

Me 375: Wildfire Science

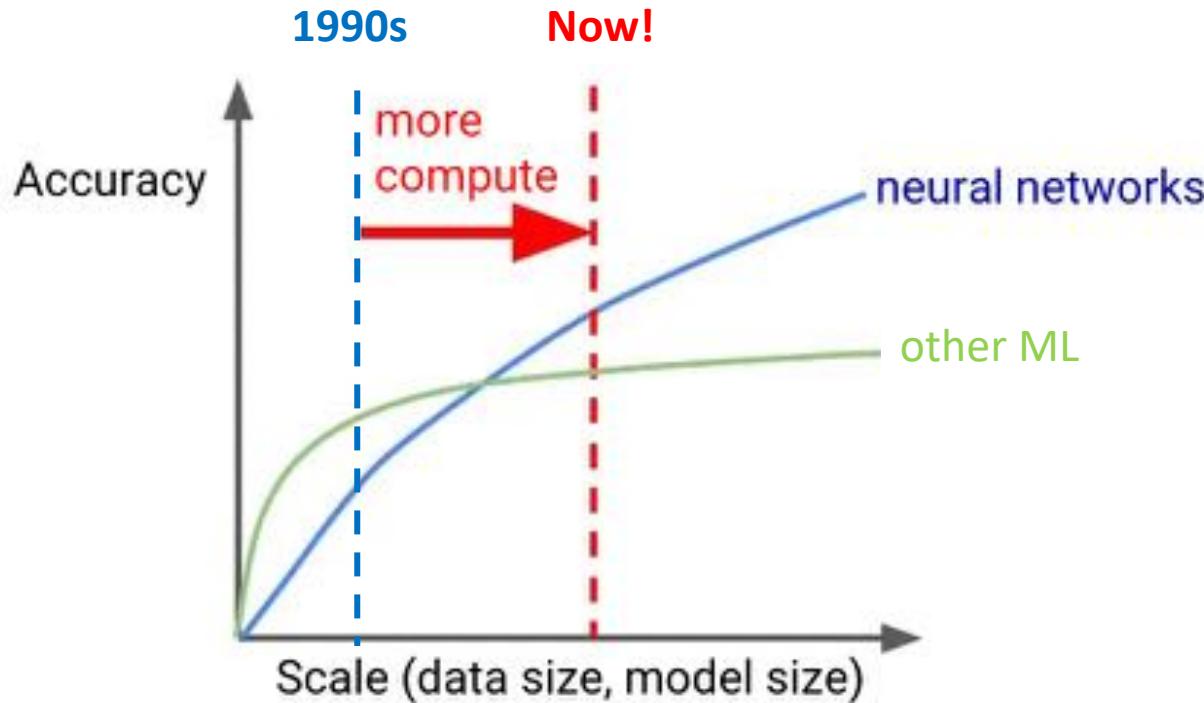
ML Fundamentals

Wai Tong Chung

Agenda

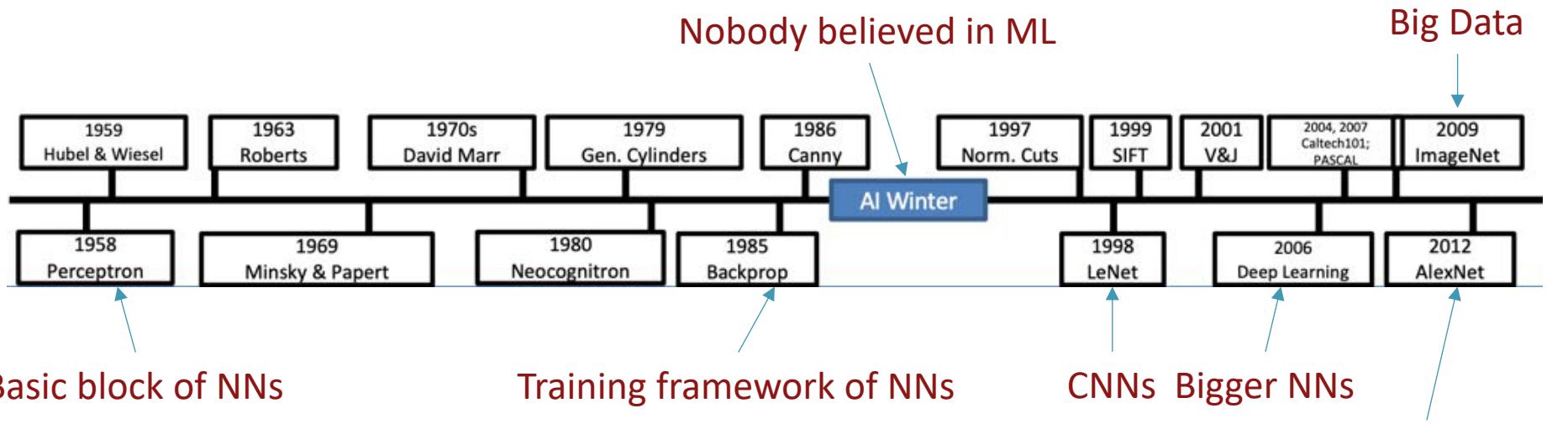
- Machine Learning
- Supervised Learning
- Logistic Regression
- Deep Learning
- ML in practice
 - Evaluation
 - Hyperparameter search
 - Train Test Split
 - Overfitting
- Classification and Regression Trees

Why is ML/Data science important (again)



Jeff Dean, Lecture for YCombinator AI (2017)

Timeline of ML

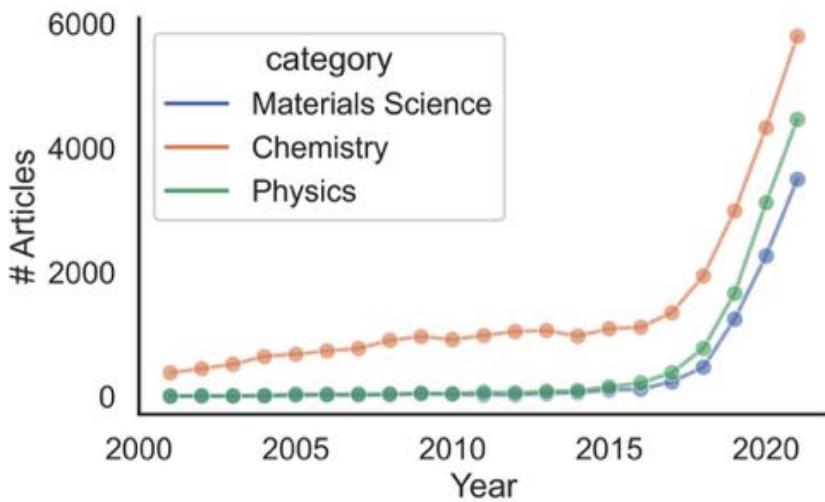


**It took 27 years to believe in a good idea!
(Because of compute, data, etc.)**

Never fear that your methods are too expensive/infeasible

AI/ML for Science

No. of ML for Science papers



One of the Biggest Problems in Biology Has Finally Been Solved

Google DeepMind CEO Demis Hassabis explains how its AlphaFold AI program predicted the 3-D structure of every known protein

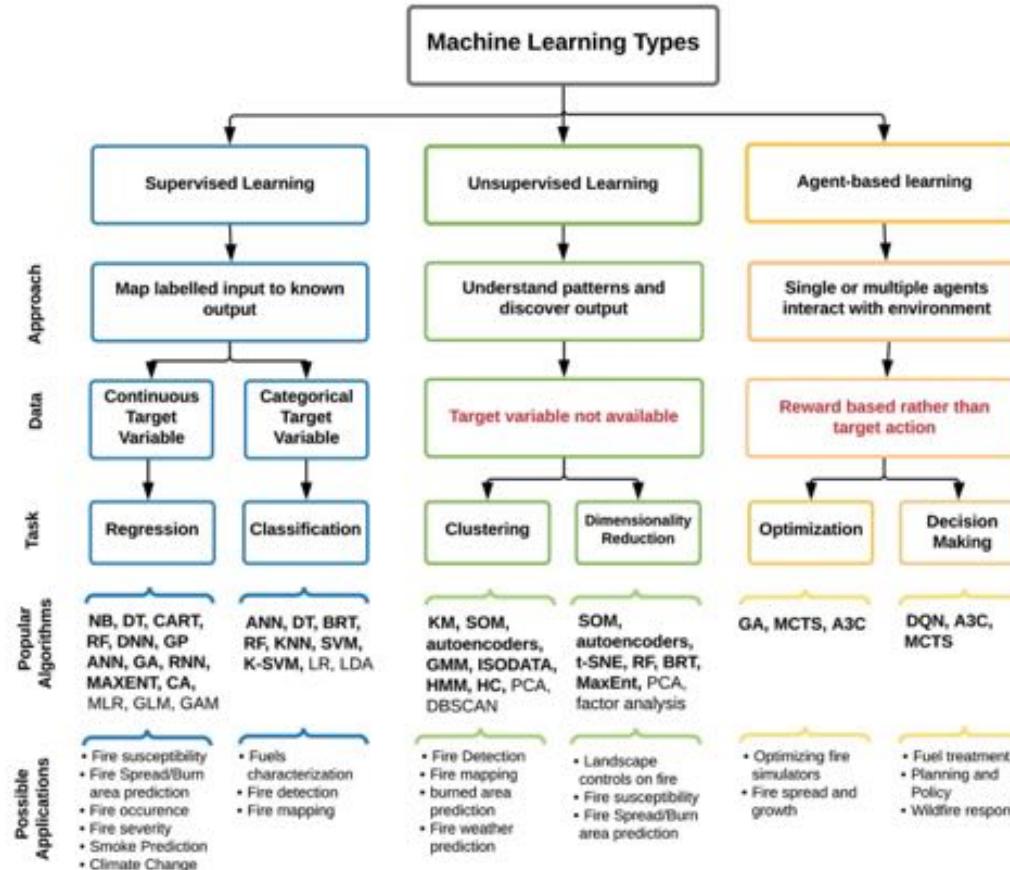
DeepMind breaks 50-year math record using AI; new record falls a week later

AlphaTensor discovers better algorithms for matrix math, inspiring another improvement from afar.

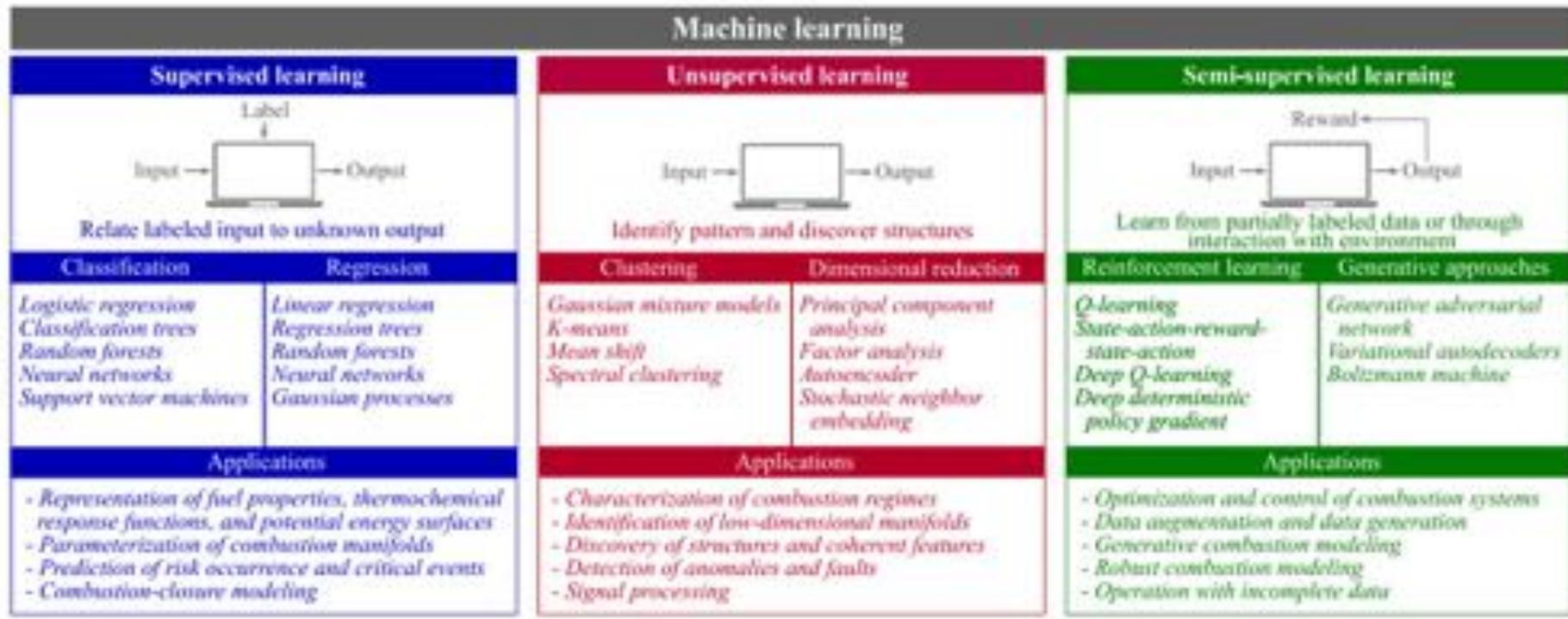
DeepMind & Google's ML-Based GraphCast Outperforms the World's Best Medium-Range Weather Forecasting System

B. Blaizik, Twitter (2022)

Machine Learning for Wildfires

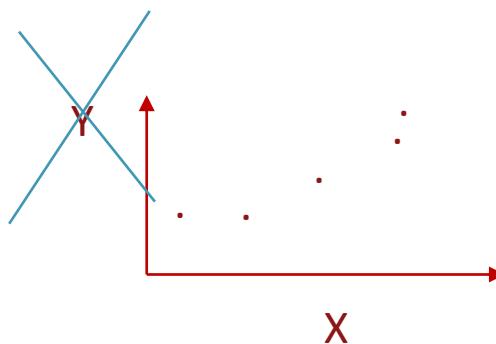


Alternatively: ML for Fundamental Combustion



Ihme et al., Prog. Energy Combust. Sci. (2022)

Unsupervised learning

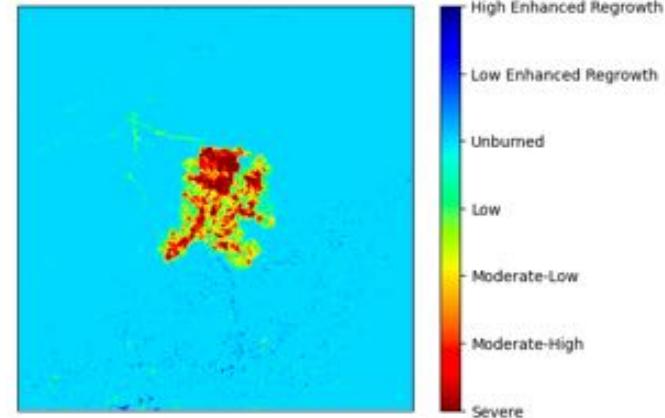


Example: Unprocessed satellite images



An Ideal Clustering
Algorithm

Fire Severity Map



Reality: Not as powerful as supervised learning

Supervised learning - Predictions

Classification



Which leaf is burning?

Ans:

Leaf 1, leaf 2, leaf 3 ...

Regression



How hot is my tree?

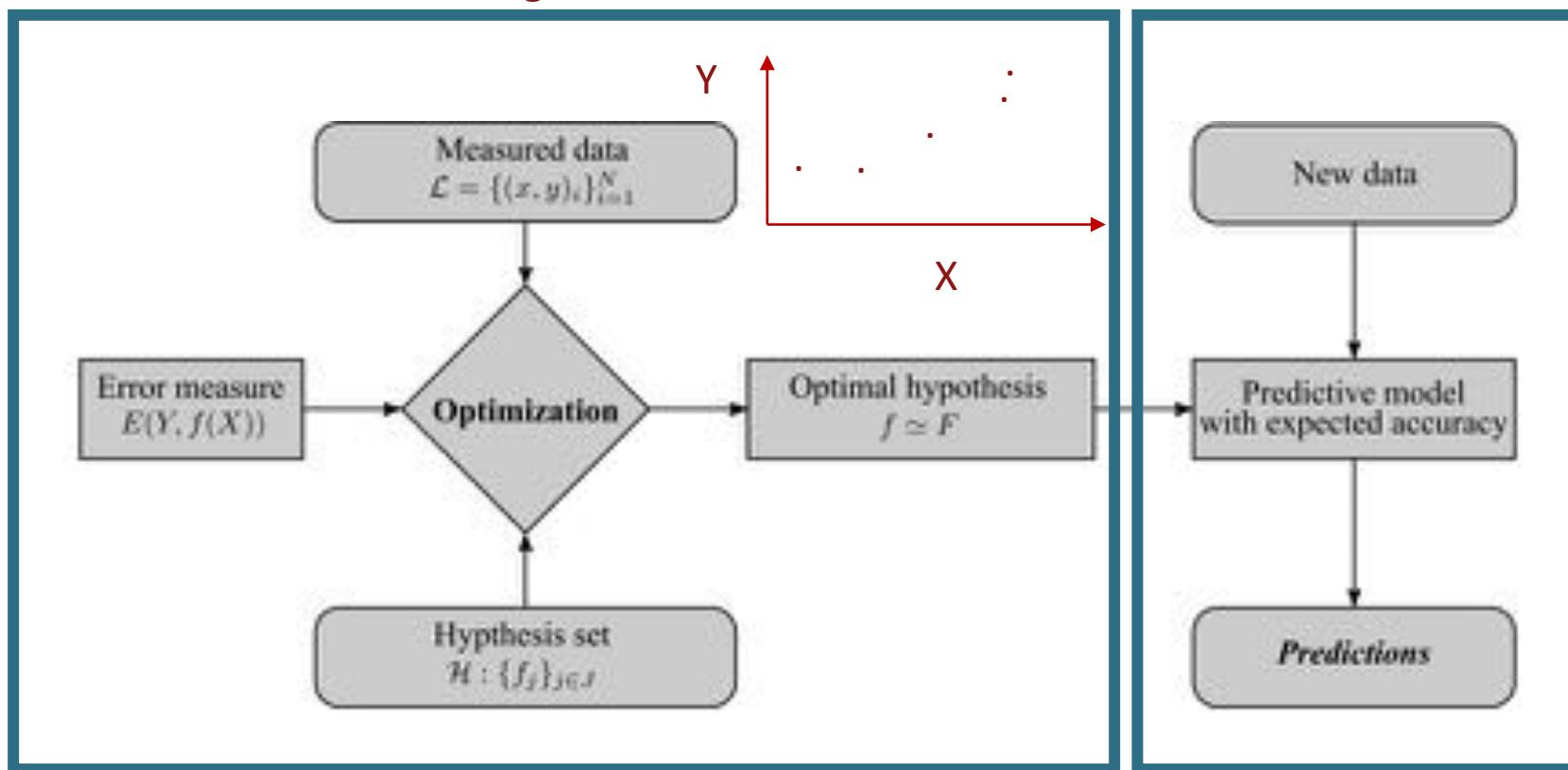
Ans:

Leaf 1 is 735.222,
Leaf 2 is 314.159...

How is this done?

