ML4CC: Lecture 2

Sit with your discussion groups!

Assignments reminder

Keep doing your weekly PMIRO+Q

Your first coding assignment will be posted after class today

It is due before the start of class on Feb 15. It involves basic data loading, plotting, and function writing.

Recap of previous class

The average global temperature of the earth is increasing

This is due to increased greenhouse gases in the atmosphere

Most human activities cause the release of greenhouse gases to some degree, particularly CO2 through the burning of fossil fuels

Increased temperature have destabilizing effects on the climate and human civilization.

We need drastic societal change to both prevent further climate change and adapt to what is happening

Climate Change in the News

AP WORLD U.S. ELECTION 2024 POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES HEALTH

Israel-Hamas war Election 2024 Cori Bush Elon Musk's pay Chita Rivera dies

WASHINGTON NEWS

Biden delays consideration of new natural gas export terminals, citing climate risk

BY MATTHEW DALY

Updated 11:43 AM EST, January 26, 2024

Share

WASHINGTON (AP) — The Biden administration is delaying consideration of new natural gas export terminals in the United States, even as gas shipments to Europe and Asia have soared since Russia's invasion of Ukraine.

The election year decision by President Joe Biden aligns with environmentalists who fear the huge increase in exports, in the form of liquefied natural gas, or LNG, is locking in potentially catastrophic planet-warming emissions when the Democratic president has pledged to cut climate pollution in half by 2030.

"While MAGA Republicans willfully deny the urgency of the climate crisis, condemning the American people to a dangerous future, my administration will not be complacent," Biden said in a statement Friday. "We will not cede to special interests. We will heed the calls of young people and frontline communities who are using their voices to demand action from those with the power to act."

The current economic and environmental analyses the Energy Department uses to evaluate LNG projects don't adequately account for potential cost hikes for American consumers and manufacturers or the impact of greenhouse gas emissions, the White House said.

Paper 1 Discussion

Applied Energy

Volume 208, 15 December 2017, Pages 889-904

Machine learning approaches for estimating commercial building energy consumption

Caleb Robinson [a], Bistra Dilkina [a], Jeffrey Hubbs [e], Wenwen Zhang [b], Subhrajit Guhathakurta [b], Marilyn A. Brown [c], Ram M. Pendyala [d]

Show more

+ Add to Mendeley Share Cite

Discussion procedure

- I post a question
- You discuss as a group
- I randomly call on groups to share their answers
- I recap the full answer with slides

At the end we will discuss the questions you included in your PMIRO+Q. Then you will have time to submit a second file to the Brightspace assignment, which updates your PMIRO as needed and provides your best answer to your Q.

Discussion Question 1

Are you a morning person?

Attendance

Select one person from the group to go to this Google Doc and write down the names of all people present in the group

https://docs.google.com/document/d/1PKhw9E2lJpAnFrFO88DOc2rscZFVlcIv47QNa5h1sGs/edit?usp=sharing (link is in Brightspace under Syllabus content)

Discussion Question 2

What is being described here and how does it relate to the work done in this paper?

From the Introduction:

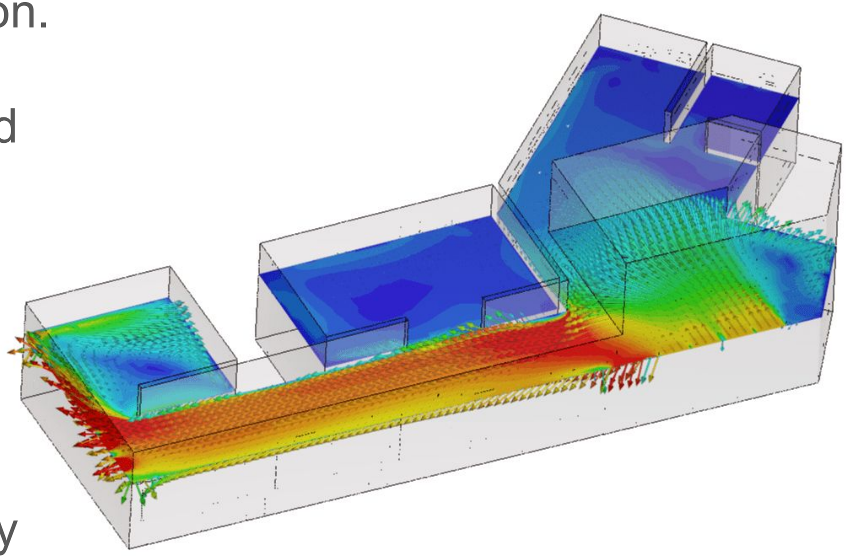
One way of estimating building energy consumption, in the absence of actual sensor data, is to create physical building models with a "template" of representative buildings, then run thermodynamic simulations to estimate the energy demands [14]. These "engineering" models of building energy consumption are computationally expensive and cannot capture the wide variety of different buildings present in cities, as modeling each type of building requires very detailed input data, which is costly to collect. Statistical models can be used to fill the gaps where resources are too limited to use physical models, or the scale of the study area makes physical modeling impractical.

Machine learning helps avoid costly physics simulations

Several questions about the physical world can be answered “from the ground up” by using a first-principles physics simulation.

The models have many parameters and long and computationally expensive runtimes. They also require a sound physical understanding of the system.

Simple machine learning models, like those used in this paper, sidestep the need for detailed physics simulations by just learning coarse associations



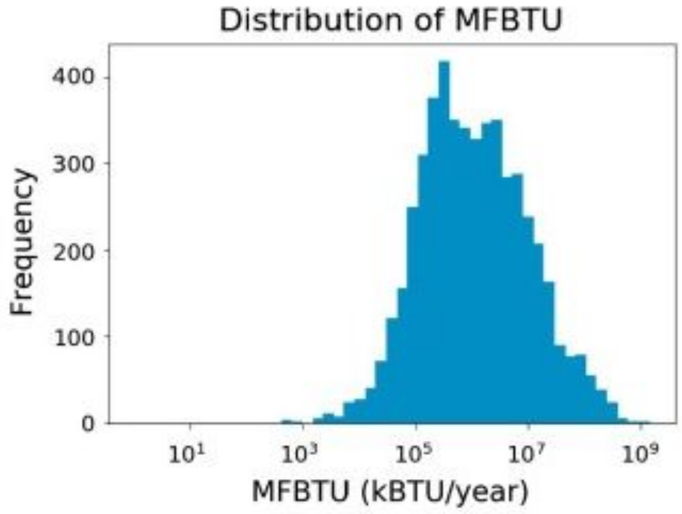
Discussion Question 3

What precisely are the models in this paper trained to predict and why?

Predicting the log of MFBTU

3.2. Modeling commercial building energy consumption

We want to predict the 'Annual Major Fuel Consumption' (the MFBTU field in CBECS) of commercial buildings by only using some features of the buildings. We express this objective in a machine learning regression format as follows: We are given $\mathbf{X}$, the features for all buildings in the CBECS dataset, and $\mathbf{e}$, the target MFBTU (energy) values for all buildings, where a row, $X_{i,:}$, represents the features for building *i*, and an entry, e_i, is the MFBTU value for building *i*. In the remainder of the paper we focus on predicting the logarithm of the actual MFBTU values as we have observed that the MFBTU values follow an approximate log-normal distribution, and some machine learning models will be able to better estimate the values in the log-transformed normal distribution [35], [36](see Fig. A.6). Specifically, we let $y_i = \log_{10}(e_i)$, and refer to $\mathbf{y}$ as our target values. We want to learn a function, $f(X_{i,:}) = \hat{y}_i$, that takes the features of a building as input, and outputs the estimated log of the energy consumption for that building, $\hat{y}_i$. From this, we predict the MFBTU value for a building as $\widehat{e_i} = 10^{\widehat{y_i}}$. To estimate *f* we will use machine learning models such as: linear regression, gradient boosting regression models, and random forest regressors. In general, these models attempt to tune their internal parameters, θ, to minimize some loss function, *L*, between the target values and values predicted by the model, i.e. solving $\min_{\theta} L(\mathbf{y}, f(\mathbf{X};\theta))$. The loss function will be a function



Discussion Question 4

Can you think of a difference between these two groups of methods that would explain their difference in performance?

And can you explain why this split doesn't show in the Extended Features model rankings?

