ML4CC: Lecture 11

Sit with your discussion groups (same as last time)

Assignments reminder

Keep doing your PMIRO+Q

Apr 18 - Project check-ins **during class**

Apr 25 - Exam II

May 2 - Project Presentations

Project reports due May 9th.

Summary of last paper

P - Want to advance methods for real-world reinforcement learning problems like EV charging

M - Create a benchmark dataset and try standard RL methods

I - Uses real world data related to sustainability problems

R - Standard methods don’t do particularly well, leaving room for improvement

O - how high can performance be on this problem?

Climate Change in the News

CarbonBrief
CLEAR ON CLIMATE

A victory for Donald Trump in November's presidential election could lead to an additional 4bn tonnes of US emissions by 2030 compared with Joe Biden's plans, Carbon Brief analysis reveals.

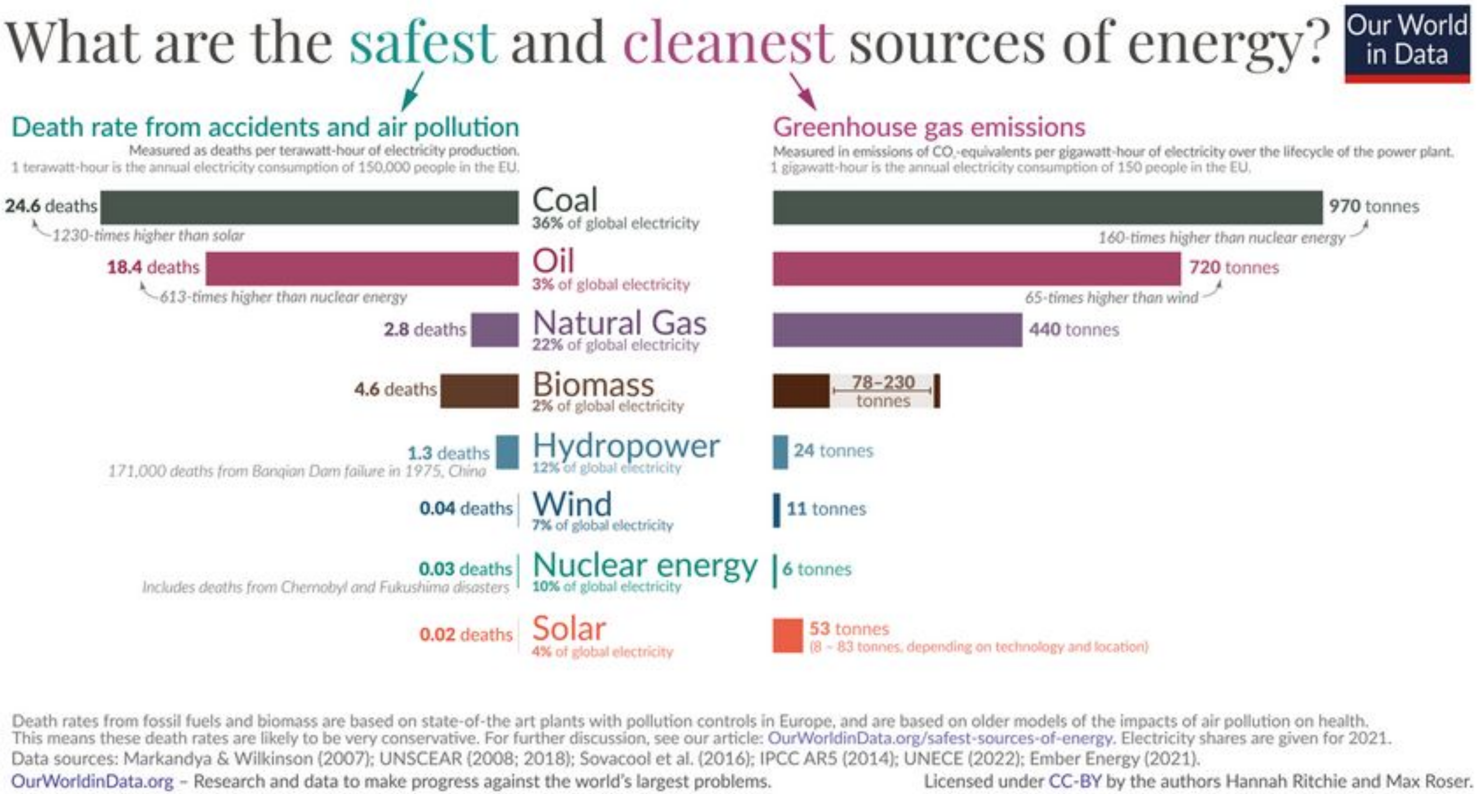
This extra 4bn tonnes of carbon dioxide equivalent ($GtCO_2e$) by 2030 would cause global climate damages worth more than $900bn, based on the latest US government valuations.

For context, $4GtCO_2e$ is equivalent to the combined annual emissions of the EU and Japan, or the combined annual total of the world's 140 lowest-emitting countries.

Put another way, the extra $4GtCO_2e$ from a second Trump term would negate – twice over – all of the savings from deploying wind, solar and other clean technologies around the world over the past five years.

If Trump secures a second term, the US would also very likely miss its global climate pledge by a wide margin, with emissions only falling to 28% below 2005 levels by 2030. The US's current target under the Paris Agreement is to achieve a 50-52% reduction by 2030.

Carbon Brief's analysis is based on an aggregation of modelling by various US research groups. It highlights the significant impact of the Biden administration's climate policies. This includes the Inflation Reduction Act – which Trump has pledged to reverse – along with several other policies.



What are the safest and cleanest sources of energy?
Our World in Data
Death rate from accidents and air pollution
Measured as deaths per terawatt-hour of electricity production.
1 terawatt-hour is the annual electricity consumption of 150,000 people in the EU.
Greenhouse gas emissions
Measured in emissions of CO2-equivalents per gigawatt-hour of electricity over the lifecycle of the power plant.
1 gigawatt-hour is the annual electricity consumption of 150 people in the EU.
24.6 deaths
1230-times higher than solar
Coal
36% of global electricity
970 tonnes
160-times higher than nuclear energy
18.4 deaths
613-times higher than nuclear energy
Oil
3% of global electricity
720 tonnes
65-times higher than wind
2.8 deaths
Natural Gas
22% of global electricity
440 tonnes
4.6 deaths
Biomass
2% of global electricity
78-230 tonnes
1.3 deaths
171,000 deaths from Banqian Dam failure in 1975, China
Hydropower
12% of global electricity
24 tonnes
0.04 deaths
Wind
7% of global electricity
11 tonnes
0.03 deaths
Includes deaths from Chernobyl and Fukushima disasters
Nuclear energy
10% of global electricity
6 tonnes
0.02 deaths
Solar
4% of global electricity
53 tonnes
(8 – 83 tonnes, depending on technology and location)
Death rates from fossil fuels and biomass are based on state-of-the art plants with pollution controls in Europe, and are based on older models of the impacts of air pollution on health.
This means these death rates are likely to be very conservative. For further discussion, see our article: OurWorldinData.org/safest-sources-of-energy. Electricity shares are given for 2021.
Data sources: Markandya & Wilkinson (2007); UNSCEAR (2008; 2018); Sovacool et al. (2016); IPCC AR5 (2014); UNECE (2022); Ember Energy (2021).
OurWorldinData.org – Research and data to make progress against the world's largest problems.
Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Paper 9 Discussion

Reduced Optimal Power Flow Using Graph Neural Network

Thuan Pham
Student Member, IEEE
Department of Electrical and Computer Engineering
University of Houston
Houston, TX, USA
tdpham7@cougarnet.uh.edu

Xingpeng Li
Senior Member, IEEE
Department of Electrical and Computer Engineering
University of Houston
Houston, TX, USA
xli82@uh.edu

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 (remember to mark who took attendance!). **If someone is virtual, mark it with a V**.

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

Discussion Question 1

What is a ROPF and why is it desirable here?

Reduced Optimal Power Flow problem

formulation of the OPF problem [13]. For this paper, we designed a GNN model that can predict overloaded and congested lines in the power network with different given load profiles. Predictions from the GNN model are used to identify a subset of critical lines for supervision. We want to lower the number of monitoring lines or number of line flow limit constraints in the OPF problem, in essence, producing a reduced OPF (ROPF) problem. With an ROPF model, it is expected that the amount of computing time for finding the optimal solution will decrease, especially for large and complex power systems.

OPF problems grow with complexity as more components and constraints are added to the grid.

By identifying a subset of the grid to focus our constraints on, we can make the OPF problem easier.