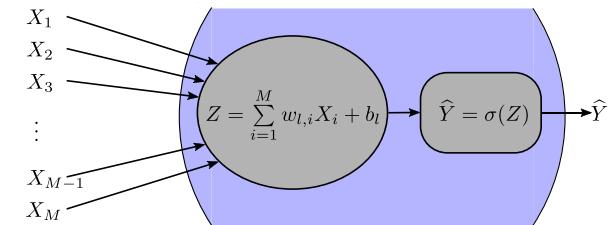
Training

Inferencing



Logistic Regression Classifiers

Simplest Classifiers



Logistic regression

1. Input/features

X_1

X_2

X_3

\vdots

X_{M-1}

X_M

2. Linear layer

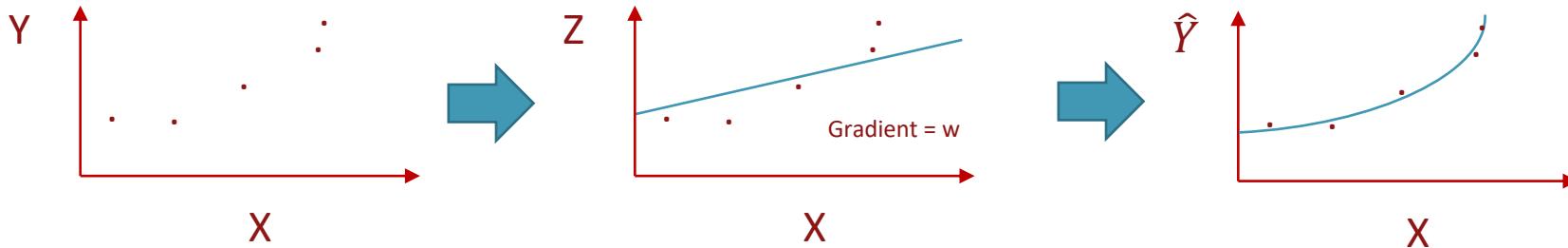
$$Z = \sum_{i=1}^M w_{l,i} X_i + b_l$$

3. Nonlinear function (activation)

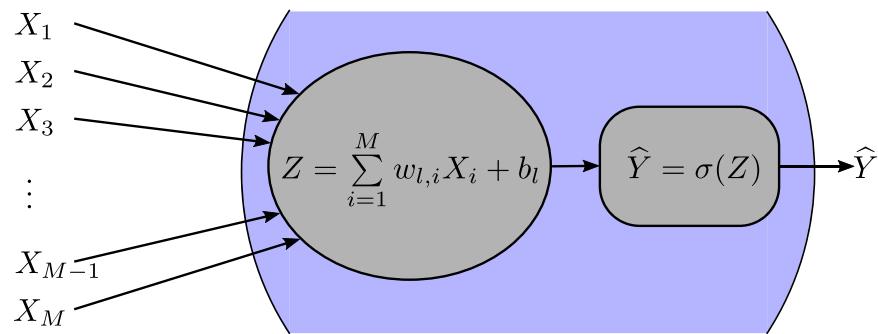
$$\hat{Y} = \sigma(Z)$$

$$\hat{Y}$$

Output:
Probability of a
sample
belonging to a
class, e.g.
 $P(\text{leaf burning})$

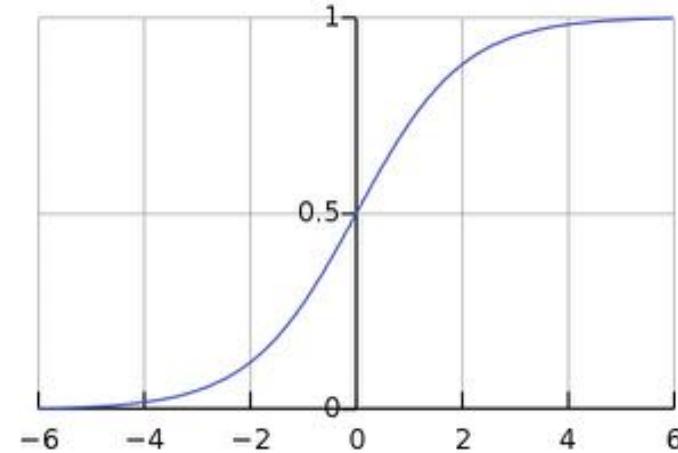


Logistic regression - classification



- Output \hat{Y} is the Probability, e.g. $P(\text{leaf burning})$.
- σ is a sigmoid function

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} = 1 - S(-x).$$



- Does 2 things:
 - Normalizes your prediction between 0 and 1 (Like probability)
 - Pushes small values to 0, big values to 1

Training logistic regression - Likelihood

We want to maximize the likelihood of an event.

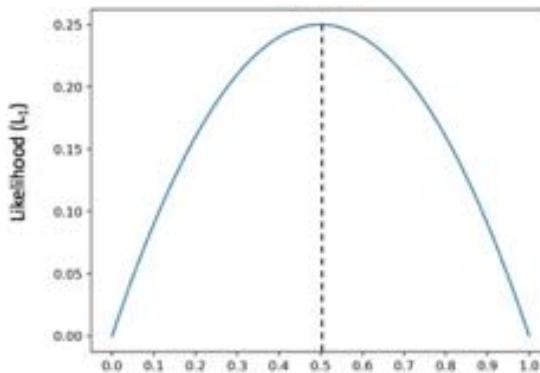
Likelihood is the product of all possible probability functions

For a 2 coin toss

Outcome = [H,T]

$$L_1 = P(H)P(T)$$

$$L_1 = Y(1 - Y)$$



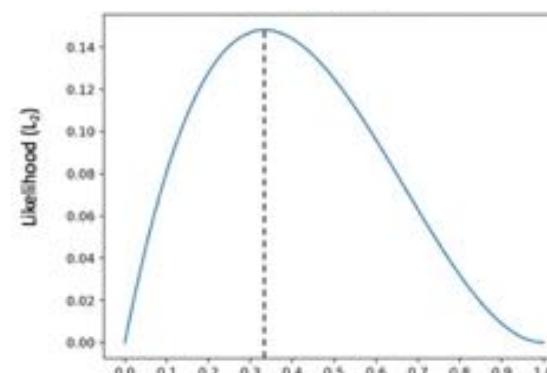
$$Y = P(H \mid \text{Outcome} = [\text{H}, \text{T}])$$

For a 3 coin toss

Outcome = [H,T,T]

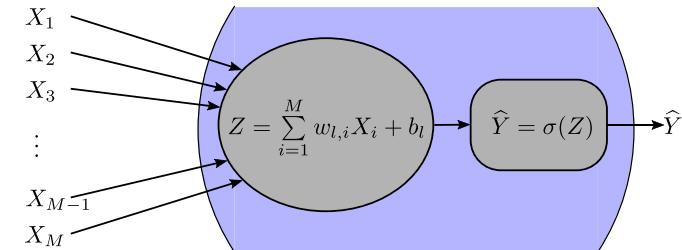
$$L_2 = P(H)P(T)P(T)$$

$$L_2 = Y(1 - Y)(1 - Y)$$



$$Y = P(H \mid \text{Outcome} = [\text{H}, \text{T}, \text{T}])$$

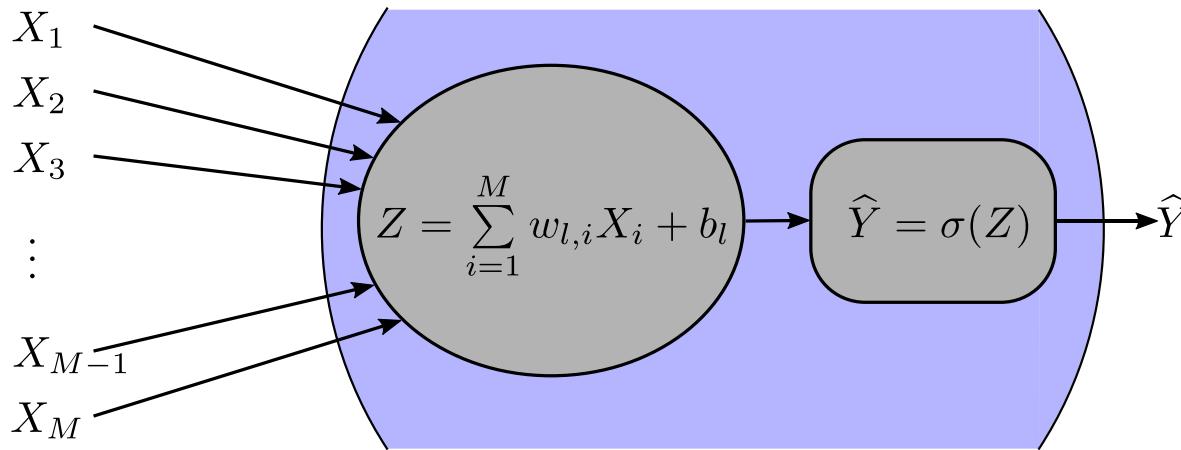
Max L \rightarrow Correct probability



Maximize this function

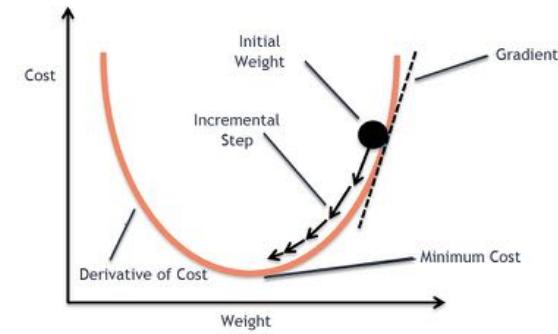
$$L(\hat{Y}(w)) = \prod_{i=1}^n \hat{Y}^{class}(1 - \hat{Y})^{1-class}$$

Training logistic regression

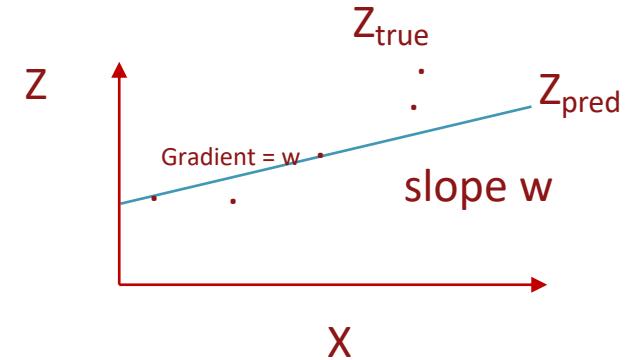
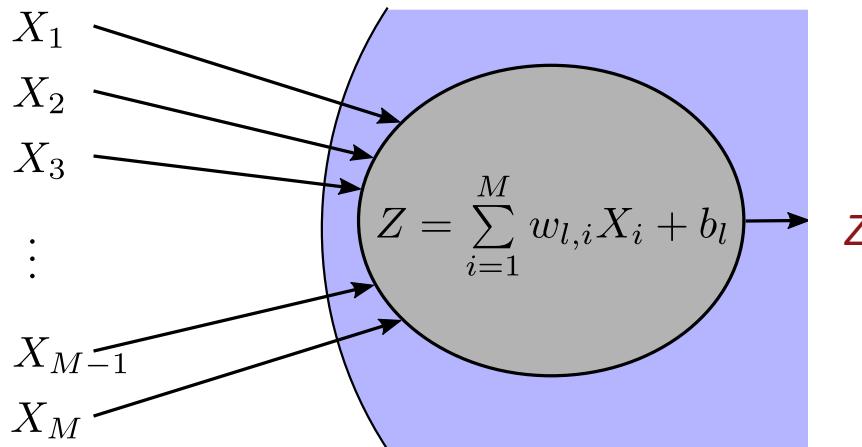


Find w that maximizes $L(\hat{Y}(w))$

- We do this via Optimization Algorithms
- The most popular is Gradient Descent-based
 - Take negative log of likelihood – minimize cross entropy
- You can do this via `model.fit` in most python ML packages

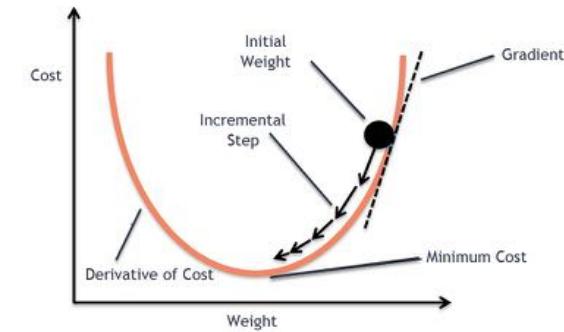


Similar to linear regression



Find w that minimizes error $\text{MSE}(Z_{\text{true}}, Z_{\text{predicted}})$ or MAE

- We do this via Optimization Algorithms
- The most popular is Gradient Descent-based
- You can do this via `model.fit` in most python ML packages

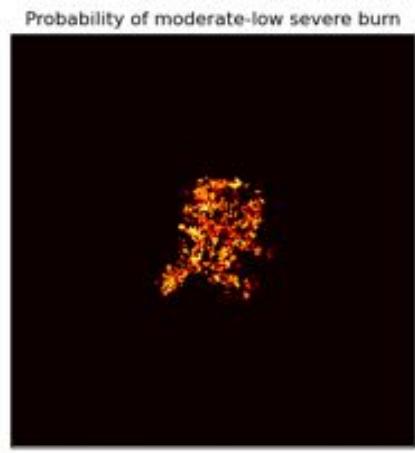


Example of Predicting Fire Severity

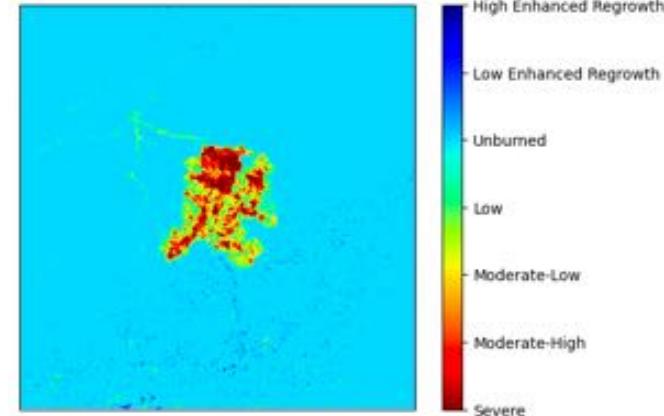
ML Input X



ML Output \hat{Y}

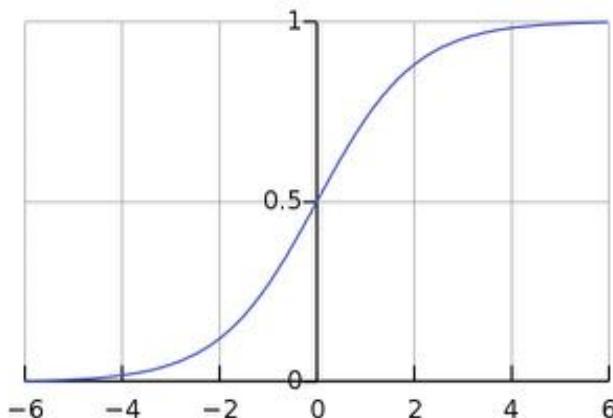
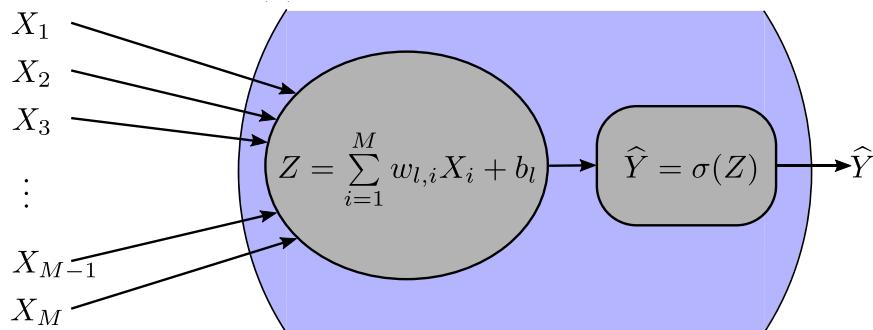


Fire Severity Map

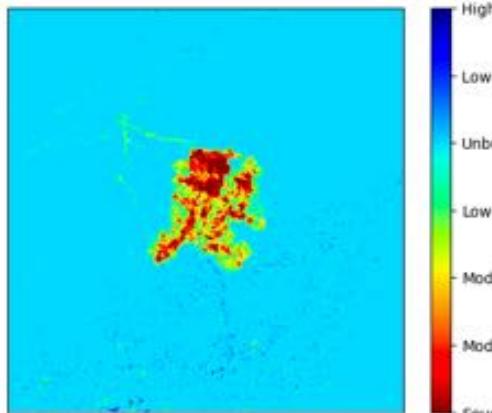


What's Missing?

Multiclass logistic regression

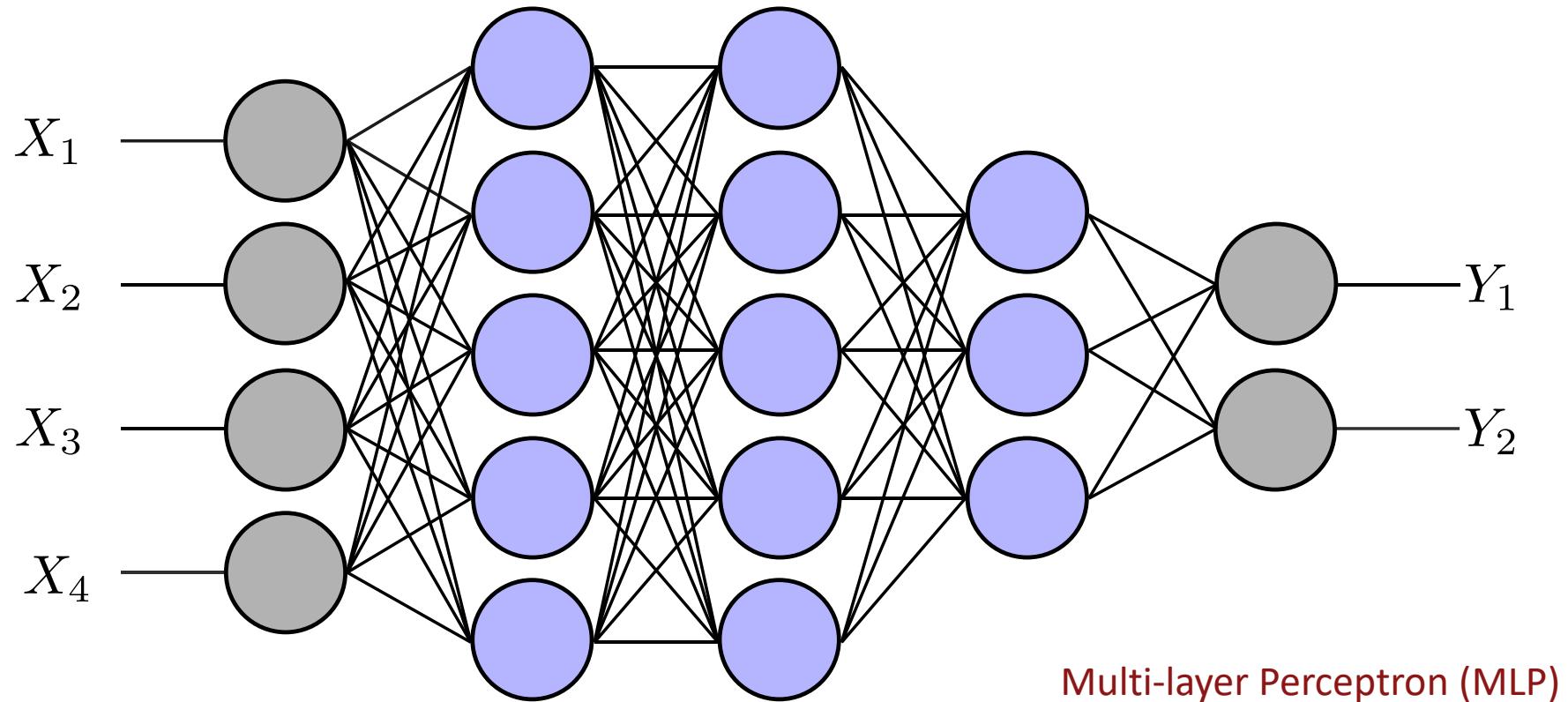


- So far we've only discussed binary classification.
- How to do multiclass classification?



Class	Labels						
High Regrowth	1	0	0	0	0	0	0
Low Regrowth	0	1	0	0	0	0	0
Unburned	0	0	1	0	0	0	0
Low	0	0	0	1	0	0	0
Moderate-Low	0	0	0	0	1	0	0
Moderate-High	0	0	0	0	0	0	1
Severe	0	0	0	0	0	0	1

Neural Networks – A bunch of connected logistic regression nodes



Deep learning

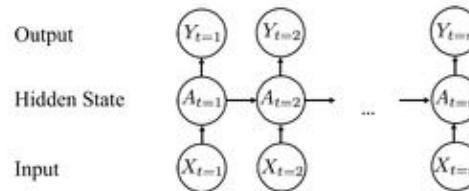
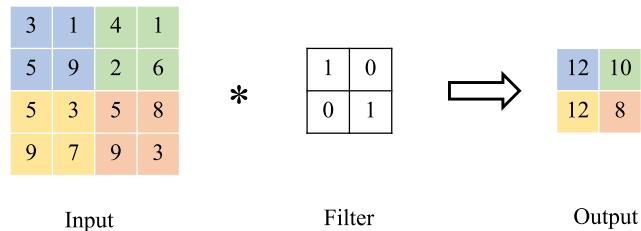
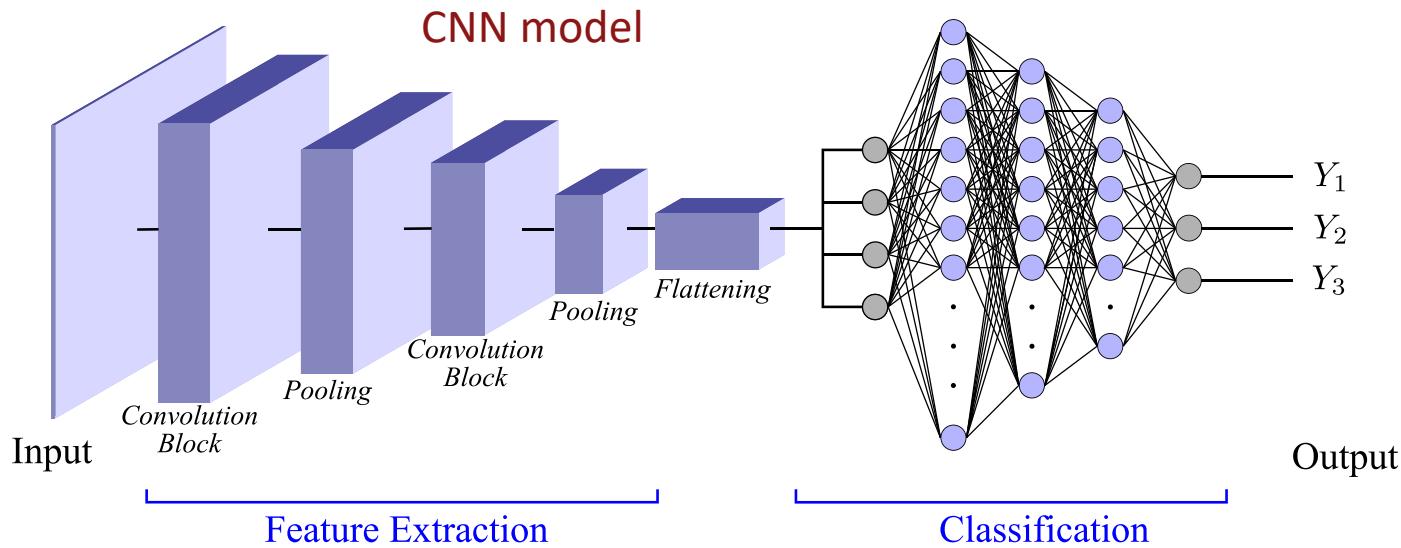


Fig. 16. Schematic of a many-to-many RNN.

