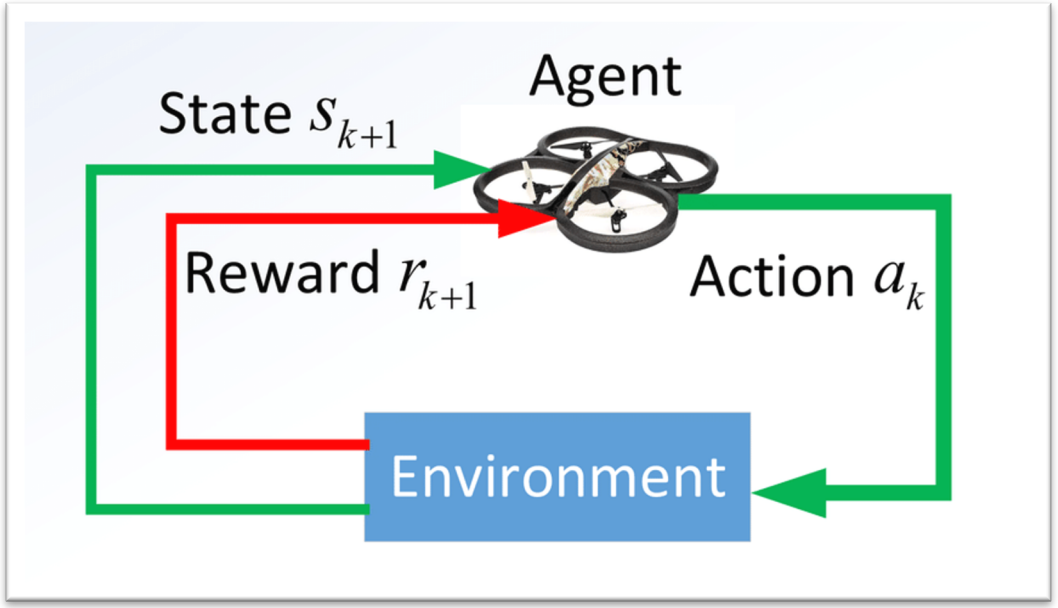
+2000 Citations!
https://arxiv.org/pdf/1112.6209.pdf

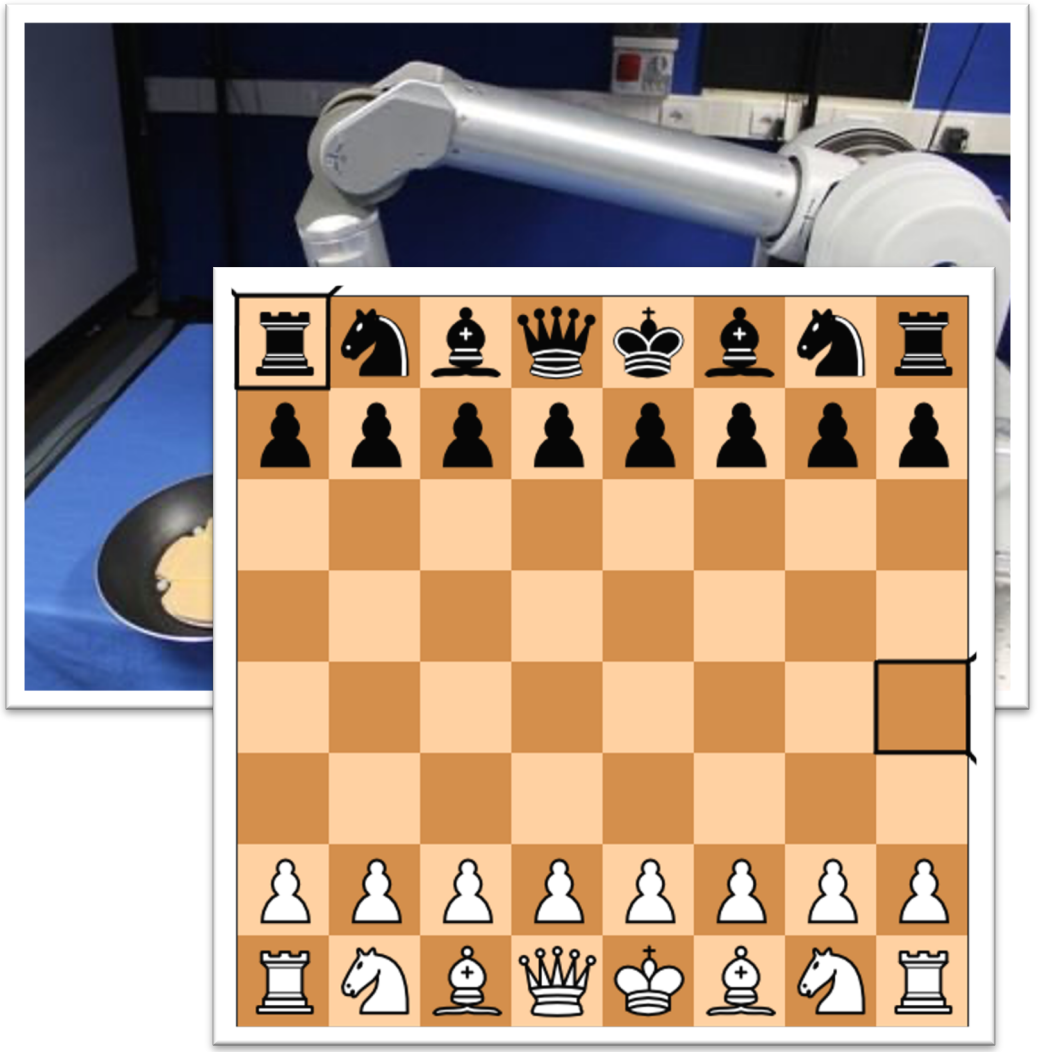
Criticism:
"Learning from Data" isn't Learning.