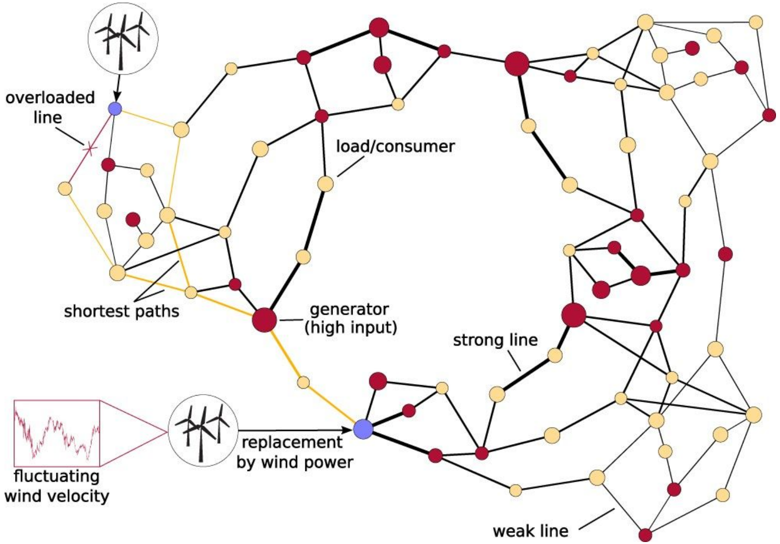
Formulation	Variables	Number of variables	Number of equations
BIM	$V_i = (v_i e^{j\delta_i})$ $S_g^G = (p_g^G + jq_g^G)$ $S_l^L = (p_l^L + jq_l^L)$	$N + G + L$ $(2N + 2G + 2L)$	N $(2N)$
BFM	$V_i = (v_i e^{j\delta_i})$ $S_g^G = (p_g^G + jq_g^G)$ $S_l^L = (p_l^L + jq_l^L)$ $I_{ij}^s = (i_{ij}^s e^{j\gamma_{ij}^s})$ $S_{ij} = (p_{ij} + jq_{ij})$ $S_{ji} = (p_{ji} + jq_{ji})$	$N + G + L + 3E$ $(2N + 2G + 2L + 6E)$	$N + 3E$ $(2N + 6E)$

Discussion Question 2

How is the data generated here?

Solving the OPF problem for a bunch of fake examples

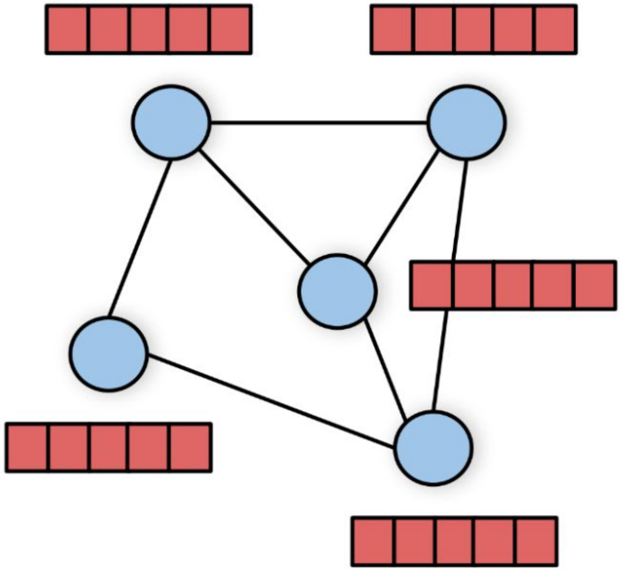
To generate a large number of samples to sufficiently train and test a GNN model, we run normal OPF simulations on the IEEE 73-bus system with different load profiles to collect 20,000 samples. The load profile of each sample is varied within ±10% of the base load profile. From the 20,000 generated samples, the data set is divided into three groups: 80% for training, 10% for validation, and 10% for testing. Then, we used Pyomo library in python to solve for the optimal solution of each sample [15] [16]. Based on the solution, we created labels for each branch based on a pre-specified loading threshold that is defined as a percentage of the line rating limit. For example, if the line rating limit is 100MW and the loading threshold is 80%, a branch will be classified/labeled as heavily loaded/congested when its flow is over 80MW. The labels are represented as one-hot encoding and used as output during the training process.



Lines that, according to the OPF solution, would need to be working at a large percentage of their capacity are labeled as congested.

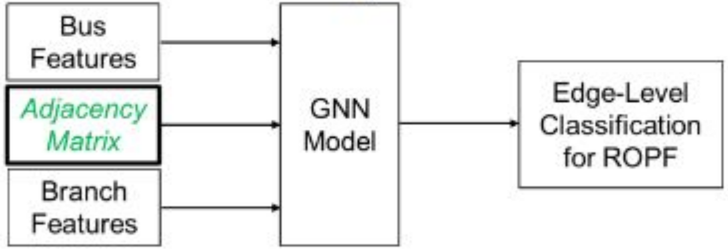
Discussion Question 3

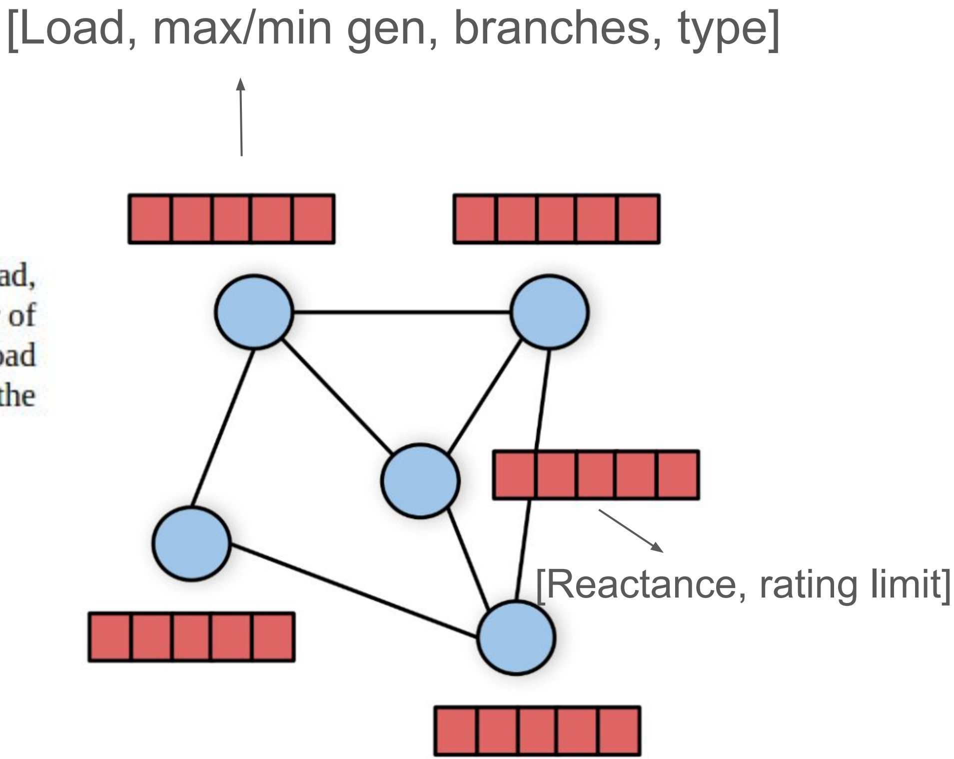
What is the input to the GNN model? What does the model output?



Input: Node and Edge Info

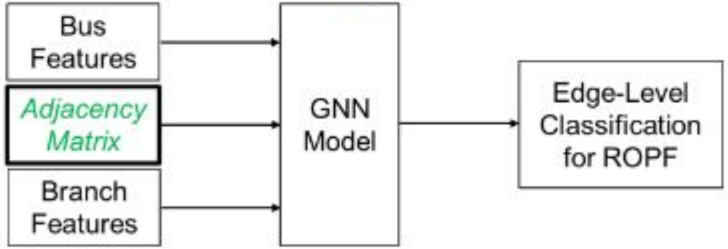
For each sample, the node features are the nodal load, maximum and minimum generation at each bus, the number of branches (edges) connected to each bus, and bus types (load bus, generator bus and slack bus). The edge features include the line reactance and the line rating limit.

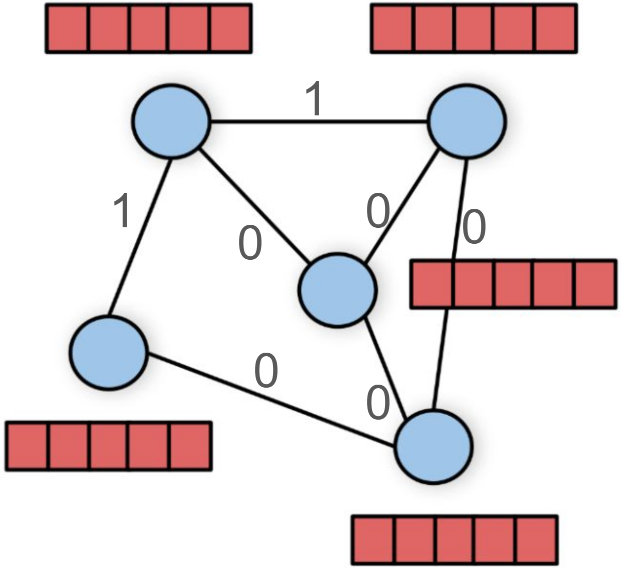




Output: Edge Classification

For each sample, the node features are the nodal load, maximum and minimum generation at each bus, the number of branches (edges) connected to each bus, and bus types (load bus, generator bus and slack bus). The edge features include the line reactance and the line rating limit.





Discussion Question 4

What other model architectures are trained? How do they differ in structure from the GNN?

Regular ANN and Convolutional NN

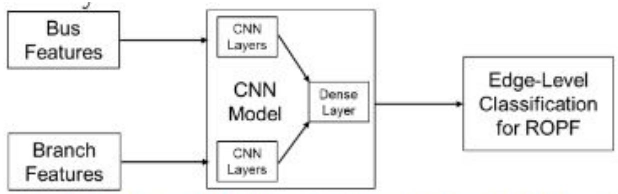


Fig. 5. Illustration of the CNN model. (4 CNN layers for bus features, 4 CNN layers for branch features, and 1 Dense layer to combine both features)

To evaluate how the GNN model perform against typical NN and CNN models, we built separate NN and CNN models that mimic the GNN model. For these models, node features and branch features are each passed and trained separately through four layers. Then, the outputs are combined in a dense layer with softmax activation. As we can see from Fig. 5 and Fig. 6, there is no adjacency matrix to keep track of network topology or global context during the training process.