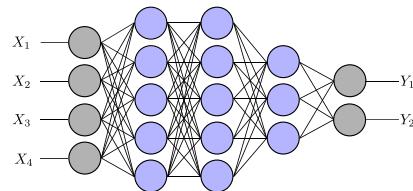
RNN model

Transformer architectures

ChatGPT



175B parameters
\$5M to train



76 param model

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usze@google.com

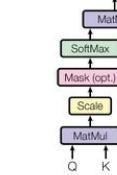
Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukasz.kaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Scaled Dot-Product Attention



Multi-Head Attention

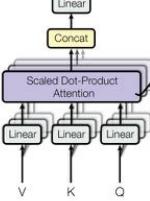


Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

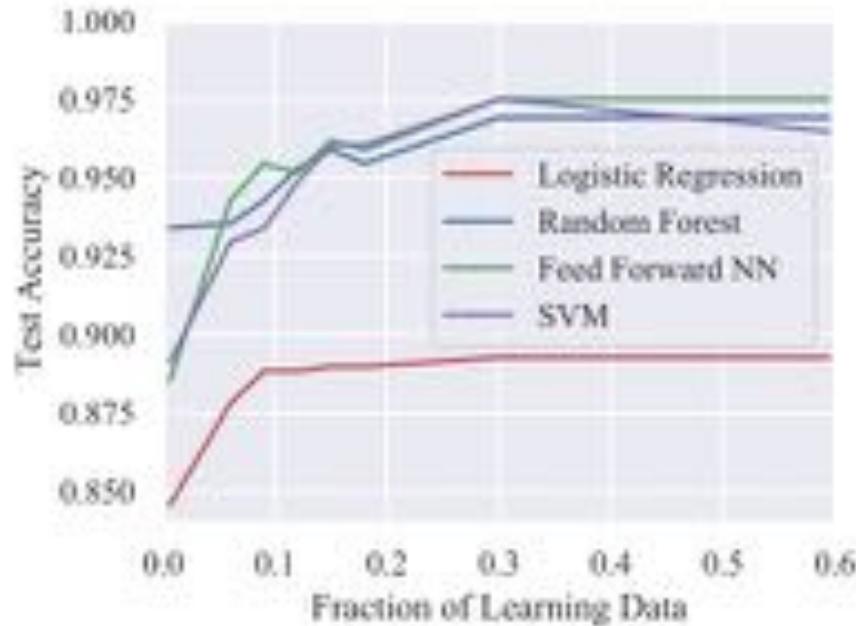
Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.8		$2.3 \cdot 10^{19}$

Algorithms

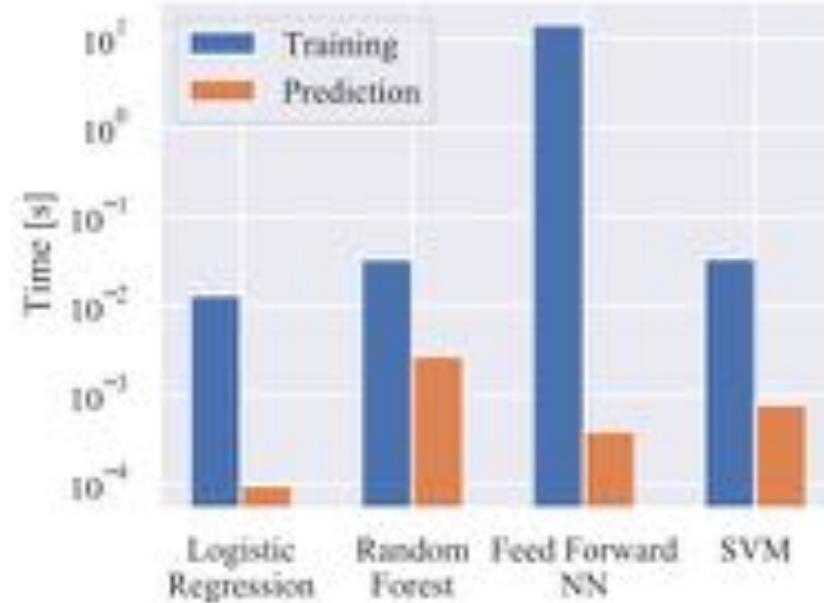
Classification	Regression
<i>Logistic regression</i> <i>Classification trees</i> <i>Random forests</i> <i>Neural networks</i> <i>Support vector machines</i>	<i>Linear regression</i> <i>Regression trees</i> <i>Random forests</i> <i>Neural networks</i> <i>Gaussian processes</i>

How do I choose the right model?

Typical trends in ML models

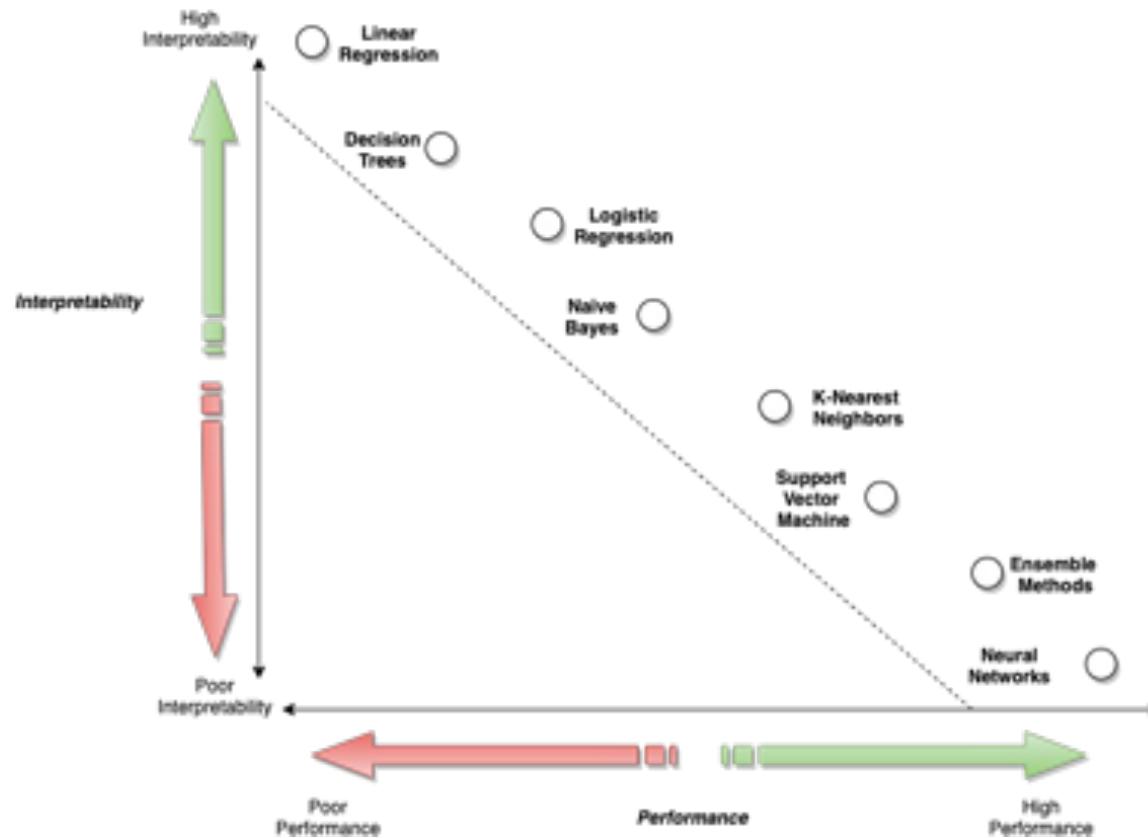


(a) Accuracy



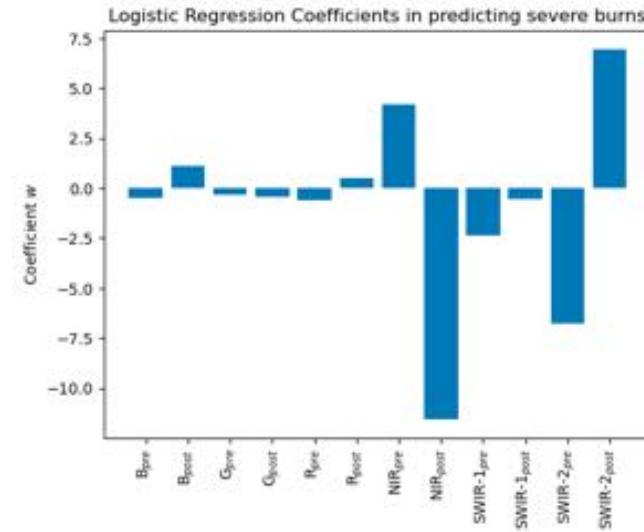
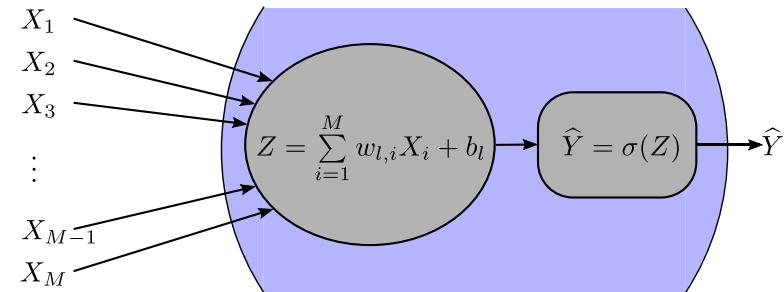
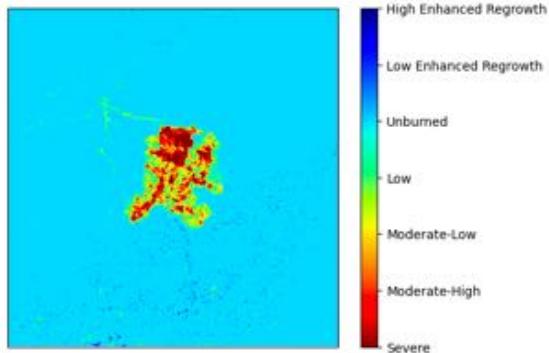
(b) Cost

Interpretability



Amazon White Paper (2023)

Logistic regression Interpretability

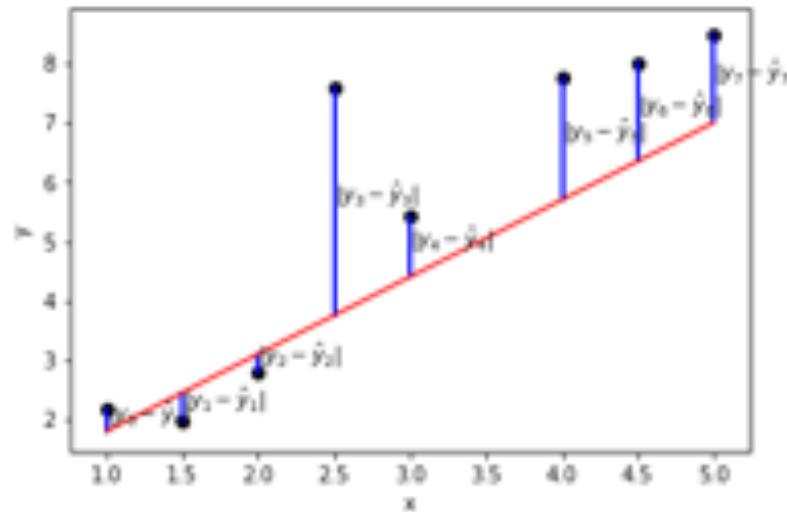




ML in Practice

Evaluation

- Regression
 - MAE
 - RMSE
- Classification
 - Accuracy
 - Precision
 - Recall
 - F1-Score



$$MAE = \frac{1}{n} \sum_{j=1}^n |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_i - \hat{y}_i)^2}$$

Gives more weight to outliers

Classification Metrics

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

Only gives info on True Predictions, not on mistakes

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

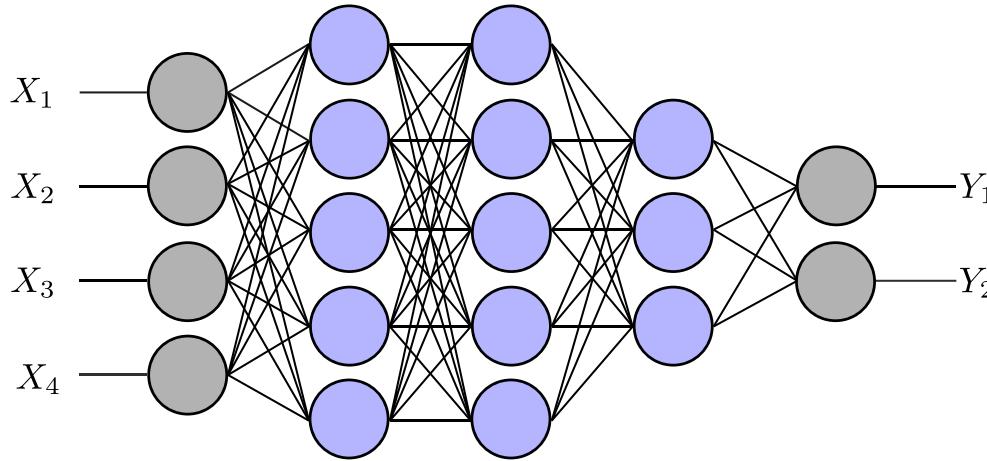
$$F1 = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Lets you know when you have False Positives. E.g. Bank Fraud

Lets you know when you have False Negatives. E.g. COVID

Gives both FP and FN
E.g. Search applications

Hyperparameter search



How many neurons in each layer?

How many layers?

What optimization scheme?

How much data?

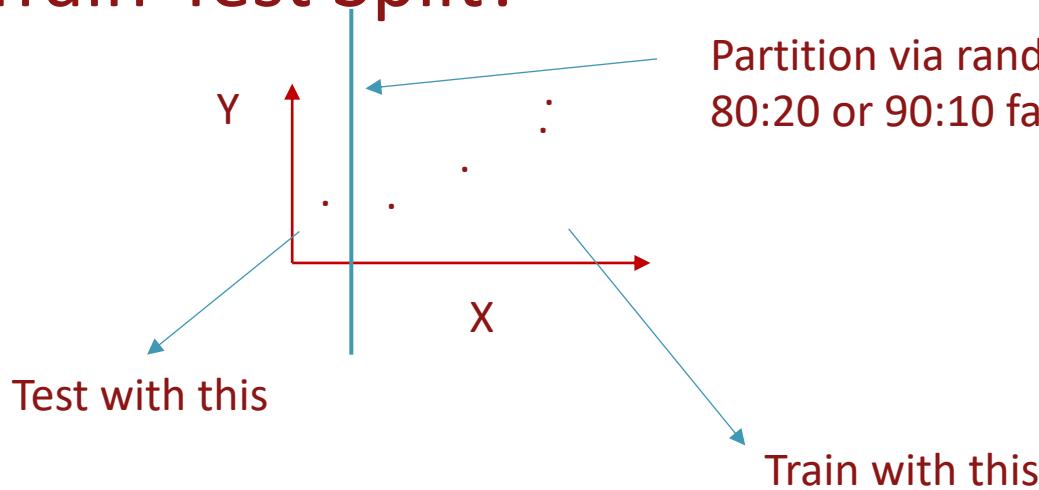
What forward operation?

What non-linearity?

NN or other method?

- 1. Intuition/Domain Knowledge
- 2. Exhaustive Search (Expensive!)
- 3. Random Search (Not efficient)
- 4. Bayesian Optimization (Gaussian Assumptions)
- 5. Reinforcement Learning
- 6. AutoML
- 7. ...

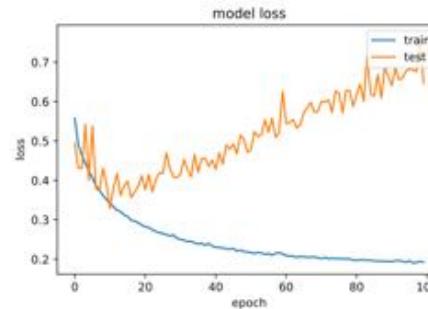
Why do a Train-Test Split?



Why? -> Can Diagnose Overfitting

Loss is your objective function

- Likelihood
- Error between points



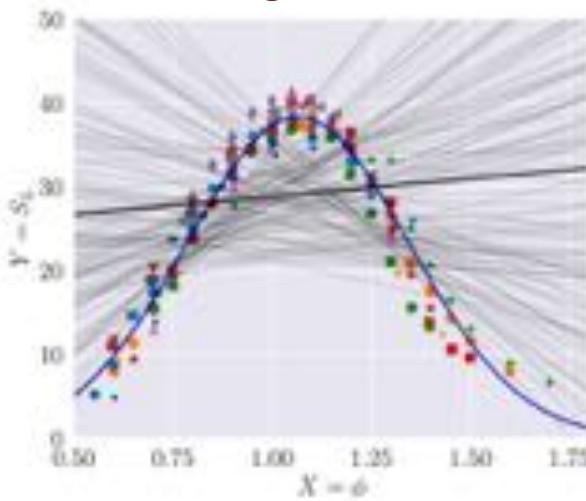
This can happen if your hyperparameter search arrives at a model that only suits your train data

Why does overfitting occur?

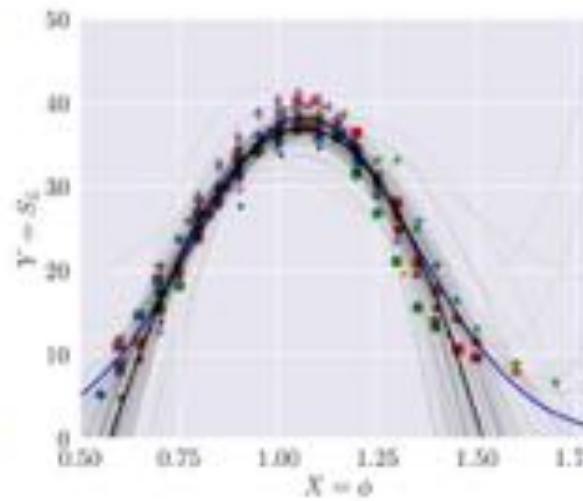
- We train 100 linear regression models of different complexity
- More complex models (more layers, etc) can arrive at high bias.

Underfit

High variance

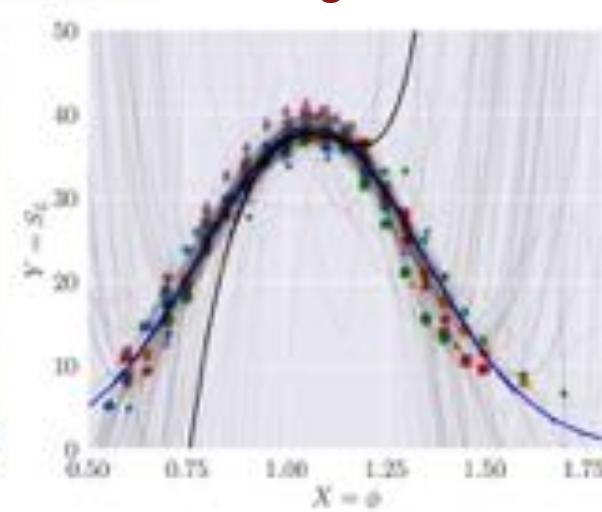
(a) 1st order regression.

