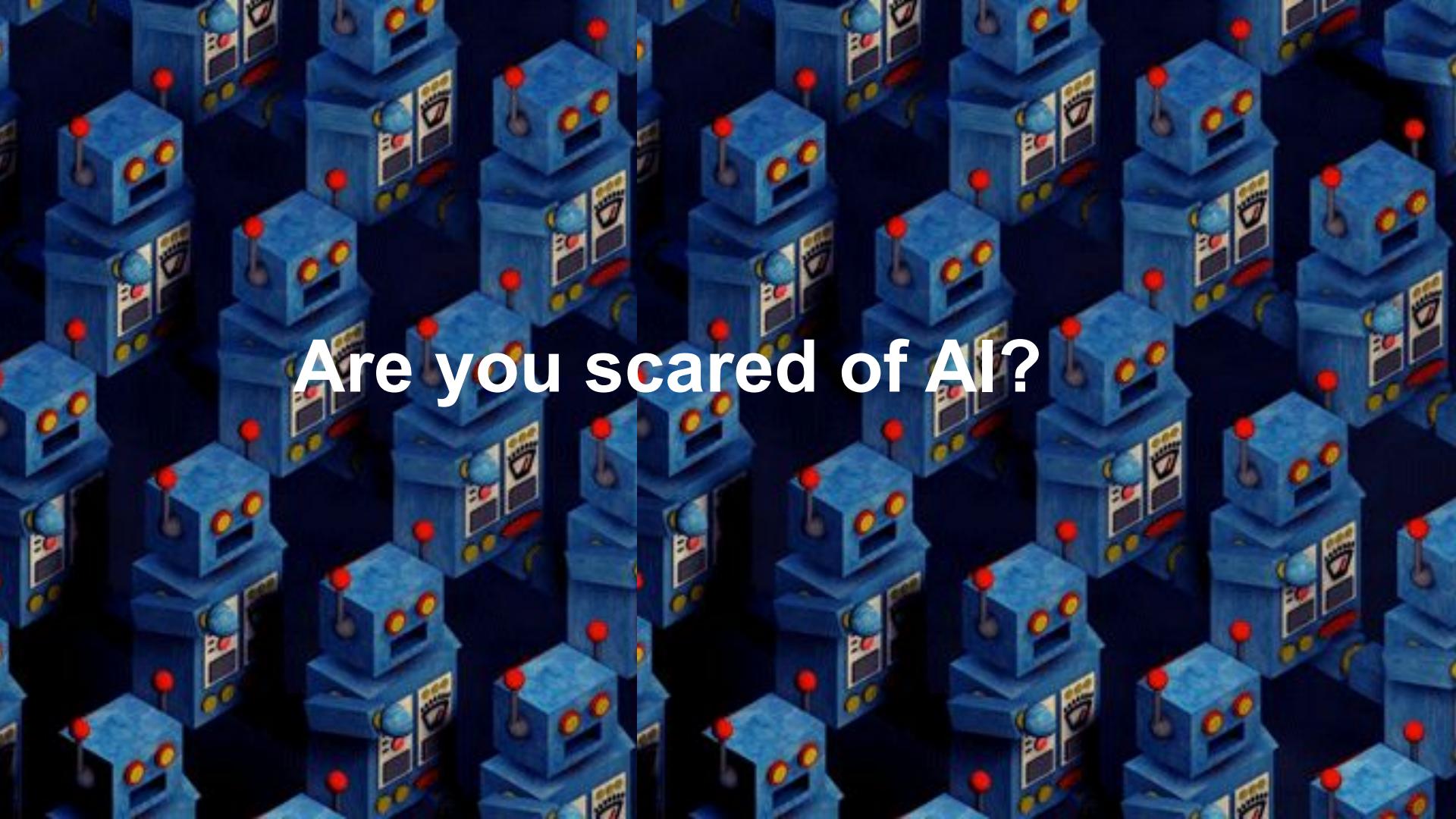


“AI” in Medicine

John R. Zech, M.D., M.A.
PGY-1 Prelim Medicine, CPMC
4/9/2019

I have no relevant financial conflicts of interest.



Are you scared of AI?

The rise of the machines



<https://www.youtube.com/watch?v=2HMPRXstSvQ>

when do we need to start cutting radiology residency spots?

Discussion in 'Radiology' started by IRrads10, Thursday at 2:05 PM.

[Previous Thread](#) [Next Thread](#)



IRrads10

2+ Year Member

Joined: Jul 20, 2016
Messages: 23
Likes Received: 9
Status: Medical Student

Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists

AI in radiology is progressing swiftly. it is obviously hard to predict how much of an impact these technologies will have.

should radiology's future start to look bleak in the coming decade, i sure hope we are proactive and not reactive.

IRrads10, Thursday at 2:05 PM

#1

Gurby likes this.



Mo991

Joined: Aug 2, 2018
Messages: 12
Likes Received: 11

This is really disturbing
I'm starting to think I made a bad decision..

Mo991, Thursday at 3:57 PM

#3

The rise of the machines

Artificial general intelligence

From Wikipedia, the free encyclopedia

Artificial general intelligence (AGI) is the intelligence of a machine that could successfully perform any intellectual task that a [human](#) being can. It is a primary goal of some [artificial intelligence](#) research and a common topic in [science fiction](#) and [future studies](#). Some researchers refer to Artificial general intelligence as "[strong AI](#)",^[1] "[full AI](#)"^[2] or as the ability of a machine to perform "general intelligent action";^[3] others reserve "strong AI" for machines capable of experiencing [consciousness](#).



The rise of the machines

DeepMind Technologies' goal is to "solve intelligence",[35] which they are trying to achieve by combining "the best techniques from machine learning and systems neuroscience to build powerful general-purpose learning algorithms"



"The search space in Go is vast... a number greater than there are atoms in the universe"

-David Silver, DeepMind

[https://en.wikipedia.org/wiki/Go_\(game\)](https://en.wikipedia.org/wiki/Go_(game))



Go is played on a grid of black lines (usually 19×19). Game pieces, called *stones*, are played on the lines' intersections.

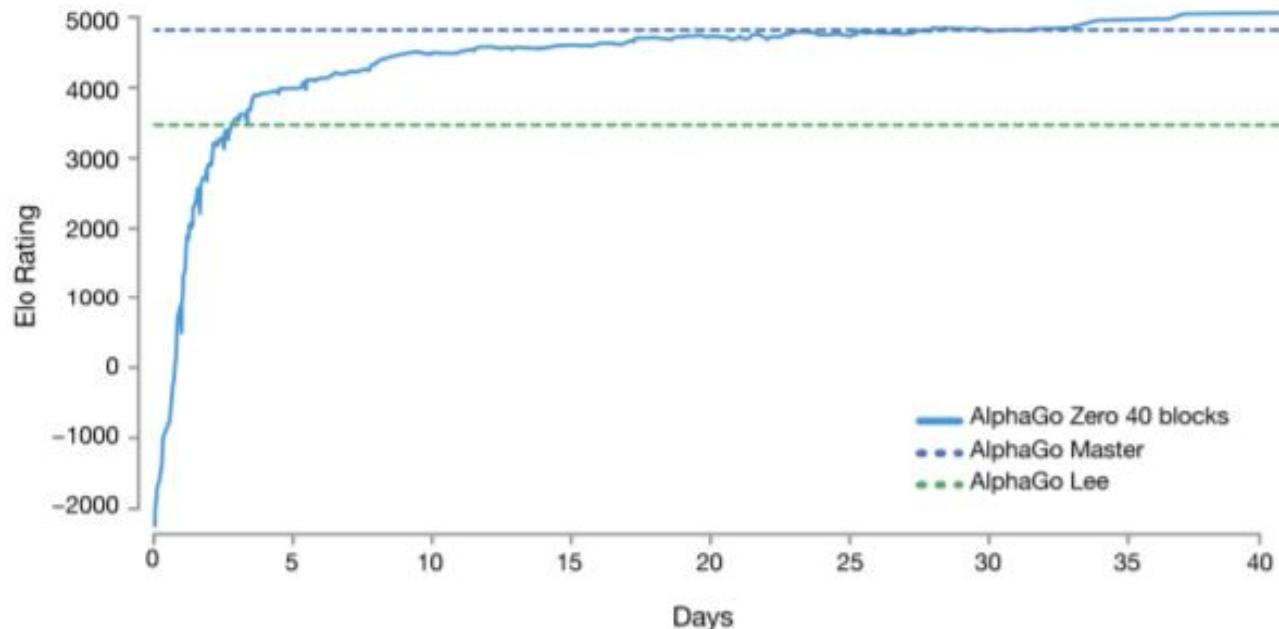
Years active Zhou dynasty (1046–256 BC) to present



https://www.youtube.com/watch?v=8tq1C8spV_g

The rise of the machines

a.



The rise of the machines



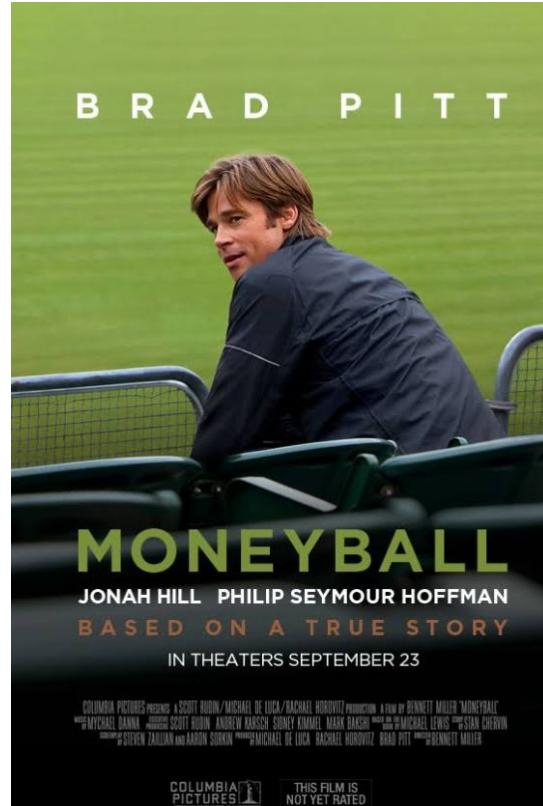
The rise of the machines



But are they truly ‘artificially intelligent’?



Or a (really intriguing) collection of models?



Overview

High level discussion of 3 main divisions of machine learning

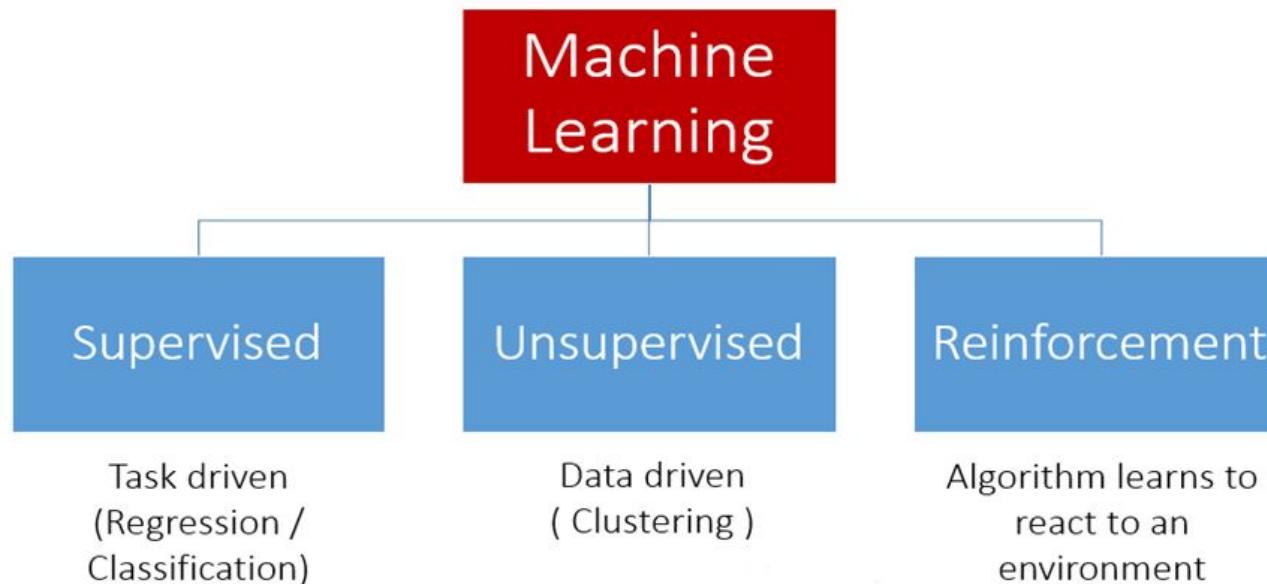
Why deep learning is particularly exciting

Challenges in applying it in medicine

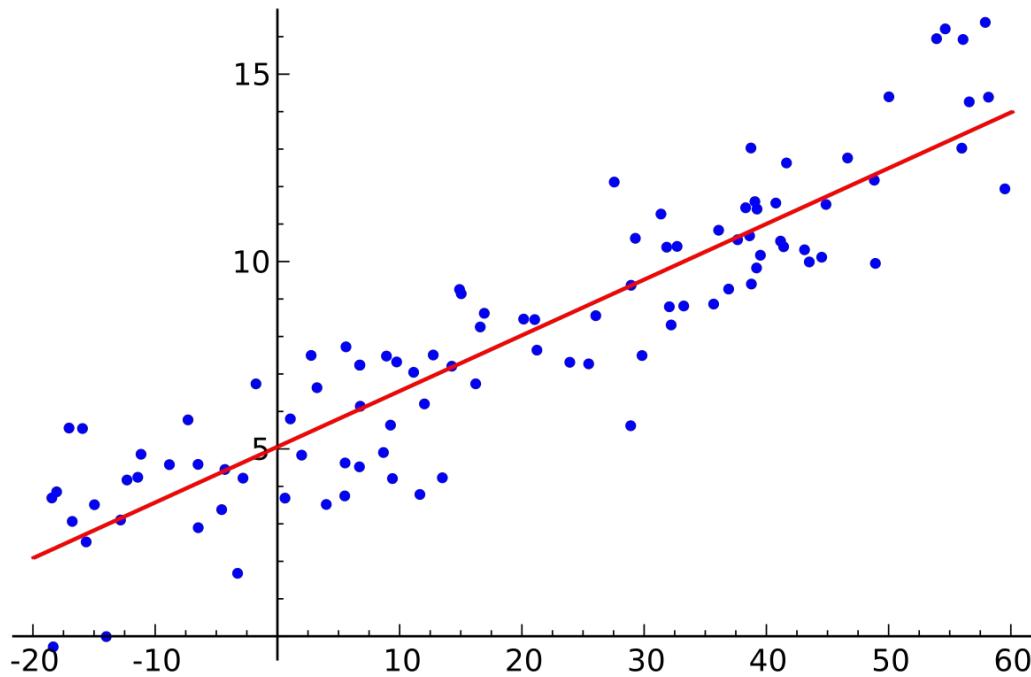
Big things to look out for going forward

What machine learning is (and isn't)

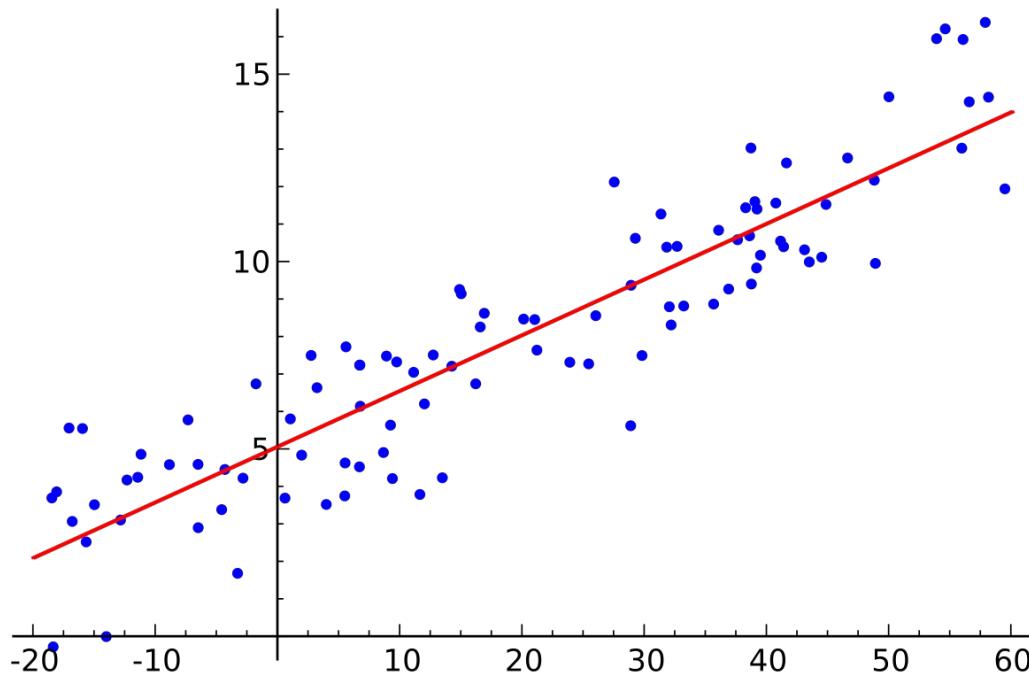
Types of Machine Learning



Supervised

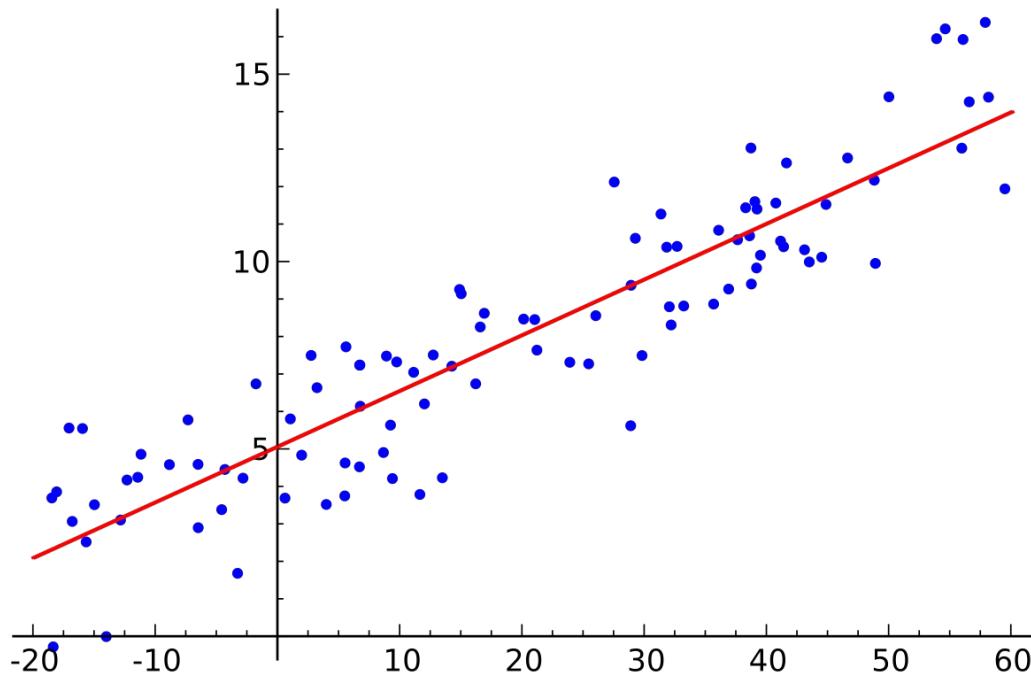


Supervised



You have INPUT data (X)