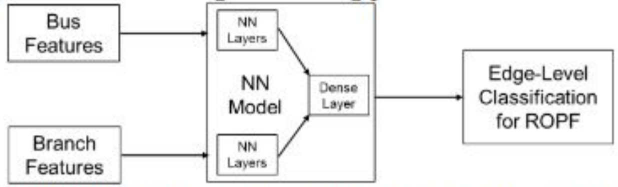
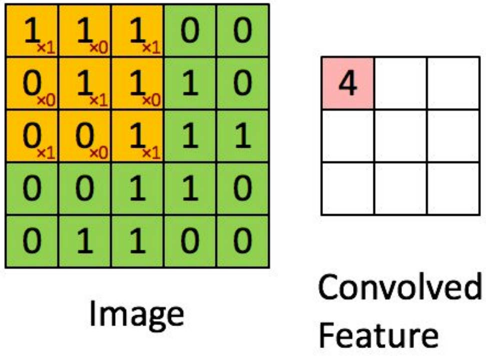


Fig. 6. Illustration of the NN model. (4 NN layers for bus features, 4 NN layers for branch features, and 1 Dense layer to combine both features)

Convolution:



These other models do not take in the adjacency matrix. They also process branch and bus features separately until the end of the network

Discussion Question 5

Describe Figure 10 in your own words

Error distribution

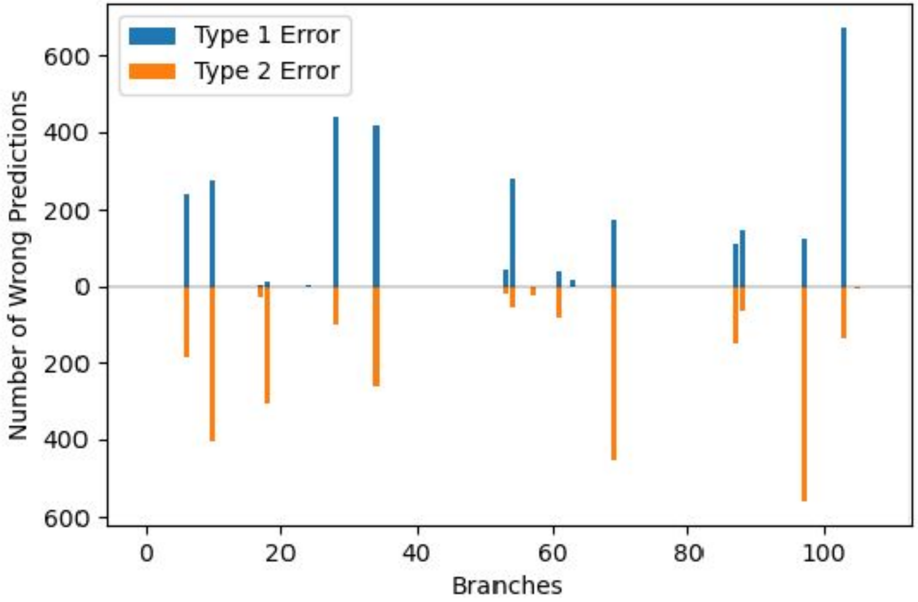


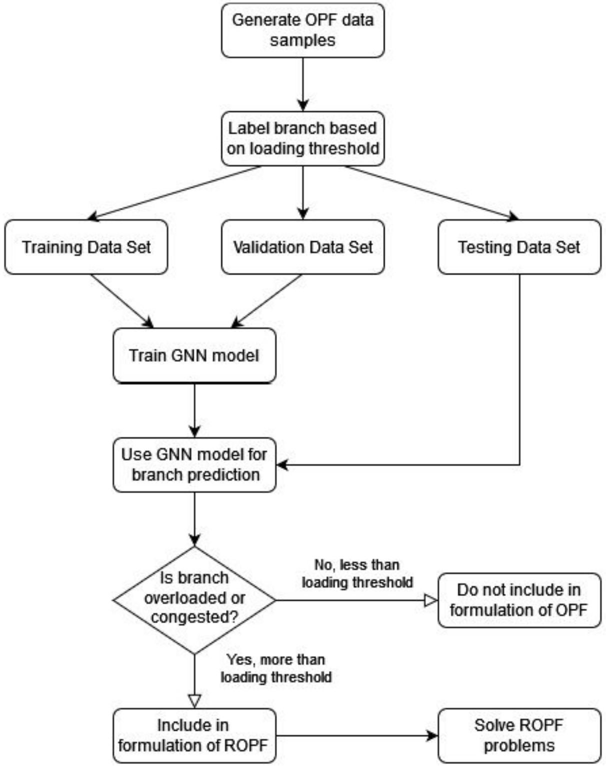
Fig. 10. Number of wrong predictions for Type 1 vs. Type 2 error per branch.

The plot shows that false positive errors (saying a line is congested when it isn't) and false negative errors (saying it is not congested when it is) tend to occur on the same lines, suggesting a small set of lines are generally hard to predict.

Discussion Question 6

Explain figure 13 in your own words. Why does this result make sense?

False negatives lead to physical violations



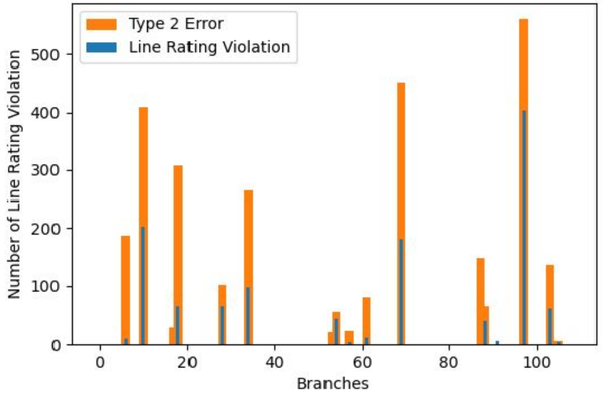


Fig. 13. Line rating violation is superimposed on top of type 2 error per branch.

When branches are incorrectly labeled as non-congested, they are not included in the ROPF problem. This means the ROPF solution is not required to obey their line limits. Therefore, the solution sometimes does indeed violate them.

Discussion Question 7

Describe the trends shown in Figures 14 and 15. Why do the changes as a function of threshold make sense?

Increasing threshold increases error and reduces compute time

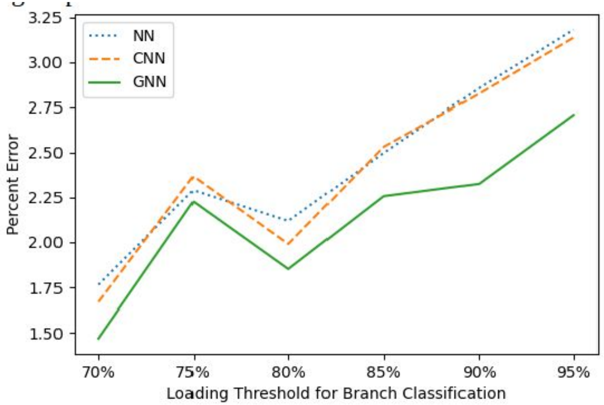
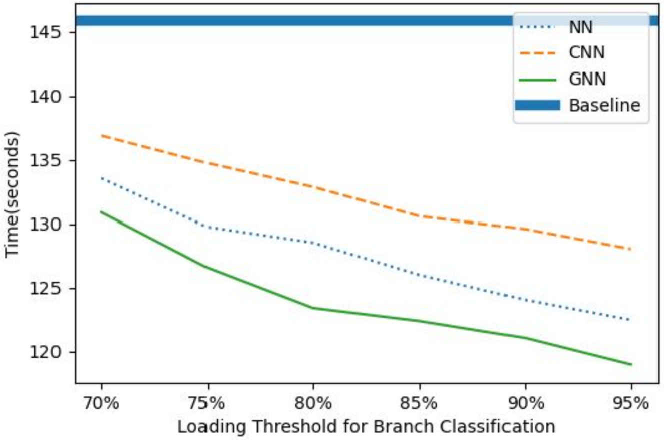


Fig. 14. Percent error of prediction at multiple loading thresholds from different ML models.



The GNN performs best and runs in the least time.

Increasing the threshold for what counts as congested reduces the number of lines monitors in the ROPF (and therefore the run time) but may involve ignoring lines that are congested.