Reinforcement Learning

Use-Case Criteria:

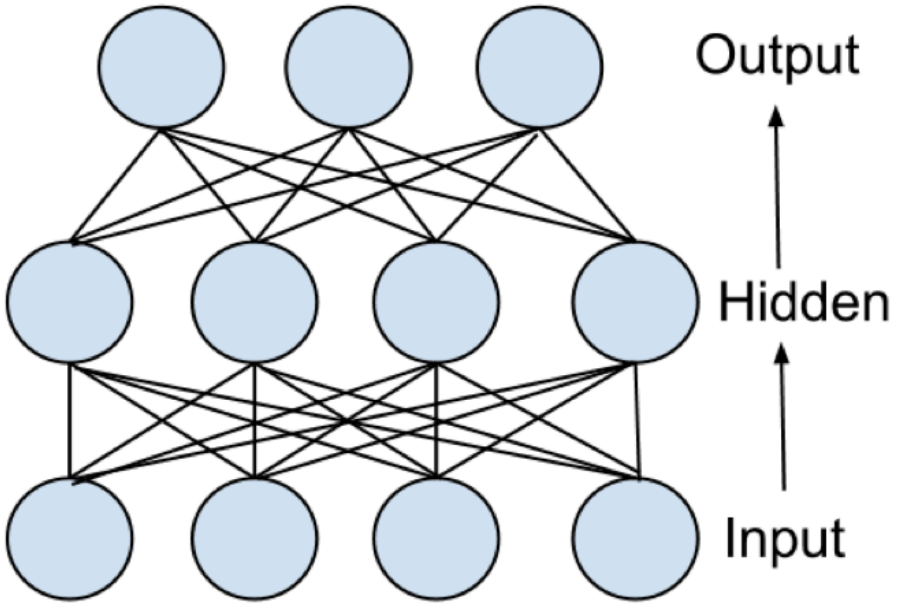
- You have a some “environment”.
- You have some notion of “good” behavior.



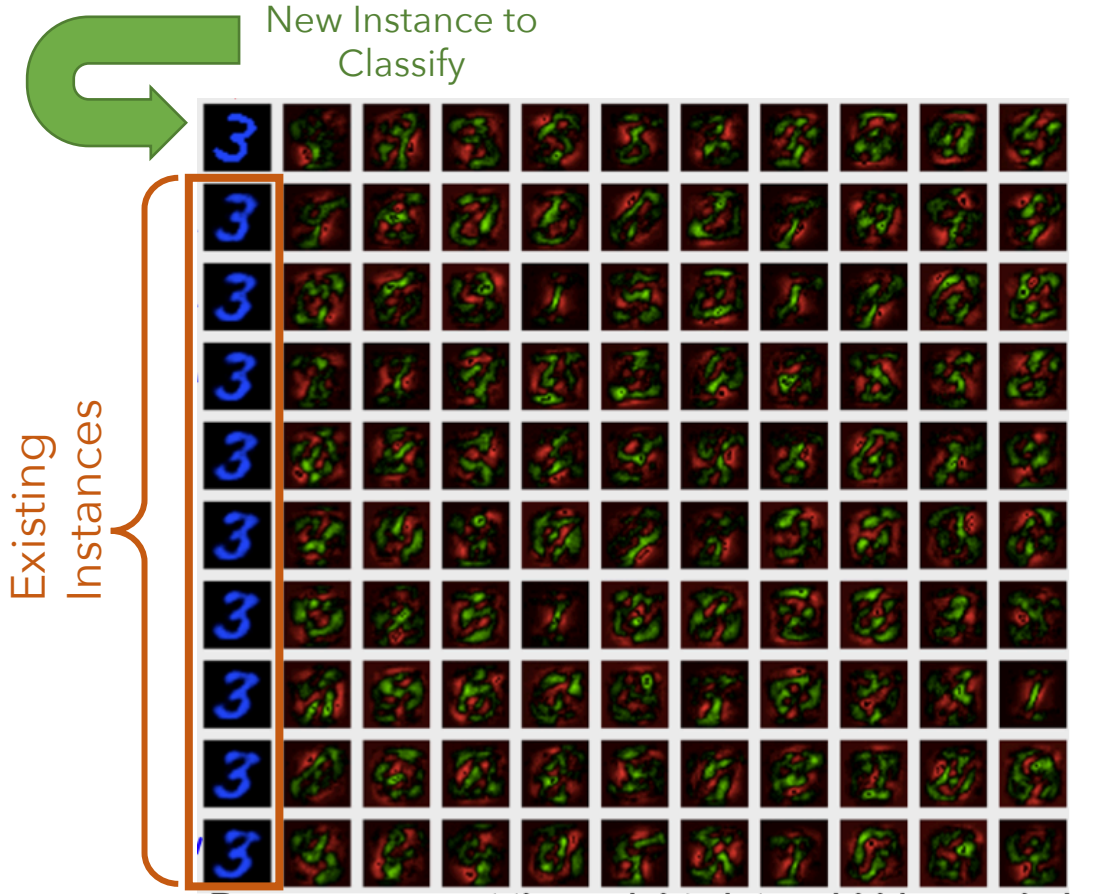


Case Studies

Case #1: OCR



A Neural Network



Boxes represent the weights into a hidden node in a neural network learner.

Case #1: OCR

Least Complex

Most Complex

Type	Classifier	Distortion	Preprocessing	Error rate (%)
Linear classifier	Pairwise linear classifier	None	Deskewing	7.6[9]
K-Nearest Neighbors	K-NN with non-linear deformation (P2DHMDM)	None	Shiftable edges	0.52[18]
Boosted Stumps	Product of stumps on Haar features	None	Haar features	0.87[19]
Non-linear classifier	40 PCA + quadratic classifier	None	None	3.3[9]
Support vector machine	Virtual SVM, deg-9 poly, 2-pixel jittered	None	Deskewing	0.56[20]
Deep Neural network	2-layer 784-800-10	None	None	1.6[21]
Deep Neural network	2-layer 784-800-10	elastic distortions	None	0.7[21]
Deep neural network	6-layer 784-2500-2000-1500-1000-500-10	elastic distortions	None	0.35[22]
Convolutional neural network	6-layer 784-40-80-500-1000-2000-10	None	Expansion of the training data	0.31[15]
Convolutional neural network	6-layer 784-50-100-500-1000-10-10	None	Expansion of the training data	0.27[23]
Convolutional neural network	Committee of 35 CNNs, 1-20-P-40-P-150-10	elastic distortions	Width normalizations	0.23[8]
Convolutional neural network	Committee of 5 CNNs, 6-layer 784-50-100-500-1000-10-10	None	Expansion of the training data	0.21[17]

https://en.wikipedia.org/wiki/MNIST_database

Case #1: OCR

Machines can be fooled!