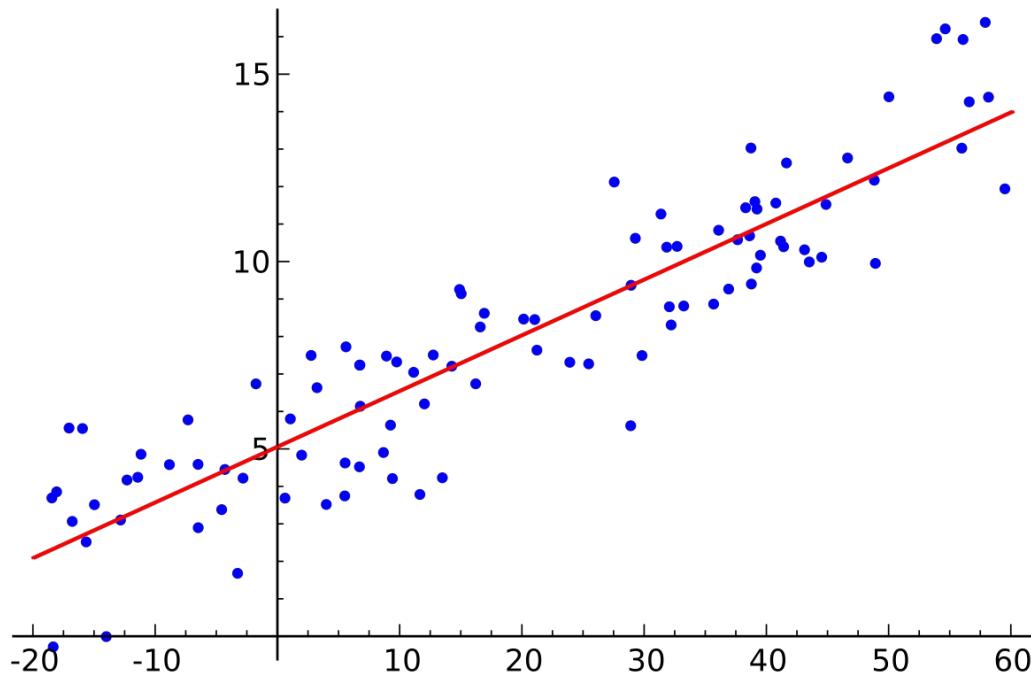
Supervised



You have INPUT data (X)

You have OUTPUT you wish to predict (Y)

Supervised

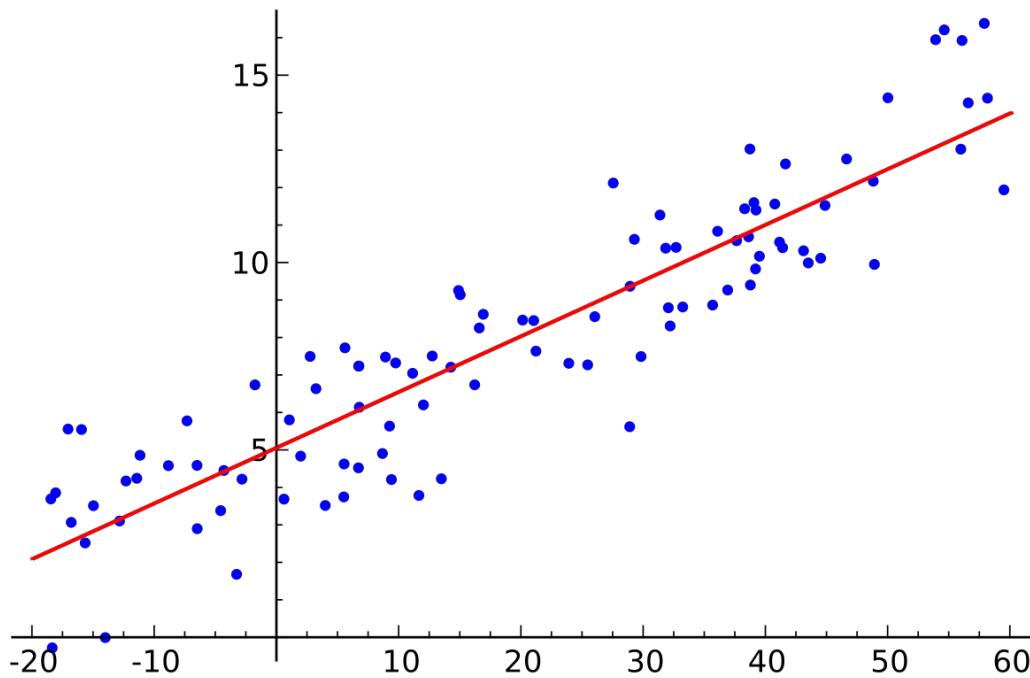


You have INPUT data (X)

You have OUTPUT you wish to predict (Y)

Build a model to predict Y given X

Supervised



You have INPUT data (X)

You have OUTPUT you wish to predict (Y)

Build a model to predict Y given X

In medicine, this is the focus

Supervised

RESPIRATORY INFECTION

Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study

W S Lim, M M van der Eerden, R Laing, W G Boersma, N Karalus, G I Town, S A Lewis,
J T Macfarlane

Thorax 2003;58:377–382

Supervised

RESPIRATORY INFECTION

Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study

W S Lim, M M van der Eerden, R Laing, W G Boersma, N Karalus, G I Town, S A Lewis, J T Macfarlane

Thorax 2003;58:377–382

a.k.a. CURB-65

See end of article for
authors' affiliations

Correspondence to:
Dr J T Macfarlane,
Respiratory Infection
Research Group,
Respiratory Medicine,
Nottingham City Hospital,
Nottingham NG5 1PB, UK;
john.macfarlane@tinyworld.co.uk

Revised version received
2 December 2002
Accepted for publication
13 December 2002

Background: In the assessment of severity in community acquired pneumonia (CAP), the modified British Thoracic Society (mBTS) rule identifies patients with severe pneumonia but not patients who might be suitable for home management. A multicentre study was conducted to derive and validate a practical severity assessment model for stratifying adults hospitalised with CAP into different management groups.

Methods: Data from three prospective studies of CAP conducted in the UK, New Zealand, and the Netherlands were combined. A derivation cohort comprising 80% of the data was used to develop the model. Prognostic variables were identified using multiple logistic regression with 30 day mortality as the outcome measure. The final model was tested against the validation cohort.

Results: 1068 patients were studied (mean age 64 years, 51.5% male, 30 day mortality 9%). Age ≥ 65 years (OR 3.5, 95% CI 1.6 to 8.0) and albumin < 30 g/dl (OR 4.7, 95% CI 2.5 to 8.7) were independently associated with mortality over and above the mBTS rule (OR 5.2, 95% CI 2.7 to 10). A six point score, one point for each of Confusion, Urea > 7 mmol/l, Respiratory rate ≥ 30 /min, low systolic(< 90 mm Hg) or diastolic (≤ 60 mm Hg) Blood pressure, age ≥ 65 years (CURB-65 score) based on information available at initial hospital assessment, enabled patients to be stratified according to increasing risk of mortality: score 0, 0.7%; score 1, 3.2%; score 2, 3%; score 3, 17%; score 4, 41.5% and score 5, 57%. The validation cohort confirmed a similar pattern.

Conclusions: A simple six point score based on confusion, urea, respiratory rate, blood pressure, and age can be used to stratify patients with CAP into different management groups.

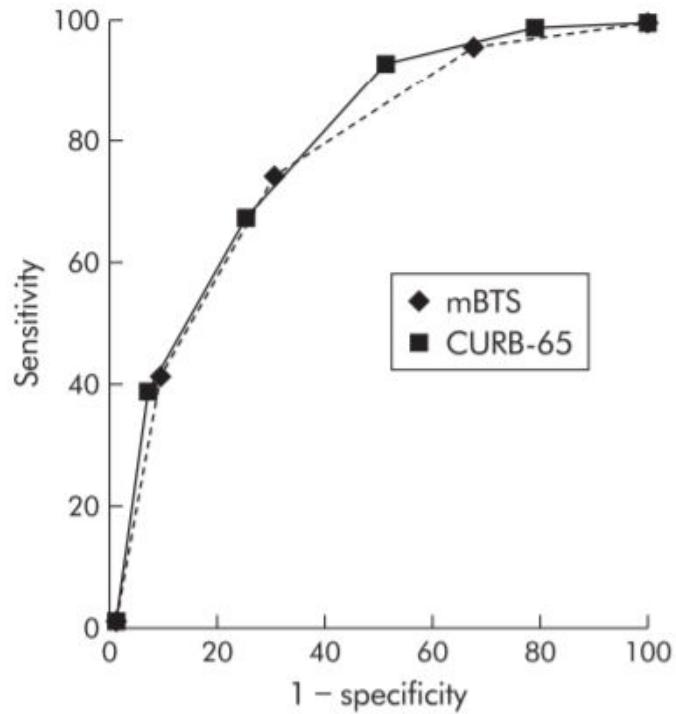


Figure 1 Receiver operating curves for modified BTS (mBTS) rule and CURB-65 score.

Unsupervised

You have INPUT data (X)

Unsupervised

You have INPUT data (X)

You DON'T have OUTPUT (no Y)

Unsupervised

You have INPUT data (X)

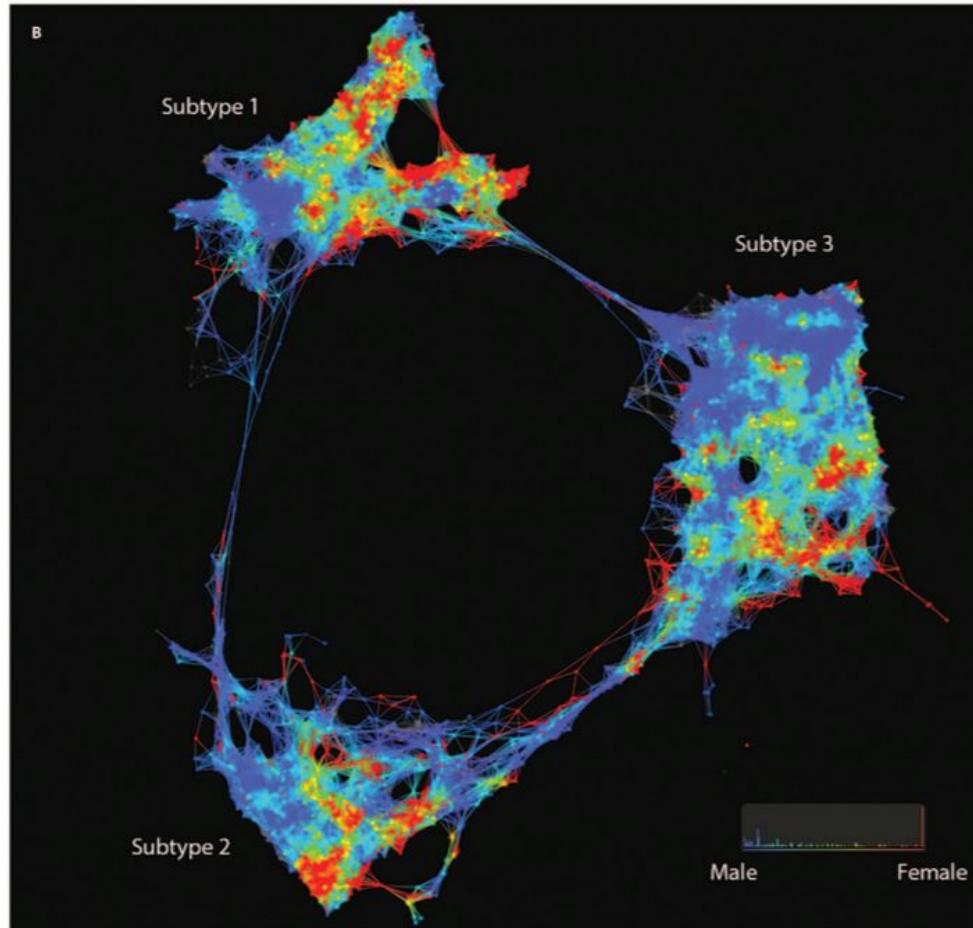
You DON'T have OUTPUT (no Y)

**Try to learn something about the
structure of X (i.e., clusters)**

Identification of type 2 diabetes subgroups through topological analysis of patient similarity

Li Li,¹ Wei-Yi Cheng,¹ Benjamin S. Glicksberg,¹ Omri Gottesman,² Ronald Tamler,³ Rong Chen,¹ Erwin P. Bottinger,² Joel T. Dudley^{1,4*}

Type 2 diabetes (T2D) is a heterogeneous complex disease affecting more than 29 million Americans alone with a rising prevalence trending toward steady increases in the coming decades. Thus, there is a pressing clinical need to improve early prevention and clinical management of T2D and its complications. Clinicians have understood that patients who carry the T2D diagnosis have a variety of phenotypes and susceptibilities to diabetes-related complications. We used a precision medicine approach to characterize the complexity of T2D patient populations based on high-dimensional electronic medical records (EMRs) and genotype data from 11,210 individuals. We successfully identified three distinct subgroups of T2D from topology-based patient-patient networks. Subtype 1 was characterized by T2D complications diabetic nephropathy and diabetic retinopathy; subtype 2 was enriched for cancer malignancy and cardiovascular diseases; and subtype 3 was associated most strongly with cardiovascular diseases, neurological diseases, allergies, and HIV infections. We performed a genetic association analysis of the emergent T2D subtypes to identify subtype-specific genetic markers and identified 1279, 1227, and 1338 single-nucleotide polymorphisms (SNPs) that mapped to 425, 322, and 437 unique genes specific to subtypes 1, 2, and 3, respectively. By assessing the human disease–SNP association for each subtype, the enriched phenotypes and biological functions at the gene level for each subtype matched with the disease comorbidities and clinical differences that we identified through EMRs. Our approach demonstrates the utility of applying the precision medicine paradigm in T2D and the promise of extending the approach to the study of other complex, multifactorial diseases.