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

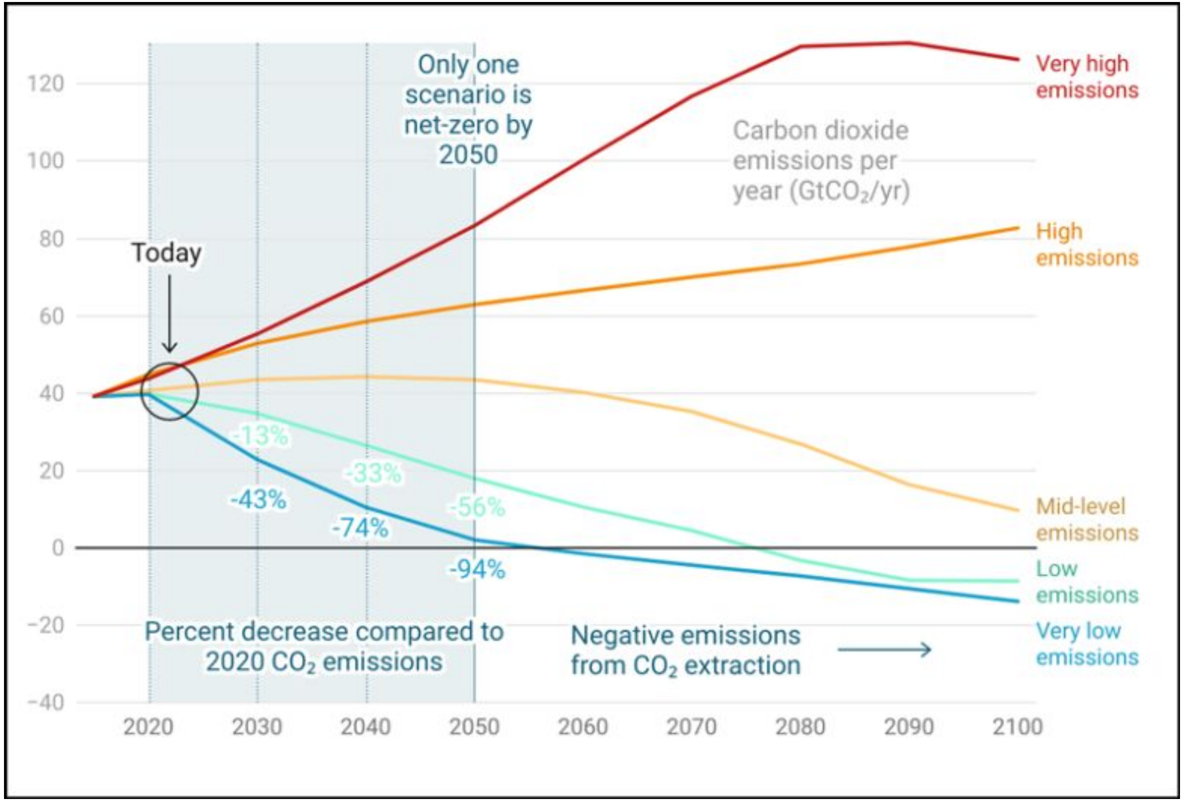
Lecture

Climate Content: Carbon dioxide removal

Machine learning: Review of many topics, Multi-task learning

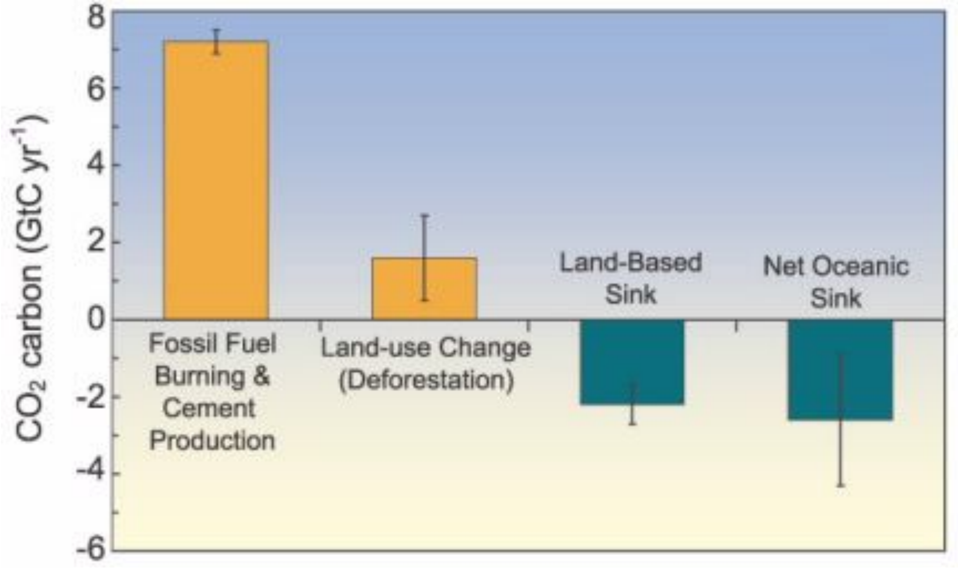
Negative emissions

Most models of how we can now stay below 2 degrees C require net negative emissions at some point.



"Net zero"

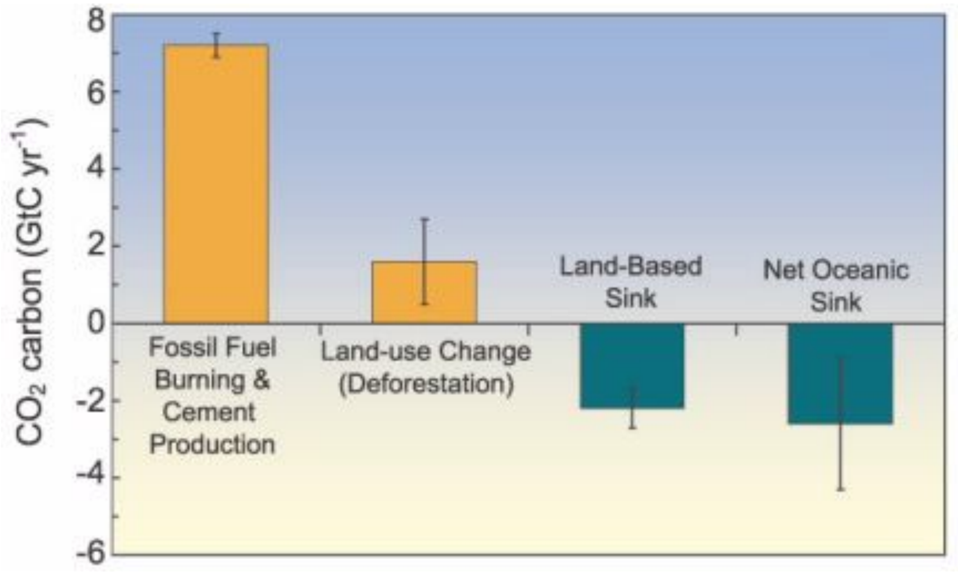
'Net-zero emissions' means that all sources of emissions are balanced out by the removal of GHGs.



“Net zero”

‘Net-zero emissions’ means that all sources of emissions are balanced out by the removal of GHGs.

Net *negative* emissions means that more GHGs are removed from the atmosphere than emitted



The IPCC on carbon dioxide removal

Methods for removing CO2 from the atmosphere are "unavoidable" if the world is to reach net-zero – both globally and nationally, the report says.

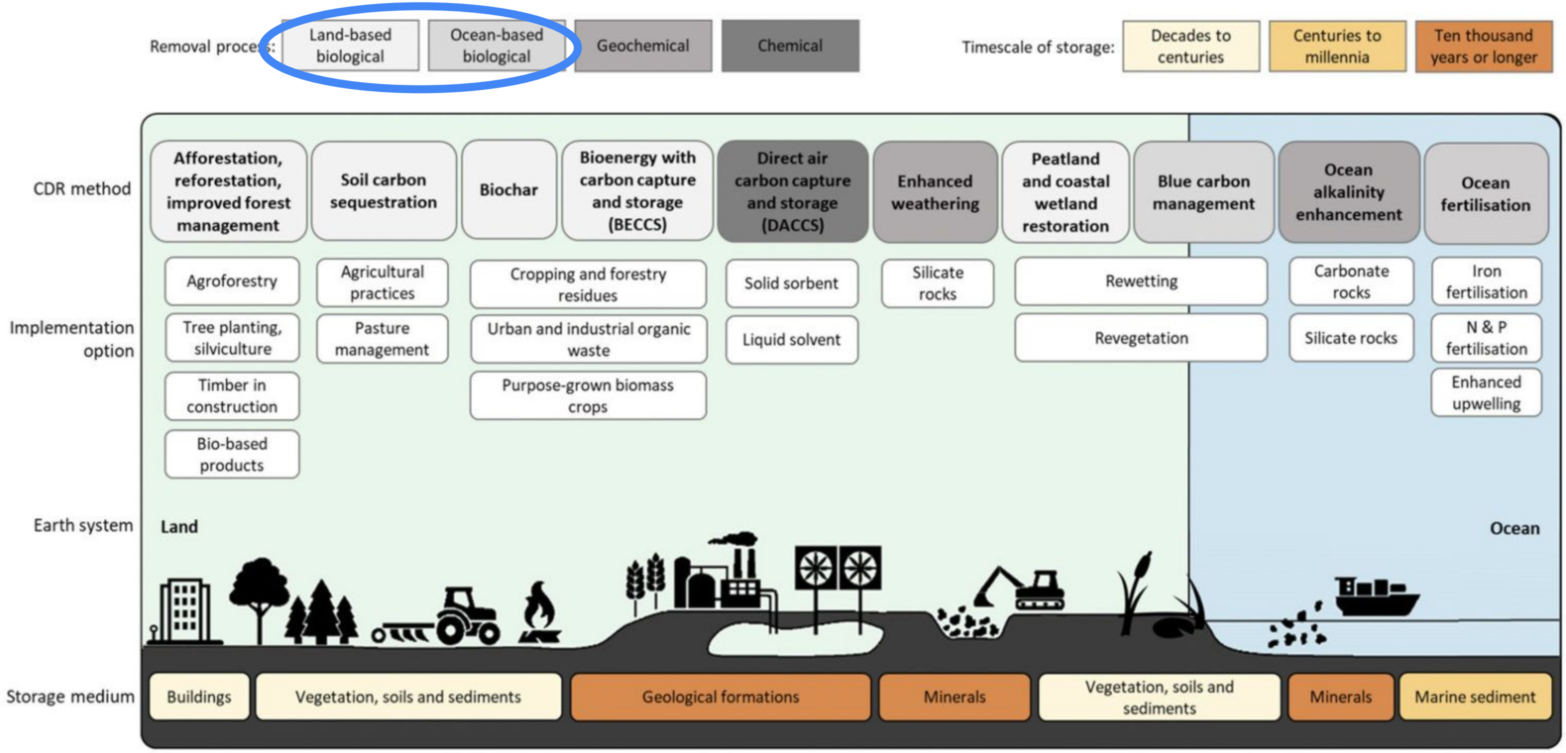
It states with *high confidence* that net-zero can only be achieved if CO2 removal is used to balance "difficult-to-abate" emissions from sectors that will find it harder to slash their climate impact, such as aviation, agriculture and some industrial processes.