Decent fit

(b) 3rd order regression.

Overfit

High-bias

(c) 5th order regression.

Better split: Train-test-validation split

- Train-test split let's you diagnose overfitting!
 - Doesn't actually do anything to stop this
- Train-val-test split can help alleviate overfitting
 - Hyper-parameter search on a validation set that is different from train and test set
- Partition via random selection in an 80:10:10 or 90:5:5 fashion.

Overfitting can happen
if your hyperparameter
search arrives at a
model that that only
suits your train data

Other ways to deal with overfitting

- Regularization
- Data augmentation
- Dropout
- Batch Normalization

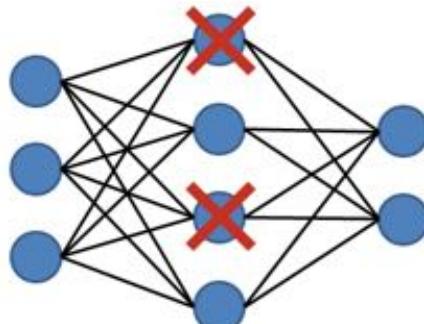
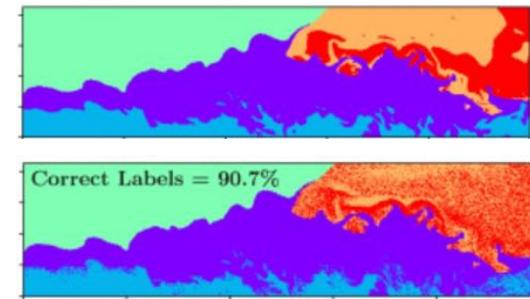
L1 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M |W_j|$$

L2 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j^2$$

Loss function Regularization Term



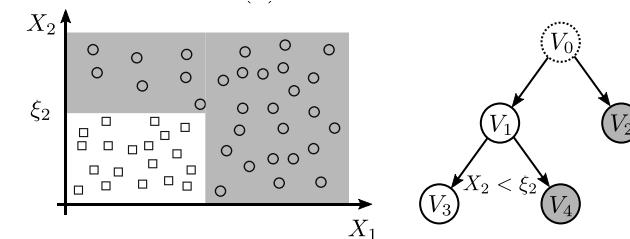
Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1 \dots m\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

```

 $\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$  // mini-batch mean
 $\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2$  // mini-batch variance
 $\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$  // normalize
 $y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)$  // scale and shift
    
```

Classification and Regression Trees (CARTs)

An Interpretable Staple for Tabular Data



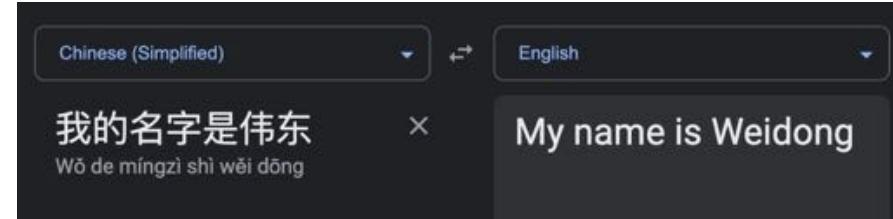
Popular Data types

Images/Spatial (N-D)



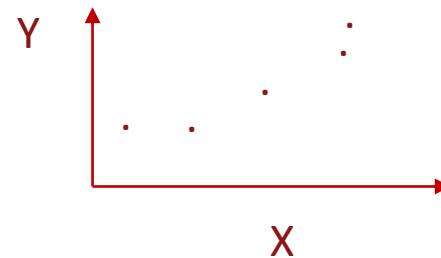
CNNs, Vision Transformers

Language/Sequential (1D)



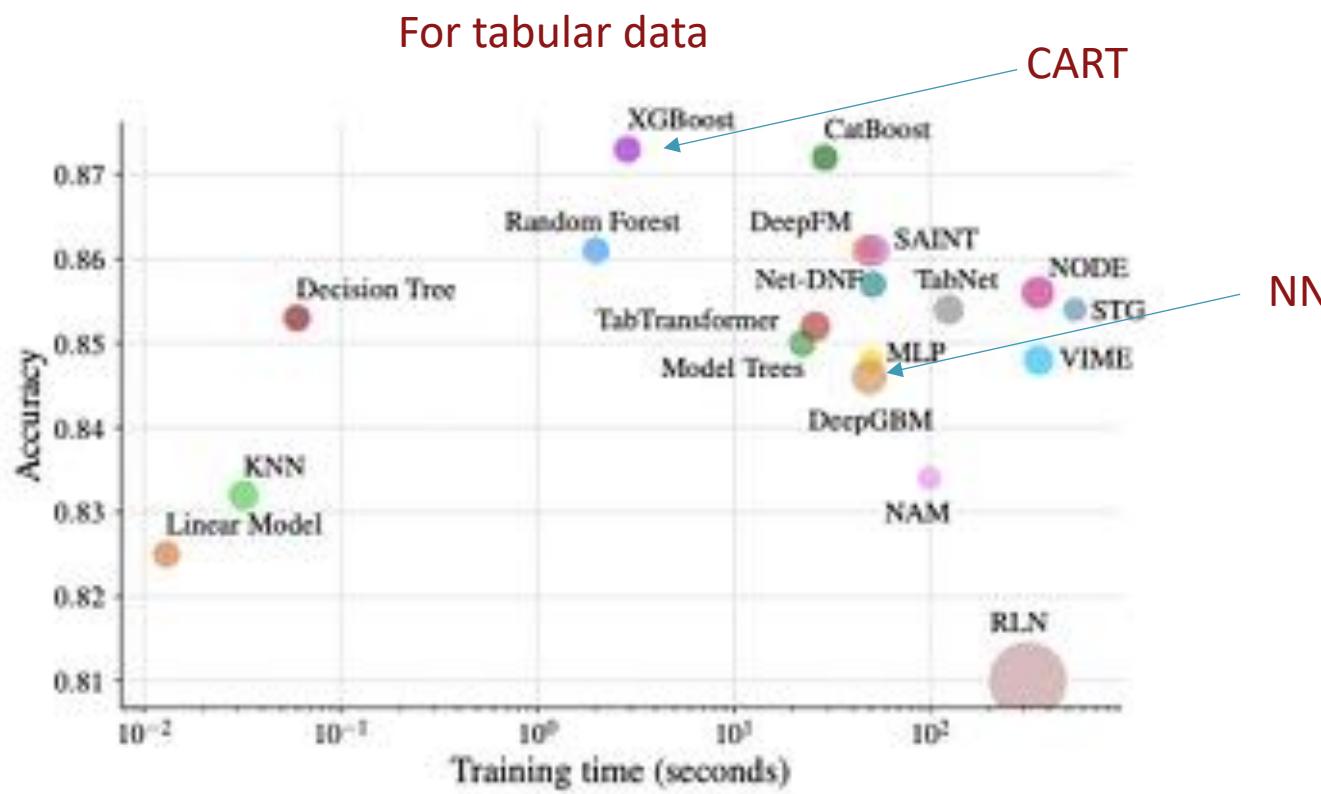
RNNs, Transformers, LSTMs

Tabular Data (0D)

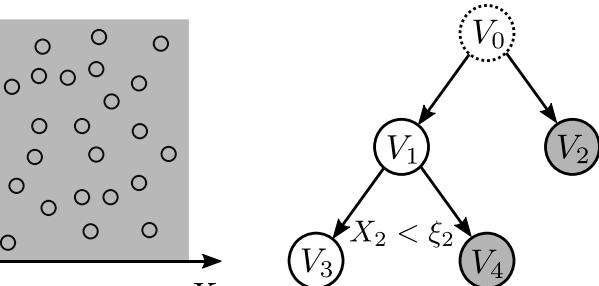
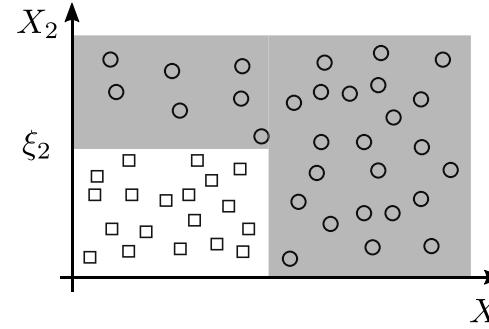
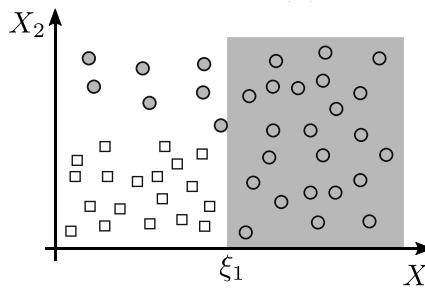
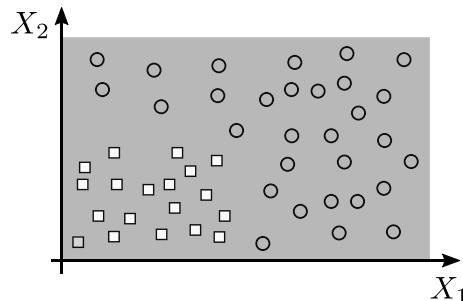


???

Why care about CARTs?



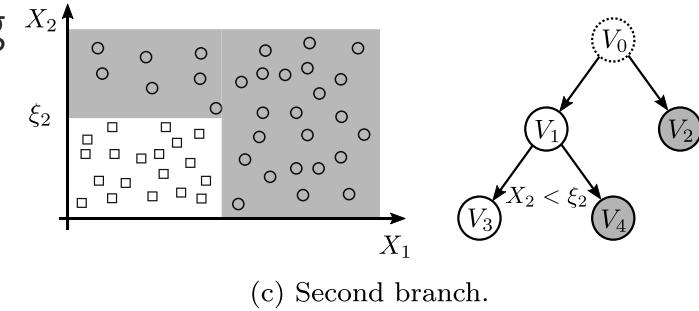
How a Decision Tree Classifier Works



Can inspect node structures to interpret model behavior -> Feature importance score (Interpretable)

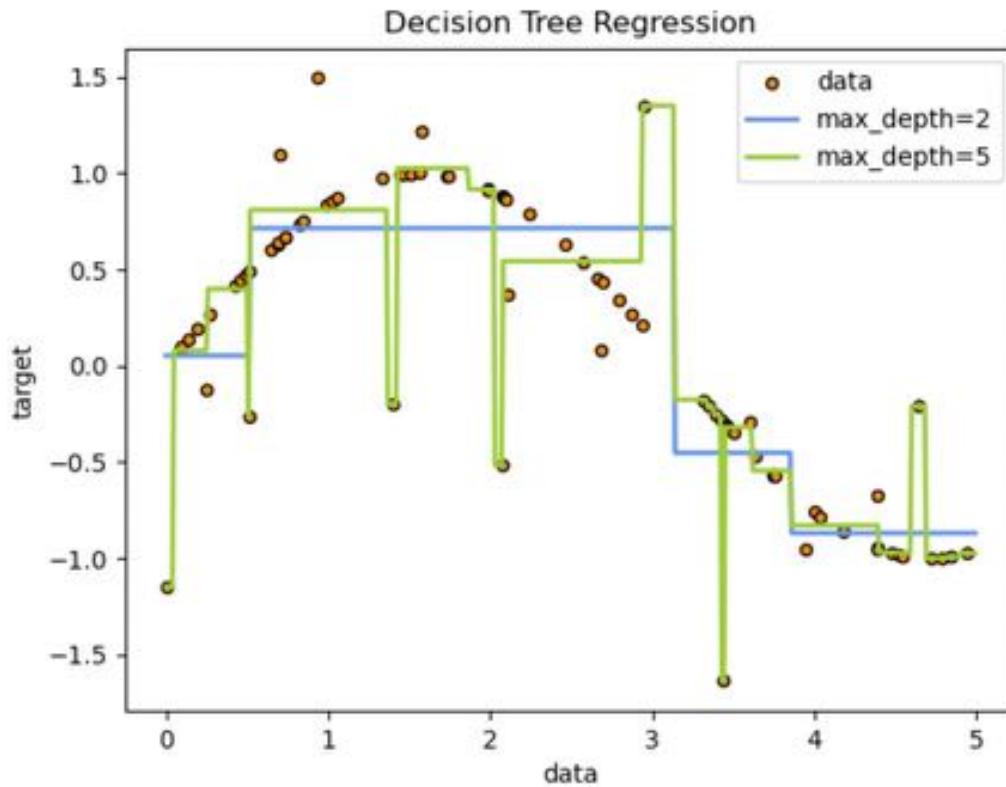
Training a decision tree

- Regression in logistic regression/deep learning
 - Find slope of function (weights) that minimize **error** between predicted and target variable.
- Classification in logistic regression/deep learning
 - Find slope of function (weights) that maximizes the likelihood of predicted variable.
- Regression in a decision tree
 - Find a split that maximizes “purity”
 - Related to variance
- Classification in a decision tree
 - Find a split that maximizes “information gain”
 - Related to entropy

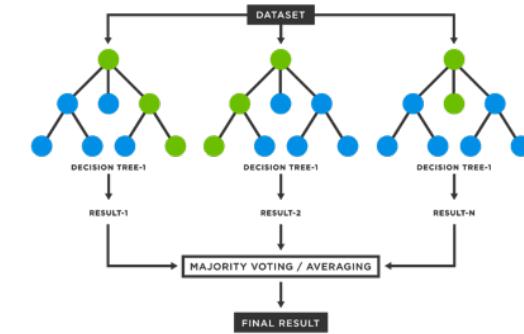


Background in information theory (EE) needed to further understand this

Decision Tree Regression



- Decision tree regressors are may be less suitable for smooth functions
- Not suitable for extrapolation
- Prone to overfitting
 - Rectify with ensemble methods



Conclusions

- Why researchers care about AI/ML?
- How does supervised learning work?
- How does a logistic regression/NN work?
- How to train/test these models in practice?
- Should we care about Non-NN Models?
- Next lecture: How have these models been applied to wildfire problems?

Thank You

Me 375: Wildfire Science

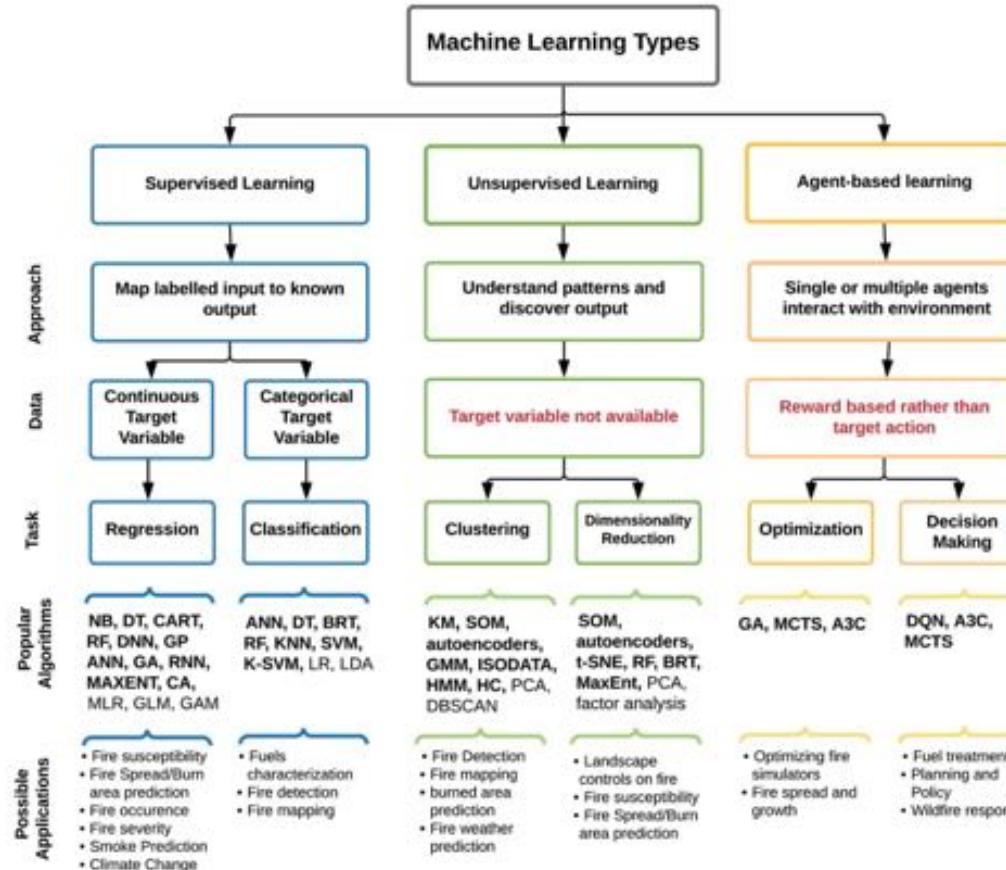
ML Applications

Wai Tong Chung

Agenda

- Last lecture: Fundamentals
- This lecture: Applications, Opportunities, Challenges
 - When to use ML
 - Past ML applications in Wildfires
 - Current ML Trends in Wildfires
 - Open Research Problems

Machine Learning for Wildfires



Possible Applications

- Fire susceptibility
- Fire Spread/Burn area prediction
- Fire occurrence
- Fire severity
- Smoke Prediction
- Climate Change
- Fuels characterization
- Fire detection
- Fire mapping

Supervised learning - Preidictions

Classification



Which leaf is burning?

Ans:

Leaf 1, leaf 2, leaf 3 ...

Regression



How hot is my tree?

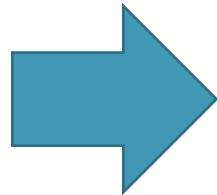
Ans:

Leaf 1 is 735.222,
Leaf 2 is 314.159...

When and how is this useful in the sciences?

When to use ML

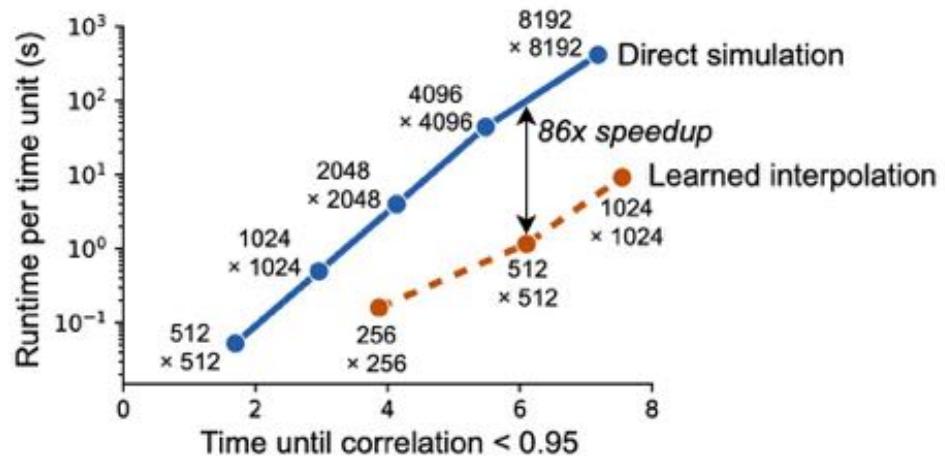
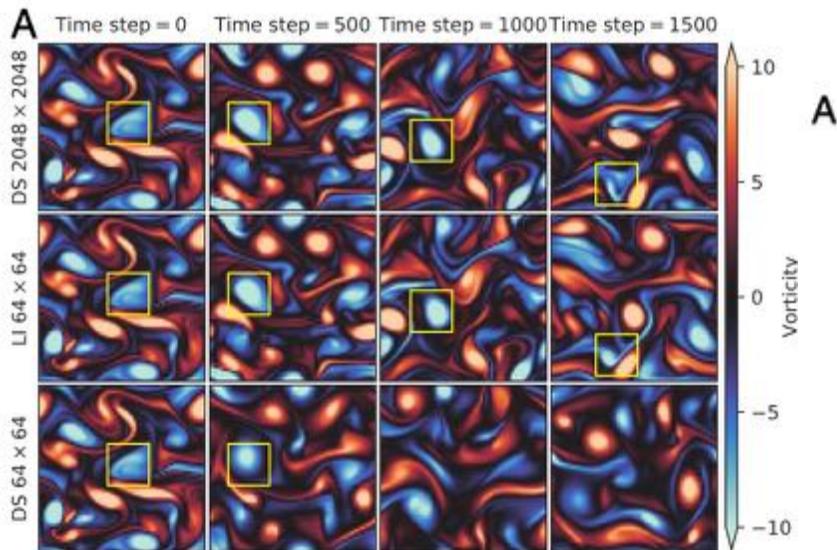
- No reliable relations



Fire spread rate from satellite images?