Common features	Mean absolute error	$10^{\text{Mean AE}}$	Median absolute error	$10^{\text{Median AE}}$	r^2
XGBoost	**0.30±0.01**	**1.99±0.06**	**0.22±0.01**	**1.66±0.03**	**0.82±0.02**
Bagging	0.33±0.01	2.13±0.07	0.24±0.01	1.73±0.05	0.78±0.03
MLP Regressor	0.33±0.01	2.16±0.05	0.25±0.01	1.77±0.04	0.78±0.02
Random Forest Regressor	0.33±0.02	2.13±0.07	0.24±0.01	1.73±0.05	0.78±0.02
Extra Trees Regressor	0.34±0.02	2.17±0.08	0.24±0.01	1.74±0.05	0.76±0.03
SVR	0.39±0.01	2.44±0.07	0.29±0.01	1.95±0.04	0.70±0.03
KNN Regressor	0.43±0.01	2.68±0.08	0.32±0.02	2.10±0.07	0.65±0.03
AdaBoost	0.43±0.03	2.71±0.16	0.36±0.03	2.29±0.17	0.68±0.03
Linear SVR	0.51±0.02	3.28±0.15	0.40±0.02	2.54±0.11	0.52±0.04
Linear Regression	0.52±0.02	3.33±0.13	0.43±0.02	2.72±0.12	0.53±0.03
Ridge Regressor	0.52±0.02	3.33±0.13	0.43±0.02	2.72±0.12	0.53±0.03
ElasticNet	0.76±0.02	5.75±0.32	0.67±0.03	4.67±0.35	0.09±0.01
Lasso	0.79±0.02	6.17±0.35	0.69±0.03	4.92±0.38	0.00±0.00

Linear Regression

$$\hat{\beta} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{X}\,\hat{\beta}$$

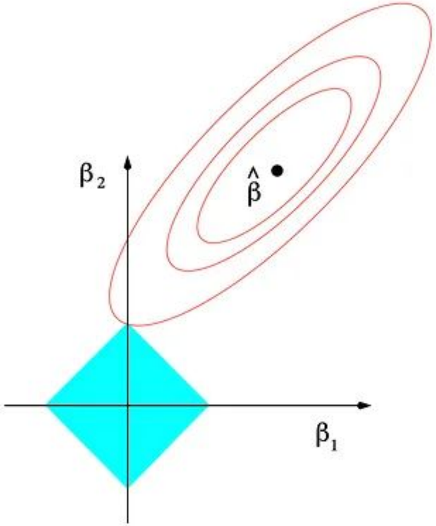
$\hat{\beta}$ = ordinary least squares estimator

X = matrix regressor variable X

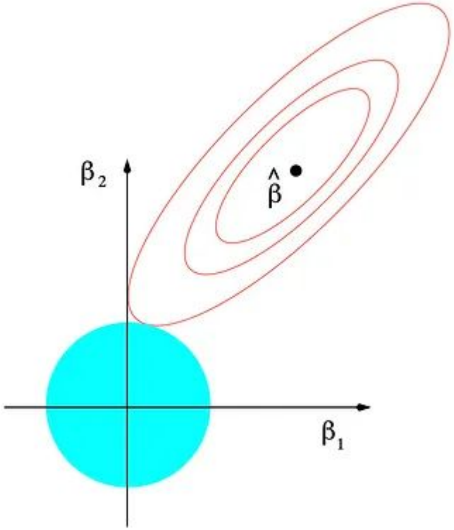
$\top$ = matrix transpose

y = vector of the value of the response variable

$\hat{y}$ = predicted values



Lasso



Ridge

Lasso

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq s \tag{6.8}$$

and

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \leq s, \tag{6.9}$$

Ridge

Lasso + Ridge = “ElasticNet”

Linear methods can only learn linear relationships

Sometimes linear relationships are all you need...

building data, but no energy consumption data. Most features in the CBECS dataset such as, 'Number of Employees', 'Number of X-ray machines', or 'Insulation upgraded', are not commonly available, and therefore should not be included when training the models. It is important, however, to determine the influence of each possible feature included in the CBECS data on predicting energy consumption, in order to determine the potential benefits of additional data collection efforts. To this end, we run two sets of experiments using the methodology described in the previous two paragraphs: one that involves training the models using only a set of features that will be commonly available, or easily obtainable, in many cities and one that includes all of the features available in CBECS. We refer to the first group of features as the "common feature set"; it includes the following features: principal building activity, square feet, number of floors, heating degree days, and cooling degree days. We refer to the second group as the "extended feature set". As the "common feature set" is the set we expect to be available when using our models in specific urban areas, Fig. 1 shows this set of features as common between the "Model Development" section and "Application" section.

Extended features	**Mean absolute error**	$10^{\text{Mean AE}}$	**Median absolute error**	$10^{\text{Median AE}}$	r^2
XGBoost with common features	0.30±0.01	1.99±0.06	0.22±0.01	1.66±0.03	0.82±0.02
XGBoost	**0.23±0.01**	**1.69±0.02**	**0.17±0.01**	**1.48±0.03**	**0.89±0.01**
Linear regression	0.24±0.01	1.75±0.02	0.19±0.01	1.53±0.04	0.88±0.01
Ridge regressor	0.24±0.01	1.75±0.02	0.19±0.01	1.53±0.04	0.88±0.01
SVR	0.25±0.01	1.79±0.04	0.19±0.01	1.53±0.03	0.87±0.01
Bagging	0.25±0.01	1.79±0.04	0.18±0.01	1.53±0.04	0.87±0.02
Random forest regressor	0.25±0.01	1.79±0.04	0.18±0.01	1.53±0.04	0.87±0.01
Extra trees regressor	0.25±0.01	1.79±0.04	0.19±0.01	1.54±0.03	0.87±0.01
Linear SVR	0.26±0.01	1.80±0.03	0.20±0.01	1.58±0.04	0.87±0.01
AdaBoost	0.32±0.01	2.07±0.05	0.26±0.01	1.80±0.05	0.82±0.01
KNN regressor	0.37±0.01	2.34±0.06	0.29±0.01	1.93±0.04	0.75±0.02
MLP regressor	0.45±0.02	2.82±0.11	0.36±0.02	2.31±0.10	0.64±0.03
ElasticNet	0.60±0.02	4.00±0.20	0.51±0.02	3.26±0.16	0.40±0.01
Lasso	0.79±0.02	6.17±0.35	0.69±0.03	4.92±0.38	0.00±0.00

Discussion Question 5

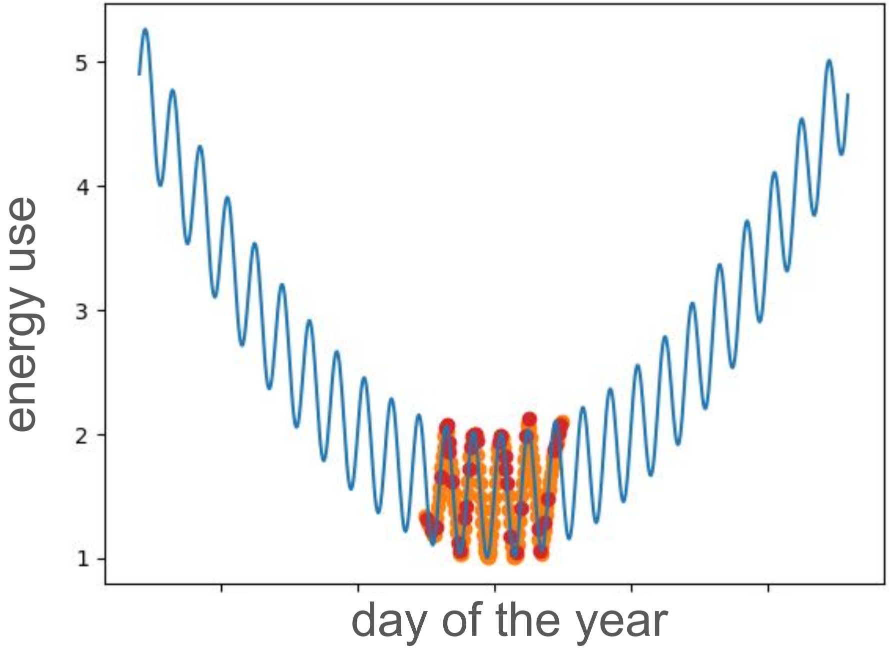
Find three different ways in which the authors tested how well their model *generalizes*

Generalize: the ability to perform well on data not seen in training

orange dots = training

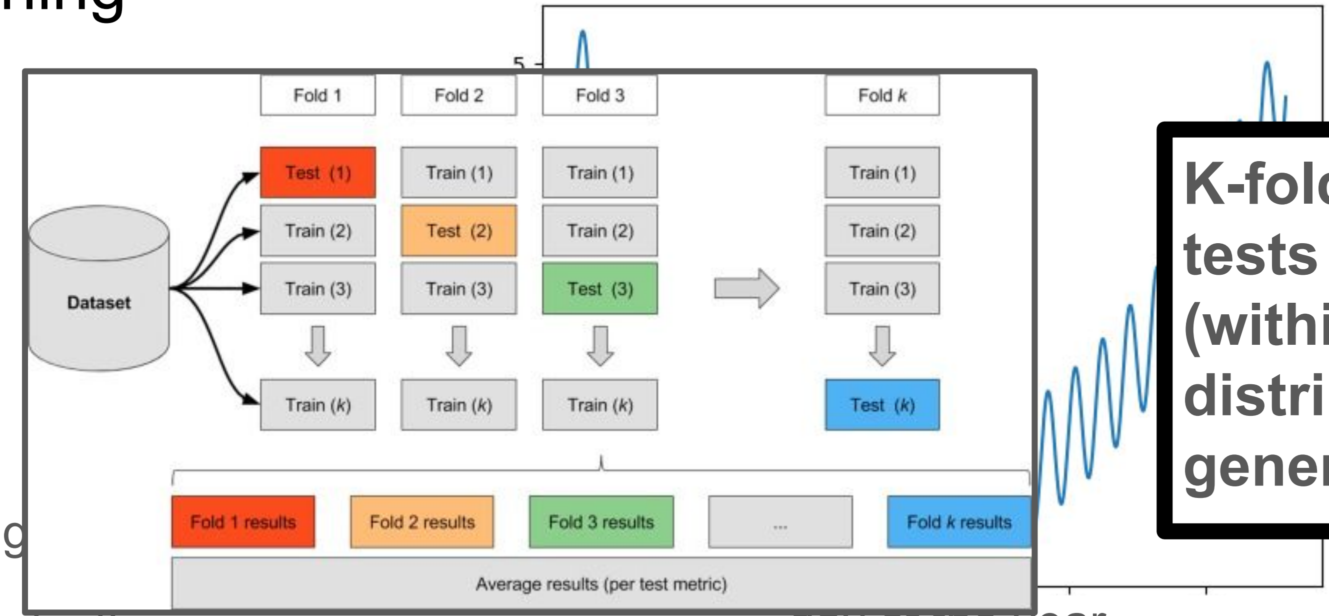
red = testing

To "validate" a model usually means to test its generalization



This model learned well and can generalize within the data distribution it was trained on