Deep Neural Networks are Easily Fooled:
High Confidence Predictions for Unrecognizable Images
https://arxiv.org/abs/1412.1897

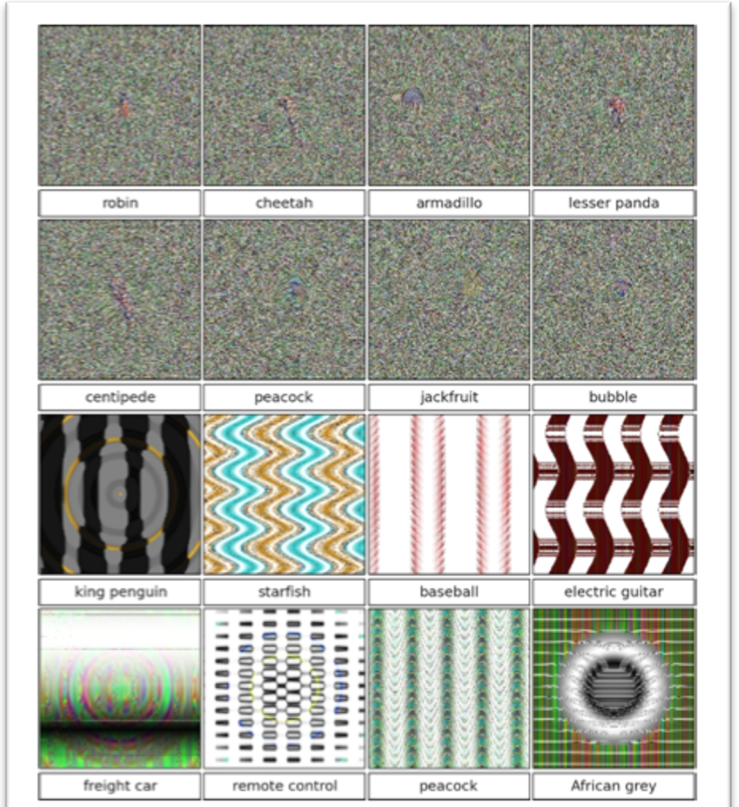
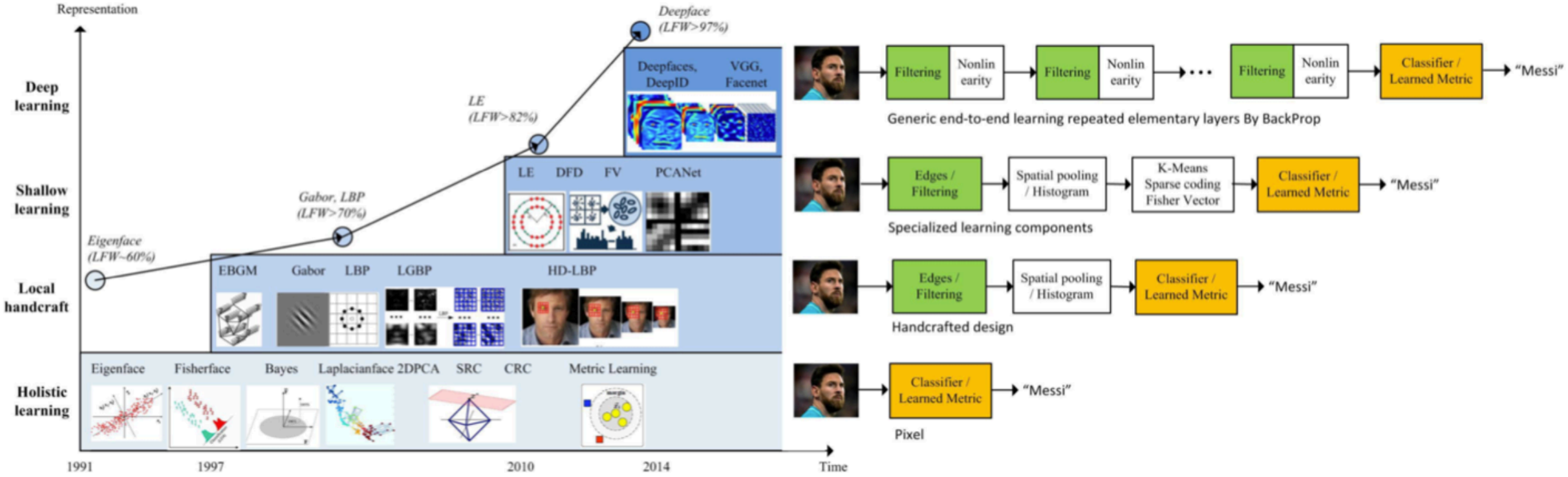