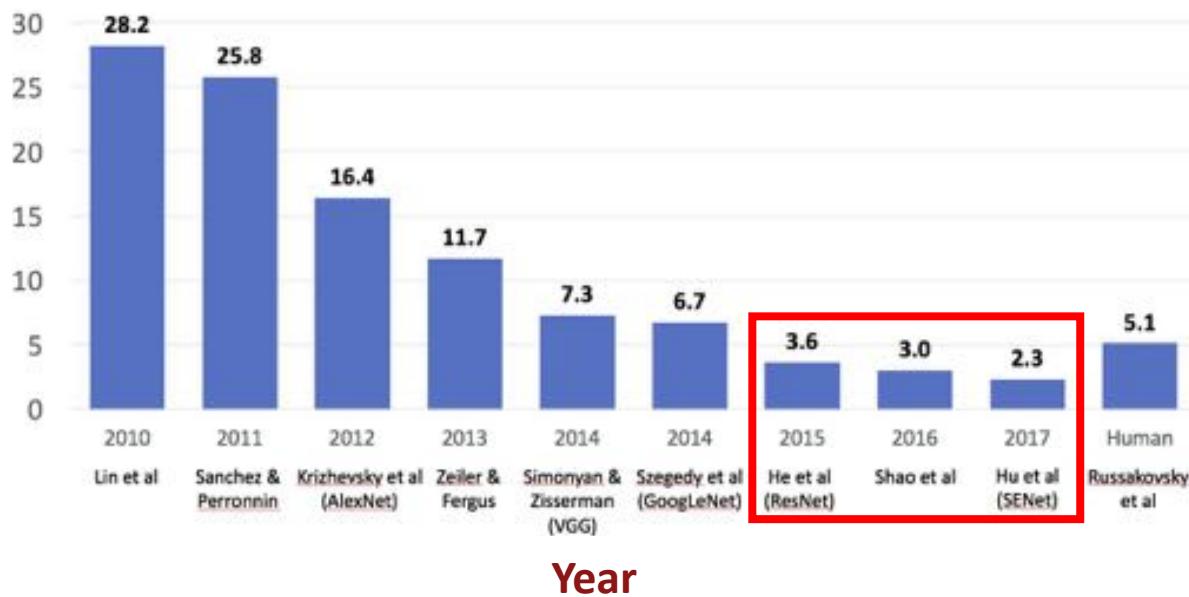
When to use ML

- Costly! -> ODEs/PDEs are expensive to calculate



When to use ML

Universal Approximators: Typical ML Error



Can I trust ~3% ML error
when driving a car?

Tesla behind eight-vehicle crash was in
'full self-driving' mode, says driver

San Francisco crash is the latest in a series of accidents blamed on
Tesla technology, which is facing regulatory scrutiny



Errors must not be too costly!

ML use cases

Not easy to scale



Errors not costly

Sufficient Data

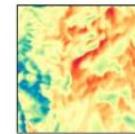
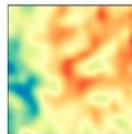


NETFLIX



My PhD research

No reliable relations



TESLA

Cost-effective alternative



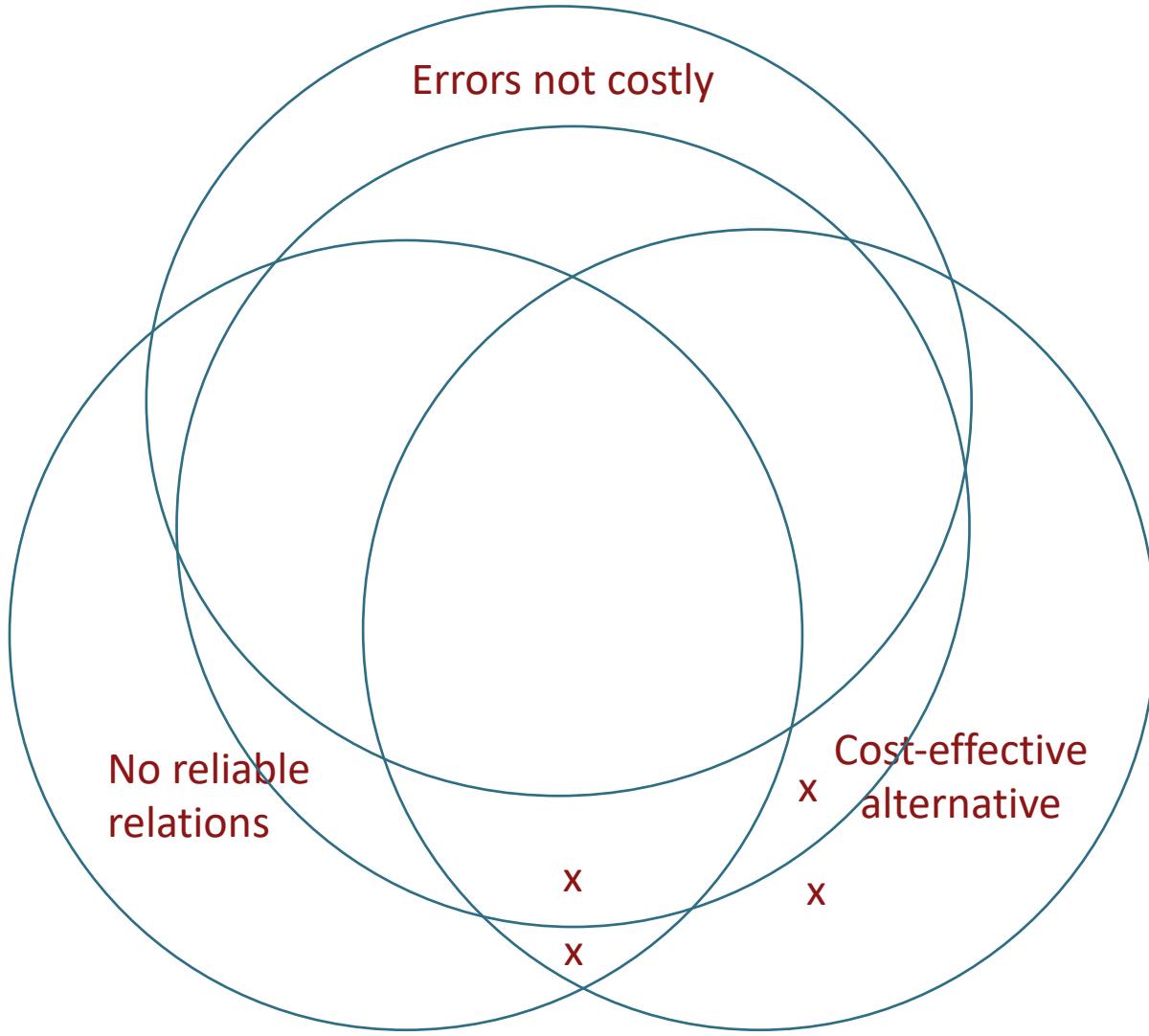
Not easy to productionize

ML use cases

Possible Applications

- Fire susceptibility
- Fire Spread/Burn area prediction
- Fire occurrence
- Fire severity
- Smoke Prediction
- Climate Change
- Fuels characterization
- Fire detection
- Fire mapping

- Sims. are expensive
- Satellite/Aerial image analysis is highly empirical
- Data sources are separate



What are challenges in applying ML to the wildfires?

- Lack of Structured Datasets
 - separate images, sensor data, observational, sim. data
- Insufficient reliability
 - Errors from ML models
 - Insufficient interpretability

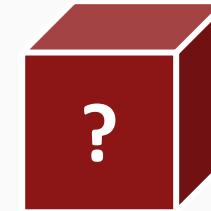
Data challenges



ML Errors

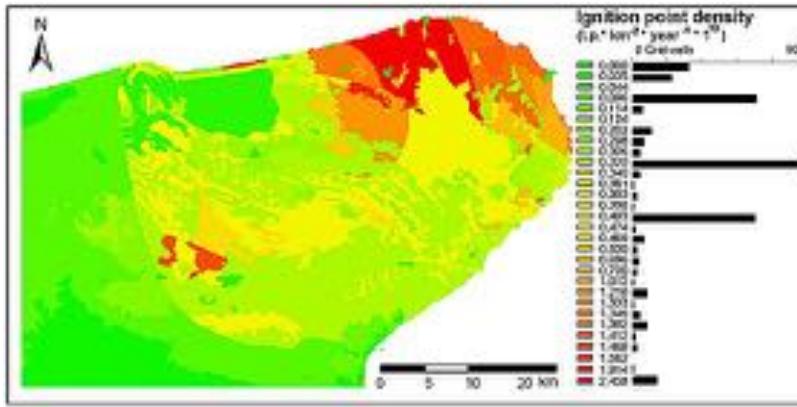


Black-box



Wildfire Risk

Fire ignition points (1997–2003), were used to derive a fire occurrence map through a kernel density approach.



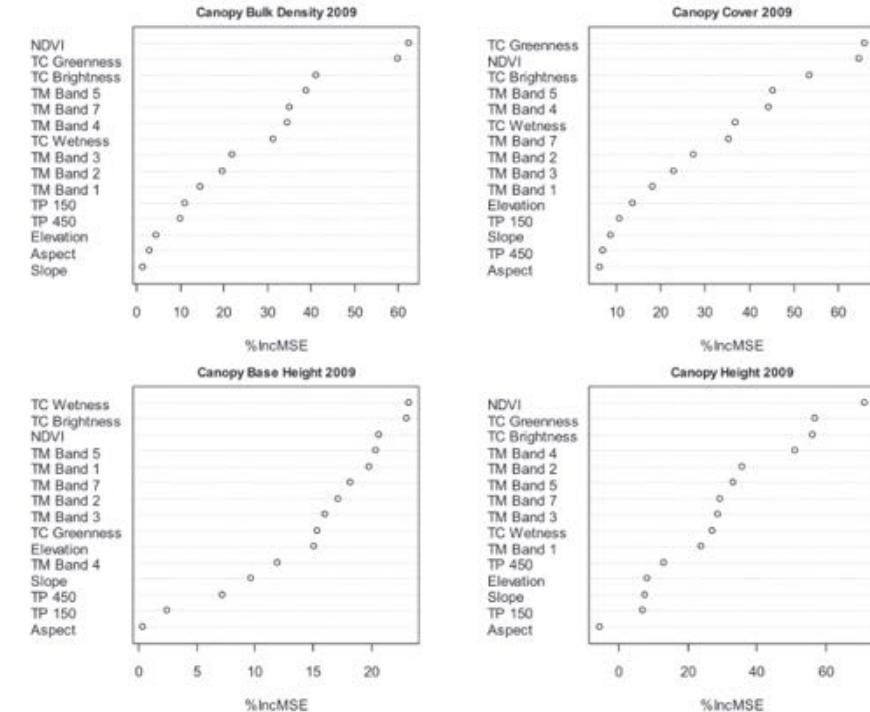
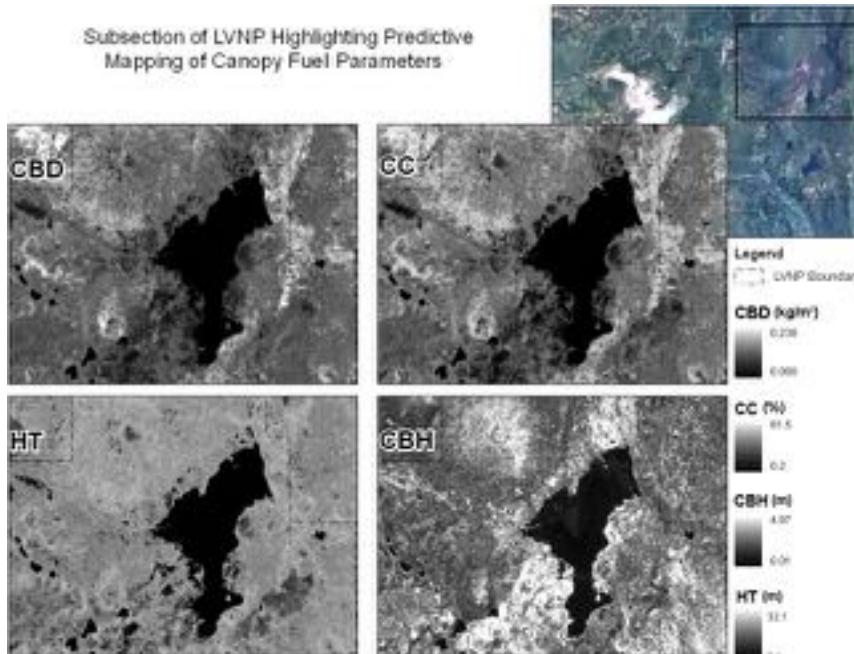
Data provided by the Italian National Database:

continuous urban fabric, discontinuous urban fabric, industrial or commercial units, port areas, mineral extraction sites, nonirrigated arable land, permanently irrigated land, vineyards, olive groves, pastures, annual crops associated with permanent crops, complex cultivation patterns, agriculture land with significant areas of natural vegetation - , broadleaf forest, coniferous forest, mixed forest, natural grasslands, Sclerophyllous vegetation, transitional woodland-shrub, bare rocks, sparsely vegetated areas, salt marshes, water bodies, average temperature of warmest month, average temperature of coldest month, annual average minimum temperature, annual average temperature, annual average rainfall, altitude, annual average maximum temperature, population distribution by CORINE (CLC90), aspect divided in 8 classes + 1 for the flat areas, Slope, mask with Gargano National Park limits, continuous grid distance from secondary roads, continuous grid distance from primary roads, illumination factor

- **Govt Datasets** are impressive
- No discussion on overfitting
- Feasibility study on ML
- Decision tree feature importance is useful
- Prior to ML popularity (2006)

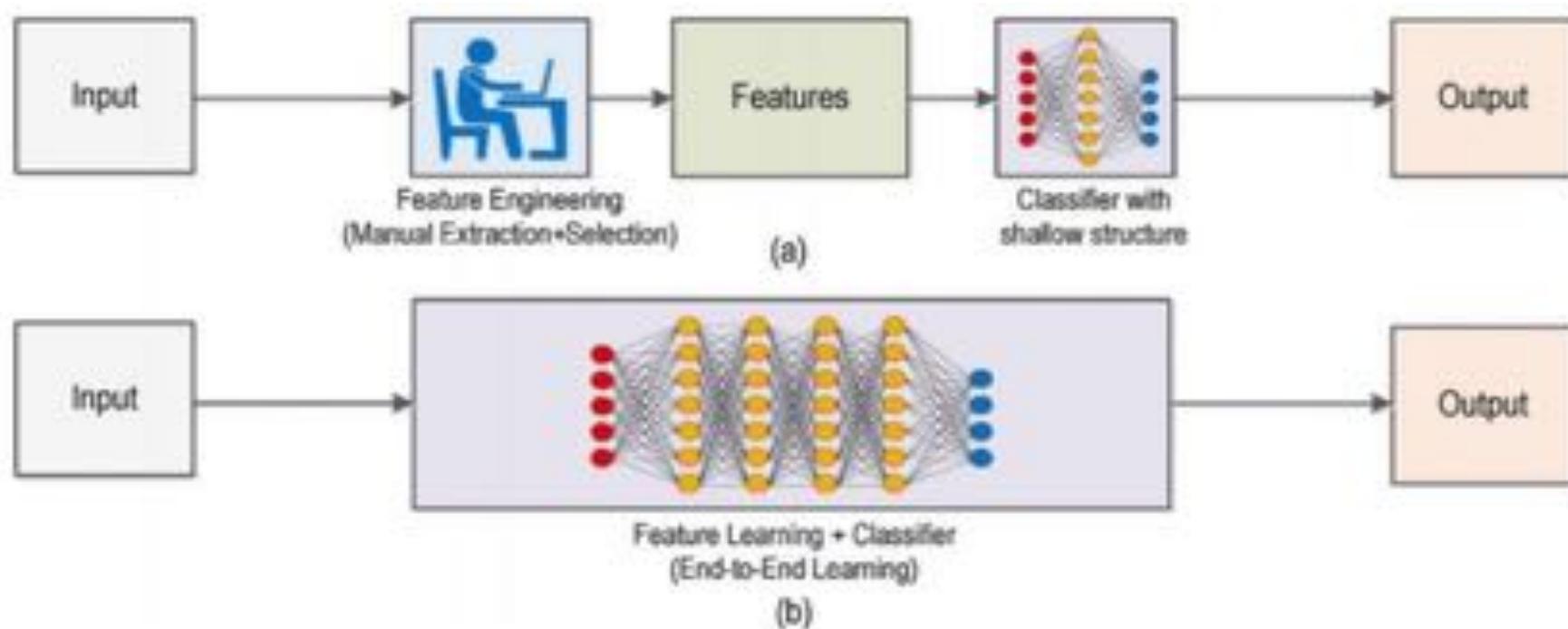
Fuel Characterization + Fire Severity

Mapping Satellite images to Canopy Bulk Density (CBD), Canopy Cover (CC), Canopy Base Height (CBH), and canopy Height (HT)



- LANDSAT data – Open-source since 2008
- r^2 values ranging from 0.55 to 0.68
- No train-test split, overfitting discussion

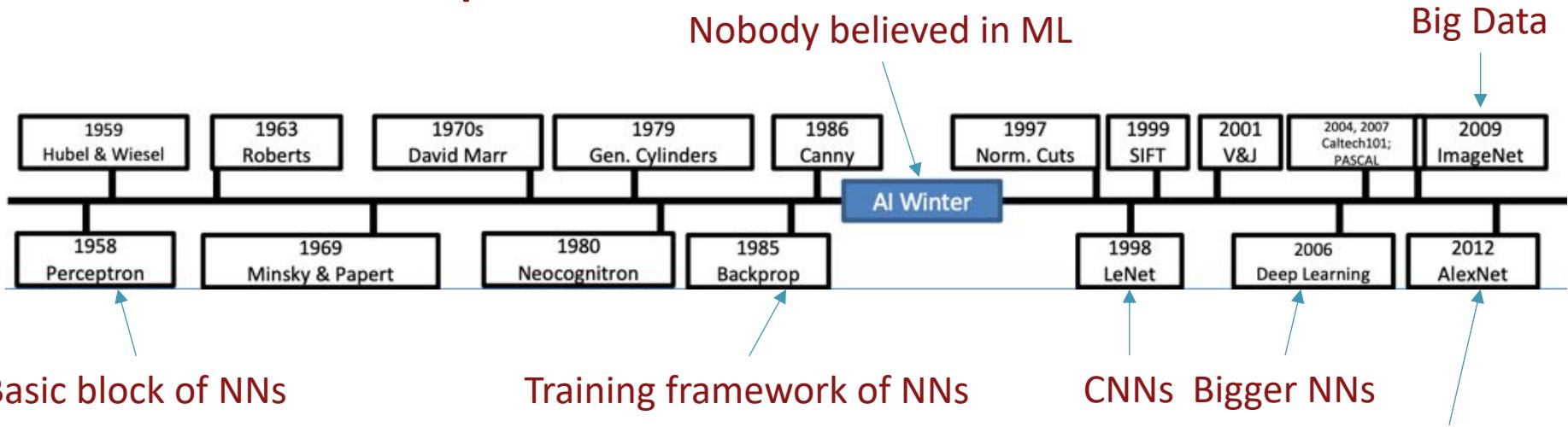
Two paradigms of ML



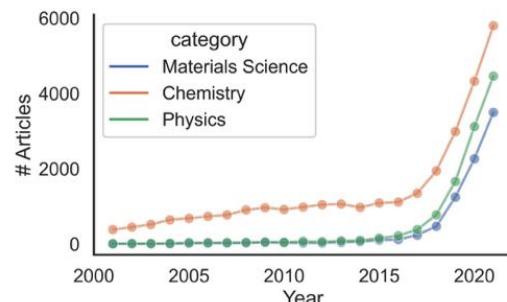
del Real et al. (2020)

Previous work predate interest in ML!

Nobody believed in ML



No. of ML for Science papers



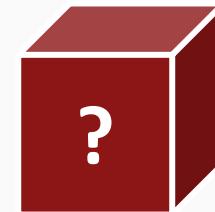
L. Fei-Fei et. al., CS231N (2023)

Early Work

- Demonstrate proof-of-concept
- Demonstrate potential use of ML models
- Not too sophisticated
 - Labels generated by Government Authorities
 - Train on Satellite images
 - Use interpretable ML
 - Find Useful features
- Probably overfitting
- What can we do with more data?