Generalize: the ability to perform well on data not seen in training



K-folds validation tests this kind of (within distribution) generalization

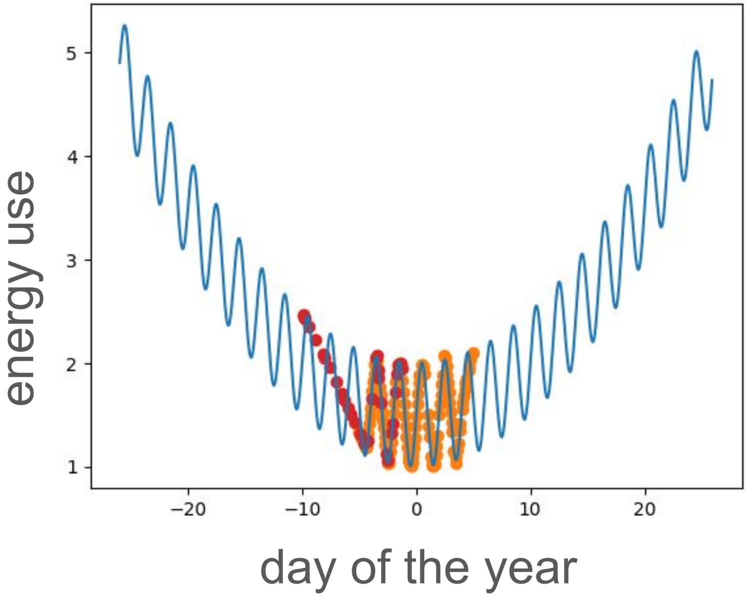
orang

red = testing

This model learned well and can generalize within the data distribution it was trained on

Out of distribution generalization

We also care about how models perform when tested on data that differs more substantially from their training data.



Out of distribution generalization

We also
perform
that diffe
from the

We aim to model commercial building energy consumption at the building level using machine learning models. This statistical approach avoids expensive physical modeling efforts, and is able to provide reasonable estimates that can be validated against existing building level energy consumption databases. Specifically, we train machine learning models on the 2012 Commercial Building Energy Consumption Survey microdata [15], then validate this approach using the Local Law 84 (LL84) dataset from New York City. We

Running the model on New York data is an “out of distribution”/”out of sample” validation test

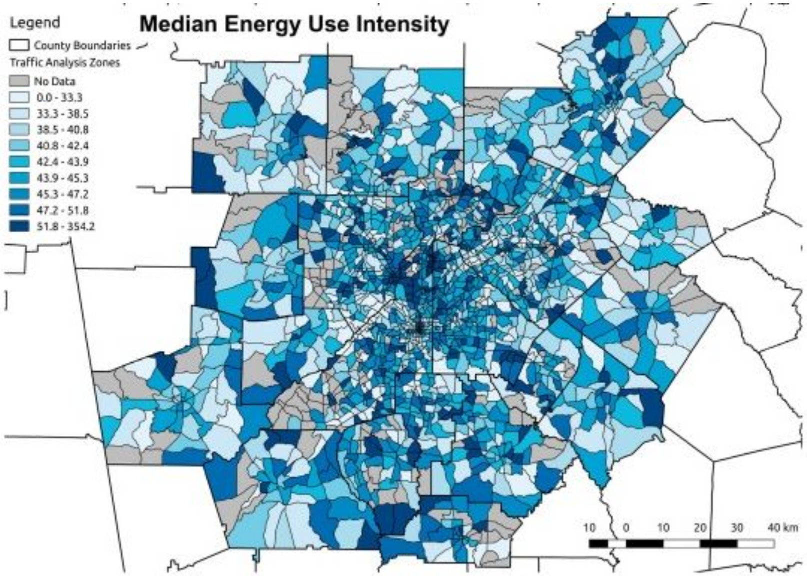
LL84	**Mean absolute error**	$10^{\text{Mean AE}}$	**Median absolute error**	$10^{\text{Median AE}}$	r^2
XGBoost - CBECS	0.25	1.78	0.15	1.41	0.51
XGBoost	**0.24±0.02**	**1.75±0.09**	**0.15±0.01**	**1.40±0.03**	**0.54±0.09**
SVR	0.25±0.02	1.77±0.10	**0.15±0.01**	**1.40±0.03**	0.51±0.11
Linear SVR	0.28±0.02	1.92±0.08	0.17±0.00	1.50±0.01	0.42±0.05
MLP regressor	0.28±0.04	1.92±0.17	0.17±0.02	1.48±0.06	0.44±0.13
Linear regression	0.29±0.02	1.96±0.10	0.19±0.01	1.56±0.05	0.44±0.08
Ridge regressor	0.29±0.02	1.96±0.10	0.19±0.01	1.56±0.05	0.44±0.08
Bagging	0.29±0.02	1.95±0.09	0.18±0.01	1.50±0.04	0.43±0.08
Random forest regressor	0.29±0.02	1.95±0.10	0.18±0.02	1.51±0.05	0.43±0.08
Extra trees regressor	0.30±0.03	2.00±0.12	0.18±0.01	1.51±0.05	0.39±0.09
KNN regressor	0.30±0.03	2.01±0.15	0.19±0.02	1.53±0.06	0.40±0.12
AdaBoost	0.42±0.07	2.67±0.43	0.30±0.04	2.01±0.20	0.14±0.22
Lasso	0.45±0.01	2.80±0.04	0.33±0.01	2.13±0.06	Negative
ElasticNet	0.45±0.01	2.80±0.04	0.33±0.01	2.13±0.06	Negative

What is the difference between these two?

Application to Atlanta

Shows ability to find necessary data and apply the model in other settings

But doesn't technically test performance



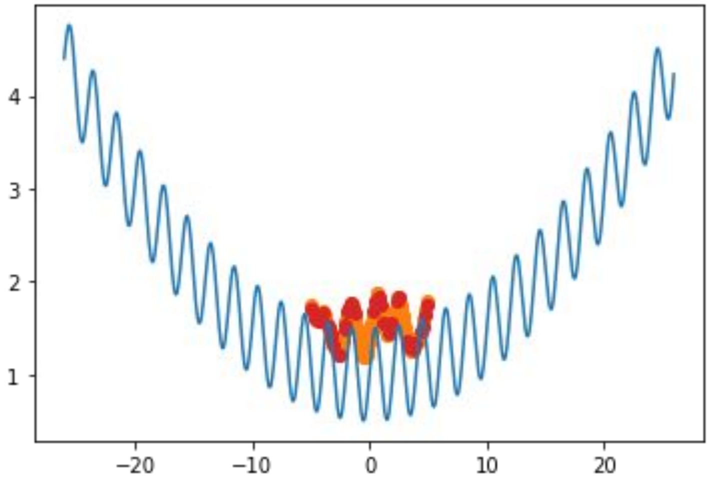
Discussion Question 6

The article states that “We note that the city of Atlanta’s new energy benchmarking ordinance for commercial buildings may change this geography of energy consumption in commercial buildings. It aims to achieve a 20% reduction of energy consumption in Atlanta’s private and City-owned buildings over 25,000 square feet, by 2030”.

Would such a change impact the accuracy of the model?

Could the model be applied to residential buildings?

If the function you are approximating changes...



Then your model likely won't do a good job approximating it anymore!

(e.g. if buildings become more efficient)

Generalizing to residential buildings

Assuming the common feature model (using principal building activity, square feet, number of floors, heating degree days, and cooling degree days), the mapping between inputs and outputs is probably not the same as for commercial buildings (e.g. it is a different function to approximate).

“Principal building activity” could have an indicator for “residential” that could help the model approximate this specific function. But right now it doesn’t, and would therefore require retraining.

If using “extended features” there are many inputs not relevant to residential buildings, and features that are relevant for residential buildings are missing.

Discussion Question 7

Based on the results in this paper, what additional data would be best to collect more broadly in order to improve prediction performance?

Extended features	Mean absolute error	$10^{\text{Mean AE}}$	Median absolute error	$10^{\text{Median AE}}$	r^2
XGBoost with common features	0.30±0.01	1.99±0.06	0.22±0.01	1.66±0.03	0.82±0.02
XGBoost	**0.23±0.01**	**1.69±0.02**	**0.17±0.01**	**1.48±0.03**	**0.89±0.01**

Common feature set			Extended feature set		
Feature name	**Feature description**	**Importance**	**Feature name**	**Feature description**	**Importance**
SQFT	Square footage	0.3634	SQFT	Square footage	0.1391
CDD65	Cooling degree days (base 65)	0.1153	NWKER	Number of employees	0.0576
HDD65	Heating degree days (base 65)	0.1125	WKHRS	Total hours open per week	0.0557
PBA 5	Non-refrigerated warehouse	0.0569	ZMFBTU	Imputed major fuels consumption	0.0312
PBA 1	Vacant	0.0524	MONUSE	Months in use	0.0299
PBA 6	Food sales	0.0412	NGUSED	Natural gas used	0.0295
PBA 15	Food service	0.0384	HDD65	Heating degree days (base 65)	0.0293
PBA 23	Strip shopping mall	0.0348	HEATP	Percent heated	0.0278
PBA 12	Religious worship	0.0345	CDD65	Cooling degree days (base 65)	0.0224
PBA 4	Laboratory	0.0282	NWKERC	Number of employees category	0.0221

Number of employees and total hours open per week are plausible additional features that could be collected more broadly, and would likely increase performance

Discussion Question 8

Share what questions you wrote in your PMIRO+Q and decide as a group what you'd like to ask.