Reinforcement Learning

You have an environment

Reinforcement Learning

You have an environment

You have actions

Reinforcement Learning

You have an environment

You have actions

**Each action has an immediate
reward and changes the
environment**

Reinforcement Learning

You have an environment

You have actions

**Each action has an immediate
reward and changes the
environment**

Maximize cumulative reward

Reinforcement Learning

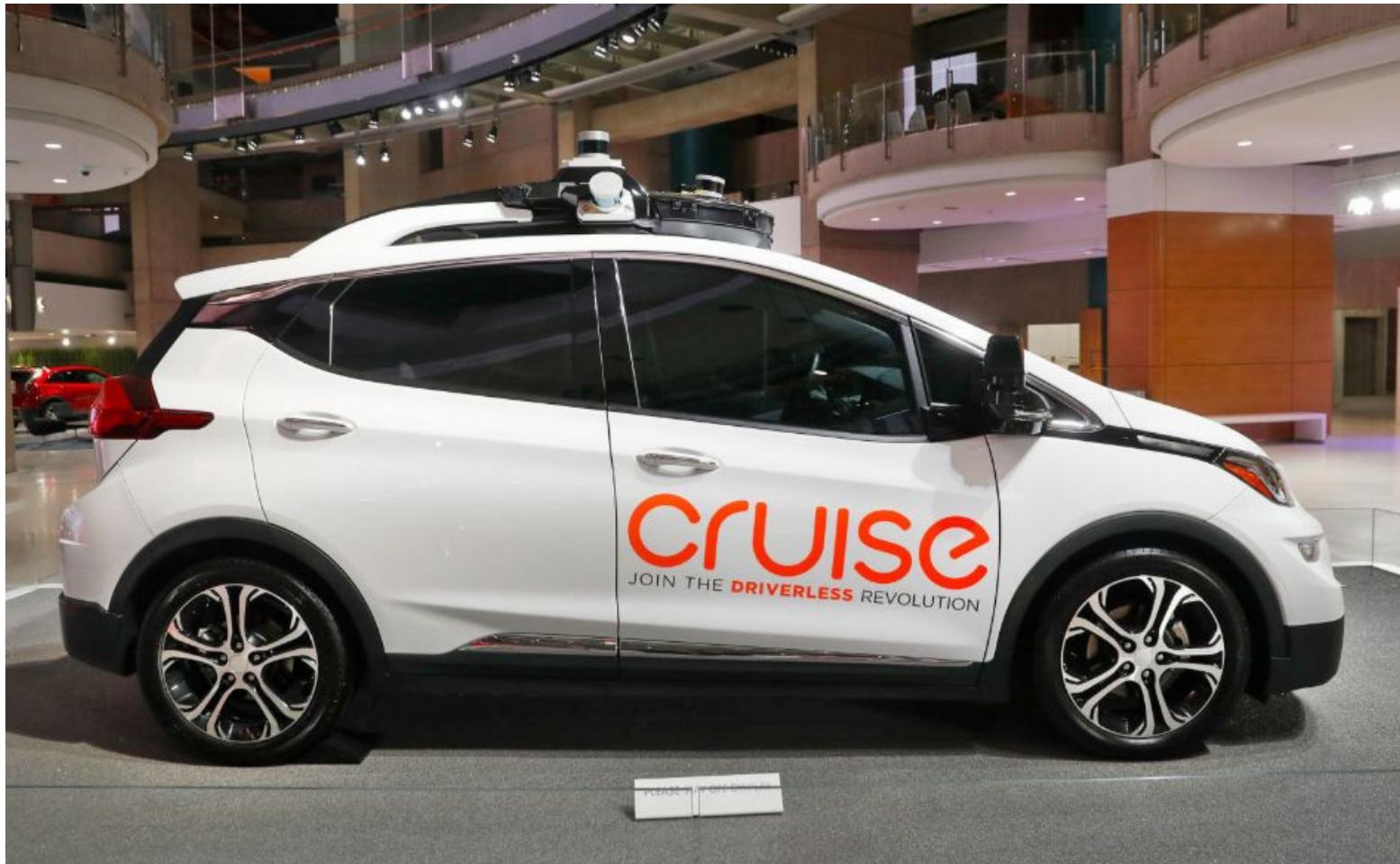
You have an environment

You have actions

Each action has an immediate reward and changes the environment

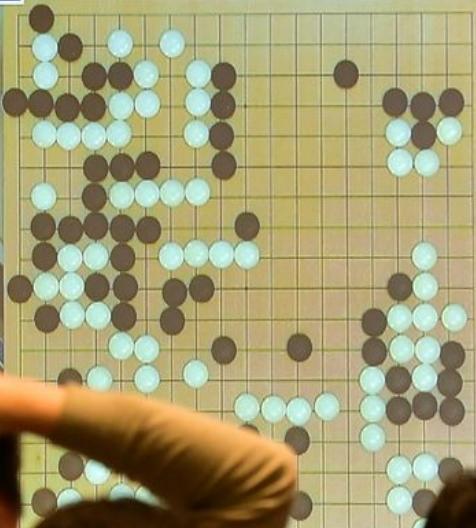
Maximize cumulative reward







ALPHAGO
00:33:46



Google DeepMind
Challenge Match

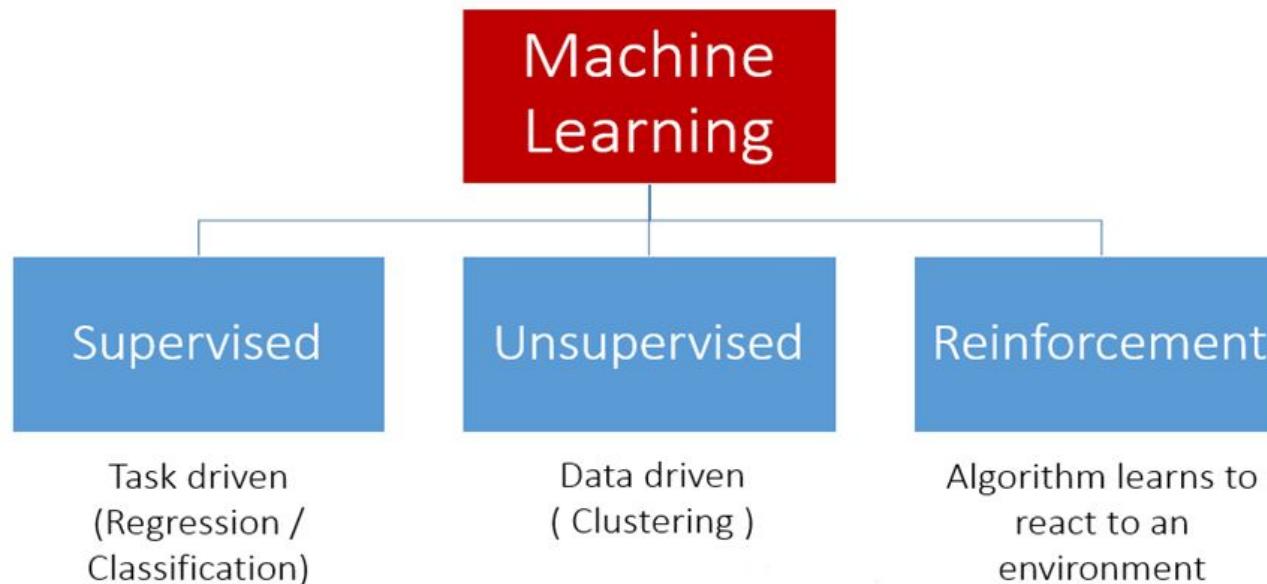


0:



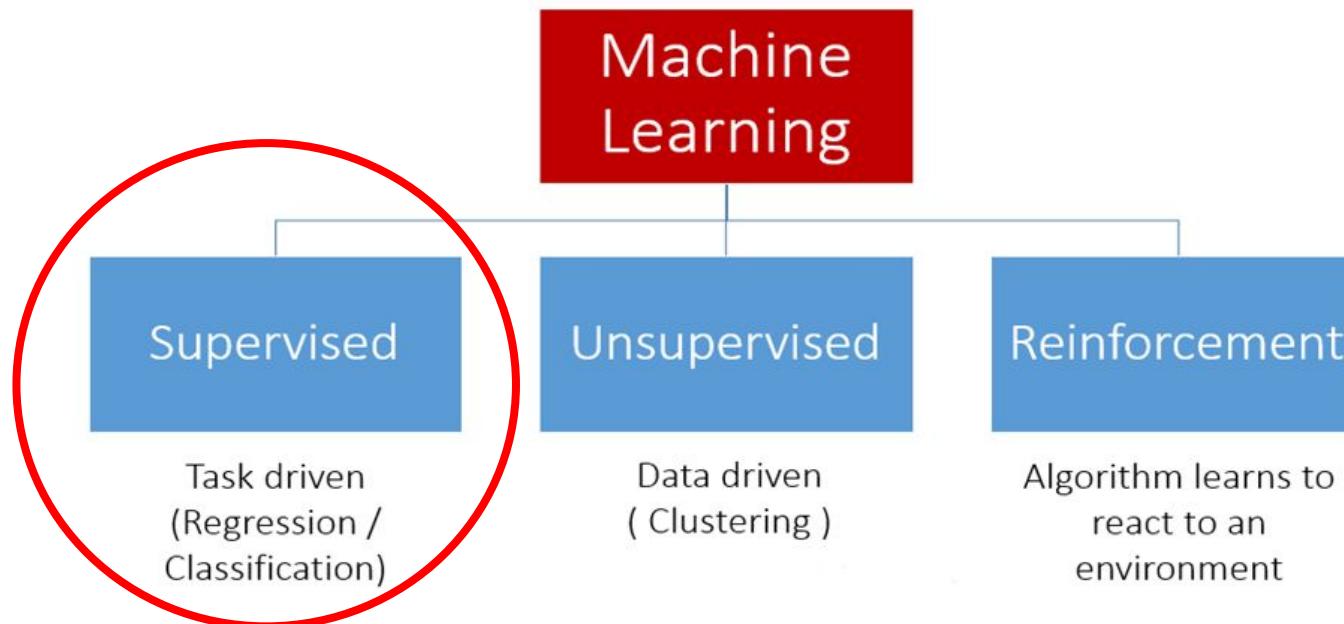
What machine learning is (and isn't)

Types of Machine Learning



What machine learning is (and isn't)

Types of Machine Learning

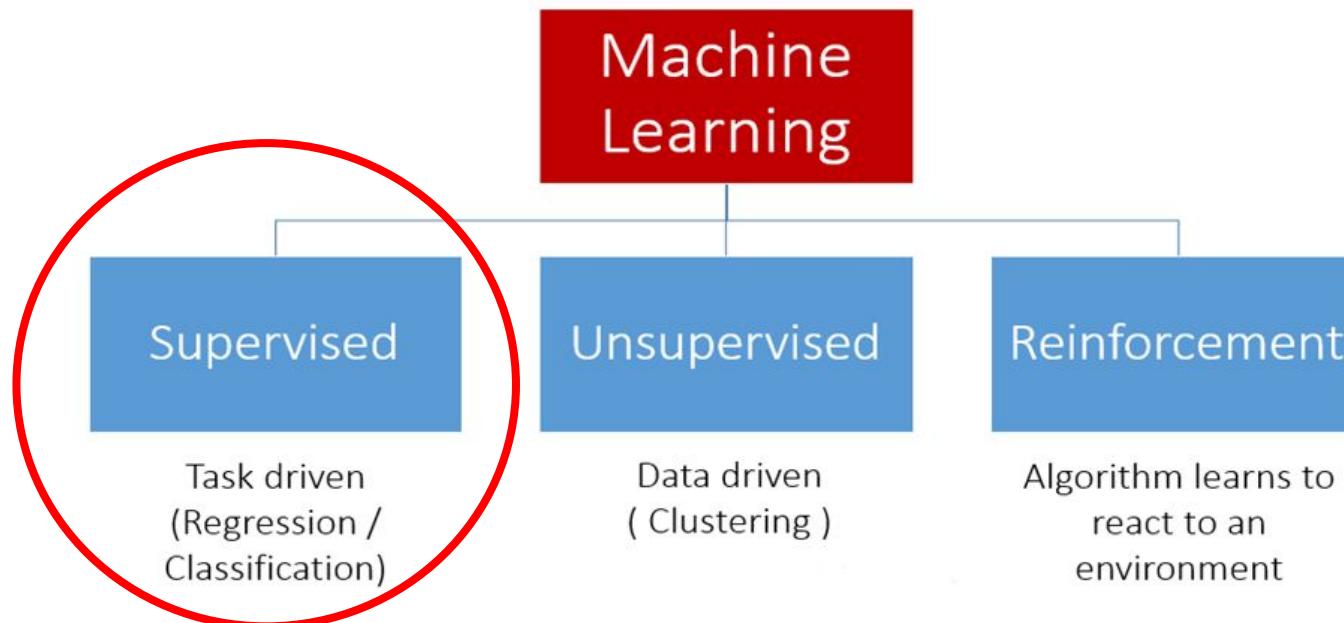


Why focus on supervised learning?

- **Unsupervised learning:** can lead to insights that drive further research, but not obvious how to use for clinical decision support
- **Reinforcement learning:** AlphaGo's success was powered by experience - experimenting with many strategies over self-play against itself for many human lifetimes. In healthcare, we don't have a realistic simulator of a patient / hospital we could use as a training environment.

What machine learning is (and isn't)

Types of Machine Learning



Linear Regression

Linear Regression

$$\mathbf{X} = \begin{pmatrix} x_1^0 & x_1^1 & \dots & x_1^n \\ x_2^0 & x_2^1 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ x_m^0 & x_m^1 & \dots & x_m^n \end{pmatrix} \quad \hat{\mathbf{c}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \hat{\mathbf{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

where $n < m$ and at least $n + 1$ of the x_i are unique

Think about the following operations purely from an algebraic standpoint
(except for the T, that is, the transpose...)

$$\mathbf{X} \hat{\mathbf{c}} = \hat{\mathbf{y}} \xrightarrow{\text{multiply both sides by } \mathbf{X}^T} \mathbf{X}^T \mathbf{X} \hat{\mathbf{c}} = \mathbf{X}^T \hat{\mathbf{y}}$$

$$\xrightarrow{\text{multiply both sides by } (\mathbf{X}^T \mathbf{X})^{-1}} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) \hat{\mathbf{c}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{y}}$$

$(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X})$ is the identity matrix so we get

$$\rightarrow \boxed{\hat{\mathbf{c}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{y}}}$$

Linear Regression

$$X = \begin{pmatrix} x_1^0 & x_1^1 & \dots & x_1^n \\ x_2^0 & \dots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ x_m^0 & \dots & x_m^1 & \dots & x_m^n \end{pmatrix} \quad \hat{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \hat{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

where $n < m$ and at least $n+1$ of the x_i are unique

Think about the following operations purely from an algebraic standpoint
(except for the T, that is, the transpose e...)

$$X \hat{c} = \hat{y} \quad \xrightarrow{\text{multiply both sides by } X^T} \quad X^T X \hat{c} = X^T \hat{y}$$

$$\xrightarrow{\text{multiply both sides by } (X^T X)^{-1}} \quad (X^T X)^{-1}(X^T X) \hat{c} = (X^T X)^{-1} X^T \hat{y}$$

$(X^T X)^{-1}(X^T X)$ is the identity matrix so we get

$$\rightarrow \boxed{\hat{c}} = (X^T X)^{-1} X^T \hat{y}$$

Linear Regression

Linear Regression

How much Lasix should I give?

Linear Regression: how much Lasix should I give?

Predictors		Target
BUN	Age	Lasix Dose
20	37	40
34	52	100
22	90	120
70	66	140
44	70	80

Linear Regression: how much Lasix should I give?

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.98							
R Square	0.97							
Adjusted R Square	0.62							
Standard Error	23.30							
Observations	5.00							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2.00	50371.73	25185.87	46.40	0.02			
Residual	3.00	1628.27	542.76					
Total	5.00	52000.00						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
BUN	0.91	0.53	1.73	0.18	-0.77	2.60	-0.77	2.60
Age	0.99	0.34	2.91	0.06	-0.09	2.07	-0.09	2.07

Linear Regression: how much Lasix should I give?

Predicted Lasix = (0.91 * BUN) + (0.99 * Age)

Linear Regression: how much Lasix should I give?

$$\text{Predicted Lasix} = (0.91 * \text{BUN}) + (0.99 * \text{Age})$$

Predictors		Target	<i>Regression Fit</i>	<i>Sq. Error</i>
BUN	Age	Lasix Dose	Lasix Dose	
20	37	40	54.8	219.9
34	52	100	82.4	309.1
22	90	120	109.1	118.4
70	66	140	129.0	120.1
44	70	80	109.3	860.8

Multiply BUN By:	0.91	Total Sq. Error	1628
Multiply Age By:	0.99		

Linear Regression: how much Lasix should I give?

Predicted Lasix = (0.91 * BUN) + (0.99 * Age)

Error Function :

Calculate Target Lasix - Predicted Lasix for each patient

Square the number

Add them all up

Example:

Target 40, Predicted 54.8: Target - Predicted = (54.8-40) = 14.8

Square it: $14.8 \times 14.8 = 219 \rightarrow$ squared error

Repeat for all examples and add them up

Linear Regression: how much Lasix should I give?

$$\text{Predicted Lasix} = (0.91 * \text{BUN}) + (0.99 * \text{Age})$$

Predictors		Target	<i>Regression Fit</i>	<i>Sq. Error</i>
BUN	Age	Lasix Dose	Lasix Dose	
20	37	40	54.8	219.9
34	52	100	82.4	309.1
22	90	120	109.1	118.4
70	66	140	129.0	120.1
44	70	80	109.3	860.8

Multiply BUN By:	0.91	Total Sq. Error	1628
Multiply Age By:	0.99		

Linear Regression: how much Lasix should I give?

“But I really thought it should be BUN + Age = Lasix Dose?”

Linear Regression: how much Lasix should I give?

Predicted Lasix = (1.00 * BUN) + (1.00 * Age)

Predictors		Target	Regression Fit	Sq. Error
BUN	Age	Lasix Dose	Lasix Dose	
20	37	40	57.0	289.0
34	52	100	86.0	196.0
22	90	120	112.0	64.0
70	66	140	136.0	16.0
44	70	80	114.0	1156.0

Multiply BUN By:	1.00	Total Sq. Error	1721
Multiply Age By:	1.00		

Linear Regression: how much Lasix should I give?

Predicted Lasix = (1.00 * BUN) + (1.00 * Age)

Predictors		Target	Regression Fit	Sq. Error
BUN	Age	Lasix Dose	Lasix Dose	
20	37	40	57.0	289.0
34	52	100	86.0	196.0
22	90	120	112.0	64.0
70	66	140	136.0	16.0
44	70	80	114.0	1156.0

Multiply BUN By:	1.00	Total Sq. Error	1721
Multiply Age By:	1.00		

Our old error was 1628 - this is worse!