In the longer term, upscaling CO2 removal could provide "net-negative CO2 emissions at the global level", allowing a "reversal of global warming", the report says with *medium confidence*.

https://www.carbonbrief.org/in-depth-qa-the-ipccs-sixth-assessment-on-how-to-tackle-climate-change/#co2removal

How can GHGs be removed from the atmosphere?

How can GHGs be removed from the atmosphere?



IPCC

When evaluating removal methods, need to consider process (including energy) and permanence.

ML for ecological restoration

collaborative earth

Global Forests

Aron Boettcher ⊙
University of Hawaii
Kyle Fisher ○ **Andre Otte** ○

We are building an accurate and global model for predicting potential rates of reforestation and resulting carbon sequestration. Such a model could have a transformational impact on global reforestation efforts by opening new streams of financing in the form of carbon credit futures.

Bison

Jason Baldes ⊙ NWF & ITBC
Gisel Booman ⊙ RegenNetwork
Emily Austin ○ **Colin Hill** ○ **Valérie Lechêne** ○ **Justin Lewis** ○ **Jens Owen** ○ **Jason Prasad** ○

Across the continent, a number of first nations are reintroducing bison to grasslands where they were once an ecologically vital species. Initial experiences and evolutionary considerations suggest that this may be ecologically beneficial in terms of biodiversity, carbon cycle, and resilience to climate change. However, these questions have not yet been studied at scale. In this lab, we will leverage remote sensing to scale up from ground measurements, establishing the large-scale patterns of bison impact.

Coastal Forested Wetlands

Elliott White Jr. ⊙
Stanford University
Nikhil Raj Deep ○ **Aaron Hirsh** ○ **Stephanie Kao** ○ **Layla Tadjpour** ○

The goal of our lab is to create a high-spatial resolution map of coastal forested wetlands at global scale. If we know precisely where these ecologically critical but fragile forests are located, we can manage freshwater flows to counteract saltwater introgression due to rising sea levels, and we can assist in their migration inland, preserving their critical function in protecting coastlines and sequestering carbon.

Beaver

Grace Lindsay ⊙
New York University
Avinash Mahech ○ **Cary Murray** ○ **Chris Norcross** ○ **Wendy Owens Rios** ○ **Quintin Tyree** ○

Beaver dams are known to result in greener, more drought-resilient waterways in semi-arid environments. We are using computer vision to spot dams in satellite imagery, generating a large dataset that we can use to train models that will tell us what the ecological effects of a dam will be at any point on a waterway. The goal is to create a tool to guide efficient restoration through the introduction of small dams.

Ganges

Anthony Acciavatti ⊙
Yale University
Sarthak Arora ○ **Markley Boyer** ○ **Nikhil Raj Deep** ○ **Jiby Matthew** ○ **James Smoot** ○ **Michael Warner** ○

In our pursuit of a successful and thriving relationship between humanity and natural systems, the Ganges river basin represents an extreme challenge. It is densely populated, remains agriculturally productive, and subject to an extremely powerful monsoon. We are mapping and analyzing a key feature of the Ganges basin—naalas—to understand how new forms of green infrastructure, such as parks, bioswales, and bioremediation, can rejuvenate this vital and sacred river.

Assisted Forest Regeneration

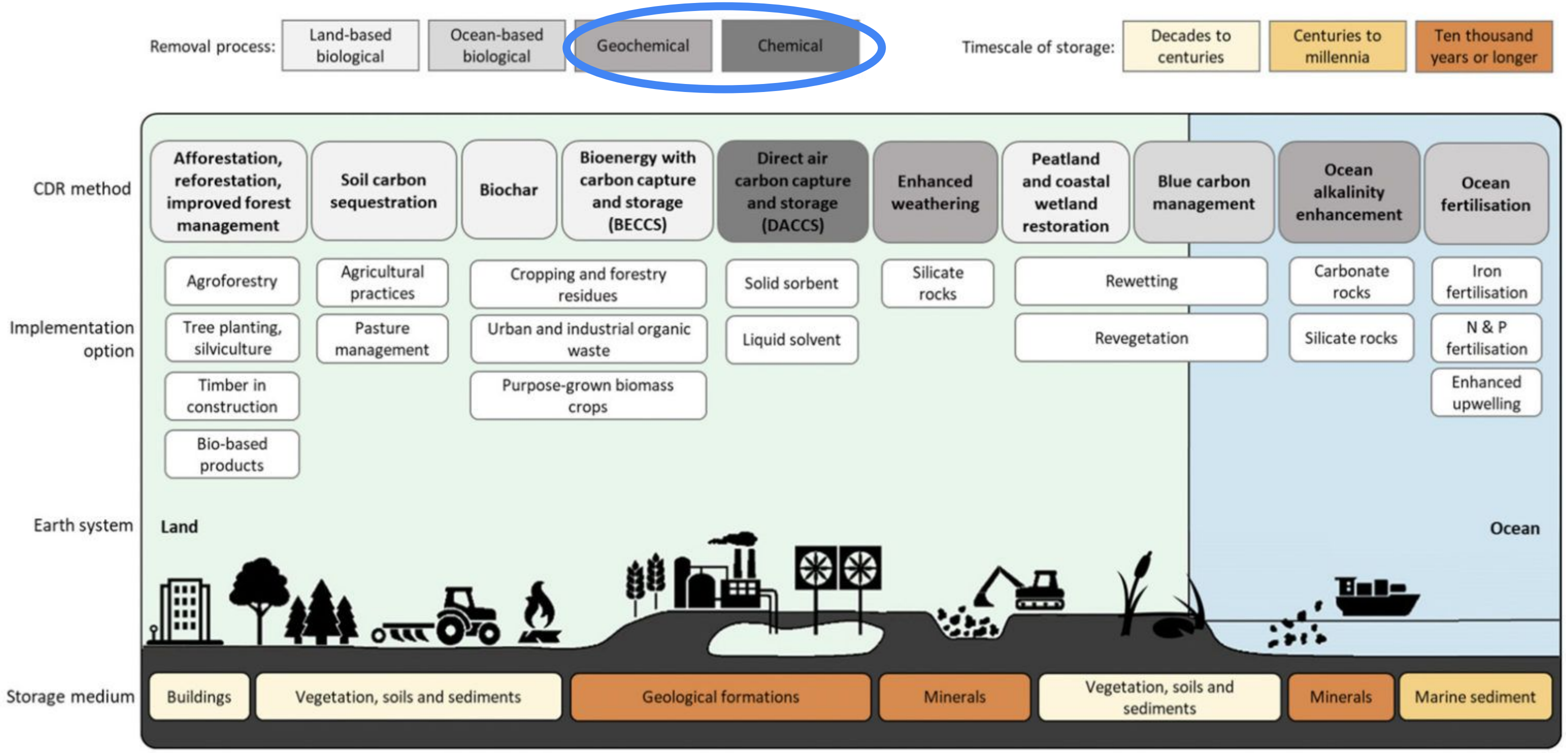
Leland Werden ⊙
ETH Zürich
Collaborative Earth AFR Data Team ○

This project is pioneering the idea of a massively open literature review. We aim to quantify potential carbon capture and plant biodiversity recovery of forest, savannah, and mangrove assisted restoration projects. To do so, we are gathering and synthesizing unpublished data from field partners, integrating as much information from non-English sources as possible. There is a wealth of wisdom that has not been published in academic journals, and we aspire to integrate these insights into our review.

According to the IPCC, tree-planting and ecosystem restoration are the only "widely deployed" forms of CO2 removal.

How can GHGs be removed from the atmosphere?

IPCC



When evaluating removal methods, need to consider process (including energy) and permanence.

(Geo)chemical carbon dioxide removal

Enhanced Weathering: A speedup of a natural process...

(Geo)chemical carbon dioxide removal

Direct air capture (DAC) aims to pull carbon dioxide out of plain air.