Challenge tackled:

Black-box

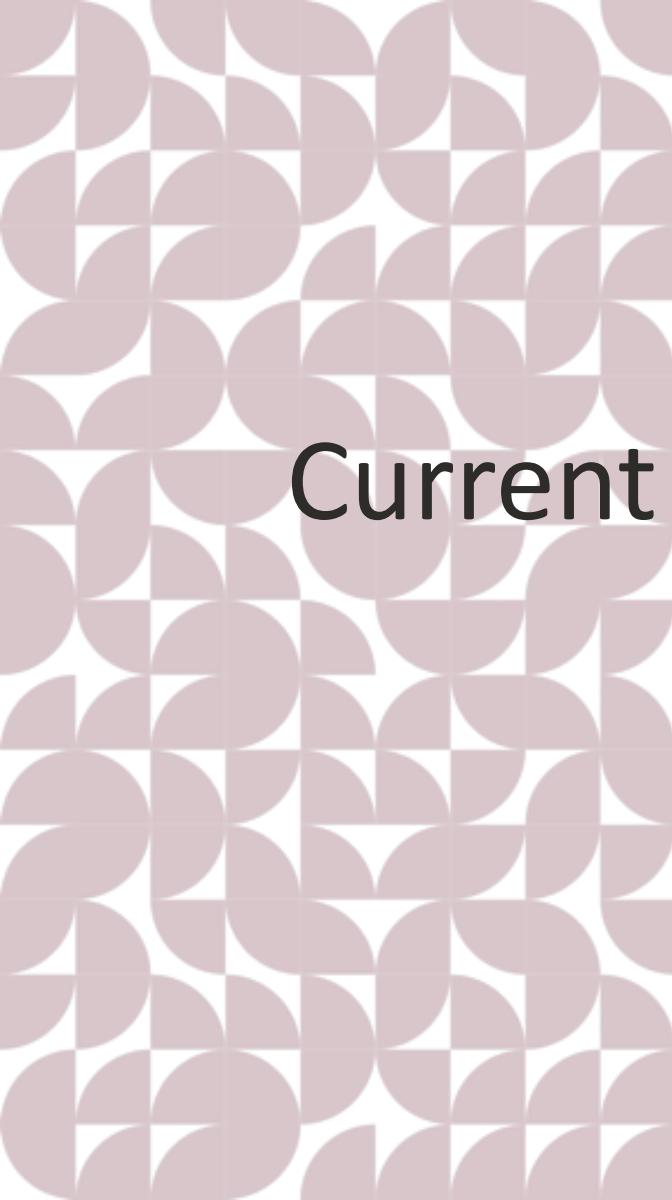


Tweet from 2022



Too many studies that apply machine learning to science & medicine employ incorrect methodologies.
Many make very basic mistakes, such as not having separate training & test sets, using the test set (not a separate validation set) for feature selection, hyperparameter tuning, etc

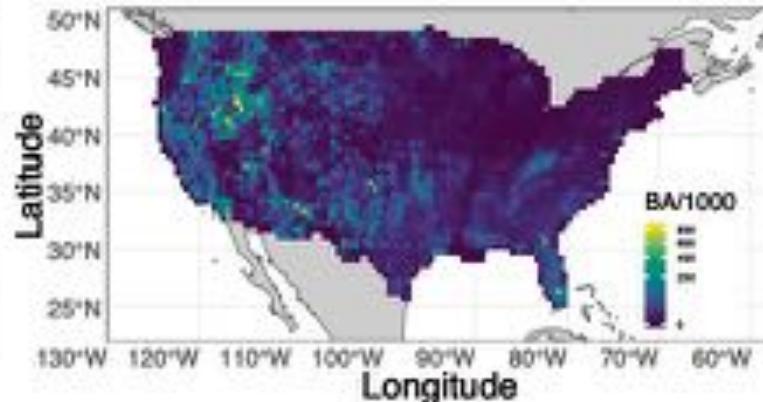
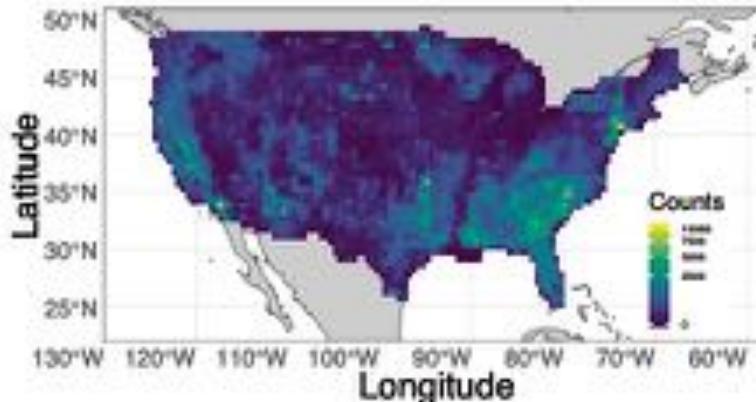
We are still early in ML adoption!



Current Trends

Data, data, data

EVA 2021 Challenge - Prediction of Fire occurrences



Global LandCover 2000
 Arable surfaces
 Bare Areas
 Cultivated and managed areas
 Herbaceous Cover, closed-open
 Irrigated Agriculture
 Cropland / Shrub-and/or grass cover

Cropland / Tree Cover / Other natural vegetation
 Tree Cover / Other natural vegetation
 Regularly flooded shrub and/or herbaceous cover
 Shrub Cover, closed-open, deciduous
 Shrub Cover, closed-open, evergreen
 Shrub Cover, closed-open, evergreen
 Shrub, burnt
 Tree, mixed leaf type

Sparse herbaceous or sparse shrub cover
 Tree, broadleaved, deciduous, closed
 Tree, broadleaved, evergreen
 Tree, broadleaved, deciduous, open
 Tree, broadleaved, evergreen
 Tree, regularly flooded, fresh water
 Tree, regularly flooded, saline water
 Water bodies

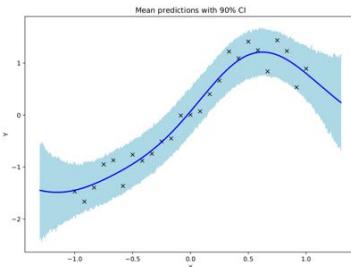
Data from various sources aggregated by U.S. Forest Service (Short, 2017)

Team name	S_{CNT}	rank _{CNT}	S_{BA}	rank _{BA}	S_{total}	rank _{total}	SI paper
BlackBox	2805	1	3316	1	6121	1	Ivek & Vlah
Kohrelation	2990	3	3446	3	6436	2	Koh
Bedouins	3146	4	3408	2	6554	3	Hazra et al.
KUNGFUPANDA	3166	5	3513	5	6679	4	Makowski
RedSea	3419	7	3467	4	6886	5	Zhang et al.
NaiveTom	3403	6	3565	7	6968	6	–
THEFIRETASTICFOUR	2979	2	4133	11	7111	7	–
EdX	3520	8	3595	8	7115	8	–
SNUBRL	4074	9	3530	6	7604	9	Kim et al.
MayLaB	4418	11	3765	10	8182	10	–
LancasterDucks	4926	12	3719	9	8645	11	D'Arcy et al.
FUFighters	4329	10	4863	13	9191	12	–
BENCHMARK	5565	13	4244	12	9810	13	–

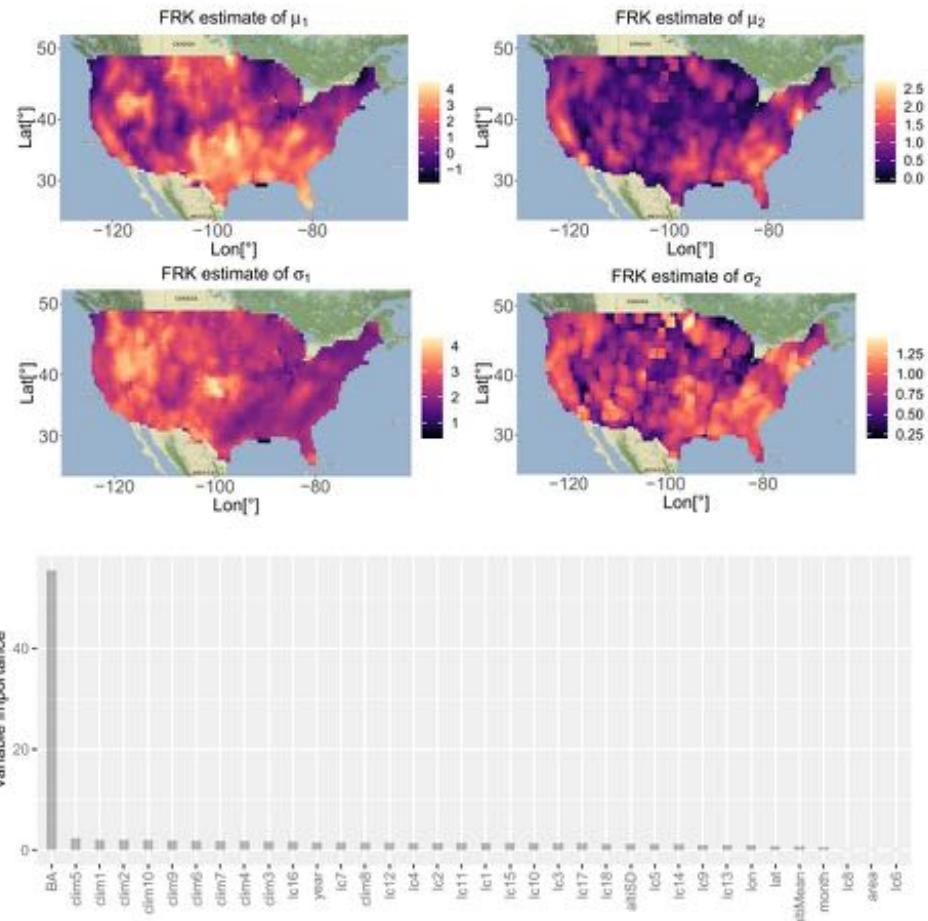
3rd Place winner

Part of a ML Competition/Challenge:

1. Use Gaussian Process Model to estimate statistics from classification
 2. Process the generated statistics
 3. Use random forests to predict fire occurrence

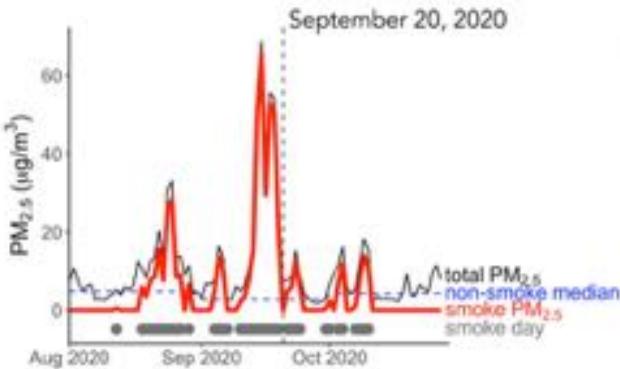


- Train-test split with cross-validation
 - Lost to 2 pure ML teams
 - Feature importance show pre-processing is somewhat flawed
 - Multicollinearity?
 - If it looks stupid but works, it isn't stupid



More modern take on regression

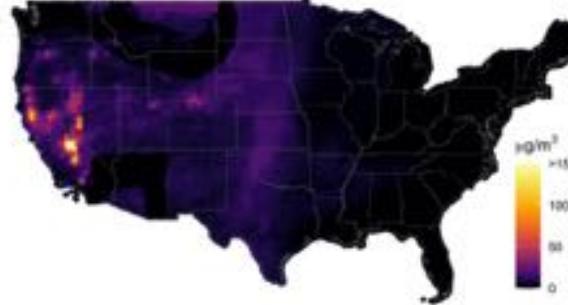
a) inferred smoke pollution from ground monitors



b) EPA monitors (test and train locations)



c) covariates

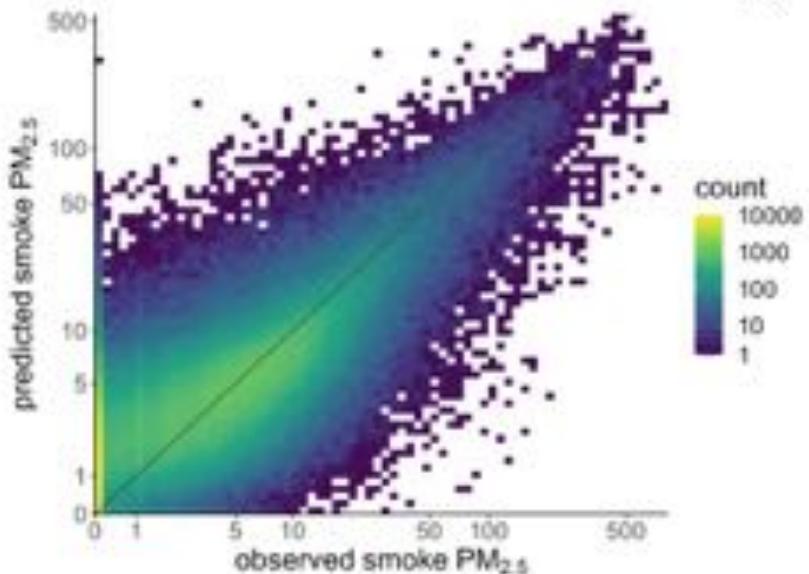
d) predicted smoke pollution
September 20, 2020

1. Curate Monitor, Sensor, Weather, Satellite data
2. Use **Gradient Boosted Trees**
3. Label small subset of data
4. Apply to 10 years of U.S geo/weather data.
5. Estimate soot emissions in U.S from past 10 years.

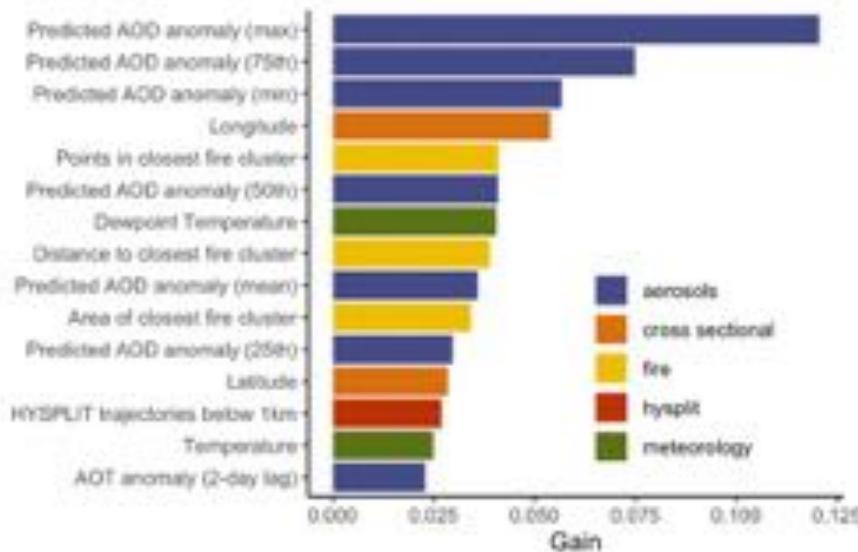
- Laborious effort in data collection
- Still using CARTs!

Results

a) predicted and observed smoke PM_{2.5}

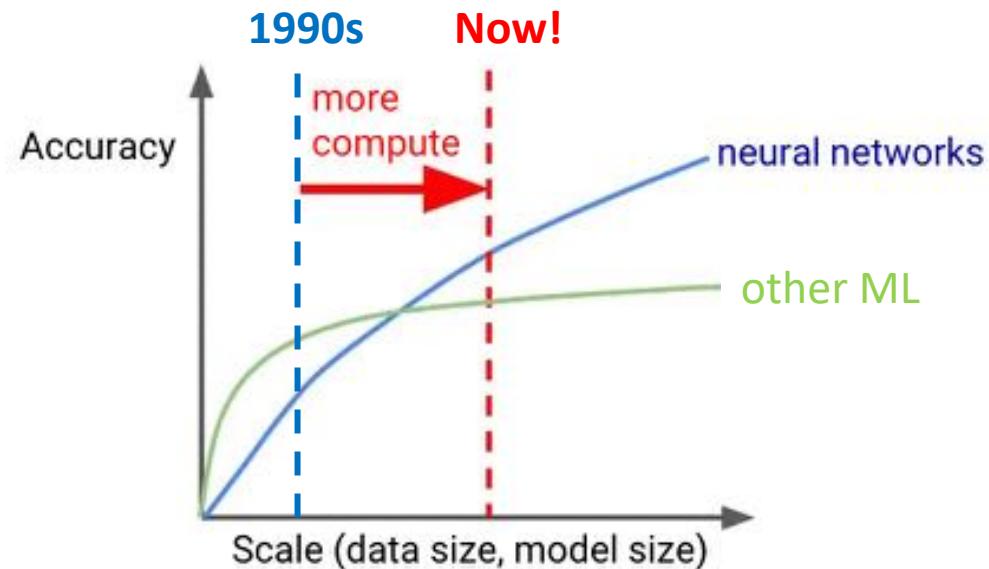
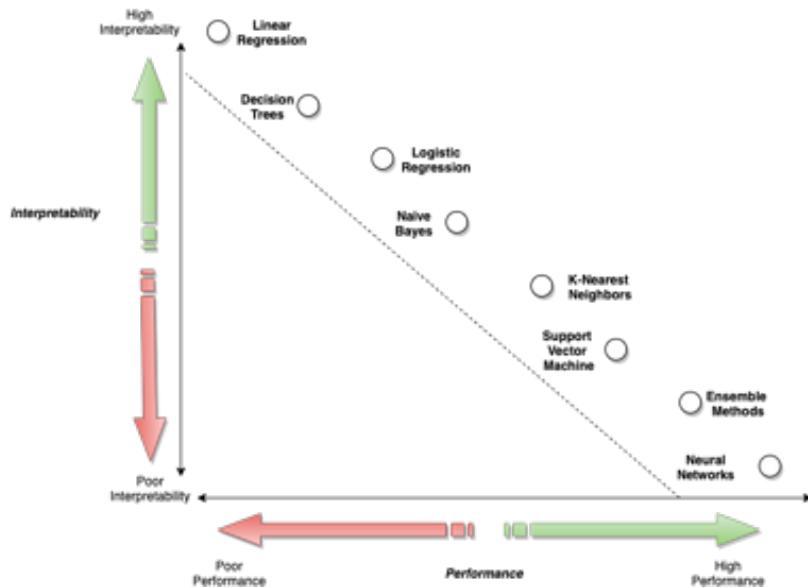


b) feature importance



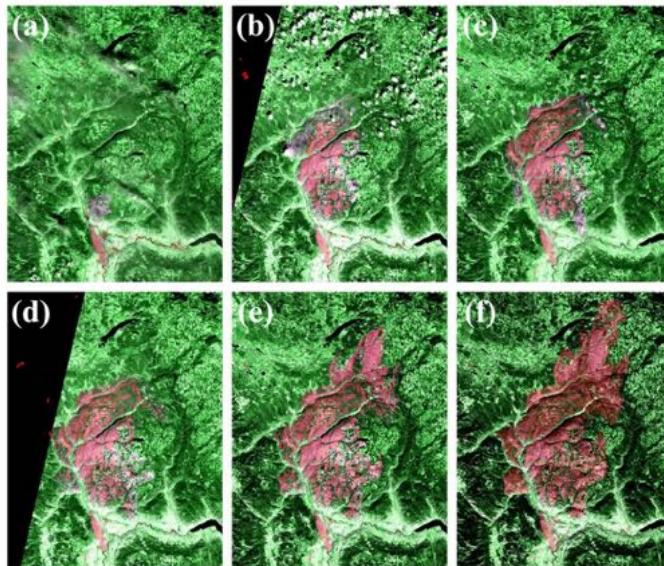
People really care about feature importance!

Two important charts driving AI for Science



Jeff Dean, Lecture for YCombinator AI (2017)
Amazon. White Paper (2021)

Real-time prediction!