Update your PMIRO+Q

Submit a second file to the Brightspace assignment (don’t overwrite the original):

It should:

Update your PMIRO as needed

Answer your own Q

You can be talking with your group during this!

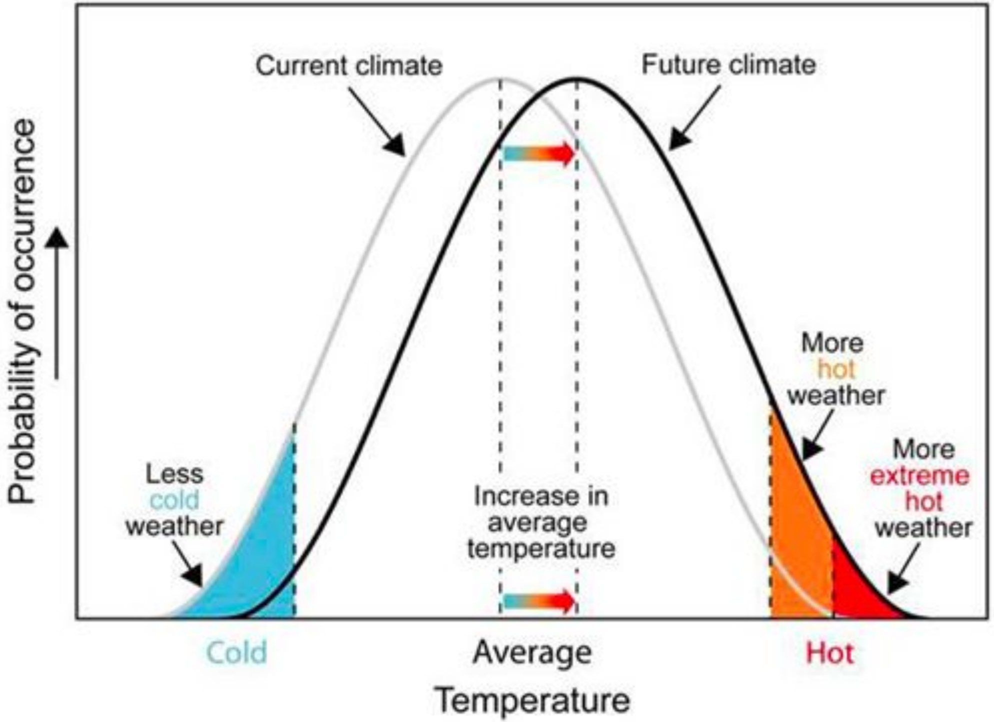
15 min break

Topics

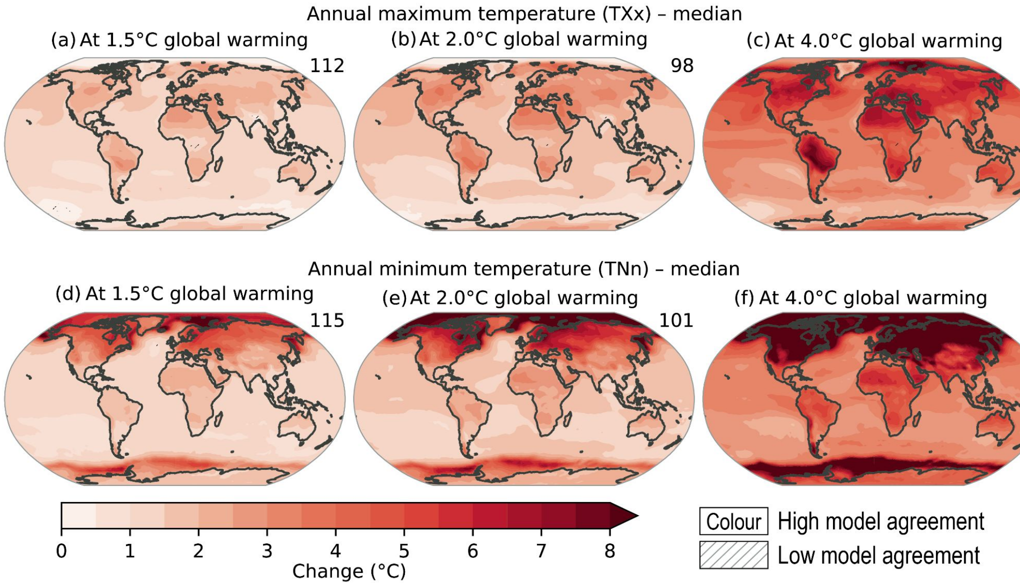
Climate Change content: Extreme Weather and Disaster Response

Machine Learning content: computer vision, artificial neural networks

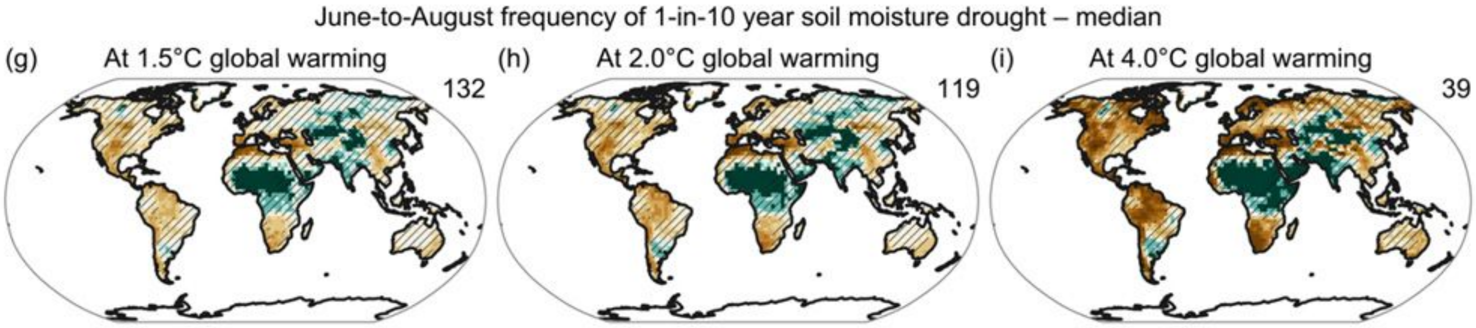
Small increases in averages can cause large changes in extremes