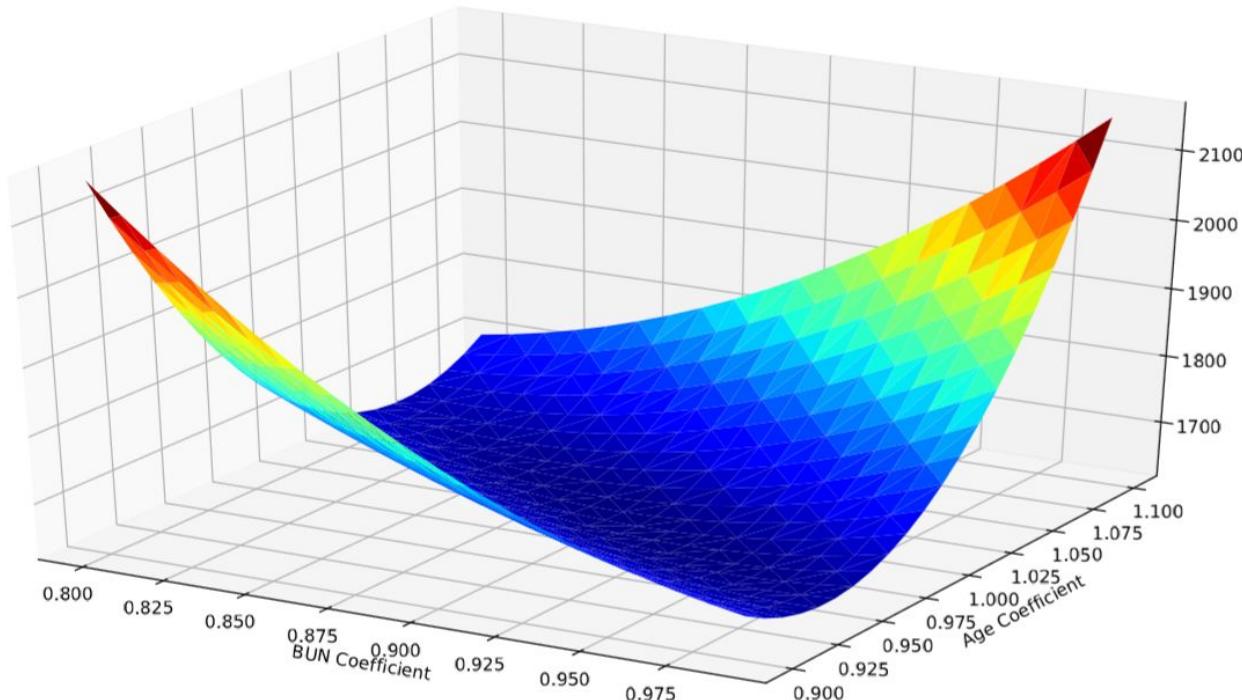
Linear Regression: how much Lasix should I give?

Predicted Lasix = ($\textcolor{red}{?} * \text{BUN}$) + ($\textcolor{red}{?} * \text{Age}$)

We can experiment by plugging in different
parameters

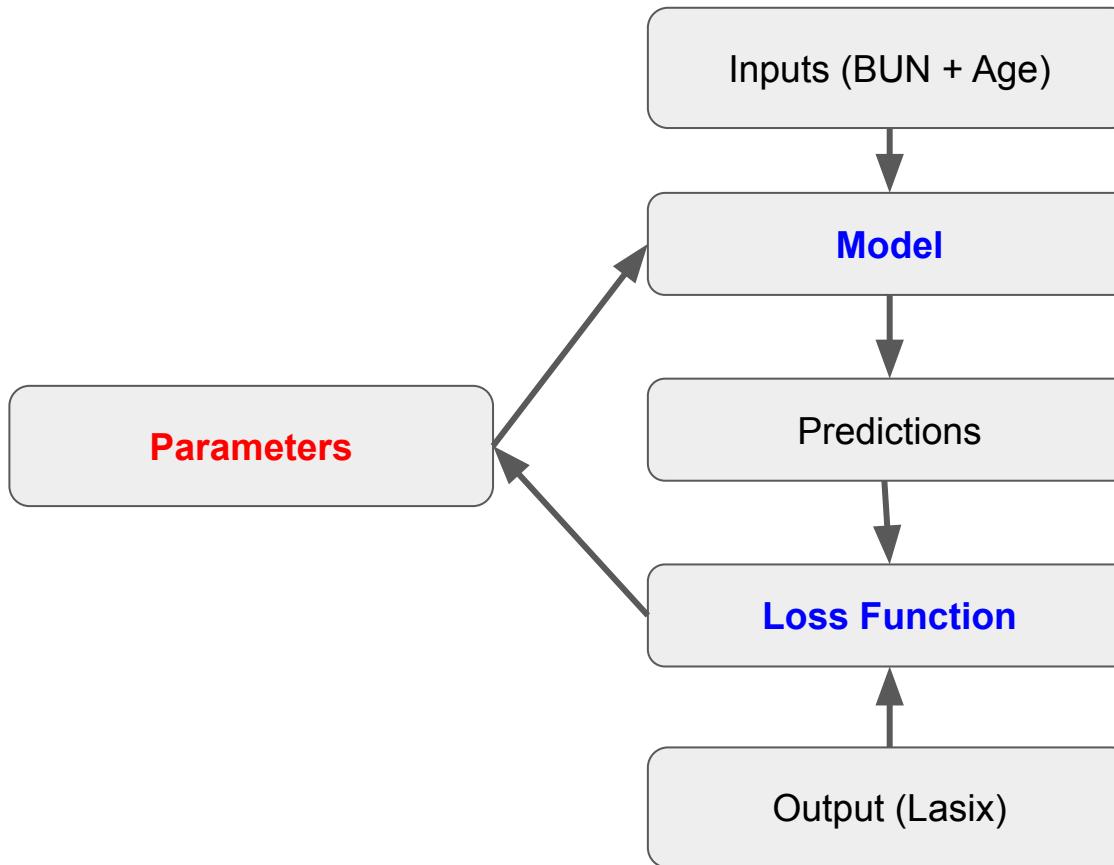
Linear Regression: how much Lasix should I give?

Predicted Lasix = ($\textcolor{red}{?} * \text{BUN}$) + ($\textcolor{red}{?} * \text{Age}$)

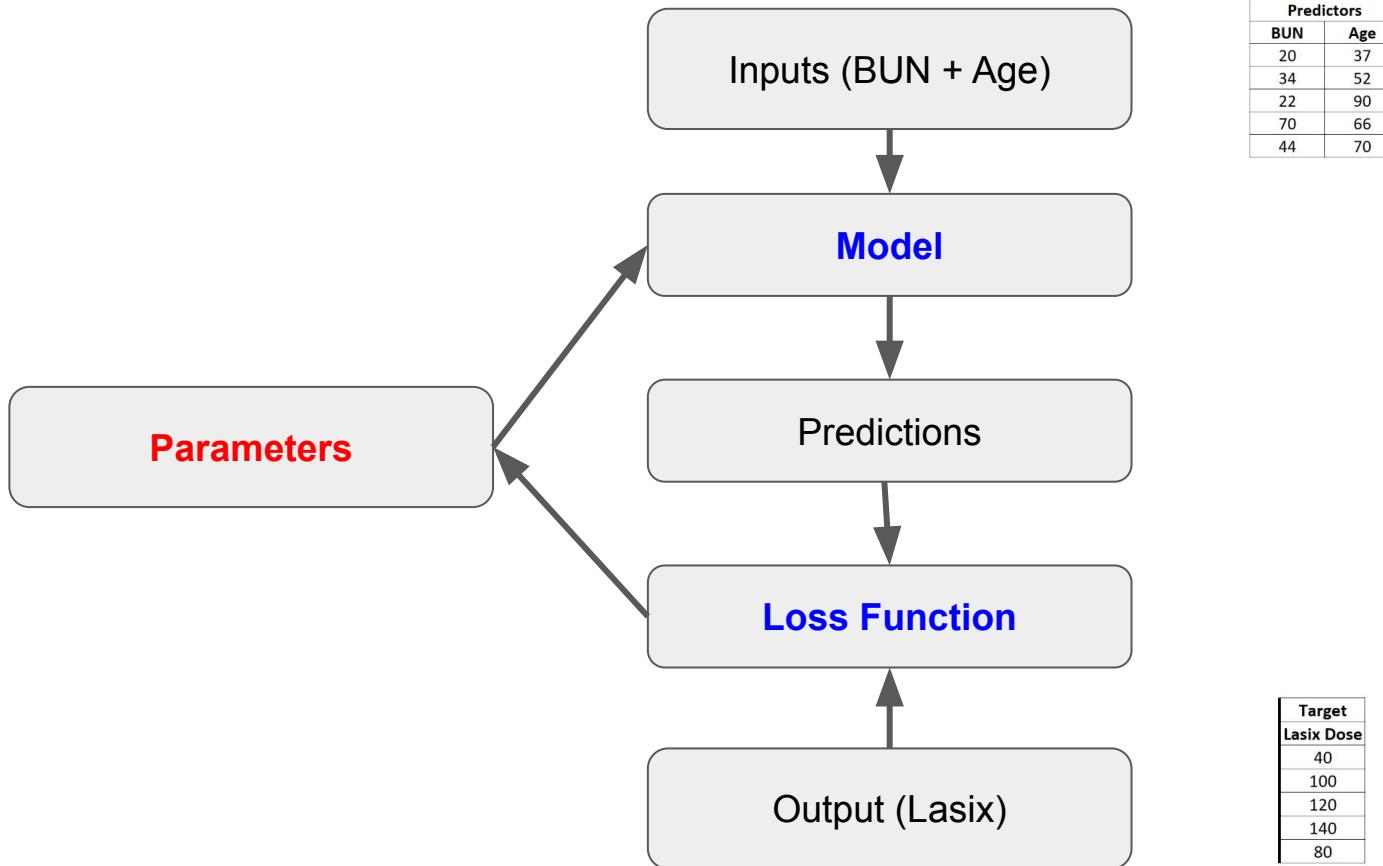


Linear Regression: a model for supervised learning

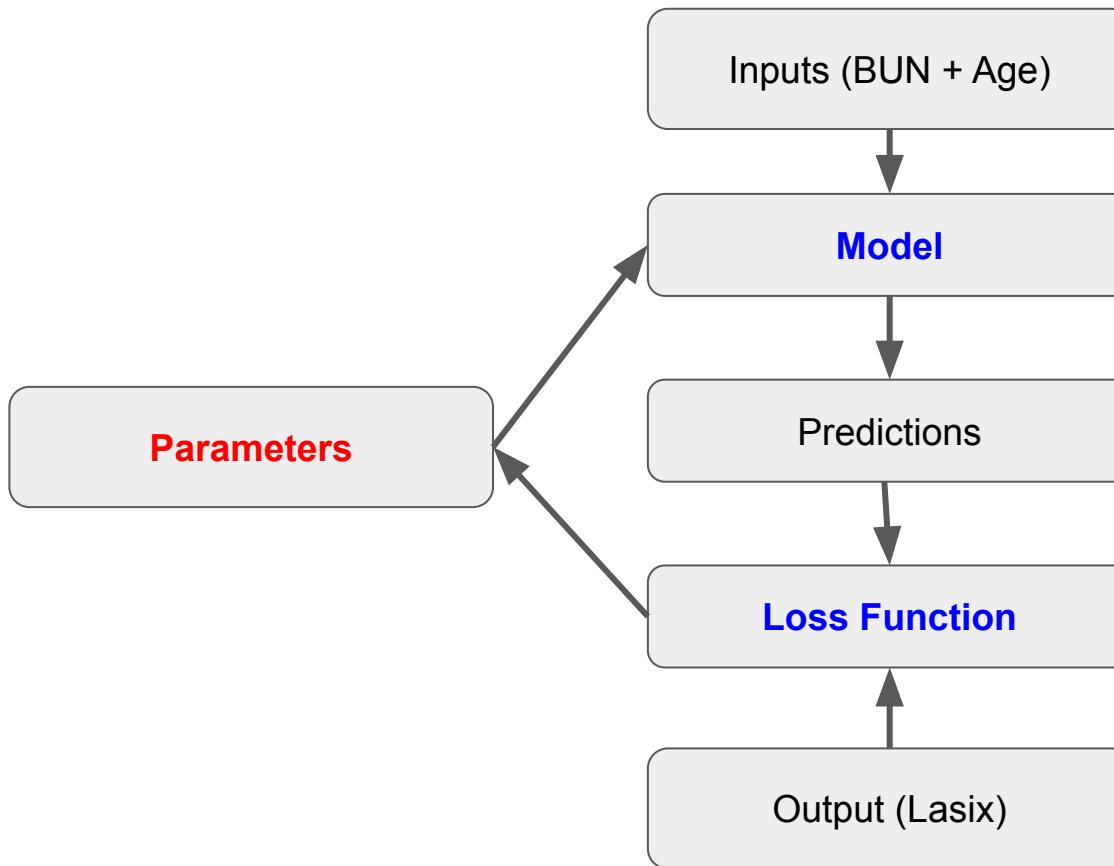
Linear Regression: a model for supervised learning



Linear Regression: a model for supervised learning



Linear Regression: a model for supervised learning

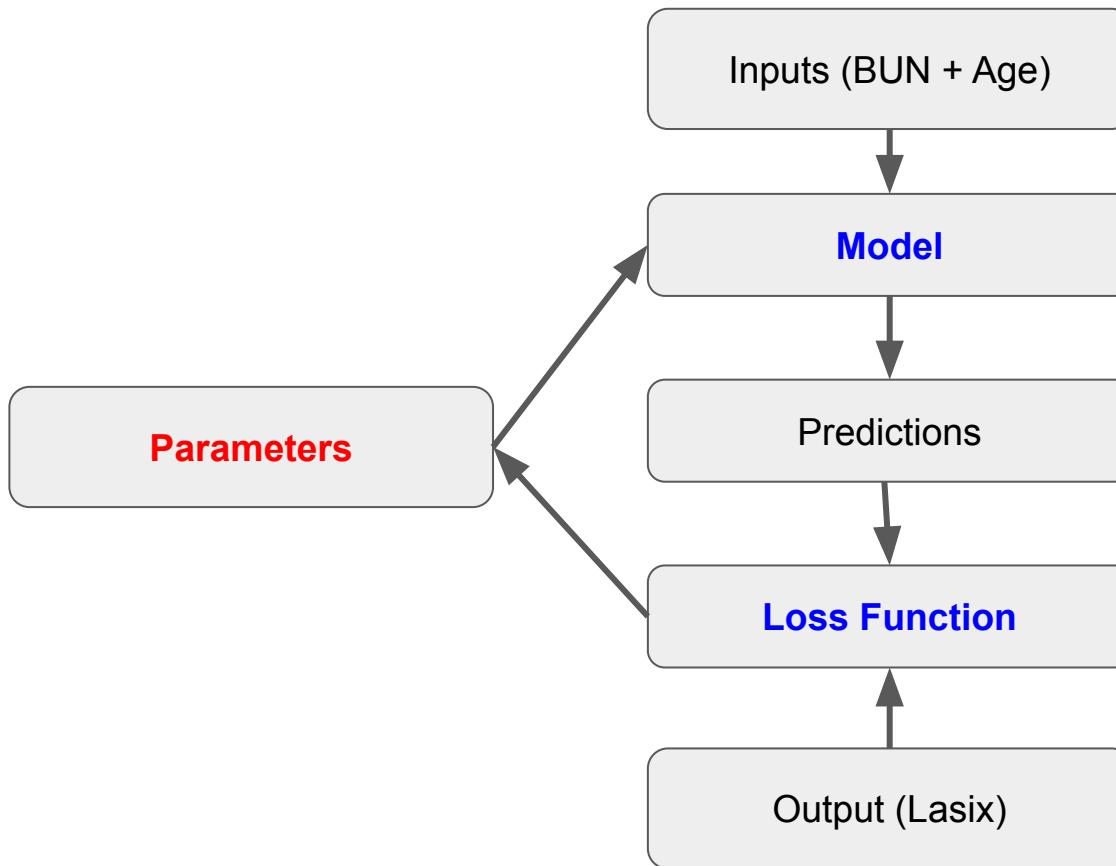


Predictors	
BUN	Age
20	37
34	52
22	90
70	66
44	70

$$\text{Predicted Lasix} = (0.91 * \text{BUN}) + (0.99 * \text{Age})$$

Target	
Lasix Dose	
40	
100	
120	
140	
80	

Linear Regression: a model for supervised learning



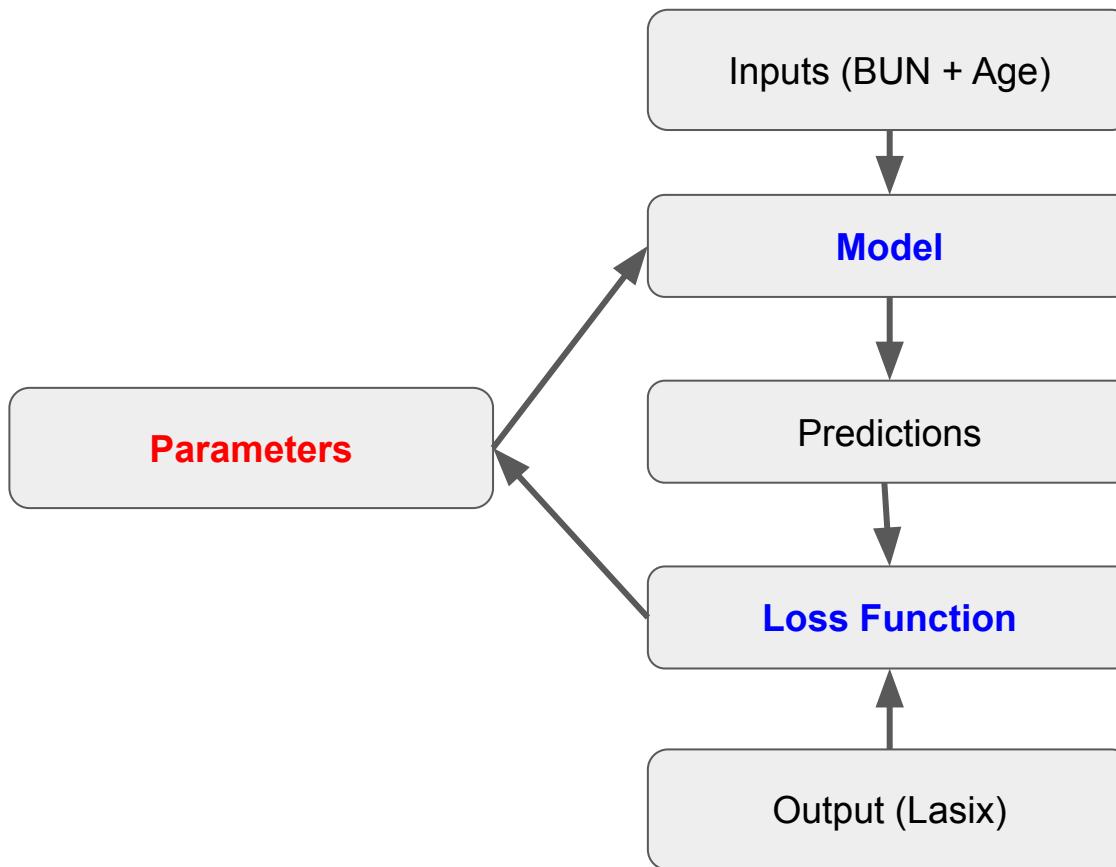
Predictors	
BUN	Age
20	37
34	52
22	90
70	66
44	70

$$\text{Predicted Lasix} = (0.91 * \text{BUN}) + (0.99 * \text{Age})$$

Lasix Dose
57.0
86.0
112.0
136.0
114.0

Target	Lasix Dose
40	57.0
100	86.0
120	112.0
140	136.0
80	114.0

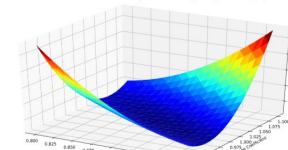
Linear Regression: a model for supervised learning



Predictors	
BUN	Age
20	37
34	52
22	90
70	66
44	70

$$\text{Predicted Lasix} = (0.91 * \text{BUN}) + (0.99 * \text{Age})$$

Lasix Dose
57.0
86.0
112.0
136.0
114.0



Target	Lasix Dose
40	40
100	100
120	120
140	140
80	80

Linear Regression: a model for supervised learning

We have data: inputs (BUN + Age) and outputs (Lasix) - they're fixed

We define a model to generate a prediction given inputs

We define an error function that penalizes predictions that are far away from outputs

We learn parameters that minimize the error function

Predicted Lasix = (0.91 * BUN) + (0.99 * Age)

Supervised

Why do we use regression?