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

Case #2: Computer Vision

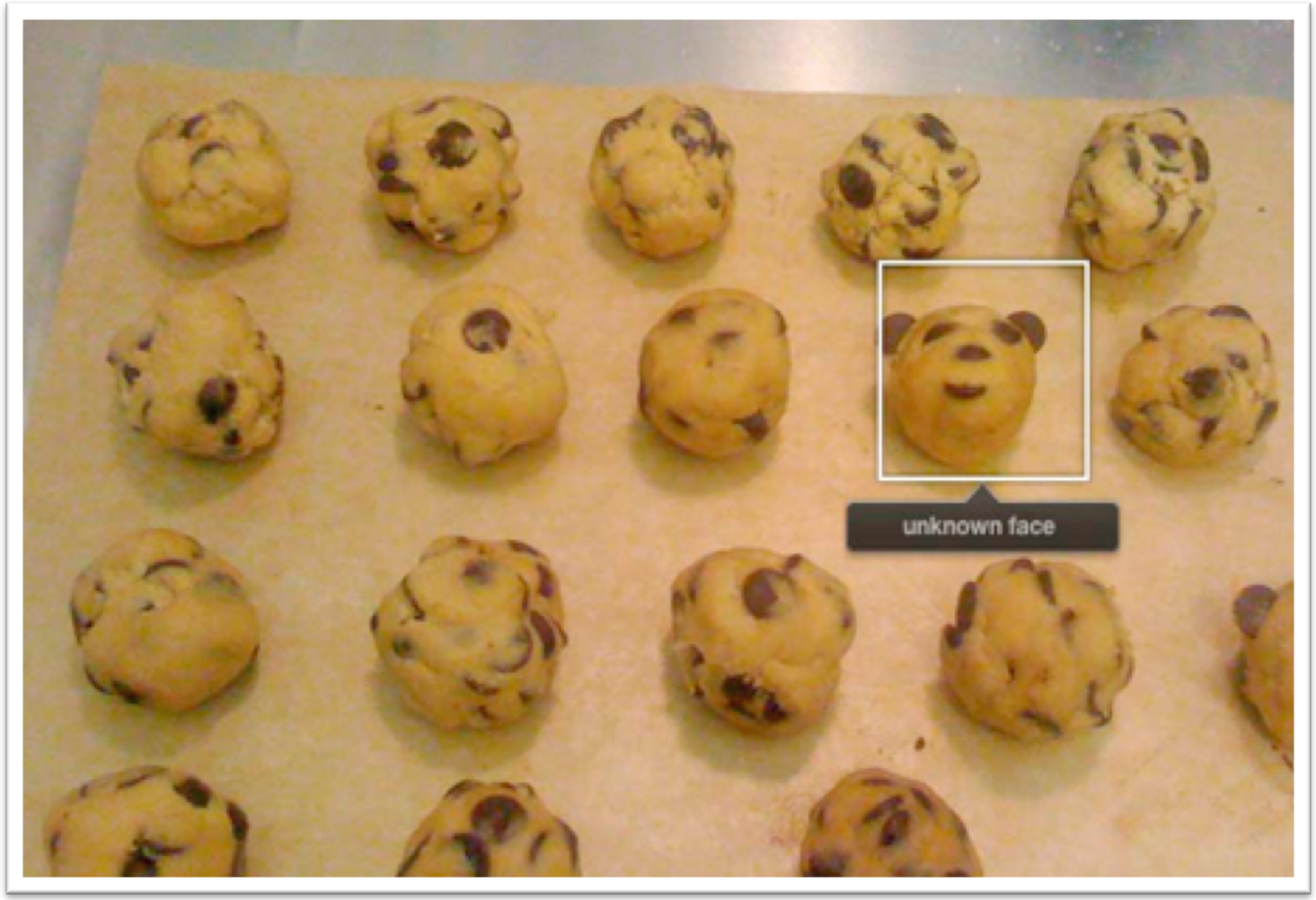
Case #2: Computer Vision



Deep Face: 97.35% vs **Human**: 97.53%

https://arxiv.org/pdf/1804.06655.pdf

Case #2: Computer Vision



Case #3: Image Captioning

"Two pizzas on a stove with wine."

"Three men playing frisbee in the grass"

Case #3: Image Captioning

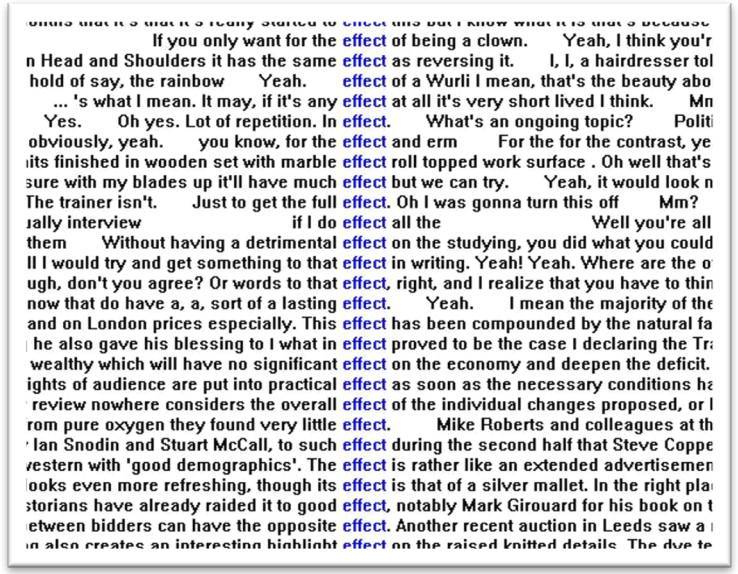
"A refrigerator filled with lots of food and drinks.

"A yellow school bus".

Case #4: Games

- **March 2016**: AlphaGo defeats Lee Sedol.
  - "AlphaGo can't beat me." - Ke Jie (World Champion)
- **May 2017**: AlphaGo Master defeats Ke Jie
  - "Last year, AlphaGo was still quite humanlike when it played. But this year, it has became like a god of Go".
- **Oct 2017**: AlphaGo Zero outperforms AlphaGo Master.
  - Key Point: No prior training based on human expertise.