Heatwaves

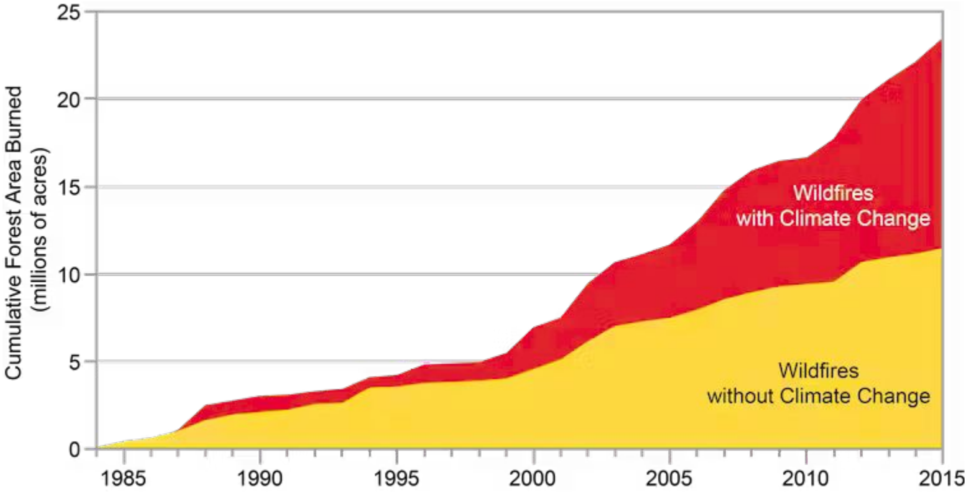


Drought



IPCC

Wildfires

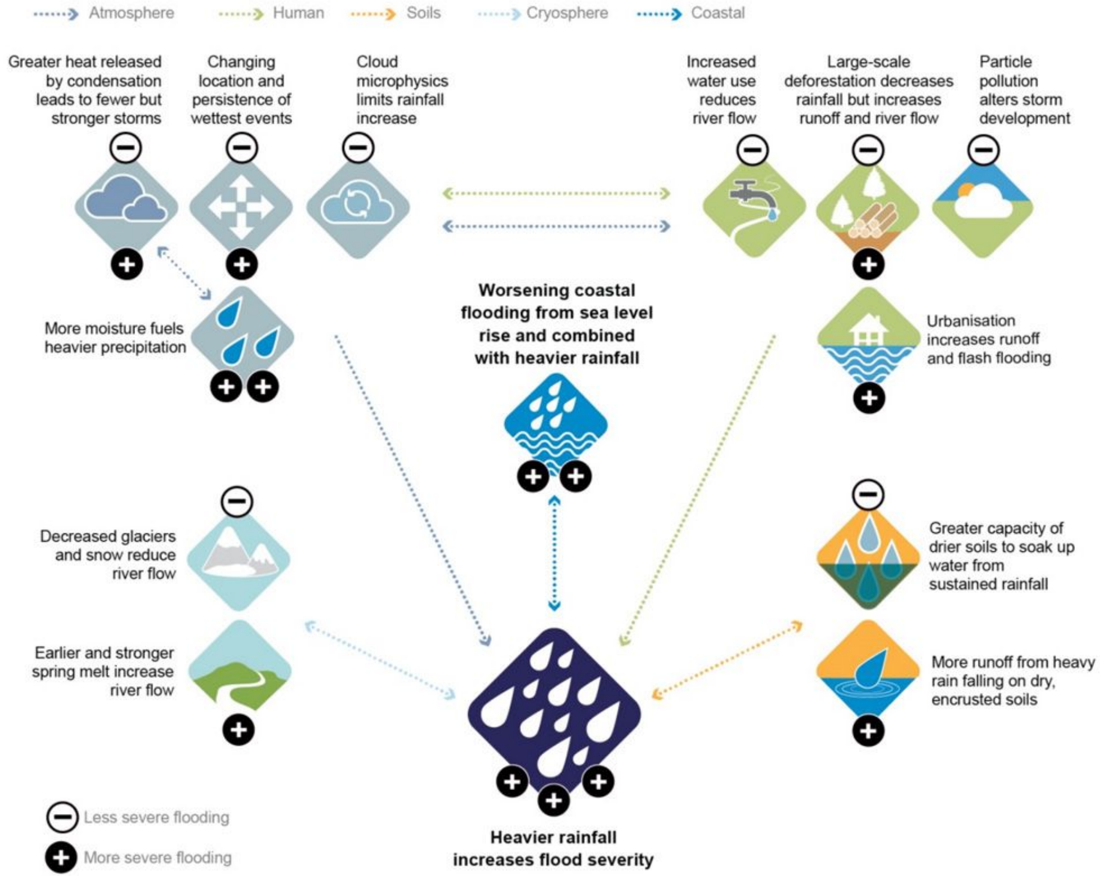


Union of Concerned Scientists

Fourth National Climate Assessment

FAQ 8.2: **Causes of more severe floods from climate change**

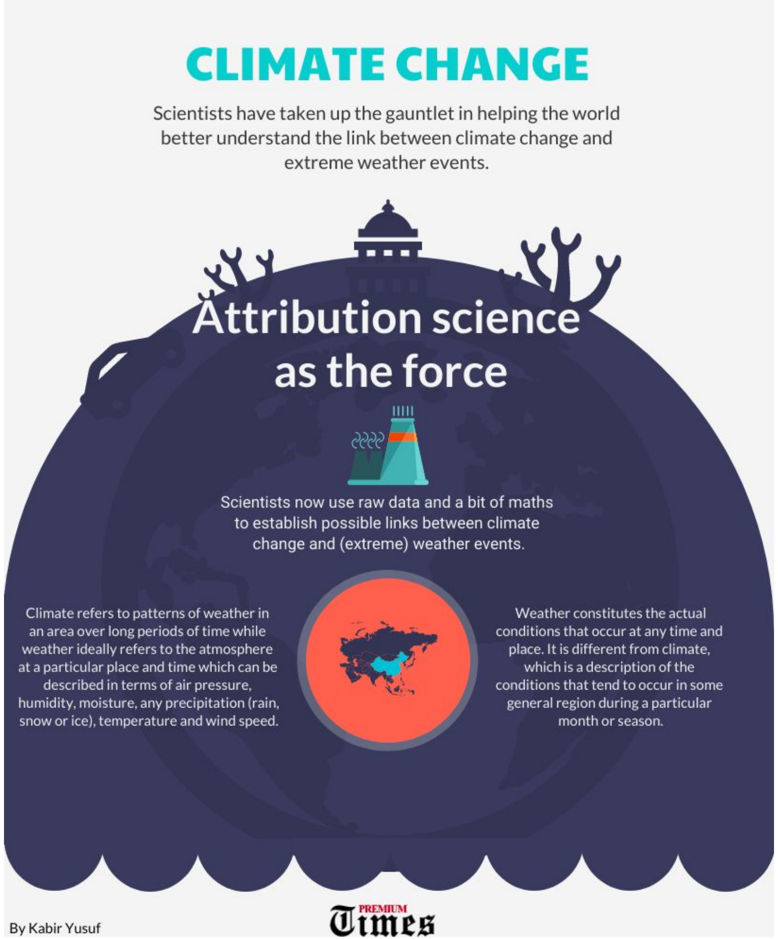
Flooding presents a hazard but the link between rainfall and flooding is not simple.
While the largest flooding events can be expected to worsen, flood occurrence may decrease in some regions.



IPCC

How climate change makes winter storms worse

1:03 to 7:16



Attribution Science helps make the argument to policy makers about how specific events were much more likely/extreme due to climate change

Disaster Response and Recovery

Response: the use of resources (including personnel, supplies and equipment) to help restore personal and environmental safety, as well as to minimize the risk of any additional property damage after the disaster

Recovery: involves stabilizing the area and restoring all essential community functions. Recovery requires prioritization: first, essential services like food, clean water, utilities, transportation and healthcare will be restored, with less-essential services being prioritized later. This can take years or decades

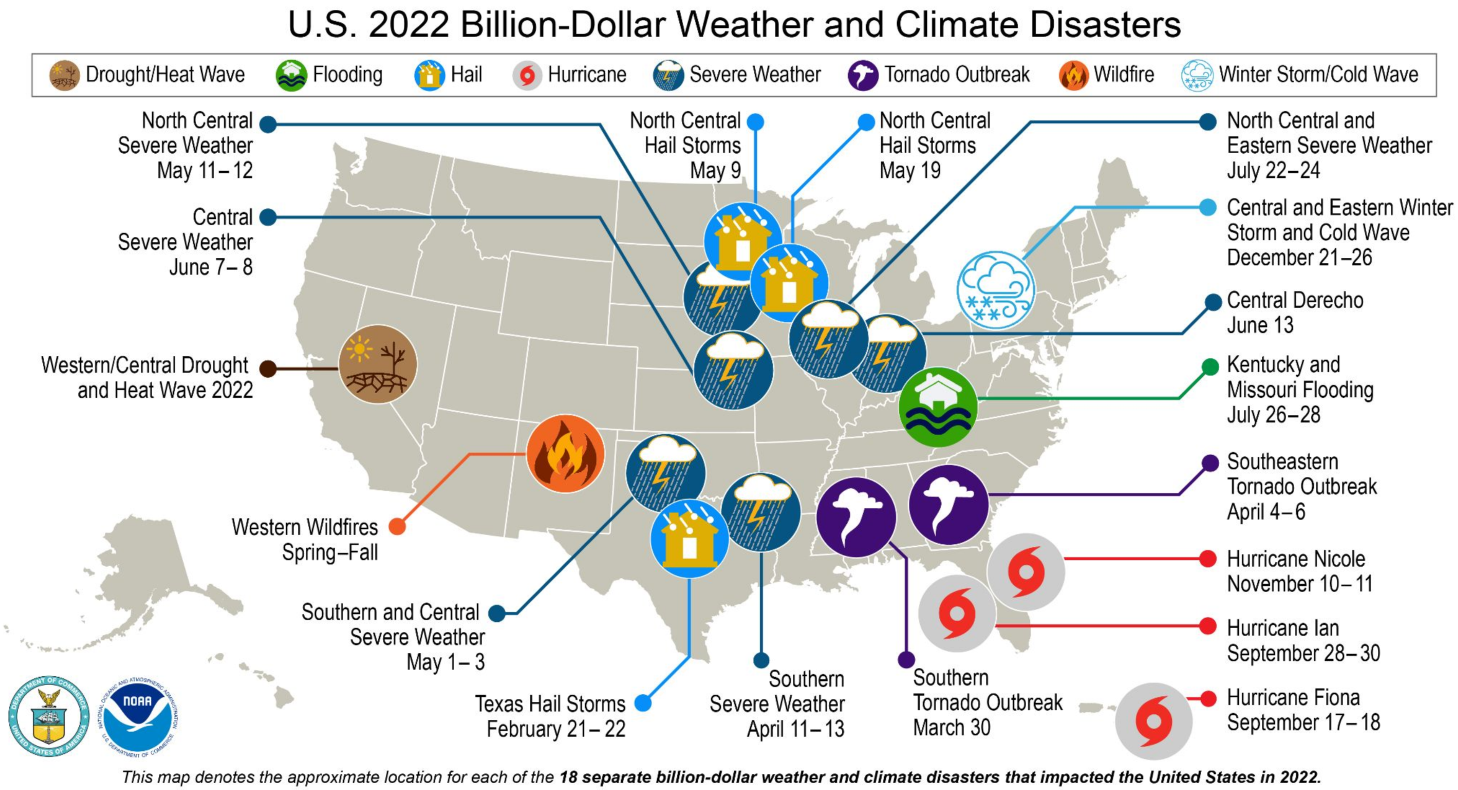
https://www.ucf.edu/online/leadership-management/news/the-disaster-management-cycle/