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

Supervised

Why do we use regression?

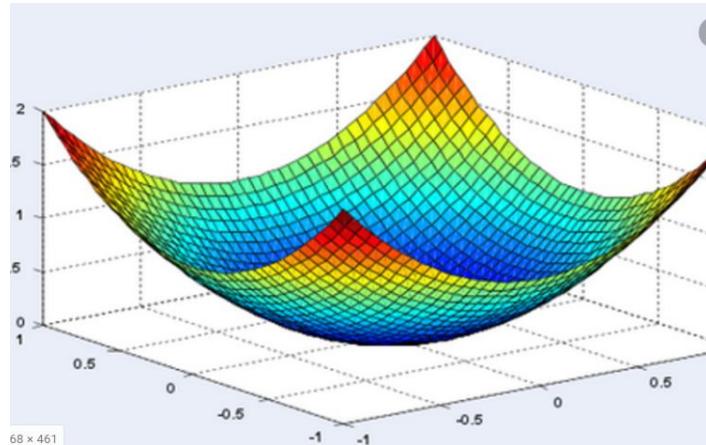
History [edit]

The earliest form of regression was the [method of least squares](#), which was published by [Legendre](#) in 1805,^[1] and by [Gauss](#) in 1809.^[7] Legendre and Gauss both applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the Sun (mostly comets, but also later the then newly discovered minor planets). Gauss published a further development of the theory of least squares in 1821,^[8] including a version of the [Gauss–Markov theorem](#).

Supervised

Why do we use regression?

It's computationally relatively easy - you can find an analytical (mathematical) formula for your parameters.



Supervised

There are lots of alternatives:

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

Supervised

What if we want to use really complicated data as an input?

- Text?
- An image?
- A patient's entire clinical record in the EHR?

Enter ‘Deep Learning’

‘Deep Learning’ = a timely rebranding of ‘neural networks’, which have been around for decades

Enter ‘Deep Learning’

‘Deep Learning’ = a timely rebranding of ‘neural networks’, which have been around for decades.

What changed?

Enter ‘Deep Learning’

RISE OF ALEXNET

The inside story of how AI got good enough to dominate Silicon Valley

By Dave Gershman • June 18, 2018



A portrait of a young man with short brown hair and glasses, wearing a dark grey hoodie. He is smiling slightly and looking directly at the camera. The background is blurred, showing what appears to be an indoor setting with warm lighting.

This Alex, behind AlexNet.

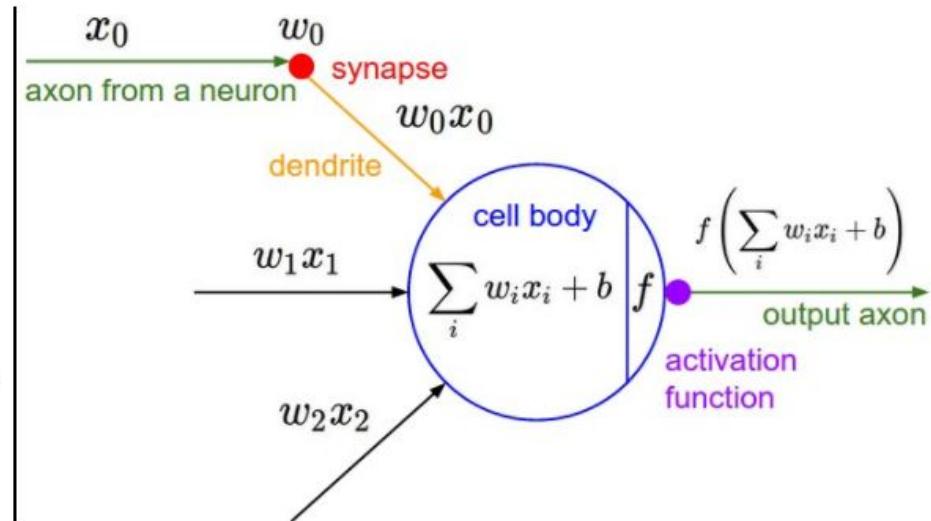
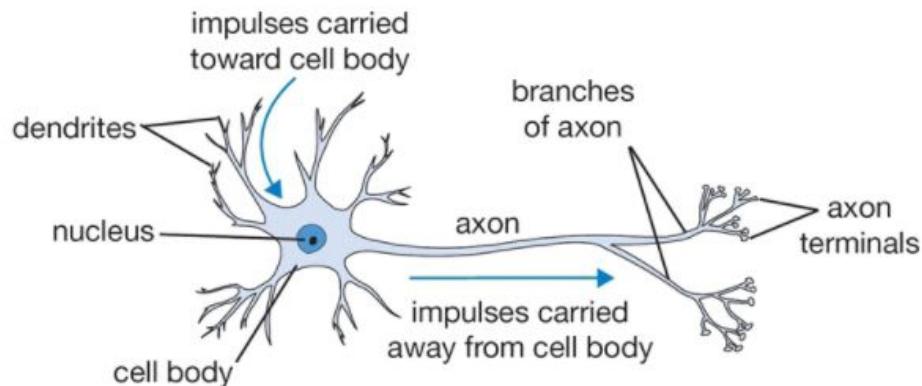
<https://qz.com/1307091/the-inside-story-of-how-ai-got-good-enough-to-dominate-silicon-valley/>

Enter ‘Deep Learning’



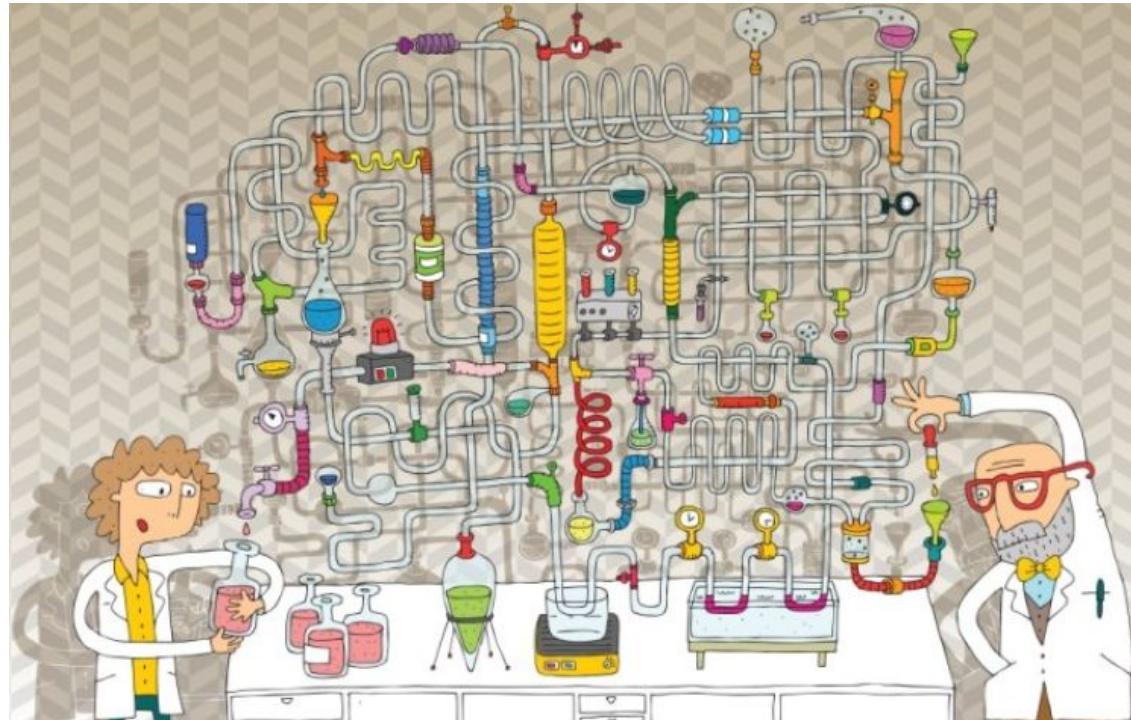
<https://www.amazon.com/EVGA-GeForce-Gaming-GDDR5X-Technology/dp/B06Y13N2B6>

Neural network: biologically inspired



A cartoon drawing of a biological neuron (left) and its mathematical model (right).

Neural network: complexity can be modeled



<https://www.wonderopolis.org/wonder/what-is-a-rube-goldberg-machine>

Neural network: complexity can be modeled

In regression, you define the input features

Neural network: complexity can be modeled

In regression, you define the input features

In deep learning, the model learns the features

Neural network: complexity can be modeled

In regression, you define the input features

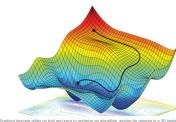
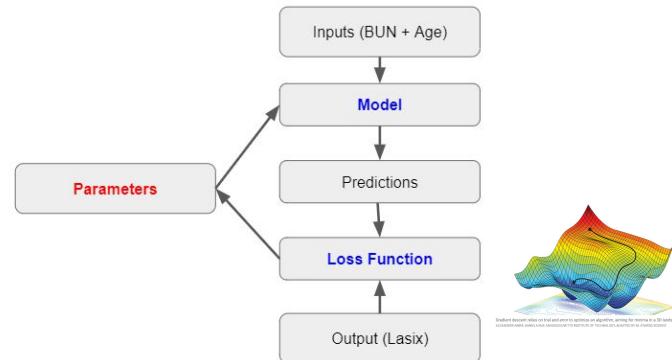
In deep learning, the model learns the features

Neural network: complexity can be modeled

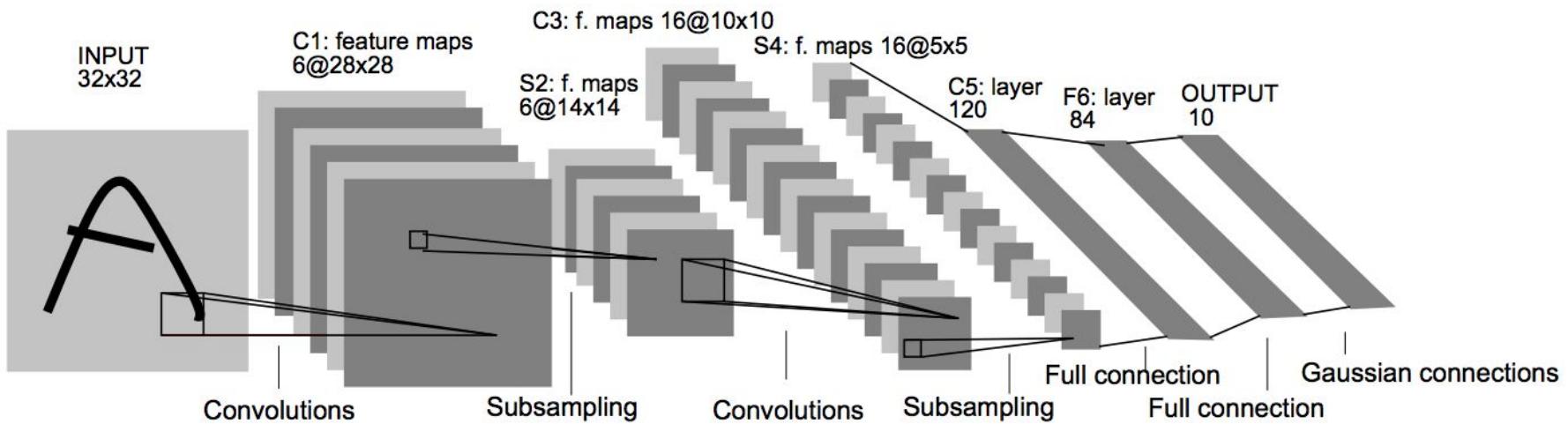
In regression, you define the input features

In deep learning, the model learns the features

Much of the high-level setup stays the same though:



Take an image as input: LeNet-5



Convolutional neural network

AVERAGING EACH PIXEL WITH ITS NEIGHBORING VALUES BLURS AN IMAGE:

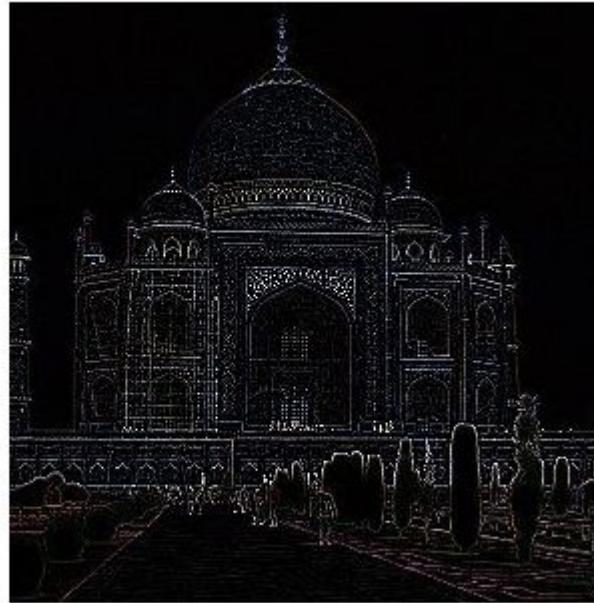
0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0



Convolutional neural network

TAKING THE DIFFERENCE BETWEEN A PIXEL AND ITS NEIGHBORS DETECTS EDGES:

0	1	0
1	-4	1
0	1	0



1	0	1
0	1	0
1	0	1

Convolutional neural network

1 <small>$\times 1$</small>	1 <small>$\times 0$</small>	1 <small>$\times 1$</small>	0	0
0 <small>$\times 0$</small>	1 <small>$\times 1$</small>	1 <small>$\times 0$</small>	1	0
0 <small>$\times 1$</small>	0 <small>$\times 0$</small>	1 <small>$\times 1$</small>	1	1
0	0	1	1	0
0	1	1	0	0