Case #5: Text Generation

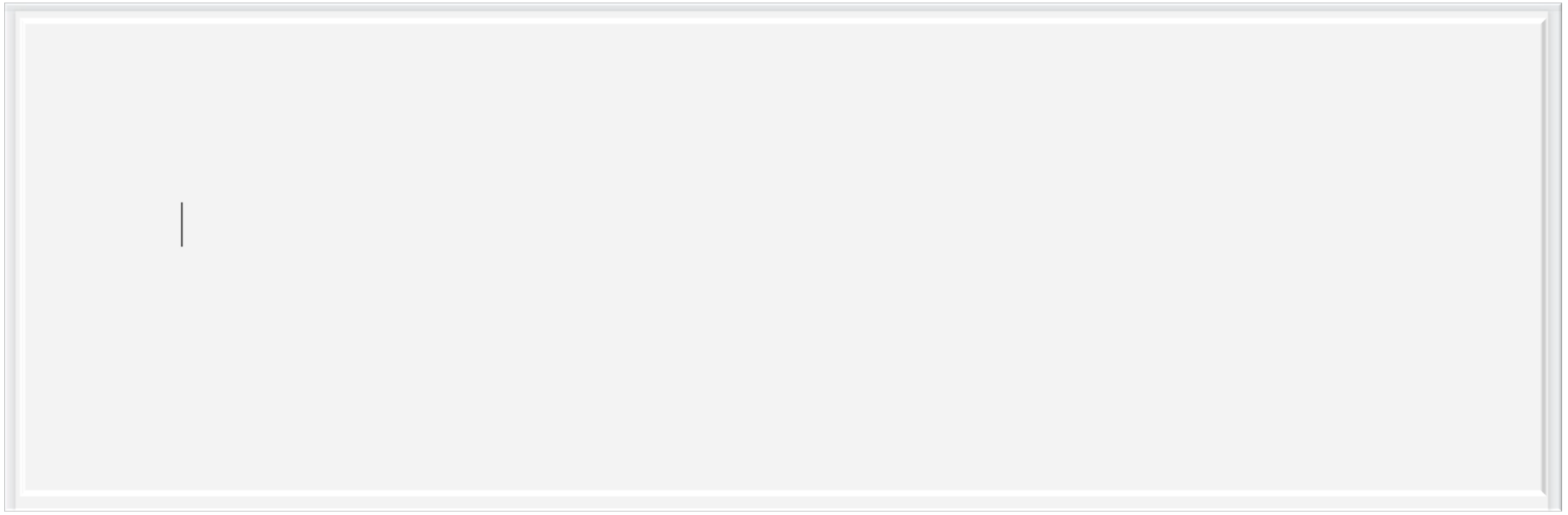


Text Corpus

$$P(W_n|W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})}$$

A Statistical Model of Language

Case #5: Text Generation



General Pre-Trained Transformer-2 (GPT-2)

This example uses arXiv-NLP's training set.

Try it here: https://transformer.huggingface.co/doc/arxiv-nlp

Case #5: Text Generation

```
// Here are the 2 description:code pairs used to give GPT-3
some context for how to provide a response

// sample 1
description: a red button that says stop
code: <button style={{color: 'white', backgroundColor:
'red'}}>Stop</button>

//sample 2
description: a blue box that contains 3 yellow circles with
red borders
code: <div style={{backgroundColor: 'blue', padding: 20}}><div
style={{backgroundColor: 'yellow', border: '5px solid red',
borderRadius: '50%', padding: 20, width: 100, height: 100}}>
</div><div style={{backgroundColor: 'yellow', borderWidth: 1,
border: '5px solid red', borderRadius: '50%', padding: 20,
width: 100, height: 100}}></div><div style={{backgroundColor:
'yellow', border: '5px solid red', borderRadius: '50%',
padding: 20, width: 100, height: 100}}></div></div>
```

carbon
carbon.now.sh

Writing HTML + CSS
… via text-commands?

GPT-3: Text Understanding

OpenAI. Beta, Summer 2020.
(Not available to the public.)

Case #5: Text Generation

Qui Gon Jinn to Jar Jar Binks.
(32 BBY)

Problem: Machine learning hinges on prior data.

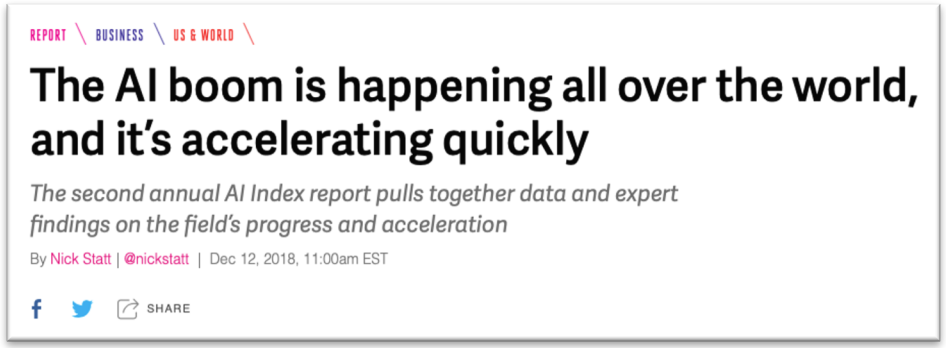
Grand Challenges

Today's Machine Learning

Machine Learning is Modern Computer Science

- Productivity Tools (e.g. Microsoft Word)
- Well-Being Toos (e.g. Woebot)
- Fraud Detection (e.g. CapitalOne, etc)
- Speech Recognition (e.g. "Hey Google")

…


REPORT \ BUSINESS \ US & WORLD \

The AI boom is happening all over the world, and it's accelerating quickly

The second annual AI Index report pulls together data and expert findings on the field's progress and acceleration

By Nick Statt | @nickstatt | Dec 12, 2018, 11:00am EST

SHARE

Why is Machine Learning Everywhere?

- Sensing + Devices → Explosion of Data
- Hardware Advances → Explosion of Processing Capabilities
- Democratized ML → Explosion of Resources, Frameworks, etc
- The Era of AI→ Companies, investors, start-ups, etc.

Grand Challenge #1: Data

$O(n^2)$ algorithms are infeasible.

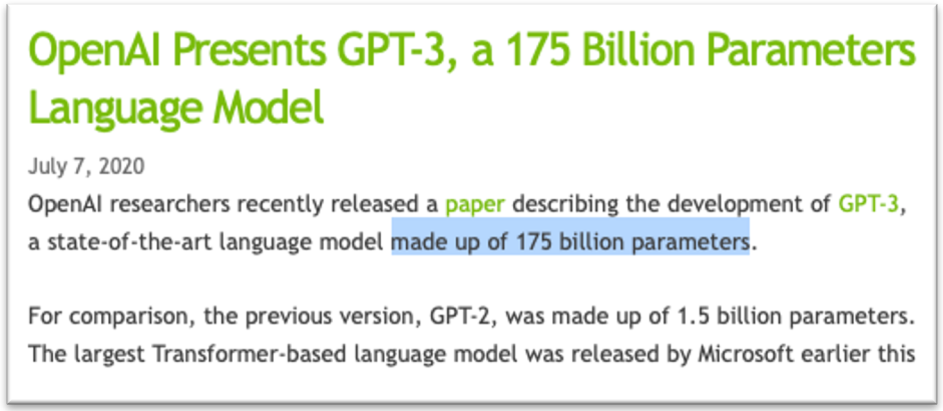
- ML has largely ignored algorithmic complexity.

A Need for Democratized Supercomputers.

- New techniques for processing large datasets.

A Need for Parallelization.

- Existing systems generally parallelize poorly (if at all).