Disaster Response and Recovery

Governments and non-profits are primarily responsible for disaster response.

https://www.ucf.edu/online/leadership-management/news/the-disaster-management-cycle/

Costs of disaster response



Disaster Response and Recovery

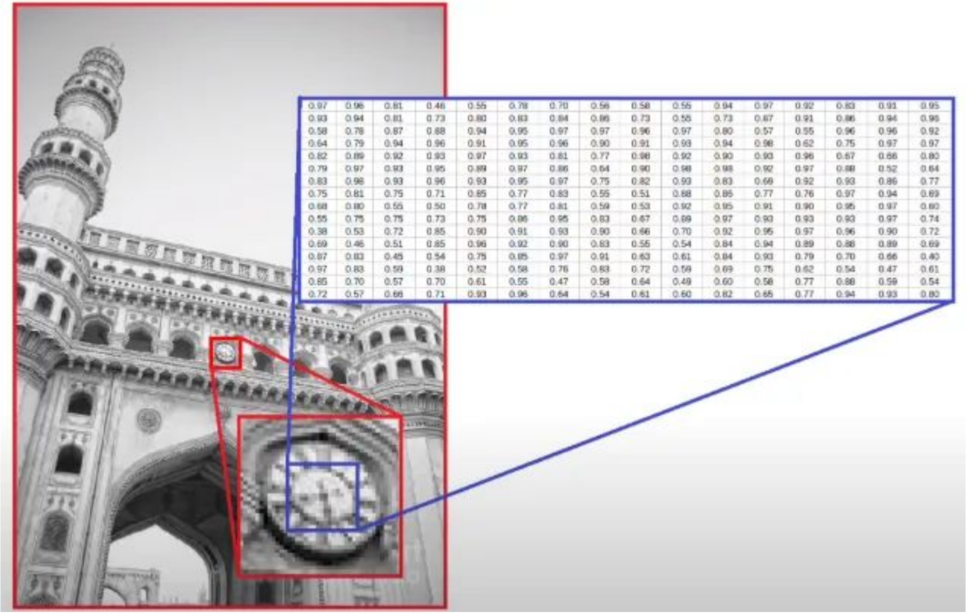
The importance of mapping damaged areas:

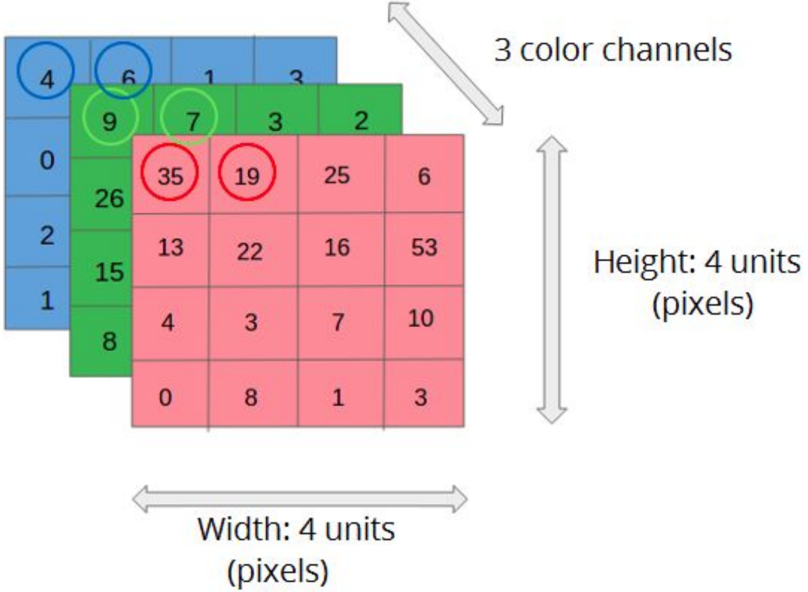
Machine learning to help disaster response

We can use computer vision techniques applied to aerial imagery to identify damaged regions after a disaster

Computer vision

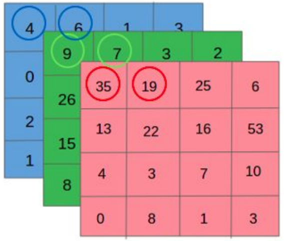
Images are data

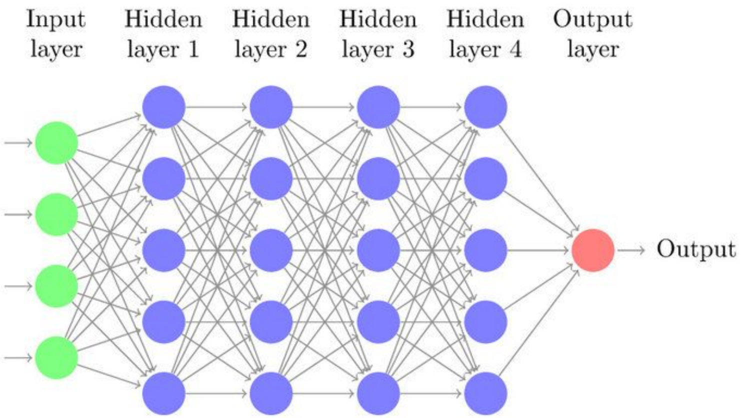




NPTEL

Images can be fed into artificial neural networks



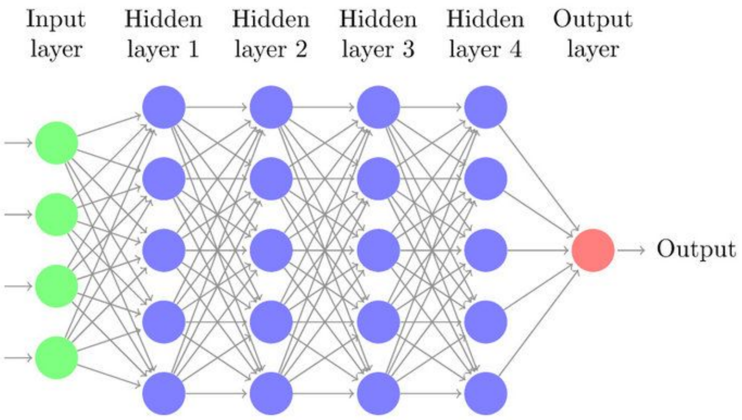


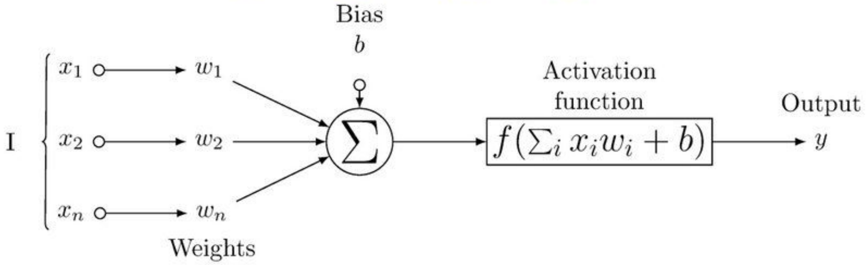
Aerial image of building

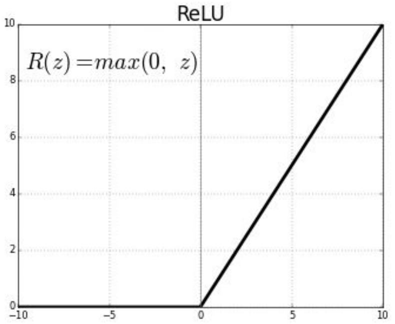
Amount of damage

Artificial neural networks are made of artificial neurons

The number of hidden layers is referred to as the “depth” of the network, hence “deep learning”.



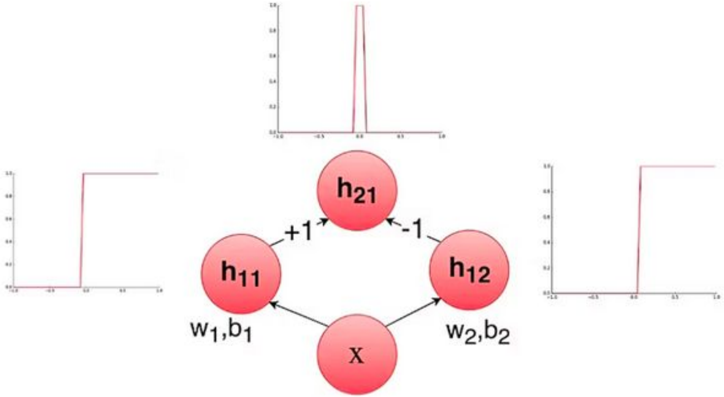


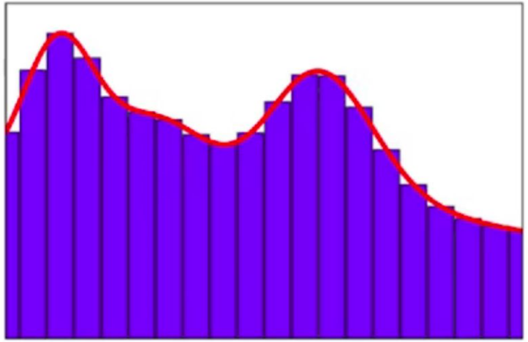


Artificial neural networks can be “universal function approximators”

By stacking many simple nonlinear functions, neural networks can approximate more complex functions.