Image

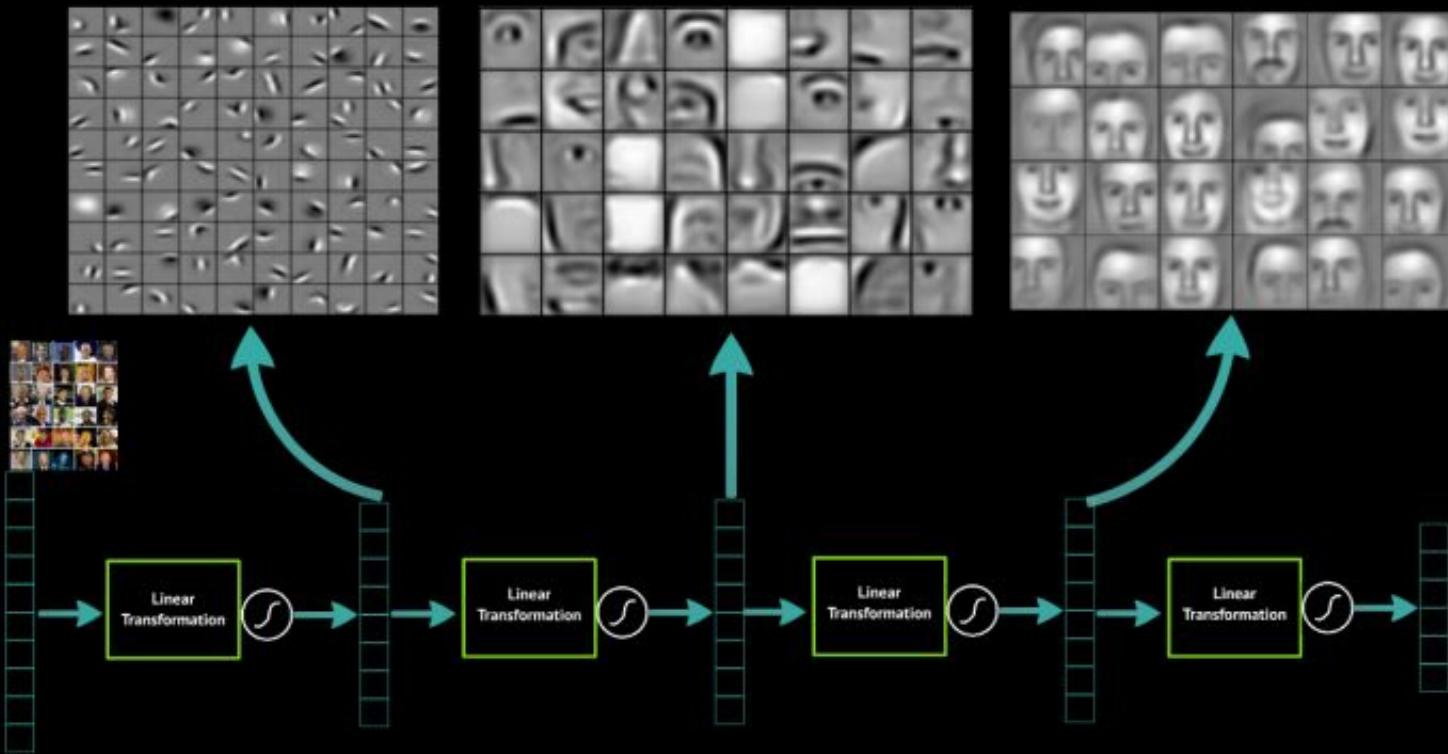
4		

Convolved Feature

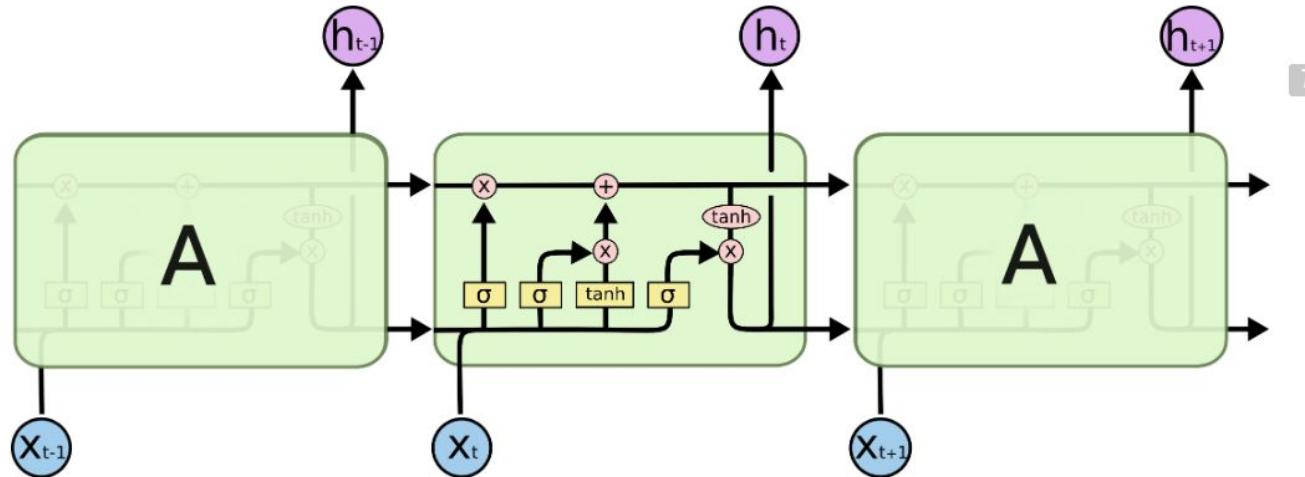


Multiplying
by this each
time

Deep Learning learns layers of features



Take text as input: Long-Short Term Memory



The repeating module in an LSTM contains four interacting layers.

Take EHR as input: Long-Short Term Memory

ARTICLE

OPEN

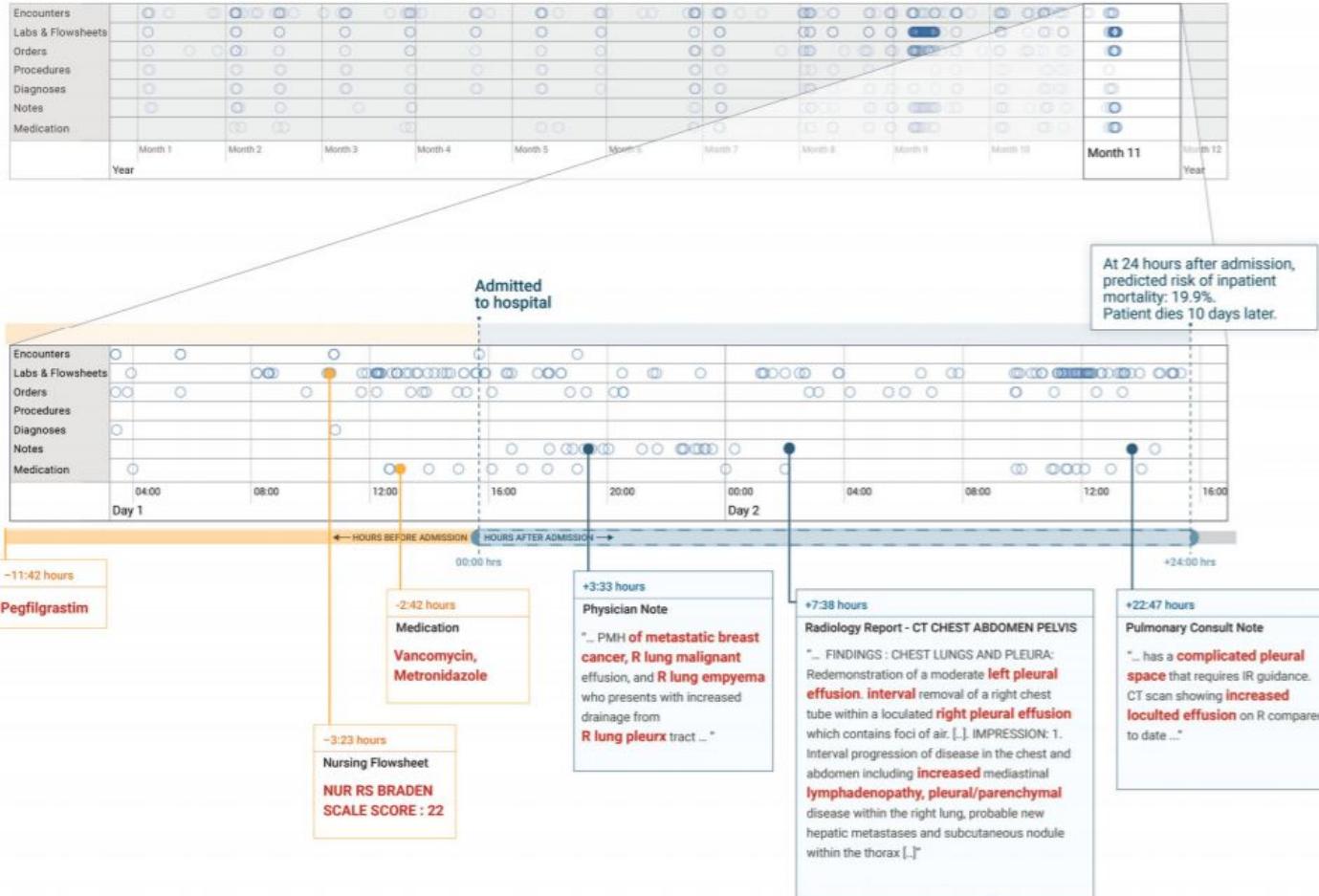
Scalable and accurate deep learning with electronic health records

Alvin Rajkomar^{1,2}, Eyal Oren¹, Kai Chen¹, Andrew M. Dai¹, Nissan Hajaj¹, Michaela Hardt¹, Peter J. Liu¹, Xiaobing Liu¹, Jake Marcus¹, Mimi Sun¹, Patrik Sundberg¹, Hector Yee¹, Kun Zhang¹, Yi Zhang¹, Gerardo Flores¹, Gavin E. Duggan¹, Jamie Irvine¹, Quoc Le¹, Kurt Litsch¹, Alexander Mossin¹, Justin Tansuwan¹, De Wang¹, James Wexler¹, Jimbo Wilson¹, Dana Ludwig², Samuel L. Volchenboum³, Katherine Chou¹, Michael Pearson¹, Srinivasan Madabushi¹, Nigam H. Shah⁴, Atul J. Butte², Michael D. Howell¹, Claire Cui¹, Greg S. Corrado¹ and Jeffrey Dean¹

Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction of curated predictor variables from normalized EHR data, a labor-intensive process that discards the vast majority of information in each patient's record. We propose a representation of patients' entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format. We demonstrate that deep learning methods using this representation are capable of accurately predicting multiple medical events from multiple centers without site-specific data harmonization. We validated our approach using de-identified EHR data from two US academic medical centers with 216,221 adult patients hospitalized for at least 24 h. In the sequential format we propose, this volume of EHR data unrolled into a total of 46,864,534,945 data points, including clinical notes. Deep learning models achieved high accuracy for tasks such as predicting: in-hospital mortality (area under the receiver operator curve [AUROC] across sites 0.93–0.94), 30-day unplanned readmission (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.86), and all of a patient's final discharge diagnoses (frequency-weighted AUROC 0.90). These models outperformed traditional, clinically-used predictive models in all cases. We believe that this approach can be used to create accurate and scalable predictions for a variety of clinical scenarios. In a case study of a particular prediction, we demonstrate that neural networks can be used to identify relevant information from the patient's chart.

npj Digital Medicine (2018)1:18; doi:10.1038/s41746-018-0029-1

Patient Timeline



REVIEW ARTICLE

FRONTIERS IN MEDICINE

Machine Learning in Medicine

Alvin Rajkomar, M.D., Jeffrey Dean, Ph.D., and Isaac Kohane, M.D., Ph.D.

A 49-year-old patient notices a painless rash on his shoulder but does not seek care. Months later, his wife asks him to see a doctor, who diagnoses a seborrheic keratosis. Later, when the patient undergoes a screening colonoscopy, a nurse notices a dark macule on his shoulder and advises him to have it evaluated. One month later, the patient sees a dermatologist, who obtains a biopsy specimen of the lesion. The findings reveal a noncancerous pigmented lesion. Still concerned, the dermatologist requests a second reading of the biopsy specimen, and invasive melanoma is diagnosed. An oncologist initiates treatment with systemic chemotherapy. A physician friend asks the patient why he is not receiving immunotherapy.

ML in Medicine: Rajkomar et. al 2019

“What if every medical decision, whether made by an intensivist or a community health worker, was instantly reviewed by a team of relevant experts who provided guidance if the decision seemed amiss?...

Such a system seems far-fetched...Yet, this is the promise of machine learning in medicine: the wisdom contained in the decisions made by nearly all clinicians and the outcomes of billions of patients should inform the care of each patient.”

Deep Learning looks awesome

What's the catch here?

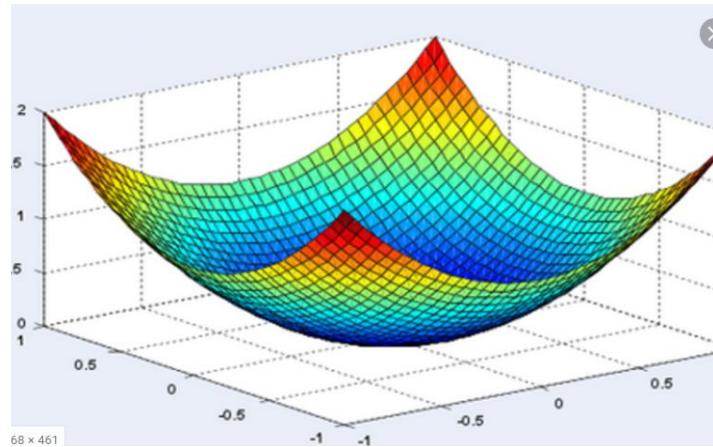
Many parameters → tricky to train

- Linear regression: 2 - 50 parameters
- Image classification neural network: millions of parameters (one example - 6,963,081)

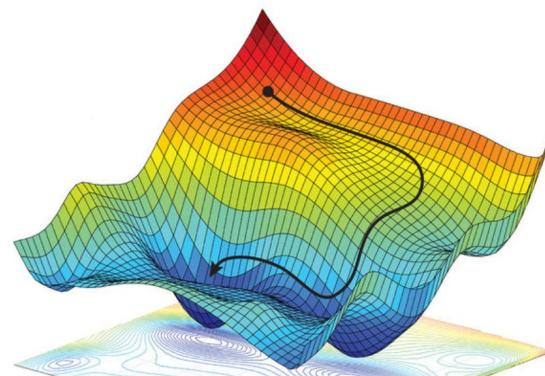
Many parameters → tricky to train

Optimization
surface w.r.t.
parameters:

Linear Regression



Deep Learning

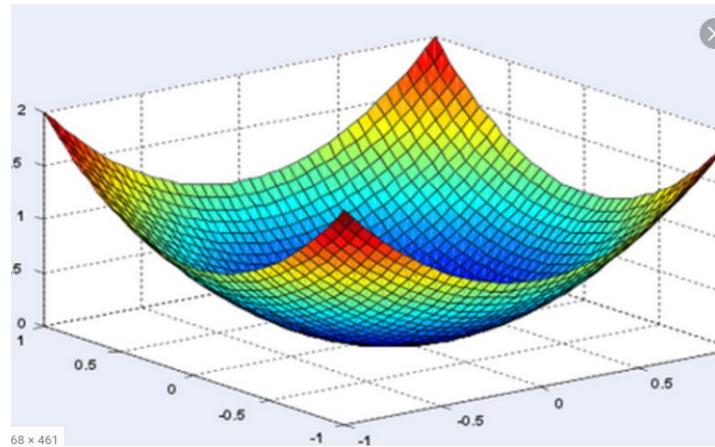


Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3D landscape.
ALEXANDER AMINI, DANIELA RUS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, ADAPTED BY M. ATAROD/SCIENCE

Many parameters → tricky to train

Optimization surface w.r.t. parameters:

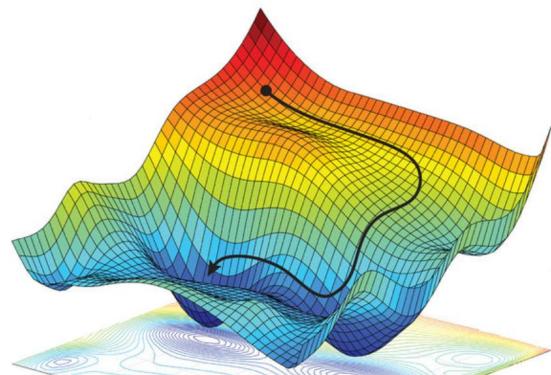
Linear Regression



Number of parameters

2 - 50

Deep Learning



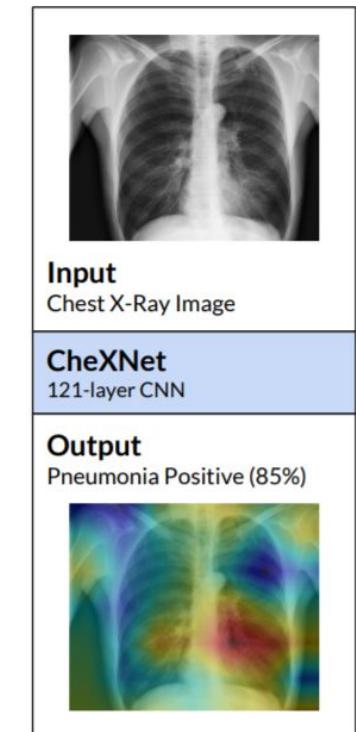
Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3D landscape.
ALEXANDER AMINI, DANIELA RUS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, ADAPTED BY M. ATAROD/SCIENCE

Millions

In Radiology: Rajpurkar et al. (2017)

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar^{*1} Jeremy Irvin^{*1} Kaylie Zhu¹ Brandon Yang¹ Hershel Mehta¹
Tony Duan¹ Daisy Ding¹ Aarti Bagul¹ Robyn L. Ball² Curtis Langlotz³ Katie Shpanskaya³
Matthew P. Lungren³ Andrew Y. Ng¹

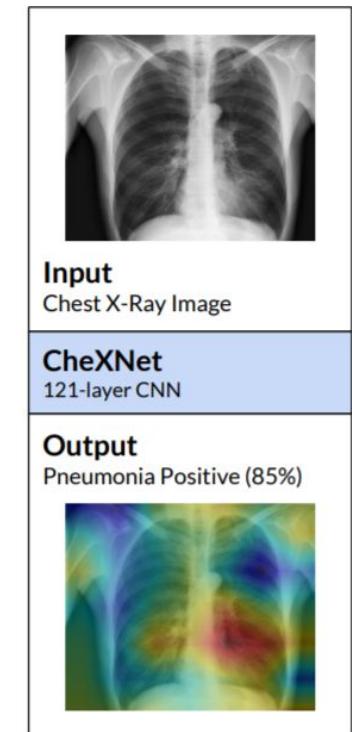


In Radiology: Rajpurkar et al. (2017)

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar^{*1} Jeremy Irvin^{*1} Kaylie Zhu¹ Brandon Yang¹ Hershel Mehta¹
Tony Duan¹ Daisy Ding¹ Aarti Bagul¹ Robyn L. Ball² Curtis Langlotz³ Katie Shpanskaya³
Matthew P. Lungren³ Andrew Y. Ng¹

185 citations in 15 months



In Radiology: Rajpurkar et al. (2017)



Andrew Ng
@AndrewYNg

Follow

Should radiologists be worried about their jobs? Breaking news: We can now diagnose pneumonia from chest X-rays better than radiologists.