OpenAI Presents GPT-3, a 175 Billion Parameters Language Model

July 7, 2020

OpenAI researchers recently released a paper describing the development of GPT-3, a state-of-the-art language model made up of 175 billion parameters.

For comparison, the previous version, GPT-2, was made up of 1.5 billion parameters. The largest Transformer-based language model was released by Microsoft earlier this

Grand Challenge #2: End-to-End Learning

The ML pipeline is substantial.

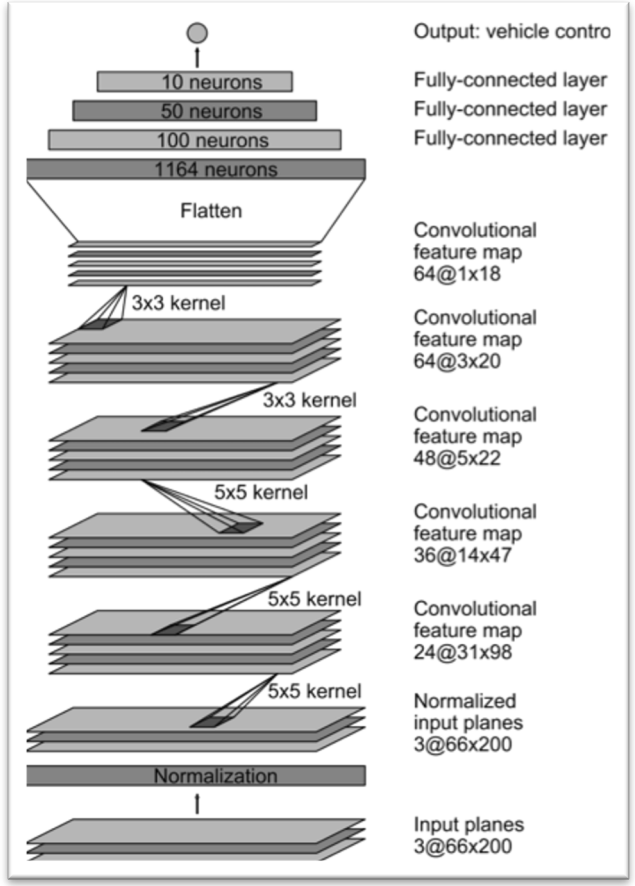
- Efforts to streamline learning.