(A separate technology, “carbon capture”, tries to recapture carbon dioxide released during fuel burning. This cannot create net-negative emissions)

According to the IPCC: Despite limited use at present, technologies such as direct air capture, are projected to make a “moderate to large” contribution to future CO2 removal

(Geo)chemical carbon dioxide removal

Deals

Occidental buys carbon air capture tech firm for $1.1 billion

By **Sabrina Valle** and **Sourasis Bose**

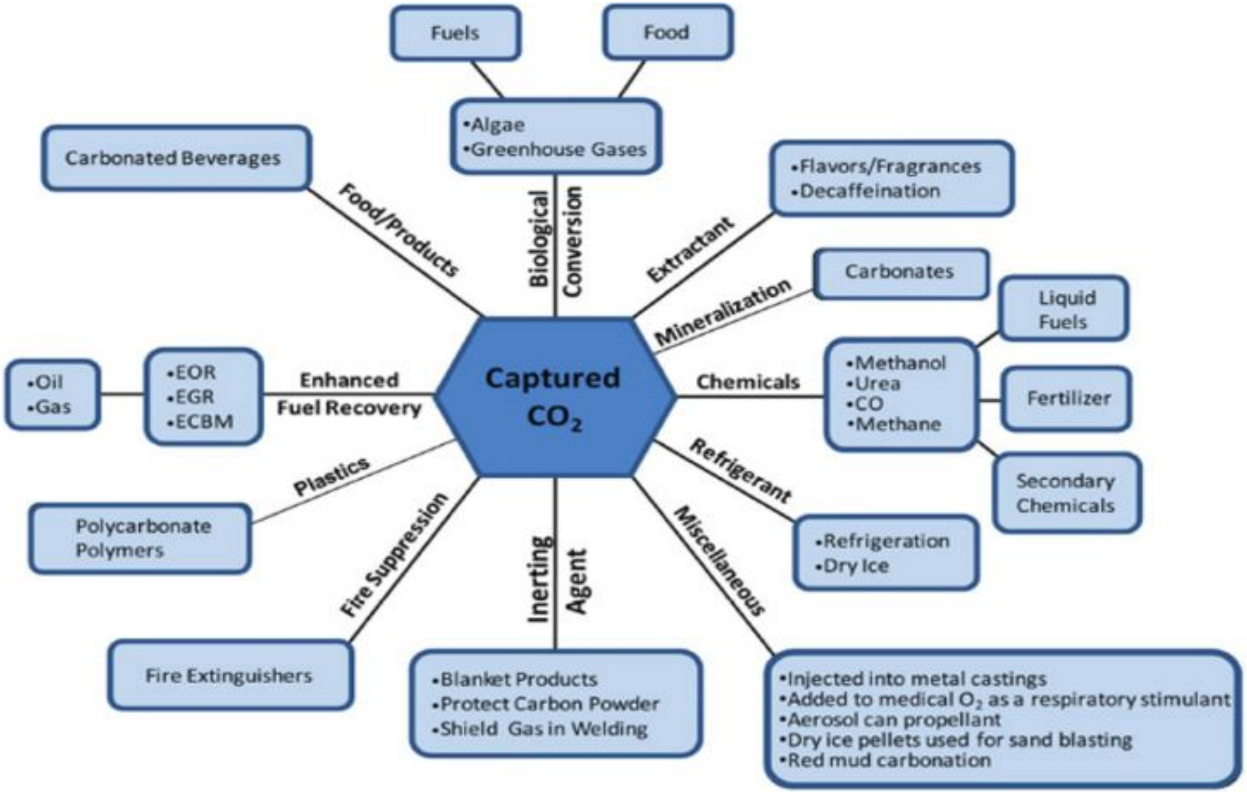
August 15, 2023 6:38 PM EDT · Updated 8 months ago

The logo for Occidental Petroleum is displayed on a screen on the floor at the New York Stock Exchange (NYSE) in New York, U.S., April 30, 2019. REUTERS/Brendan McDermid/File Photo *Purchase Licensing Rights*

Aug 15 (Reuters) - U.S. oil and gas producer Occidental Petroleum (OXY.N) on Tuesday agreed to pay $1.1 billion for technology supplier Carbon Engineering Ltd to help it develop a string of carbon-capture sites it hopes will profit from tackling climate change.

The U.S. oil producer aims to build about 100 plants using direct air capture (DAC) technology that strips

Uses of captured carbon



Aspen Institute

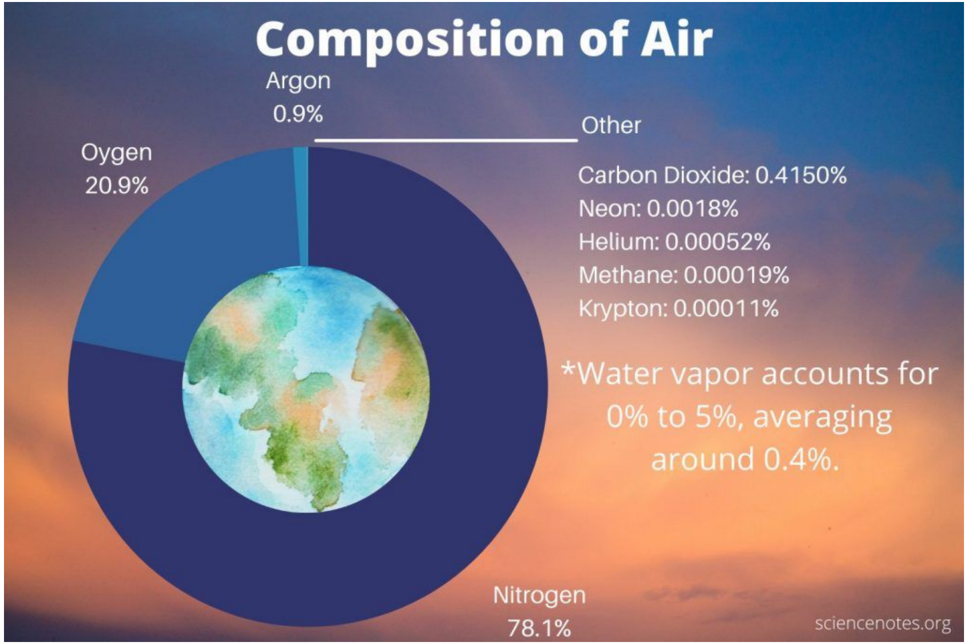
According to the IPCC...

The upscaling of many CO2 removal methods faces “various feasibility and sustainability constraints”

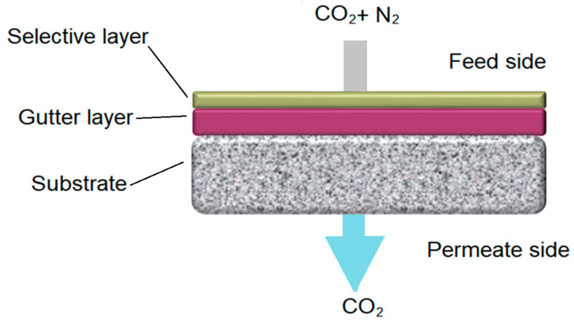
Enhanced weathering, meanwhile, “has been demonstrated in the laboratory and in small scale field trials, but has yet to be demonstrated at scale”.

Direct air capture and storage is currently limited by its large energy requirements and by cost; the technology is at a “medium readiness level”.

Direct Air Capture is hard: Can ML make it better?



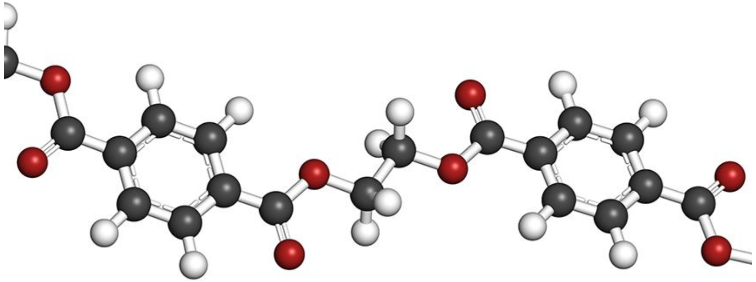
We need materials that can cheaply filter out CO2.

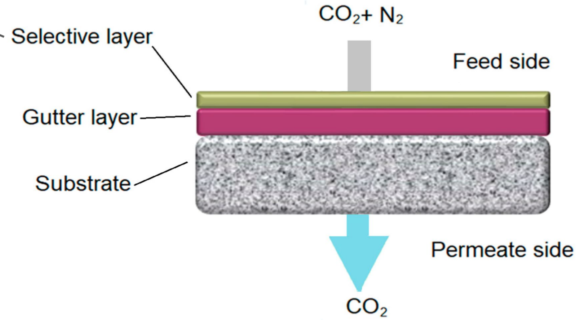


Direct Air Capture is hard: Can ML make it better?

We need materials that can cheaply filter out CO2.

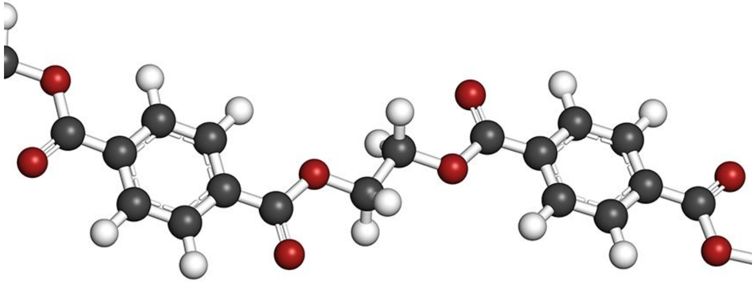
“Polymer membranes” need to be selective (only let CO2 pass) and permeable (let a lot of CO2 pass)





Direct Air Capture is hard: Can ML make it better?

"Polymer membranes" need to be selective (only let CO2 pass) and permeable (let a lot of CO2 pass)



Can ML design better polymers?