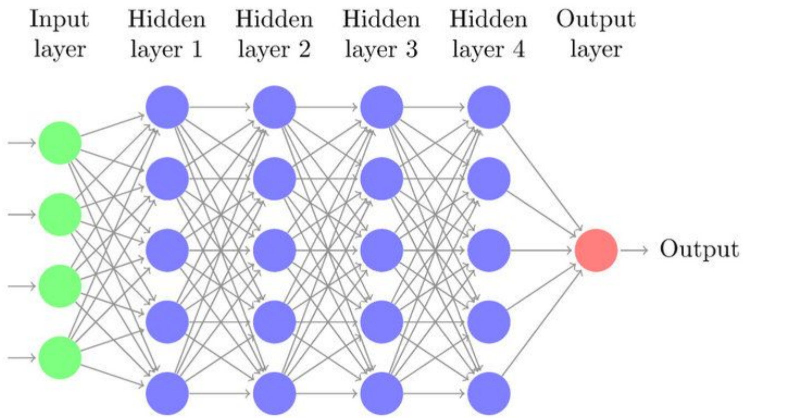
The exact functions they approximate is controlled by the weights.





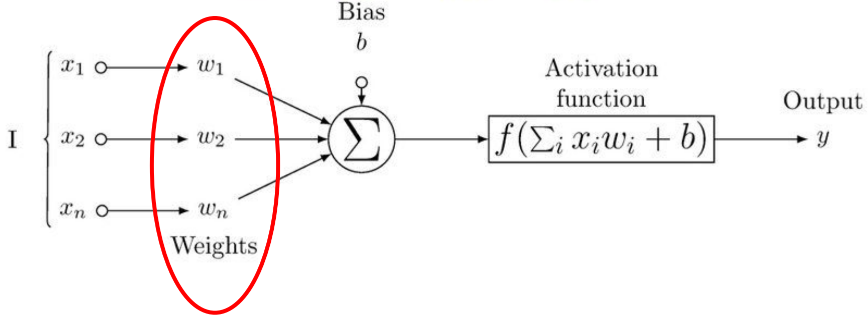
hackernoon

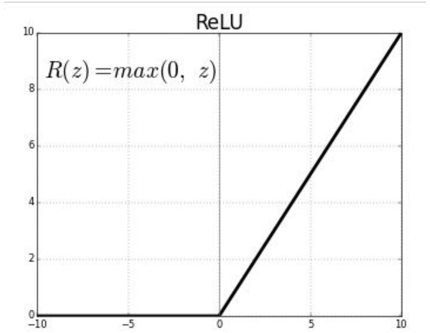
Artificial neural networks are made of artificial neurons



How do we set these weights?

Finding the best weights is known as “training” the network





The loss function tells the network what we want it to do

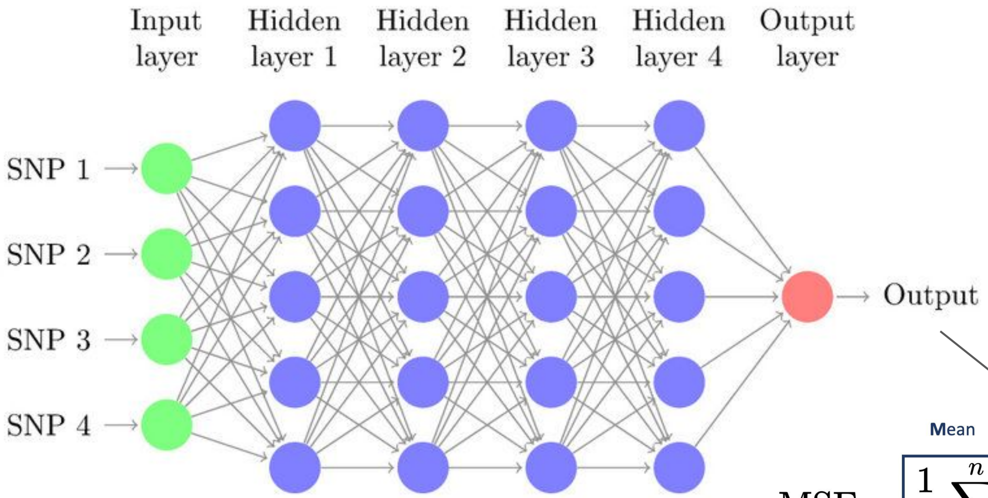
If we want to train a model on a regression problem, for example, we may use Mean Squared Error as the loss function.

Also known as “cost” or “objective” function. Higher values mean the model is performing poorly.

When we have the “correct answer” that we can train the network with, this is known as “supervised learning”

$$\mathrm{MSE} = \overbrace{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}^{\text{Mean}}\Big(\overbrace{\boxed{\underbrace{Y_i}_{\text{Correct answer}} - \underbrace{\hat{Y}_i}_{\text{Output of the model}}}}^{\text{Error}}\Big)^{\overbrace{\boxed{2}}^{\text{Squared}}}$$

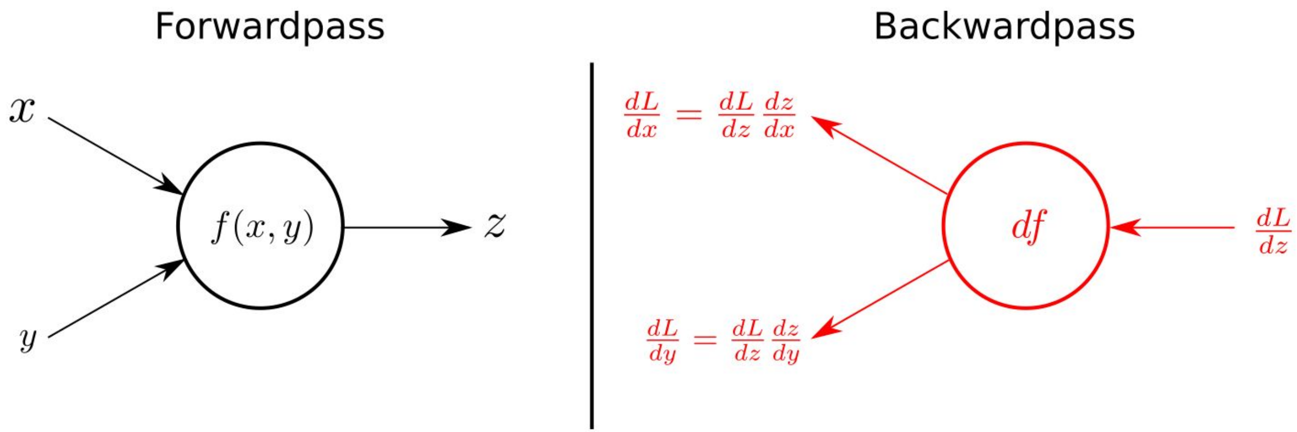
How do we use the loss function to learn the right weights?



$$\text{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n}\sum_{i=1}^{n}}} \overset{\text{Error}}{\boxed{(Y_i - \hat{Y}_i)}}^{\overset{\text{Squared}}{\boxed{2}}}$$

“Backpropagate
the error”

Backpropagation algorithm



Kratzert

By applying the chain rule for derivatives, we can calculate exactly in which direction a weight should change in order to make the loss function decrease

Backpropagation algorithm

Derivative of Error with respect to weight

Old weight

$$ {}^{*}W_x = W_x - a\left(\frac{\partial Error}{\partial W_x}\right) $$

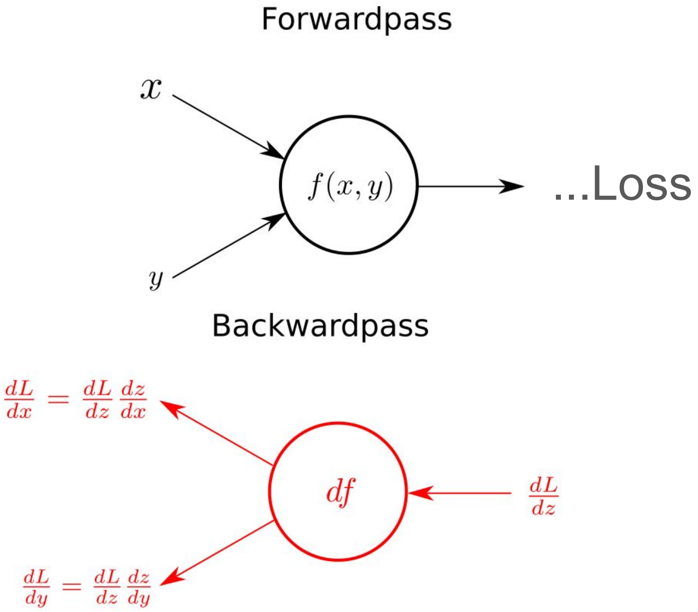
New weight

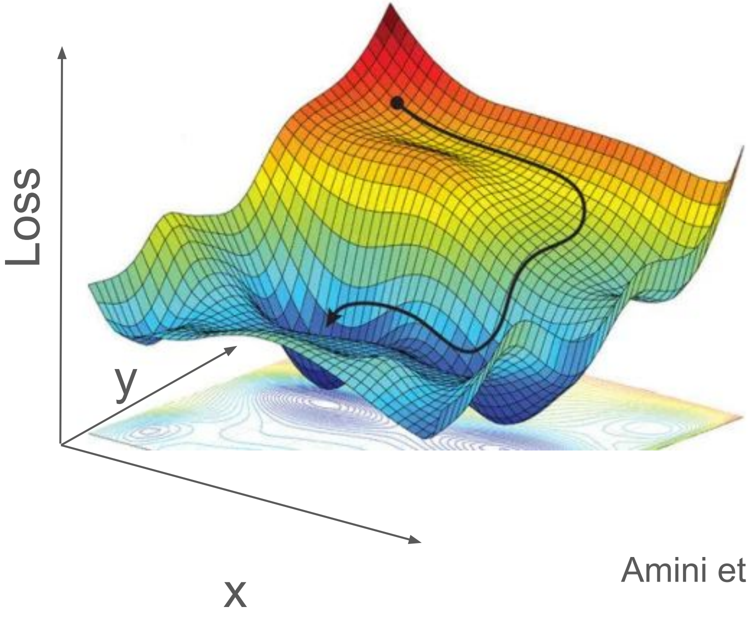
Learning rate

Weights are updated in this direction, with a magnitude dependent on the learning rate parameter.