Single characters → Text Classification

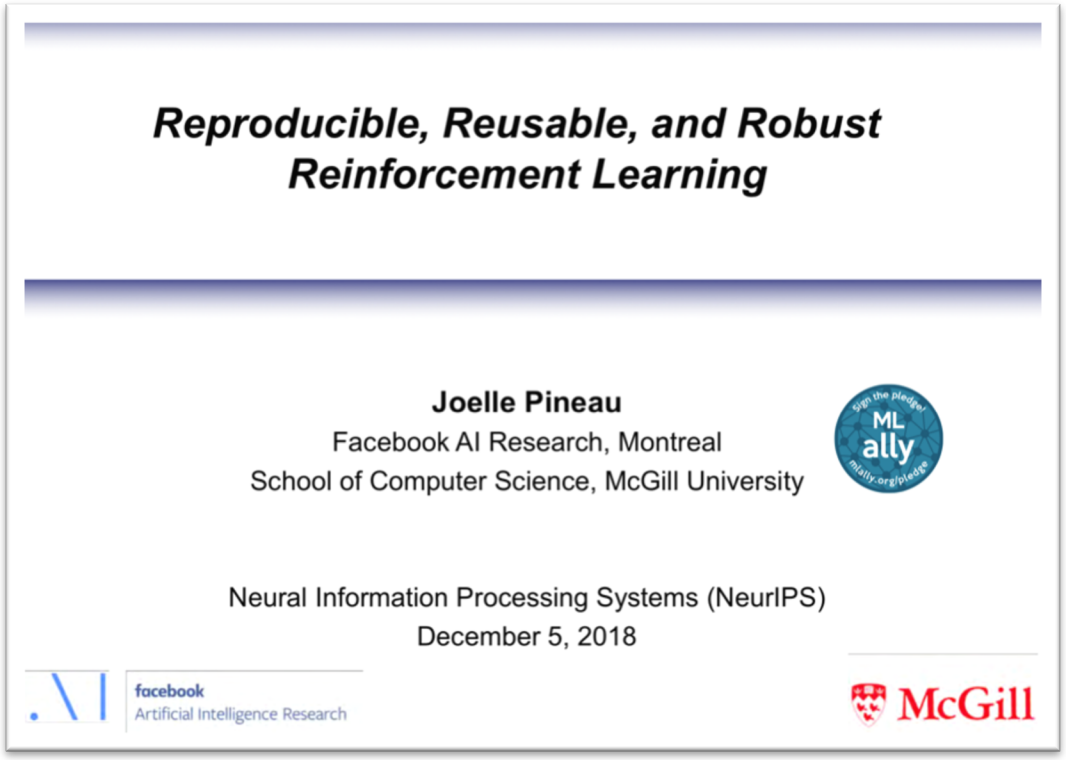
- https://arxiv.org/abs/1509.01626

Pixels → Autonomous Steering

- https://arxiv.org/pdf/1604.07316v1.pdf

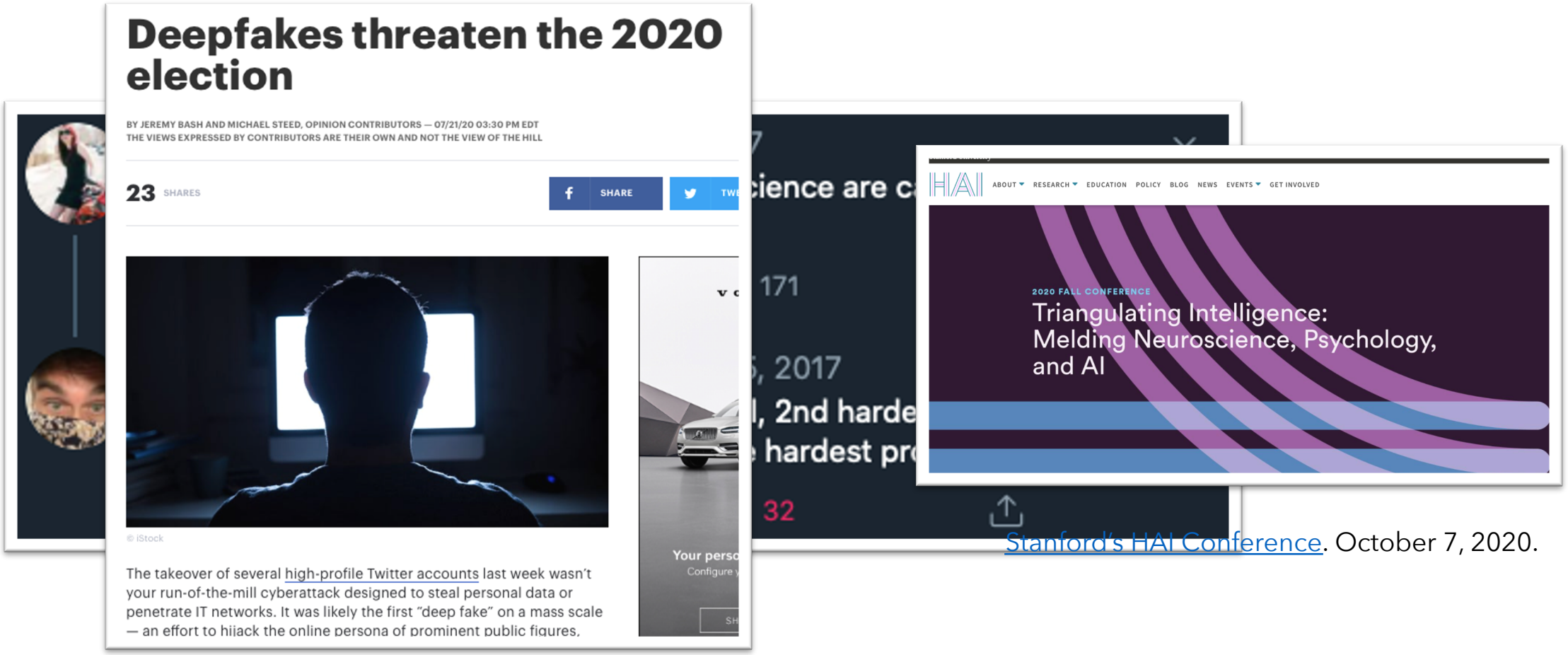


Grand Challenge #3: ML Research



https://www.youtube.com/watch?v=-0G98MYUtjI

Grand Challenge #4: People



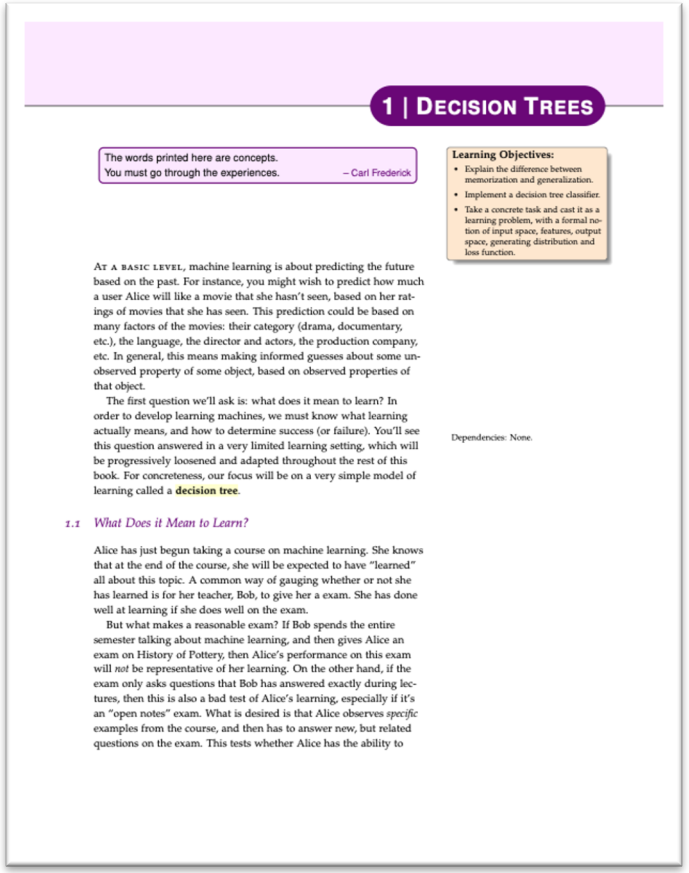
Stanford's HAI Conference. October 7, 2020.

Today's Agenda

You should now have answers to:

1. What is "*Machine Learning*" (ML)?
2. How is ML operationalized?
3. What are the grand challenge of modern ML?

Reading



1 | DECISION TREES

The words printed here are concepts.
You must go through the experiences. – Carl Frederick

Learning Objectives:
- Explain the difference between memorization and generalization.
- Implement a decision tree classifier.
- Take a concrete task and cast it as a learning problem, with a formal notion of input space, features, output space, generating distribution and loss function.

Dependencies: None.

AT A BASIC LEVEL, machine learning is about predicting the future based on the past. For instance, you might wish to predict how much a user Alice will like a movie that she hasn't seen, based on her ratings of movies that she has seen. This prediction could be based on many factors of the movies: their category (drama, documentary, etc.), the language, the director and actors, the production company, etc. In general, this means making informed guesses about some unobserved property of some object, based on observed properties of that object.