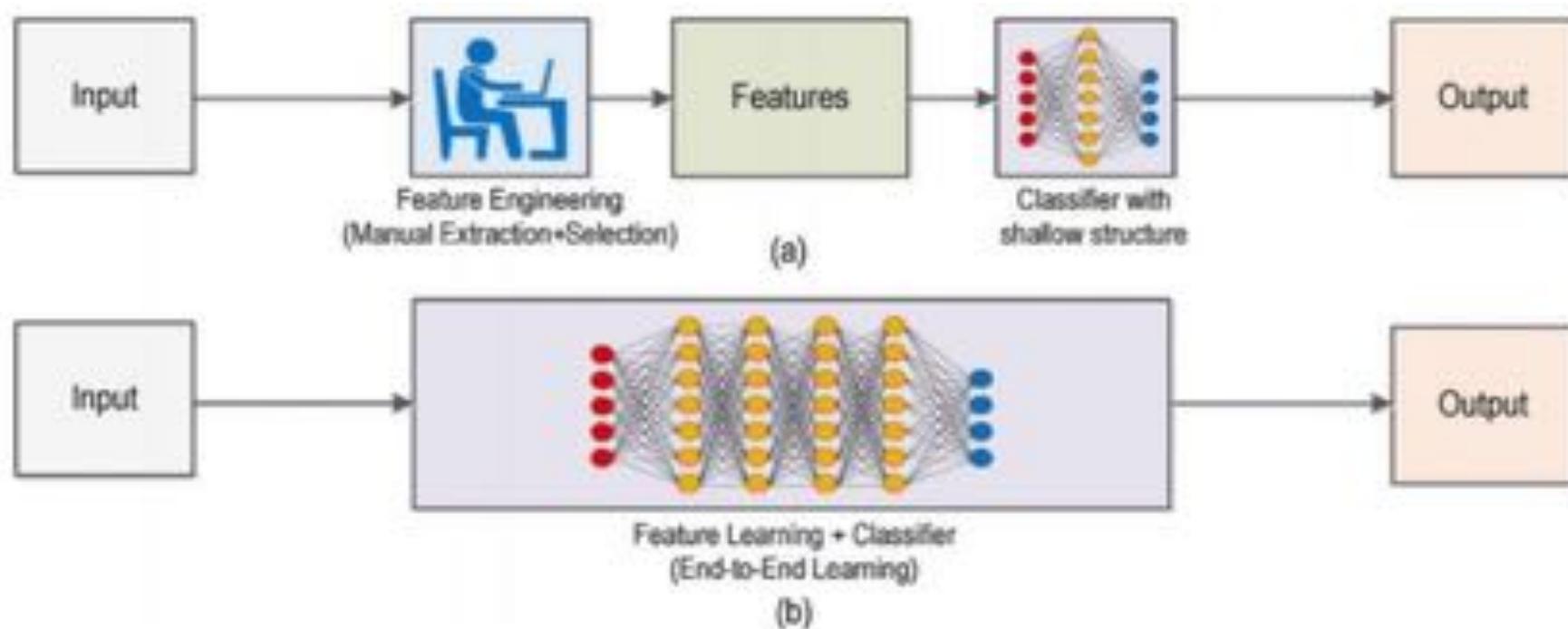


Can beat traditional stats methods

- deep learning can leverage large satellite datasets.
- Compute powerful enough for real-time prediction

Sat.	Data	Method	Seg.	Precision	Recall	OA	Kappa	F_1
S1	VH	kmap	>2	21.82%	65.13%	55.07%	0.1014	0.3269
	VV			45.21%	80.19%	67.02%	0.3404	0.5782
	VH, VV	kmap	>2	51.17%	82.11%	70.01%	0.4002	0.6305
	VH, VV	CNN_mrg	Otsu	71.82%	93.91%	83.58%	0.6716	0.8139
		CNN_tsc_mrg		70.60%	95.07%	83.47%	0.6694	0.8103

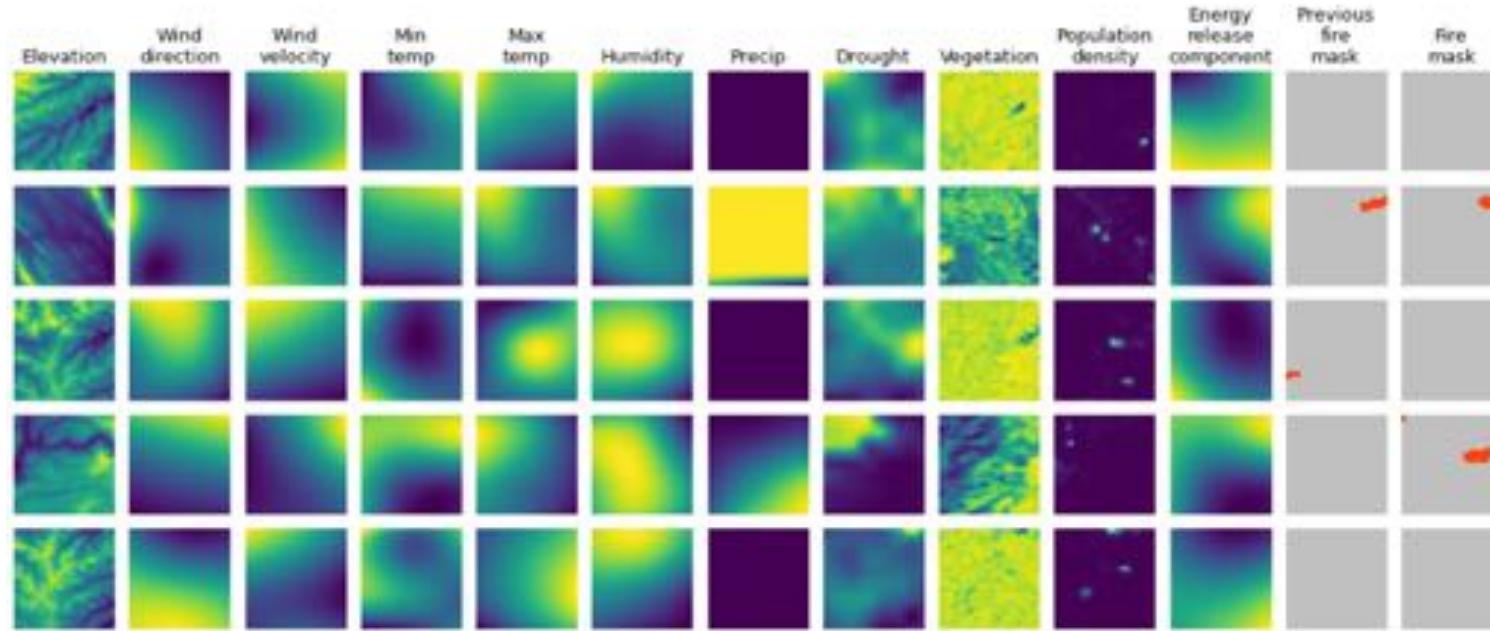
Two paradigms of ML



del Real et al. (2020)

Partial Govt Open-source wildfire datasets

- Much of the previous work driven by data availability
- We are starting to see more data from all parties



Data taken from Google Earth Engine + the Moderate Resolution Imaging Spectroradiometer (MODIS), the Visible Infrared Imaging Radiometer Suite (VIIRS), and the Shuttle Radar Topography Mission (SRTM)

Current State-of-the-art

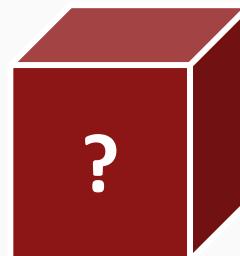
- Open-source resources enable a wider range of data access
- Researchers are starting to curate and structure unstructured data
- What's next?

Focus of studies so far:

Data challenges



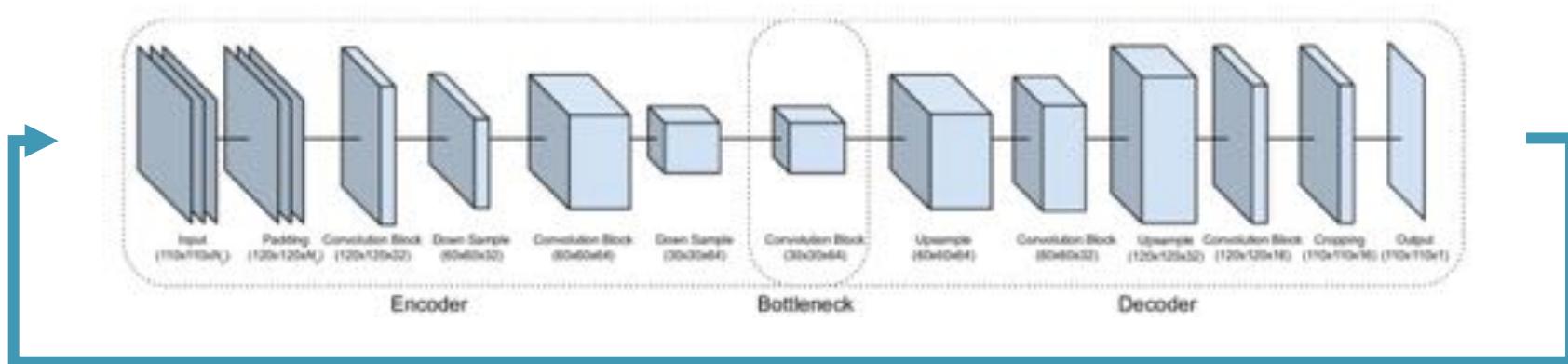
Black-box



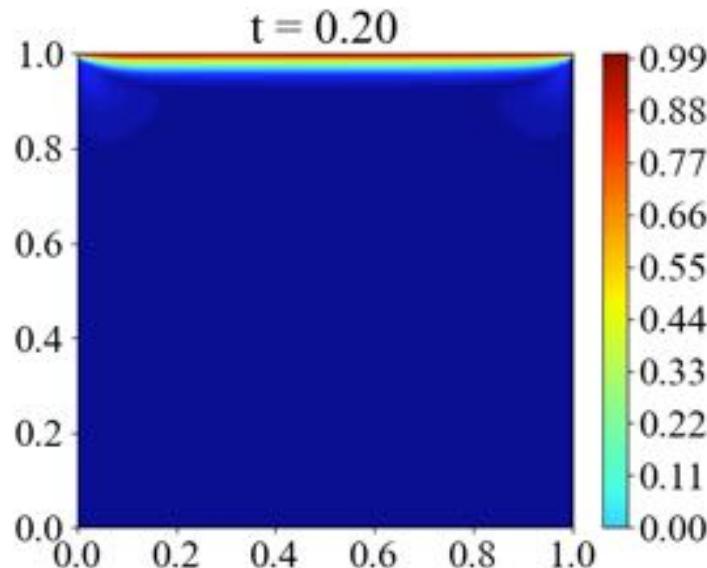
Reliability!!

How do we get people to trust
ML models?

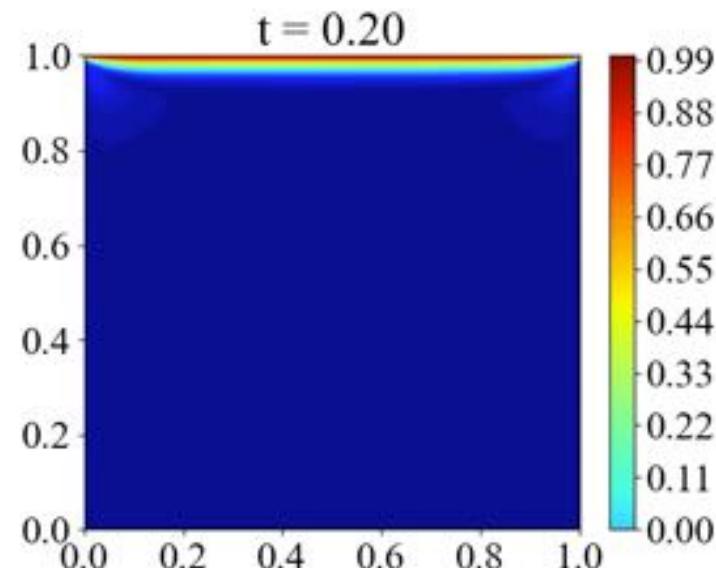
Autoregressive modeling



Problem with Autoregressive modeling



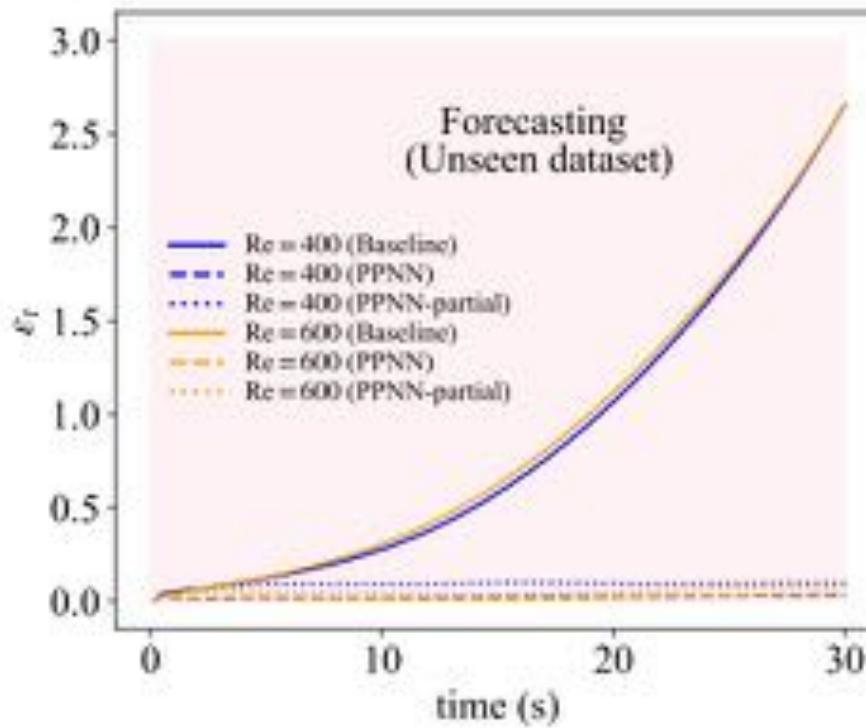
**Incompressible
Finite Difference Solver**



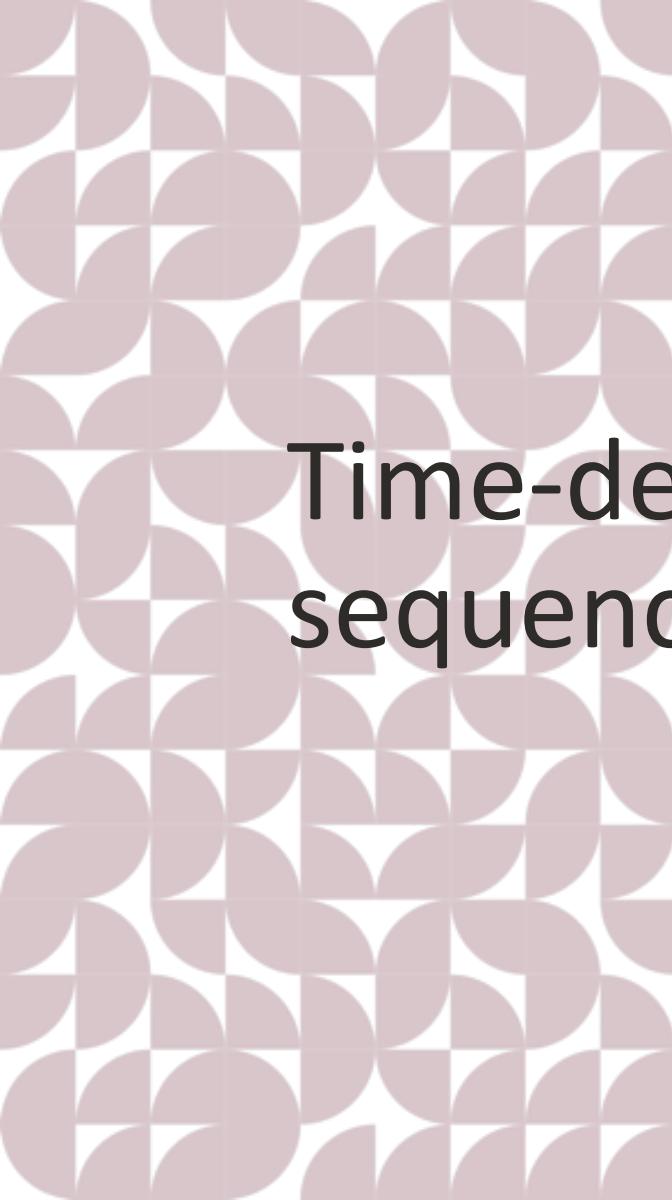
**ConvResNet (ML)
Prediction**

Why does this happen?

Problem with Autoregressive modeling



Some top researchers believe in better alternatives to LLMs for this reason

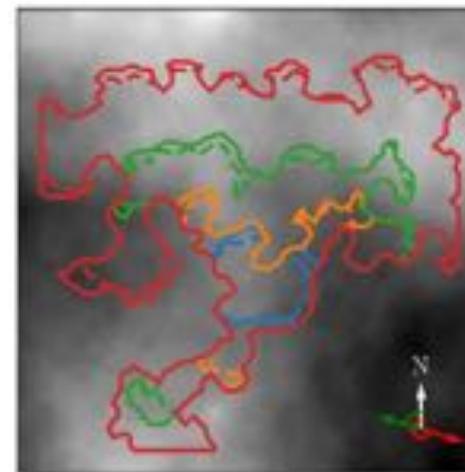
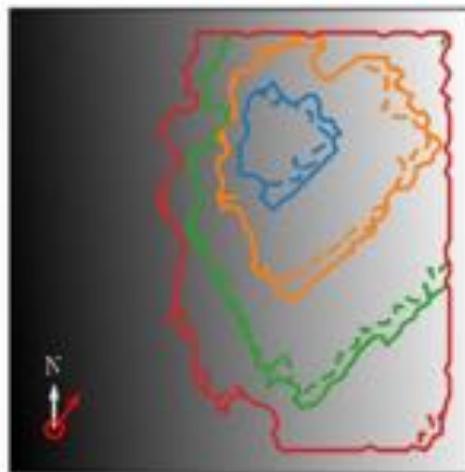
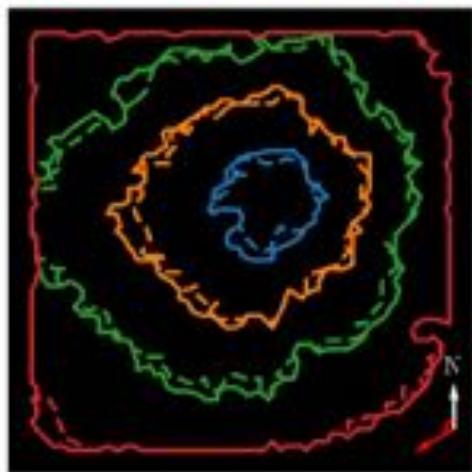
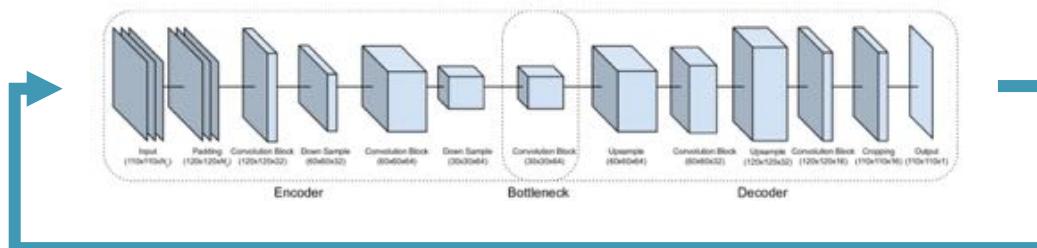
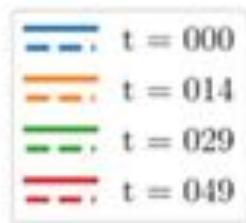


Time-dependent/Long-sequence predictions are hard

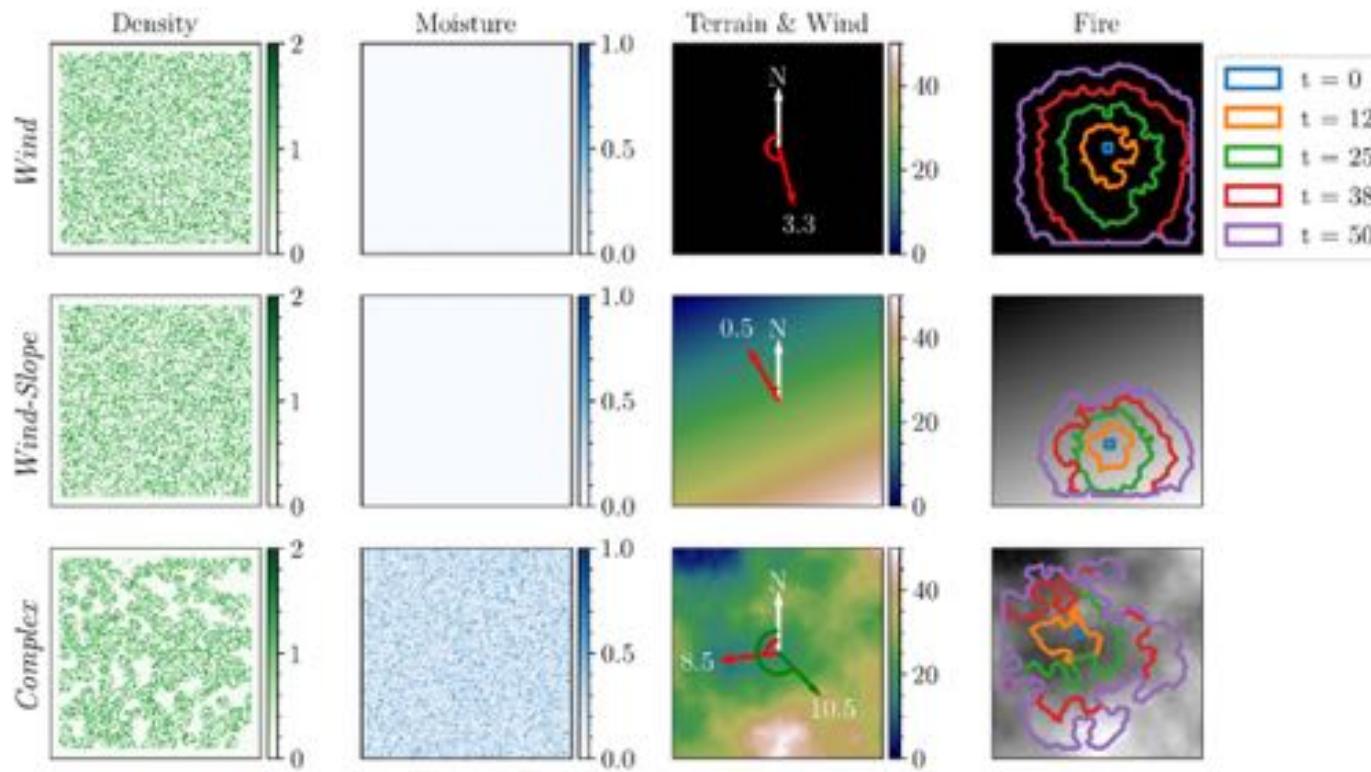
How do we address this?