[stanfordmlgroup.github.io/projects/chexn...](https://stanfordmlgroup.github.io/projects/chexnet/)

3:20 PM - 15 Nov 2017 from Mountain View, CA

1,406 Retweets 2,371 Likes



113

1.4K

2.4K



Input
Chest X-Ray Image

CheXNet
121-layer CNN

Output
Pneumonia Positive (85%)



when do we need to start cutting radiology residency spots?

Discussion in 'Radiology' started by IRrads10, Thursday at 2:05 PM.

[Previous Thread](#) [Next Thread](#)



IRrads10

2+ Year Member

Joined: Jul 20, 2016
Messages: 23
Likes Received: 9
Status: Medical Student

Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists

AI in radiology is progressing swiftly. it is obviously hard to predict how much of an impact these technologies will have.

should radiology's future start to look bleak in the coming decade, i sure hope we are proactive and not reactive.

IRrads10, Thursday at 2:05 PM

#1

Gurby likes this.



Mo991

Joined: Aug 2, 2018
Messages: 12
Likes Received: 11

This is really disturbing
I'm starting to think I made a bad decision..

Mo991, Thursday at 3:57 PM

#3

when do we need to start cutting radiology residency spots?

Discussion in 'Radiology' started by IRrads10, Thursday at 2:05 PM.

[Previous Thread](#) [Next Thread](#)



IRrads10

2+ Year Member

Joined: Jul 20, 2016
Messages: 23
Likes Received: 9
Status: Medical Student

Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists

AI in radiology is progressing swiftly. it is obviously hard to predict how much of an impact these technologies will have.

should radiology's future start to look bleak in the coming decade, i sure hope we are proactive and not reactive.

IRrads10, Thursday at 2:05 PM

#1

Gurby likes this.



Mo991

Joined: Aug 2, 2018
Messages: 12
Likes Received: 11

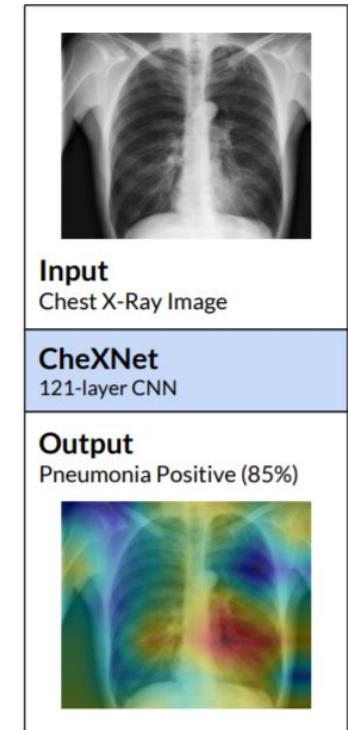
This is really disturbing
I'm starting to think I made a bad decision..

Mo991, Thursday at 3:57 PM

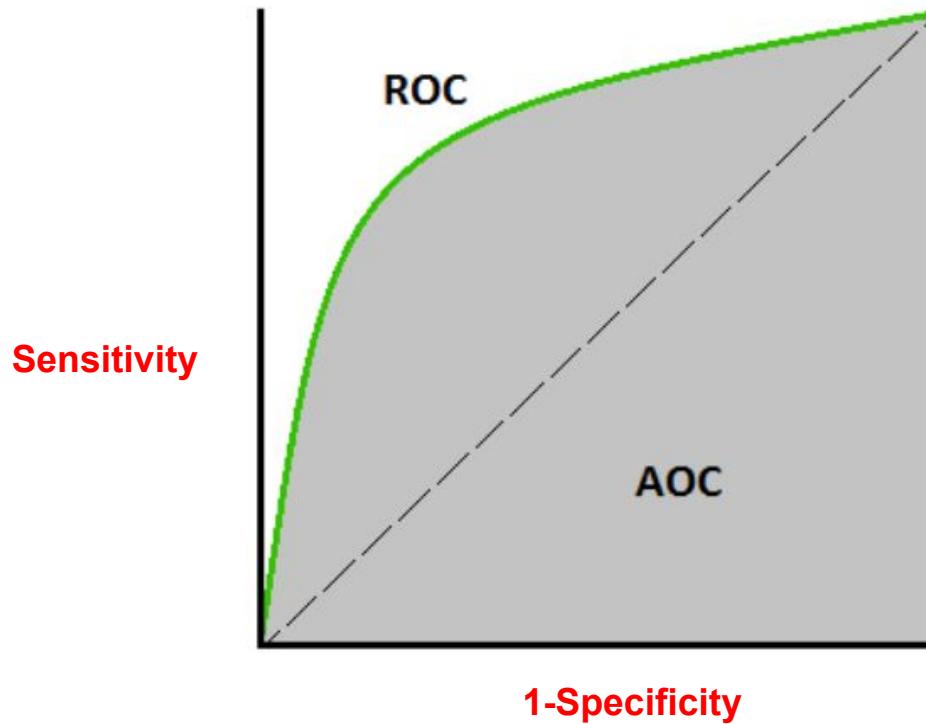
#3

In Radiology: Rajpurkar et al. (2017)

- 112,120 NIH chest x-rays -- 70% train, 10% tune, 20% test
- 14 diagnoses, including pneumonia
- Pre-trained DenseNet-121 (224 x 224 pixels)
- **AUC for pneumonia: 0.7680**

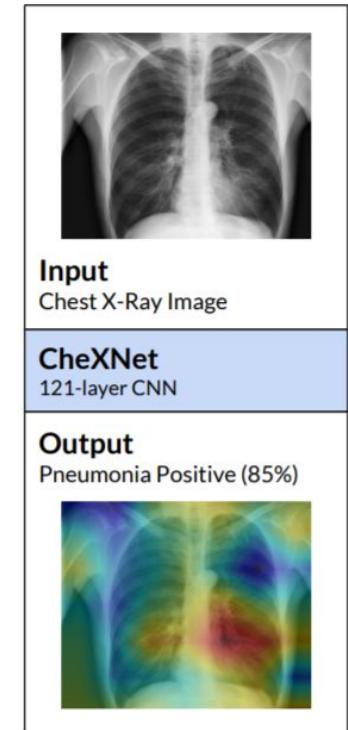


Quick review: what's an AUC curve?



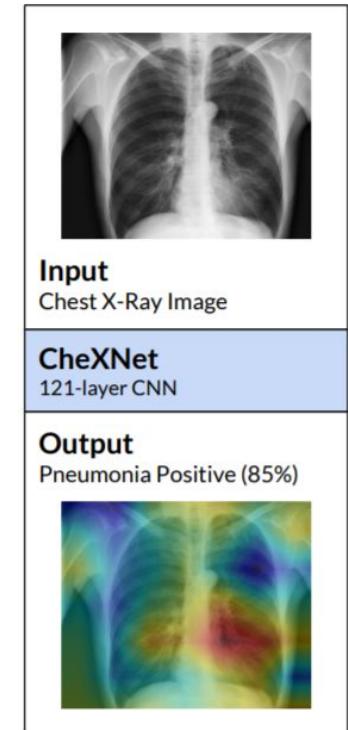
In Radiology: Rajpurkar et al. (2017)

- 112,120 NIH chest x-rays -- 70% train, 10% tune, 20% test
- 14 diagnoses, including pneumonia
- Pre-trained DenseNet-121 (224 x 224 pixels)
- **AUC for pneumonia: 0.7680**



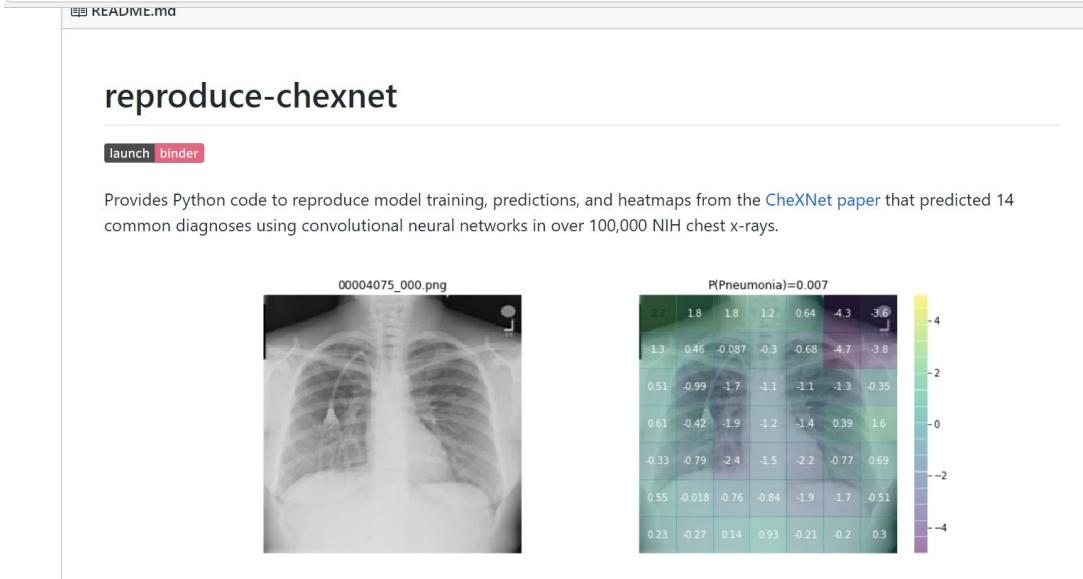
In Radiology: Rajpurkar et al. (2017)

- **Low AUC:** 0.7680
- **Remarkably low agreement among radiologists -**
even only 50% sensitivity 50% PPV would get you to
 $F1=0.5$ vs reported 0.387
→ if true, should radiographic findings be required for dx?
- **How does the algorithm perform relatively well?**



Confounders in Radiology: Zech et al. 2018

GitHub, Inc. [US] | <https://github.com/jrzech/reproduce-chexnet>



```
62     def train_model(
63         model,
64         criterion,
65         optimizer,
66         LR,
67         num_epochs,
68         dataloaders,
69         dataset_sizes,
70         weight_decay):
71     """
72     Fine tunes torchvision model to NIH CXR data.
73
74     Args:
75         model: torchvision model to be finetuned (densenet-121 in this case)
76         criterion: loss criterion (binary cross entropy loss, BCELoss)
77         optimizer: optimizer to use in training (SGD)
78         LR: learning rate
79         num_epochs: continue training up to this many epochs
80         dataloaders: pytorch train and val dataloaders
81         dataset_sizes: length of train and val datasets
82         weight_decay: weight decay parameter we use in SGD with momentum
83
84     Returns:
85         model: trained torchvision model
86         best_epoch: epoch on which best model val loss was obtained
87
88     """
89     since = time.time()
90
91     start_epoch = 1
```

Confounders in Radiology: Zech et al. 2018



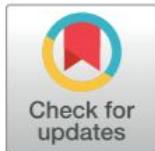
RESEARCH ARTICLE

Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study

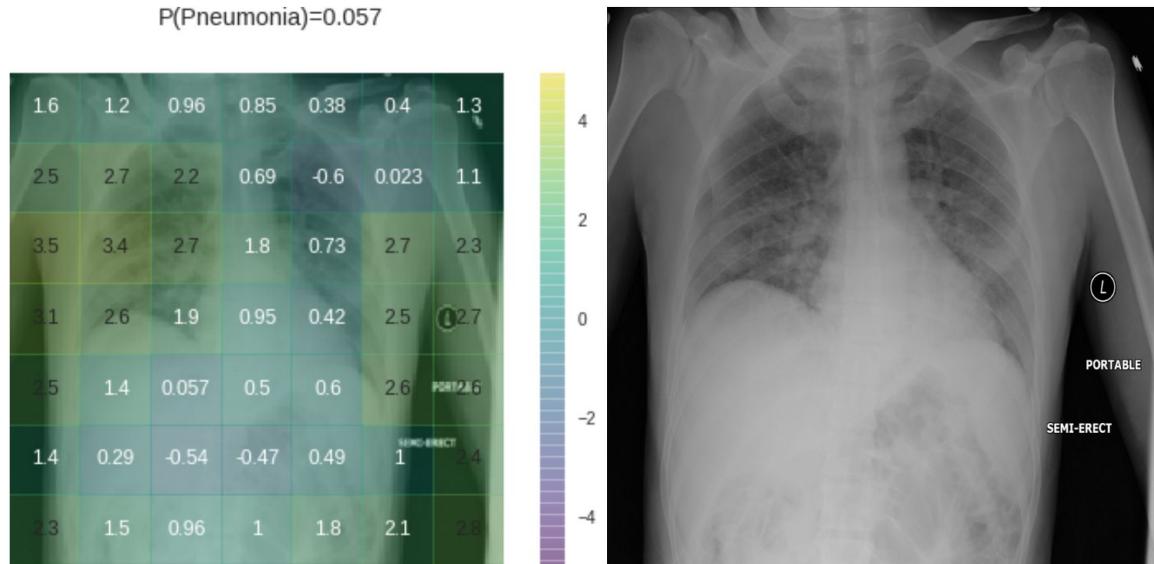
John R. Zech¹*, Marcus A. Badgeley², Manway Liu², Anthony B. Costa³, Joseph J. Titano⁴, Eric Karl Oermann³*

1 Department of Medicine, California Pacific Medical Center, San Francisco, California, United States of America, **2** Verily Life Sciences, South San Francisco, California, United States of America, **3** Department of Neurological Surgery, Icahn School of Medicine, New York, New York, United States of America, **4** Department of Radiology, Icahn School of Medicine, New York, New York, United States of America

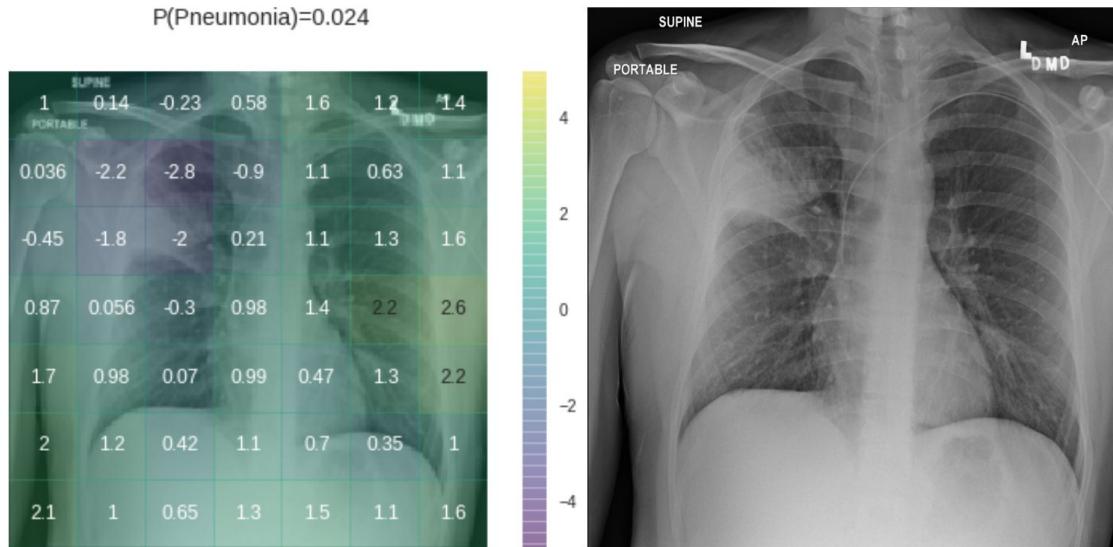
• These authors contributed equally to this work.
* eric.oermann@mountsinai.org



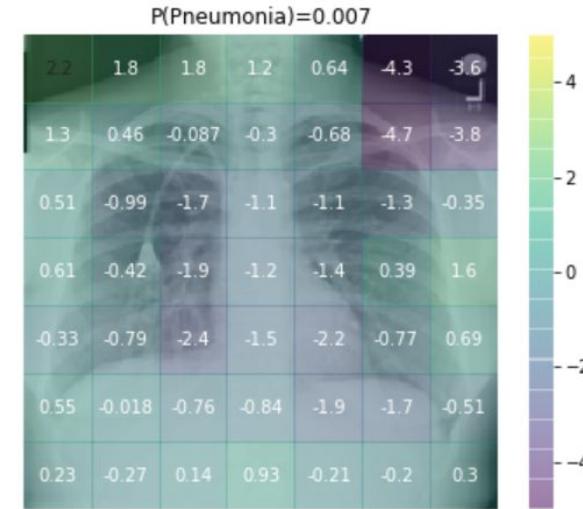
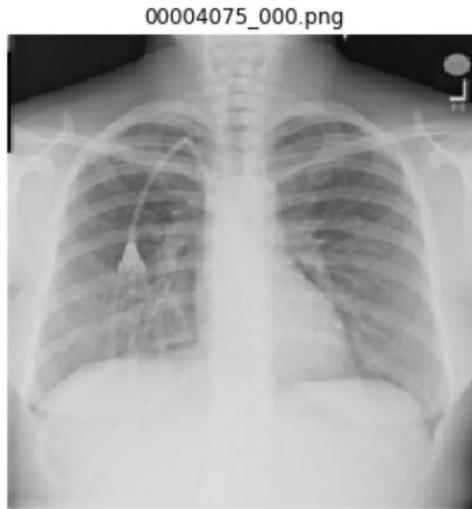
Confounders in Radiology: Zech et al. 2018



Confounders in Radiology: Zech et al. 2018



Confounders in Radiology: Zech et al. 2018



Confounders in Radiology: Zech et al. 2018

- **CNNs appear to exploit information beyond specific disease-related imaging findings on x-rays to calibrate their disease predictions.**
- Scanner type (especially portable vs regular PA/lateral) is easily exploited.
- Hard to reconcile this with CheXNet claims w.r.t. human radiologists.