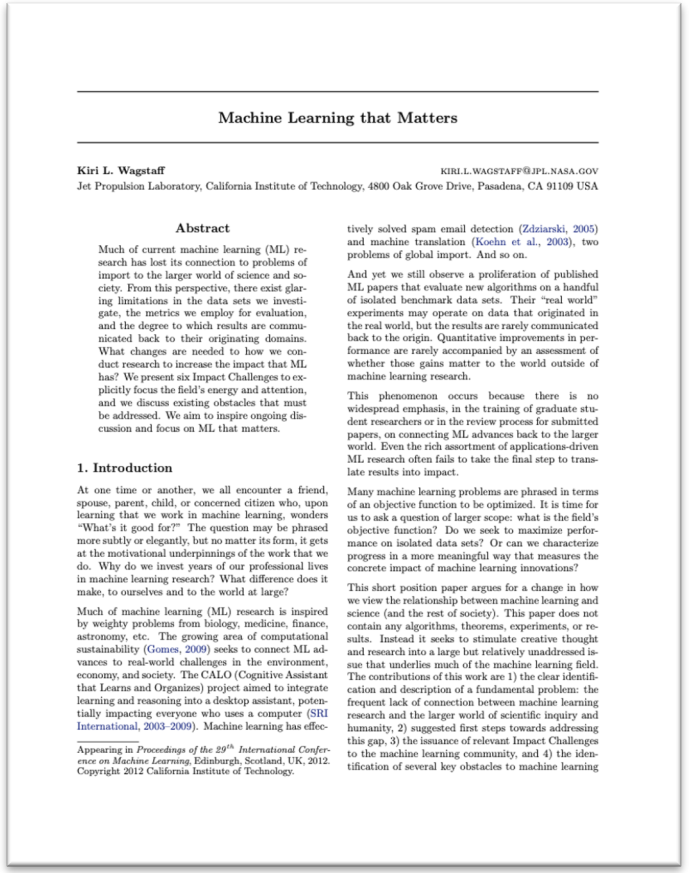
The first question we'll ask is: what does it mean to learn? In order to develop learning machines, we must know what learning actually means, and how to determine success (or failure). You'll see this question answered in a very limited learning setting, which will be progressively loosened and adapted throughout the rest of this book. For concreteness, our focus will be on a very simple model of learning called a **decision tree**.

1.1 What Does it Mean to Learn?

Alice has just begun taking a course on machine learning. She knows that at the end of the course, she will be expected to have "learned" all about this topic. A common way of gauging whether or not she has learned is for her teacher, Bob, to give her a exam. She has done well at learning if she does well on the exam.

But what makes a reasonable exam? If Bob spends the entire semester talking about machine learning, and then gives Alice an exam on History of Pottery, then Alice's performance on this exam will *not* be representative of her learning. On the other hand, if the exam only asks questions that Bob has answered exactly during lectures, then this is also a bad test of Alice's learning, especially if it's an "open notes" exam. What is desired is that Alice observes *specific* examples from the course, and then has to answer new, but related questions on the exam. This tests whether Alice has the ability to

Daume. Sec 1.1 + 1.2



Machine Learning that Matters

Kiri L. Wagstaff KIRI.L.WAGSTAFF@JPL.NASA.GOV
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109 USA

Abstract

Much of current machine learning (ML) research has lost its connection to problems of import to the larger world of science and society. From this perspective, there exist glaring limitations in the data sets we investigate, the metrics we employ for evaluation, and the degree to which results are communicated back to their originating domains. What changes are needed to how we conduct research to increase the impact that ML has? We present six Impact Challenges to explicitly focus the field's energy and attention, and we discuss existing obstacles that must be addressed. We aim to inspire ongoing discussion and focus on ML that matters.

1. Introduction

At one time or another, we all encounter a friend, spouse, parent, child, or concerned citizen who, upon learning that we work in machine learning, wonders "What's it good for?" The question may be phrased more subtly or elegantly, but no matter its form, it gets at the motivational underpinnings of the work that we do. Why do we invest years of our professional lives in machine learning research? What difference does it make, to ourselves and to the world at large?

Much of machine learning (ML) research is inspired by weighty problems from biology, medicine, finance, astronomy, etc. The growing area of computational sustainability (Gomes, 2009) seeks to connect ML advances to real-world challenges in the environment, economy, and society. The CALO (Cognitive Assistant that Learns and Organizes) project aimed to integrate learning and reasoning into a desktop assistant, potentially impacting everyone who uses a computer (SRI International, 2003–2009). Machine learning has effectively solved spam email detection (Zdziarski, 2005) and machine translation (Koehn et al., 2003), two problems of global import. And so on.

And yet we still observe a proliferation of published ML papers that evaluate new algorithms on a handful of isolated benchmark data sets. Their "real world" experiments may operate on data that originated in the real world, but the results are rarely communicated back to the origin. Quantitative improvements in performance are rarely accompanied by an assessment of whether those gains matter to the world outside of machine learning research.

This phenomenon occurs because there is no widespread emphasis, in the training of graduate student researchers or in the review process for submitted papers, on connecting ML advances back to the larger world. Even the rich assortment of applications-driven ML research often fails to take the final step to translate results into impact.

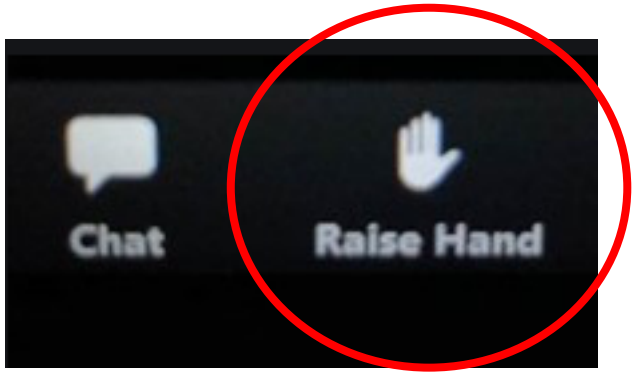
Many machine learning problems are phrased in terms of an objective function to be optimized. It is time for us to ask a question of larger scope: what is the field's objective function? Do we seek to maximize performance on isolated data sets? Or can we characterize progress in a more meaningful way that measures the concrete impact of machine learning innovations?

This short position paper argues for a change in how we view the relationship between machine learning and science (and the rest of society). This paper does not contain any algorithms, theorems, experiments, or results. Instead it seeks to stimulate creative thought and research into a large but relatively unaddressed issue that underlies much of the machine learning field. The contributions of this work are 1) the clear identification and description of a fundamental problem: the frequent lack of connection between machine learning research and the larger world of scientific inquiry and humanity, 2) suggested first steps towards addressing this gap, 3) the issuance of relevant Impact Challenges to the machine learning community, and 4) the identification of several key obstacles to machine learning

Appearing in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.
Copyright 2012 California Institute of Technology.

Wagstaff. All of it!

Any Questions?



Use Zoom's "*Raise Hand*" feature, and I'll un-mute you.

Next Week
** All Asynchronous **

August 24th: *Decision Trees*
August 26th: *Decision Trees (continued)*
August 28th: *The Limits of Learning*