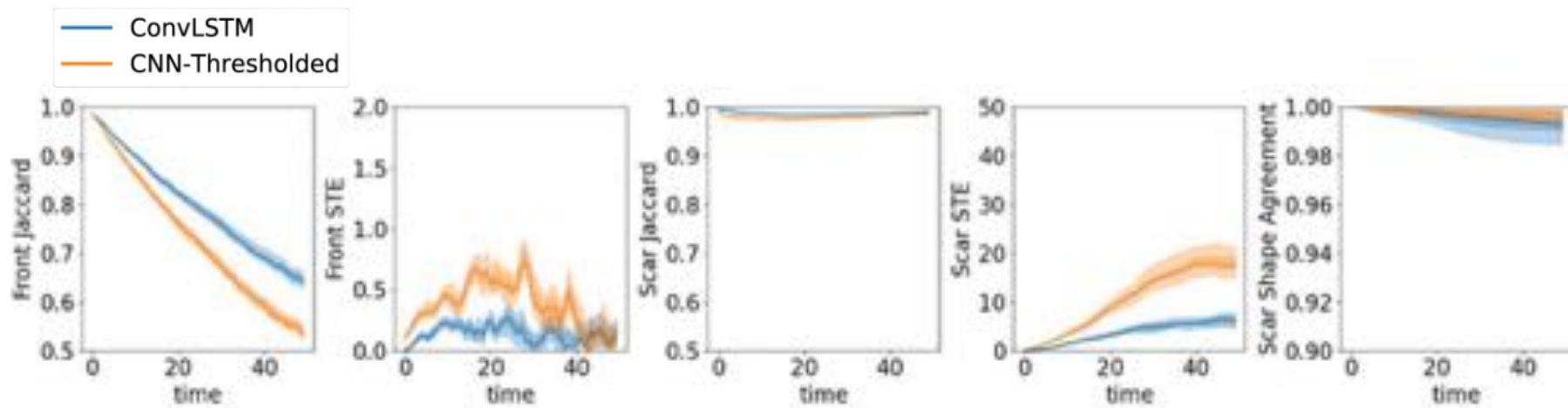
Next-frame fire-front prediction



Fire Front Simulation Data



Conv LSTM

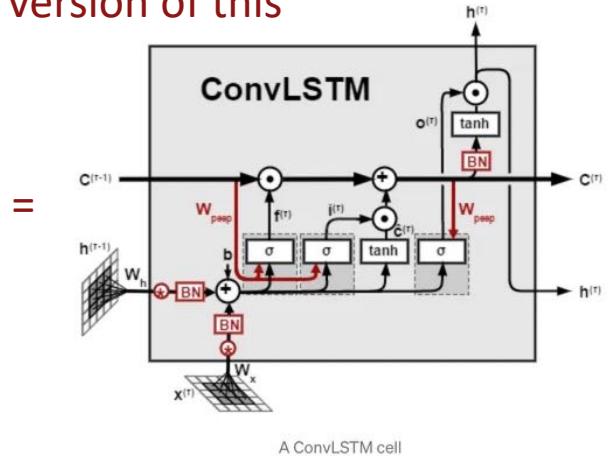
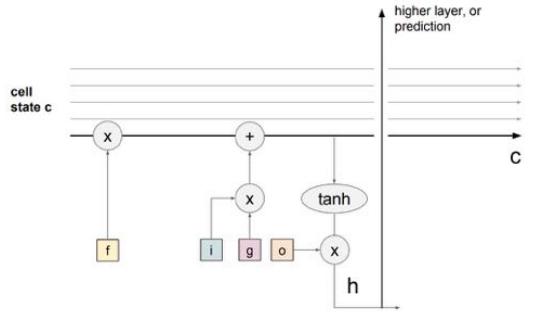
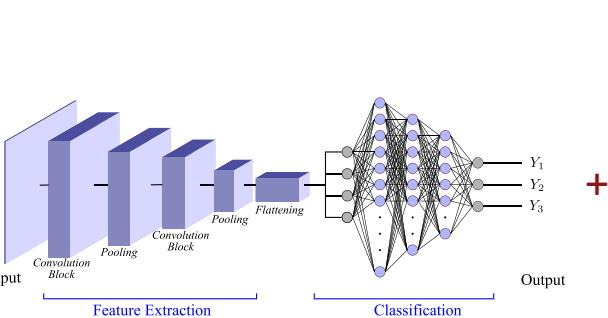


Why does the Conv LSTM work better?

Larger context windows!

ConvLSTM

Vision transformers are more parallelizable version of this



More stable version of this

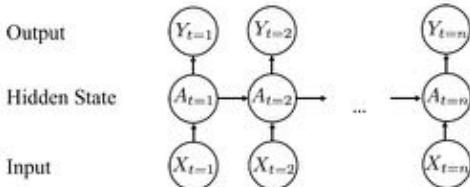
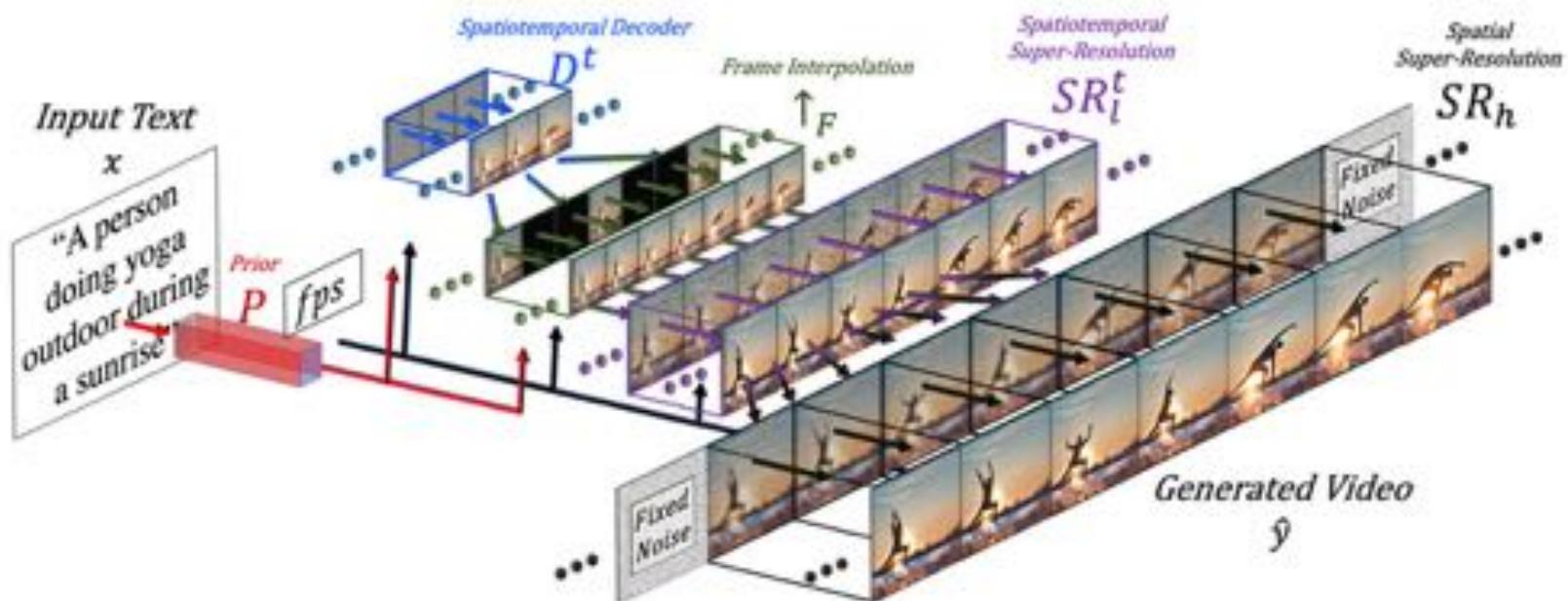


Fig. 16. Schematic of a many-to-many RNN.

- Has memory from the previous time steps
- Only 10 snapshots - Still using autoregression
- Why?

Limited memory!
A100 GPUs have 40GB

Taking context windows to the asymptote

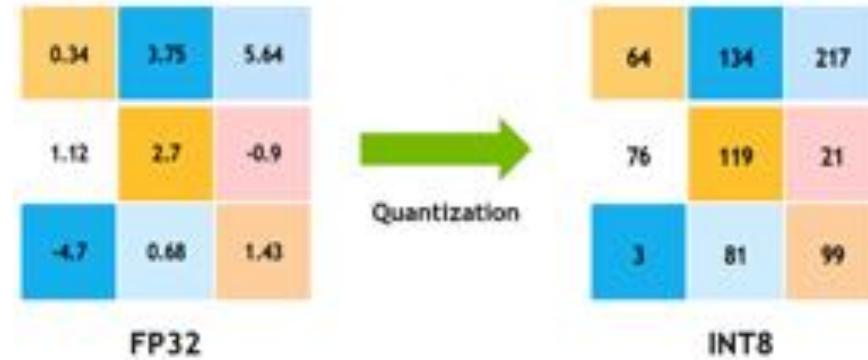


Hu (2022) <https://towardsdatascience.com/generative-ai-878909fb7868>

Turn 2D + time into a 3D block

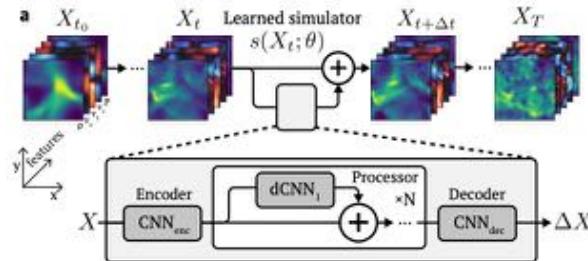
Still have issues with memory

Dealing with memory

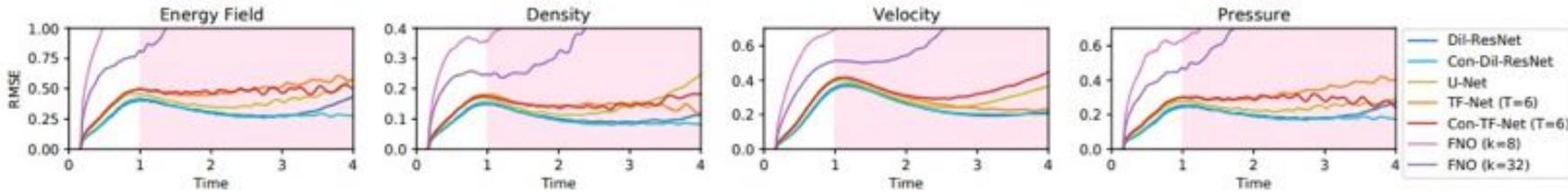
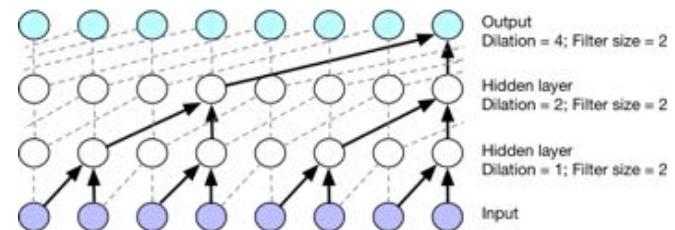
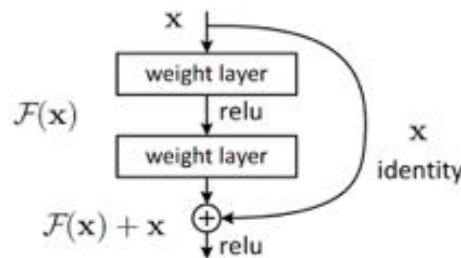


<https://hackernoon.com/using-ai-to-super-compress-images-5a948cf09489>

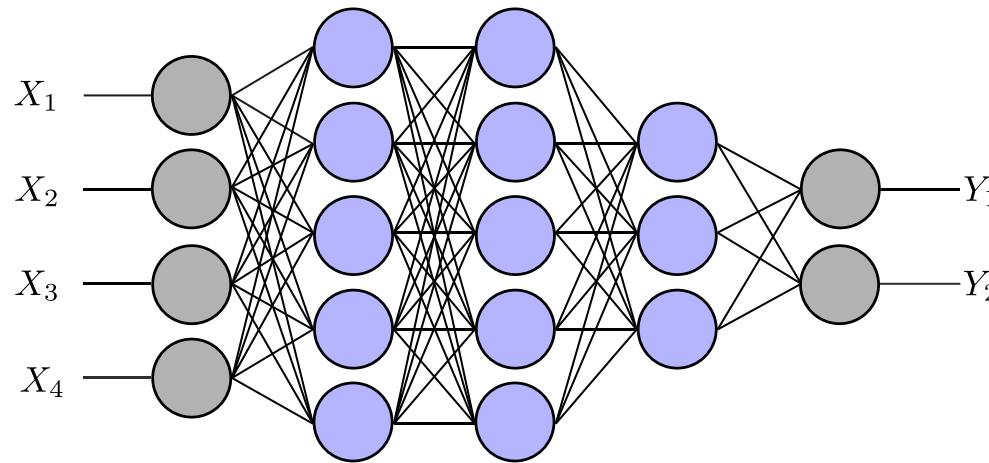
Another Pure ML Approach



Use of residual blocks, dilated layers, optimal timestep help



Basically Hyperparameter Search



How many neurons in each layer?

How many layers?

What optimization scheme?

How much data?

What forward operation?

What non-linearity?

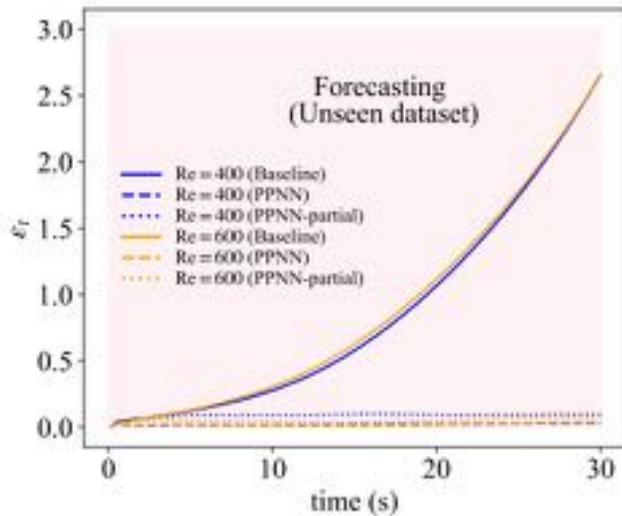
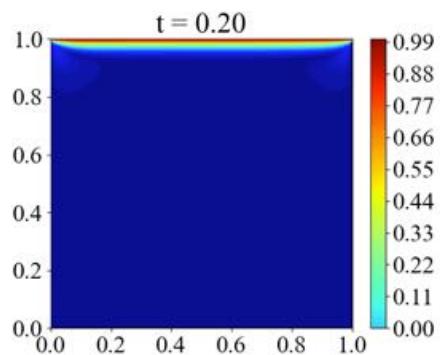
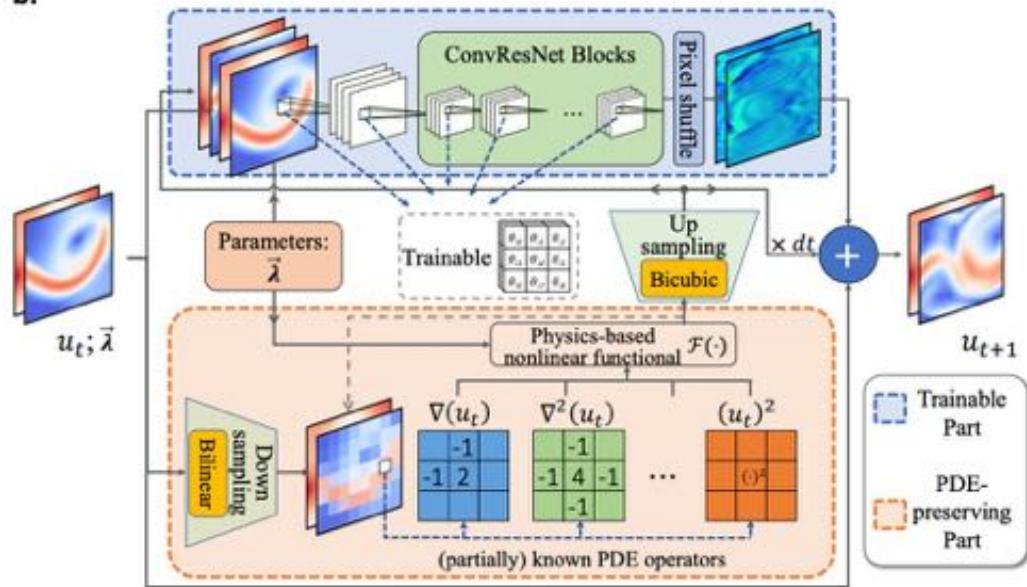
NN or other method?

- 1. Intuition/Domain Knowledge
- 2. Exhaustive Search (Expensive!)
- 3. Random Search (Not efficient)
- 4. Bayesian Optimization (Gaussian Assumptions)
- 5. Reinforcement Learning
- 6. AutoML
- 7. ...

Physics informed ML

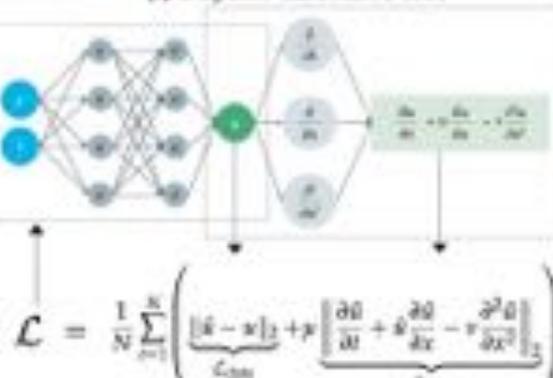
$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \nu \nabla^2 \mathbf{u}, \quad t \in [0, T],$$

$$\nabla \cdot \mathbf{u} = 0,$$

b.

Physics-informed ML in Fluid Mechanics

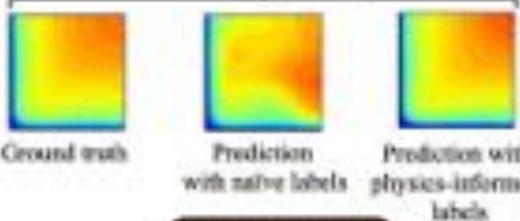
(i) Physics-informed loss



Loss function

(ii) Physics-informed data

Mean axial velocity



Data

Training

(iii) Physics-informed architecture

Enforced functional form

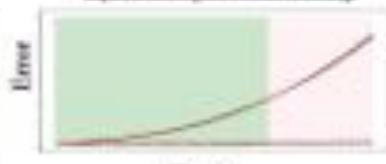
$$\mathbf{b} = \sum_{i=1}^{10} g^{(i)}(\lambda_1, \dots, \lambda_5) \mathbf{T}^{(i)}$$



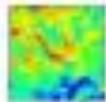
ML. model

(iv) Examples of Applications

Spatiotemporal modeling

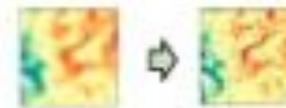


Model discovery



$$\tau_{ij}^{(p)} \approx \bar{\rho} C_p \Delta^2 \frac{\partial \tilde{u}_i}{\partial x_k} \frac{\partial \tilde{u}_j}{\partial x_k}$$

Turbulence modeling



Predictions for flows

Conclusions

- This lecture: Applications, Opportunities, Challenges
 - When should I use ML?
 - Data
 - Forgiveable Errors
 - No better alternative
 - Low cost
 - How was ML done prior to current interest in ML?
 - How is ML research being done now in Wildfires?
 - What's next?

Thank You