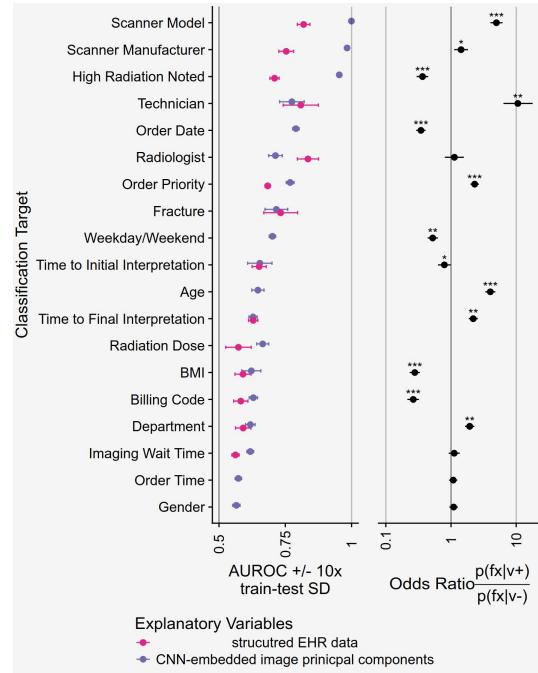
A problem with deep learning

It's hard to know what's driving predictions: our model for pneumonia had 6,963,081 parameters

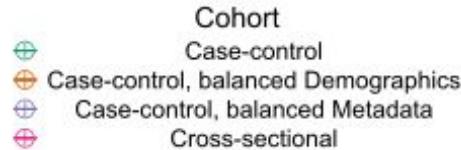
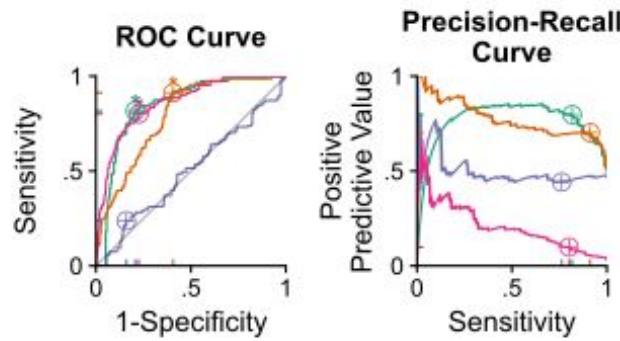
Compare to CURB-65

Confounders in Radiology: Badgeley et al. 2018

- Identify hip fracture (and many other covariates) in 25,000 x-rays from Mount Sinai Hospital using Inception v3 features (299 x 299).



Confounders in Radiology: Badgeley et al. 2018



-->all predictive power of fracture model disappears when you control for confounders (AUC 0.80 goes to no better than chance)

Big limitations in these models

- Limited resolution (224 x 224, 299 x 299)
- “Weakly supervised”
- Suboptimal labels
- **Stronger models have been demonstrated in the meantime...**

Hip fracture: Gale et al. (2017)

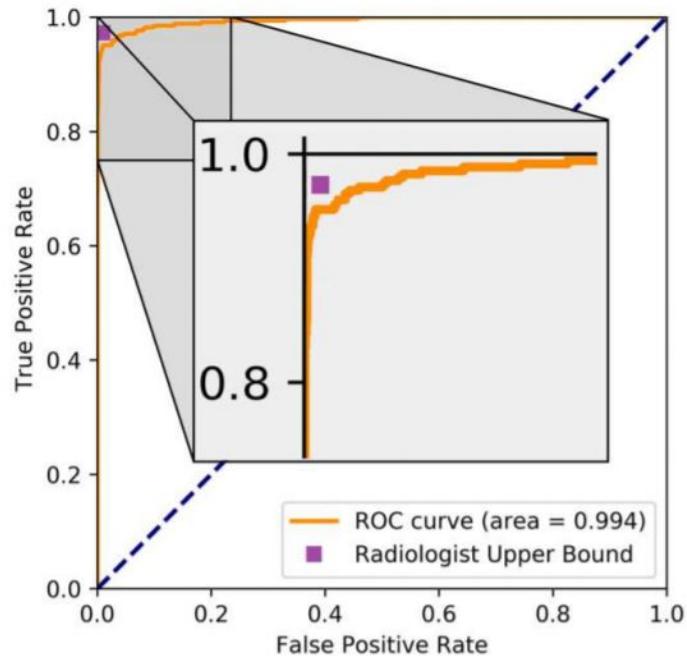
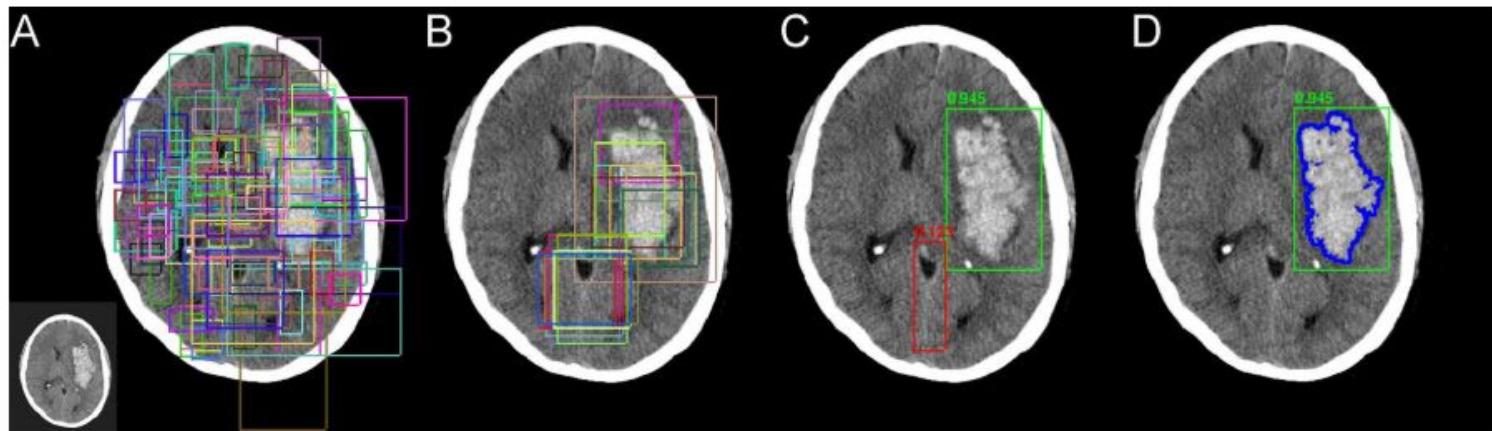


Figure 1: ROC curve showing the performance of the model with AUC 0.994, with a point reflecting the optimistic upper bound of human performance.

Intracranial hemorrhage: Chang et al. (2018)



AUC
0.989

Mask residual CNN architectures can provide a framework for parallel evaluation of region proposal (attention), object detection (classification), and instance segmentation. In this approach, (A) preconfigured bounding boxes at various shapes and resolutions are tested for the presence of a potential abnormality. (B) The highest ranking bounding boxes are identified and used to generate region proposals that focus algorithm attention. (C) Composite region proposals are pruned using nonmaximum suppression and used as input into a classifier to determine presence or absence of hemorrhage. (D) Segmentation masks are generated for positive cases of hemorrhage. All images courtesy of Dr. Peter Chang.

Will stronger models generalize? Lee et. al (2018)

Performance for detecting intracranial hemorrhage		
	AI performance on test dataset	AI performance on real-world dataset
Sensitivity	98%	87.1%
Specificity	95%	58.3%
Area under the curve	0.993	0.834

The problem with deep learning

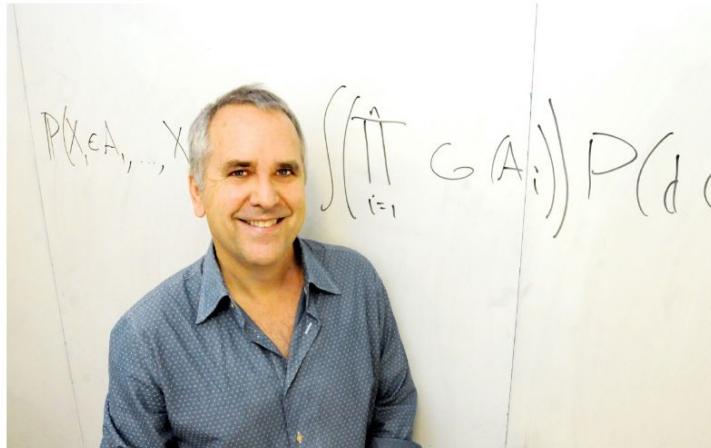


Photo credit: Peg Skorpinski

Artificial Intelligence—The Revolution Hasn't Happened Yet



Michael Jordan [Follow](#)
Apr 18, 2018 • 16 min read

<https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7>

The problem with deep learning

The idea that our era is somehow seeing the emergence of an intelligence in silicon that rivals our own entertains all of us—enthraling us and frightening us in equal measure. And, unfortunately, it distracts us.

The problem with deep learning

When my spouse was pregnant 14 years ago, we had an ultrasound. There was a geneticist in the room, and she pointed out some white spots around the heart of the fetus. “Those are markers for Down syndrome,” she noted, “and your risk has now gone up to 1 in 20.” She further let us know that we could learn whether the fetus in fact had the genetic modification underlying Down syndrome via an amniocentesis. But amniocentesis was risky—the risk of killing the fetus during the procedure was roughly 1 in 300.

The problem with deep learning

Being a statistician, I determined to find out where these numbers were coming from. To cut a long story short, I discovered that a statistical analysis had been done a decade previously in the UK, where these white spots, which reflect calcium buildup, were indeed established as a predictor of Down syndrome. But I also noticed that the imaging machine used in our test had a few hundred more pixels per square inch than the machine used in the UK study. I went back to tell the geneticist that I believed that the white spots were likely false positives—that they were literally “white noise.”

She said “Ah, that explains why we started seeing an uptick in Down syndrome diagnoses a few years ago; it’s when the new machine arrived.”

The problem with deep learning

We do not want to build systems that help us with medical treatments, transportation options and commercial opportunities to find out after the fact that these systems don't really work—that they make errors that take their toll in terms of human lives and happiness.

In this regard, as I have emphasized, there is an engineering discipline yet to emerge for the data-focused and learning-focused fields.

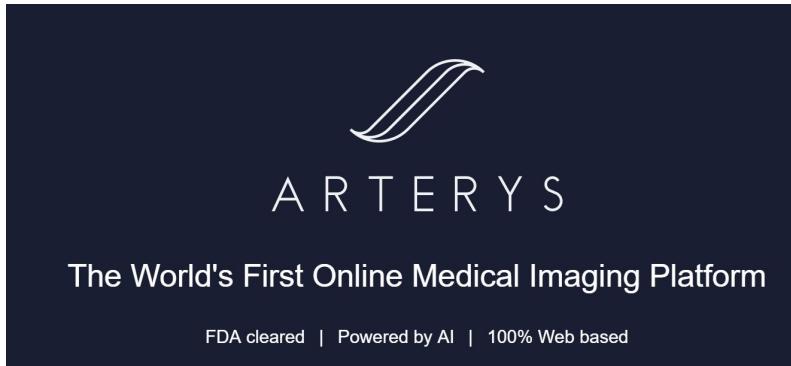
The problem with deep learning

I will resist giving this emerging discipline a name, **but if the acronym “AI” continues to be used as placeholder nomenclature going forward, let’s be aware of the very real limitations of this placeholder.** Let’s broaden our scope, tone down the hype and recognize the serious challenges ahead.

But ... companies are offering this service **now**

The HeartFlow® Analysis

Using data from a standard CT scan, the non-invasive HeartFlow Analysis creates a **personalized 3D model** of the coronary arteries and analyzes the impact that blockages have on blood flow.



Bottom line: how will we use ML in medicine?

- Managing information overload by processing a wider array of data than humans can (i.e. differential diagnosis; precision medicine)

Bottom line: how will we use ML in medicine?

- Managing information overload by processing a wider array of data than humans can (i.e. differential diagnosis; precision medicine)
- More consistently than humans (scanning lungs for nodules, segmenting lesions, checking screening)

Bottom line: how will we use ML in medicine?

- Managing information overload by processing a wider array of data than humans can (i.e. differential diagnosis; precision medicine)
- More consistently than humans (scanning lungs for nodules, segmenting lesions, checking screening)
- Faster than humans

Bottom line: how will we use ML in medicine?

- Avoiding stupid mistakes is much easier than enhancing the best clinicians: whatever happened to Watson (2011)?
- Parsing the oncological literature is harder than Wikipedia, which strives for NPOV: much of the art of interpreting studies is in critically evaluating their methods. Can't uncritically take claims at face value.



Bottom line: how will we use ML in medicine?

- Bringing up the bottom >>> enhancing the best



vs



Bottom line: how will we use ML in medicine?

- The ‘last mile’ is very very hard for machine learning in general
 - Integrating different sources of information (medicine isn’t a board game)
 - Shifts in the distribution (Michael Jordan’s fetal ultrasound example)
 - One-off cases (‘zero shot learning’)

Bottom line: how will we use ML in medicine?

- Think hard about the problem an algorithm solves - is it worth solving?

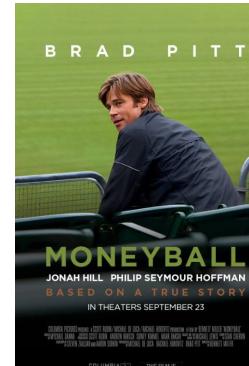
Bottom line: how will we use ML in medicine?

- Think hard about the problem an algorithm solves - is it worth solving?



Bottom line: how will we use ML in medicine?

- You should be excited, but not afraid
- It's "AI", not AI
- Figuring out how to use it - a really monumental undertaking
- But its potential is real. Clinicians who get good at using it will be empowered to make better clinical decisions than those who don't.



Bottom line: how will we use ML in medicine?

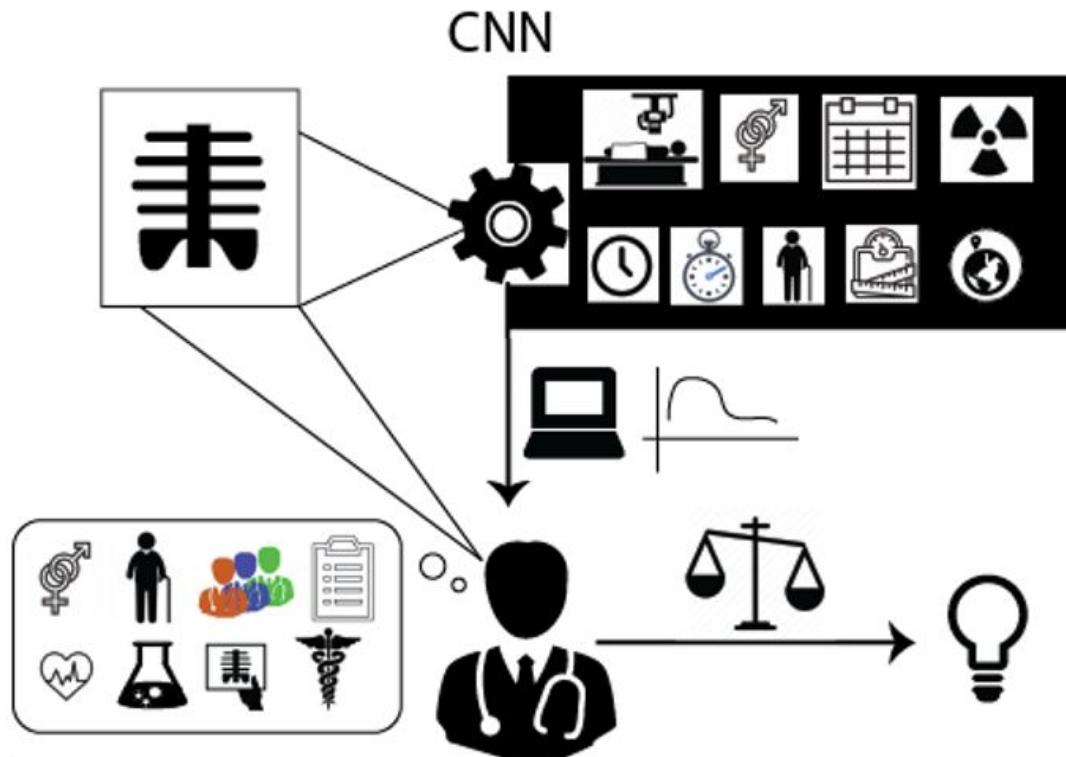


Figure courtesy Marcus Badgeley

Takeaways

- Beware the hype about “artificial generalized intelligence.” Progress in machine learning in last decade has been huge, particularly due to deep learning, but we are a long way from Skynet. Doctors (even radiologists!) aren’t disappearing anytime soon.
- Medicine is a very promising domain to apply these methods because of information overload, but it’s tricky to build models that learn what you want.
- If you think this is cool, take an interest in it - your expertise is needed to drive progress in the “emerging engineering discipline” of applying these methods to medicine.

Thank you!

Further reading:

Rajkomar et al, “Machine Learning in Medicine”, NEJM

Aston Zhang et al, “Dive Into Deep Learning”

More info on my projects:

www.johnzech.com

@johnrzech