Gradient descent

One way to think of training is as a “downhill walk” through parameter space, according to the loss function. The derivatives with respect to the loss function are known as gradients.



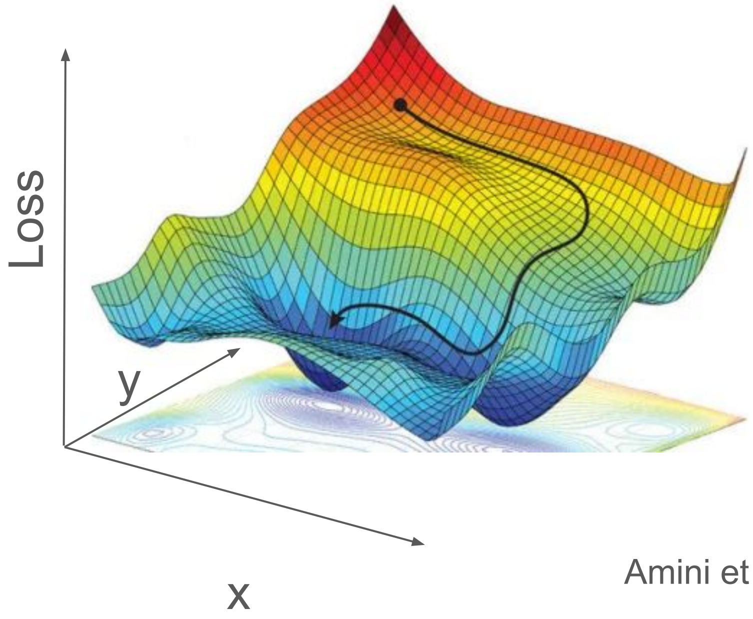


Amini et al

“Stochastic” Gradient descent

These steps are taken one per “batch” of inputs (batches are made by dividing the full training data into many smaller sets).

A full pass through all batches in the training data is known as an “epoch”

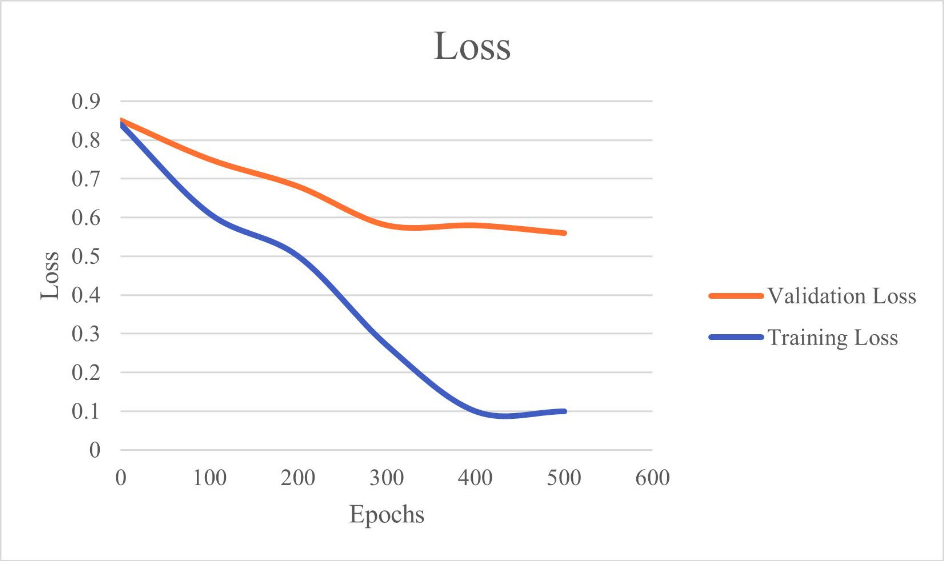


Amini et al

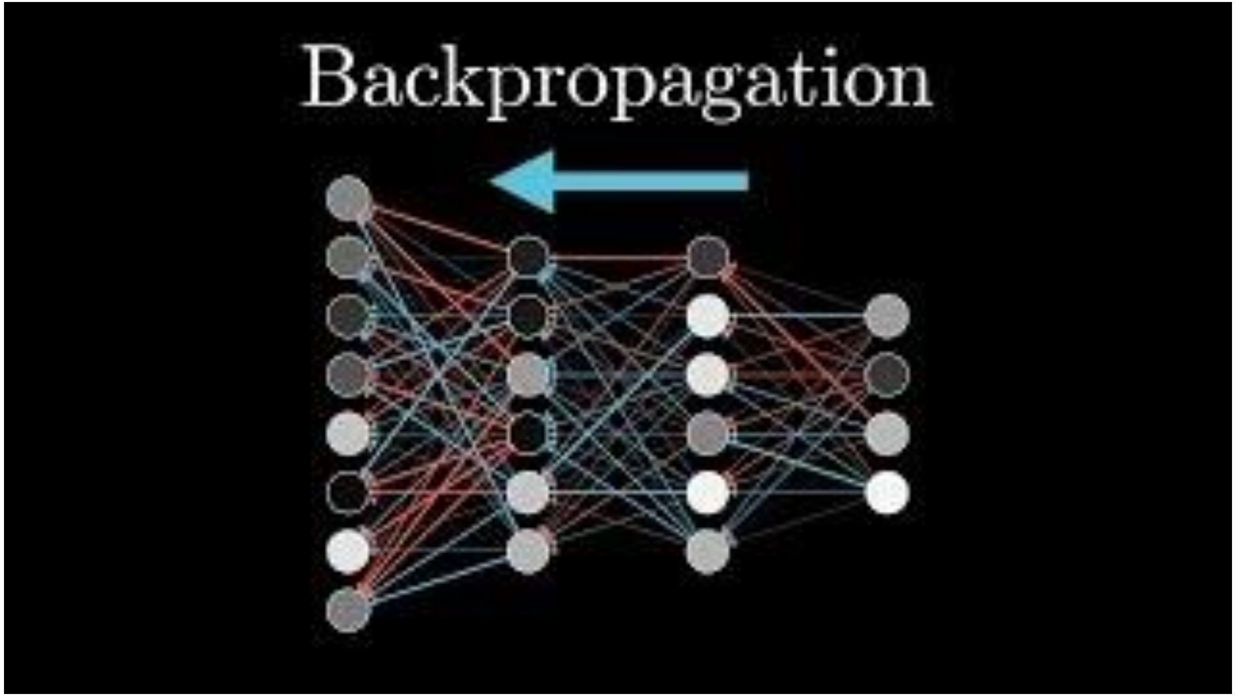
So, to train a network you:

1. Pass a batch of images through it and get the model’s output
2. Calculate the value of the loss function on those images using the correct answers
3. Calculate the average of the gradient of the loss function for that batch with respect to the weights
4. Move the weights in the direction the gradients tell you to
5. Repeat until you’ve gone through at least one full epoch of your training data (but usually many more)

If we train on a set of training data, we need to...



Validate how well the model generalizes to a held out test set


Backpropagation

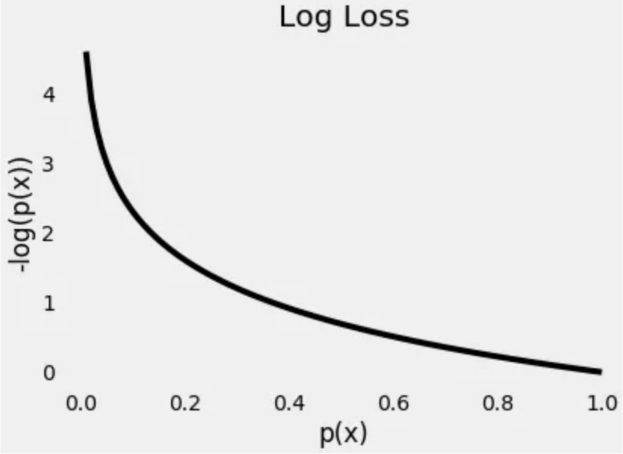
For your reading:

The paper uses a specific type of artificial neural network commonly used in computer vision called a “convolutional neural network”. We will get into the details of this kind of model next week.

For your reading:

The paper also explores “cross entropy” loss functions. These are normally used when you are trying to categorize your inputs rather than map them to a continuous value (regression). In this case, the model outputs a probability distribution over categories and the loss is:

$$H(P^*|\,P) = -\sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTRIBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTRIBUTION}}$$



Reminders for your PMIRO+Q

Keep it short! Get to the essence!

Use your own words, don’t copy-paste