Topics we've already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

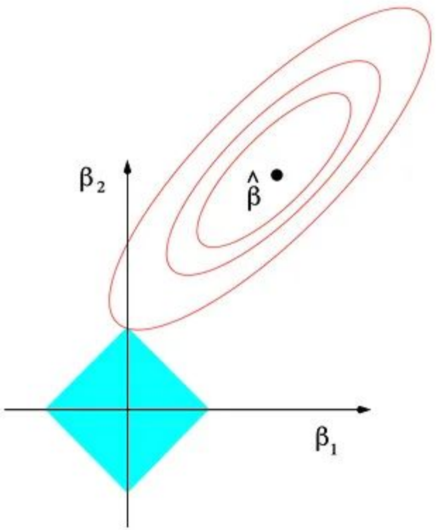
Transformers

Unsupervised learning

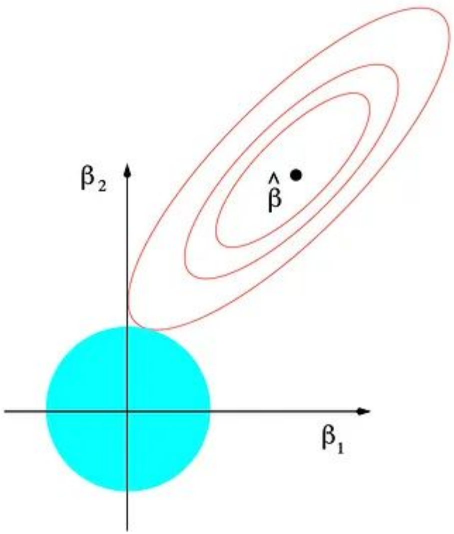
Cross validation

Hyper parameter tuning

ElasticNet: Regularized Linear Regression



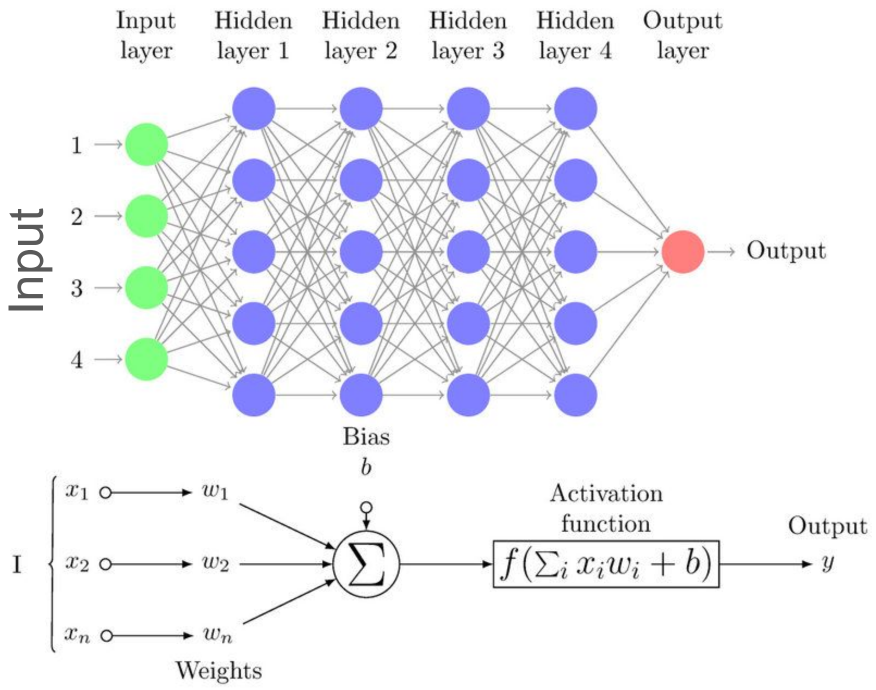
Lasso



Ridge

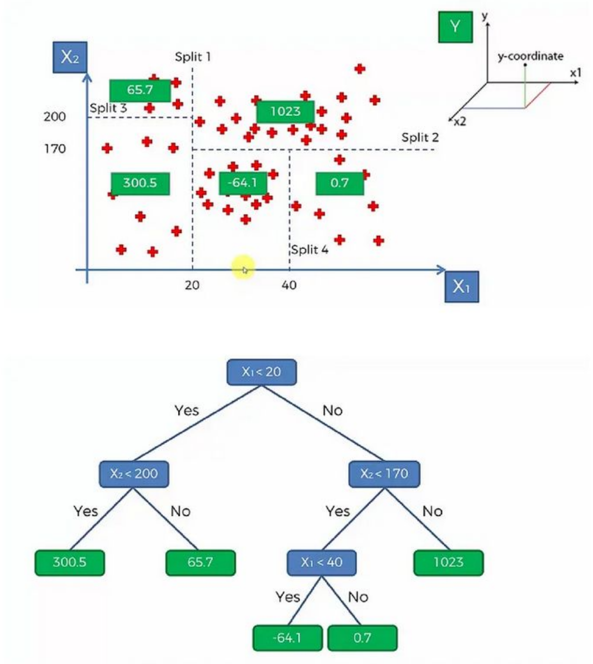
Lasso + Ridge = “ElasticNet”

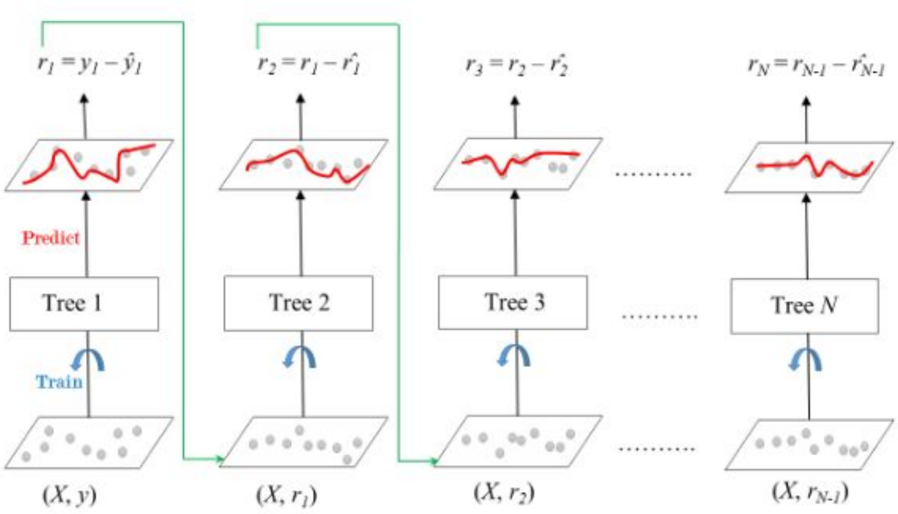
Basic Feedforward Neural Network



XGBoost

Gradient Boosted Trees (i.e. an ensemble method)





Each tree predicts and subtracts the errors of the one before

Topics we’ve already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

Transformers

Unsupervised learning

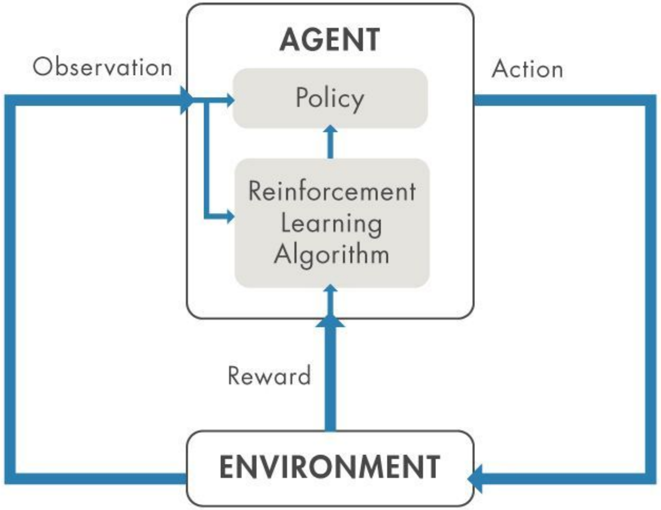
Cross validation

Hyper parameter tuning

Reinforcement learning

When using reinforcement learning to solve a control problem, the aim is to develop a **policy** that controls an **agent** to maximize **reward**.

The policy is a function that takes in a **state observation** and produces an **action**.



https://www.mathworks.com/company/newsletters/articles/reinforcement-learning-a-brief-guide.html

Topics we've already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

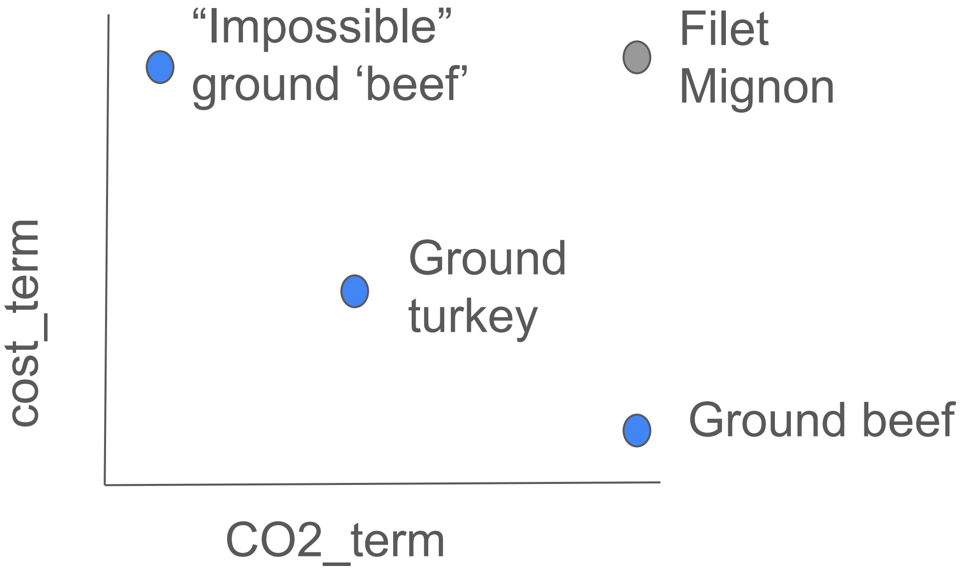
Transformers

Unsupervised learning

Cross validation

Hyper parameter tuning

The pareto front is the result of trade-offs in the various objective terms



Trying to minimize:

Topics we've already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

Transformers

Unsupervised learning

Cross validation

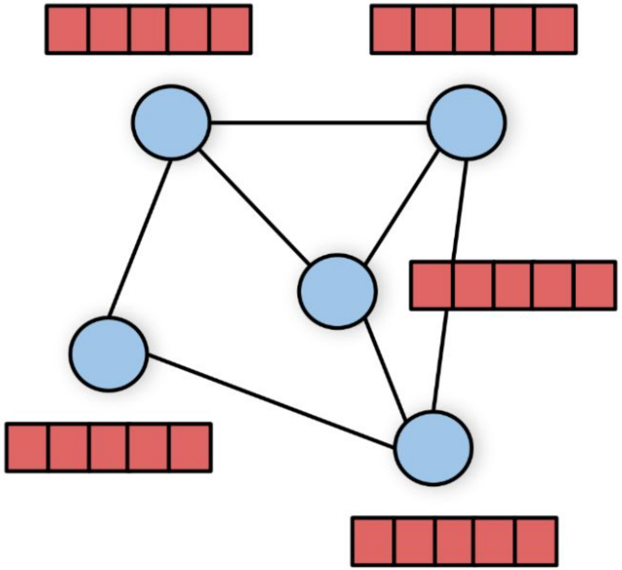
Hyper parameter tuning

Graph Neural Networks (GNNs)

GNNs are artificial neural networks that can take *graphs* as input.

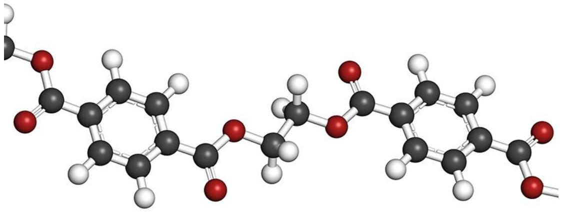
The graphs are represented by their adjacency matrix and any values needed to provide information about each node or edge.

The neural network learns how to combine information across nodes using a *message passing* algorithm.



Graph Neural Networks for molecules

Molecules are naturally represented as graphs!



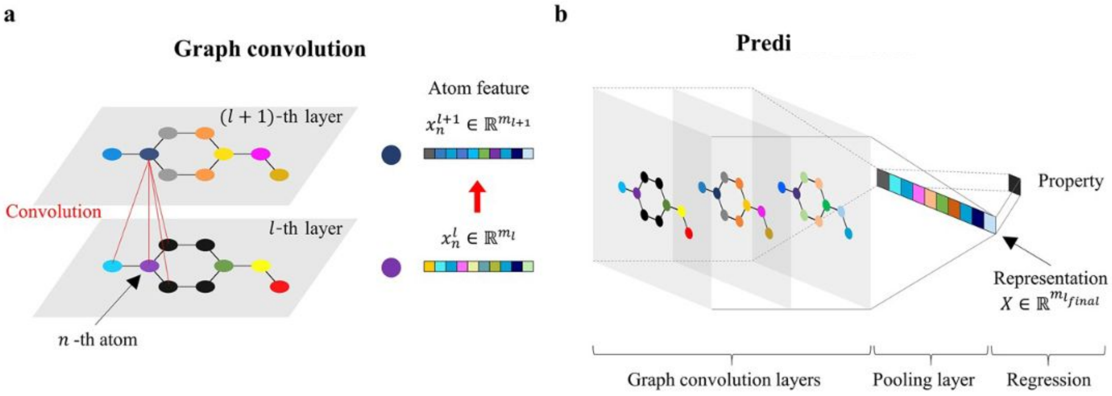


Figure 2. Property prediction model based on the GCN. (a) Feature vector of n-th atom x_n^l is updated iteratively through the l-th convolutional layer by the graph convolution that aggregates the features of neighboring atoms. (b) Graph-level molecular representation vector X is obtained by the pooling layer that sums up the feature vectors of all the atoms at the final convolution layer. The vector X is input to the regression algorithms (linear regression (LR) only shown here) for the property prediction.

Park et al, 2022

Topics we’ve already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

$\boxed{\text{Transformers}}$

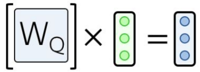
Unsupervised learning

Cross validation

Hyper parameter tuning

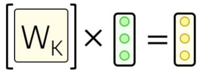
Transformer architecture

Each vector receives three representations ("roles")



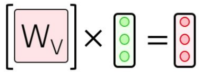
Query: vector **from** which the attention is looking

"Hey there, do you have this information?"



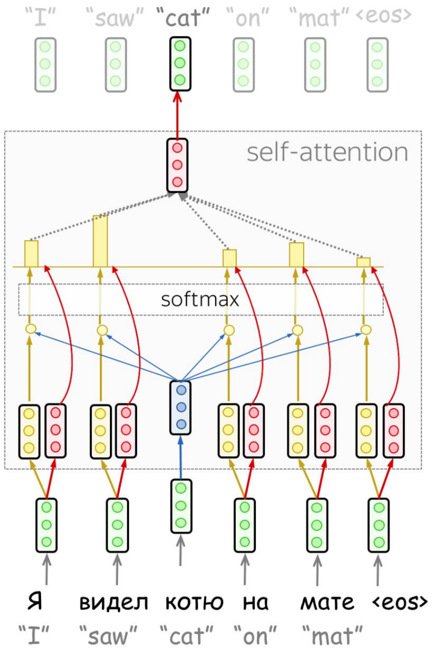
Key: vector **at** which the query looks to compute weights

"Hi, I have this information – give me a large weight!"



Value: their weighted sum is attention output

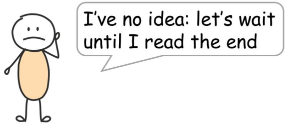
"Here's the information I have!"



Key insight: combine information across words. This is known as "self-attention".

I arrived at the bank after crossing the ... ...street? ...river?

What does **bank** mean in this sentence?



RNNs

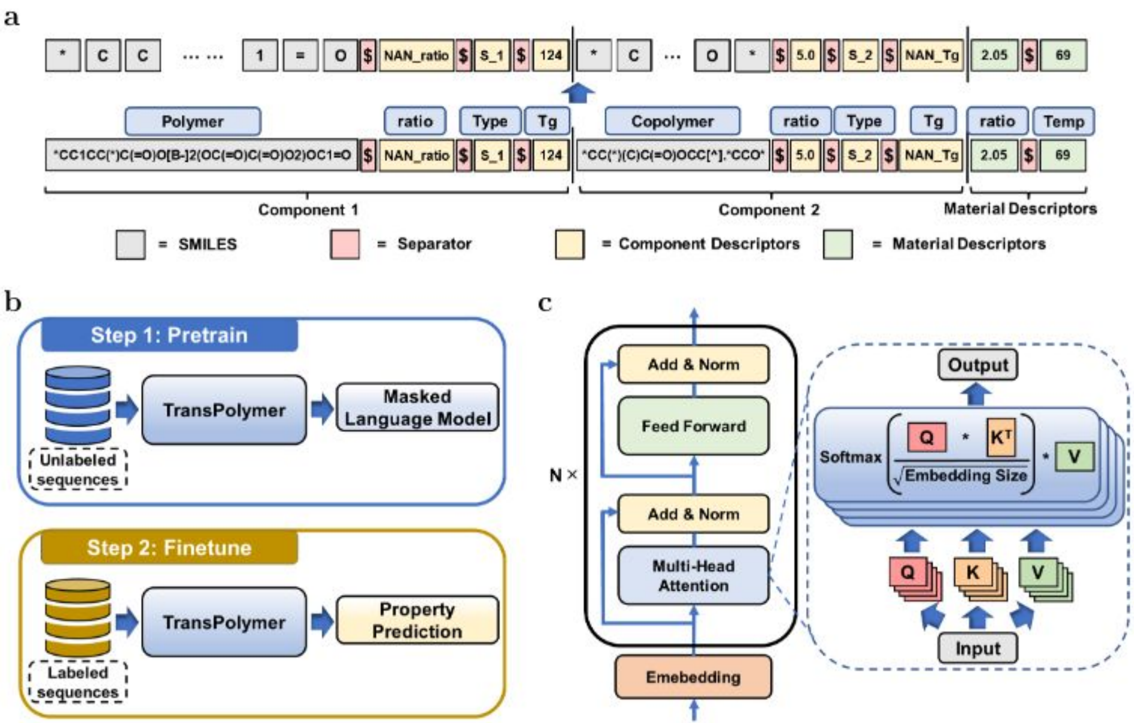
O(N) steps to process a sentence with length N

Transformer

Constant number of steps to process any sentence

Transformers for molecules

Need to represent the molecular structure as a sequence of characters



Xu et al, 2022

Topics we've already talked about that are mentioned in this paper:

Elasticnet

Ensemble methods

XGBoost

Feed forward neural networks

Emulators

Reinforcement Learning

Pareto fronts

Graph neural networks

Transformers

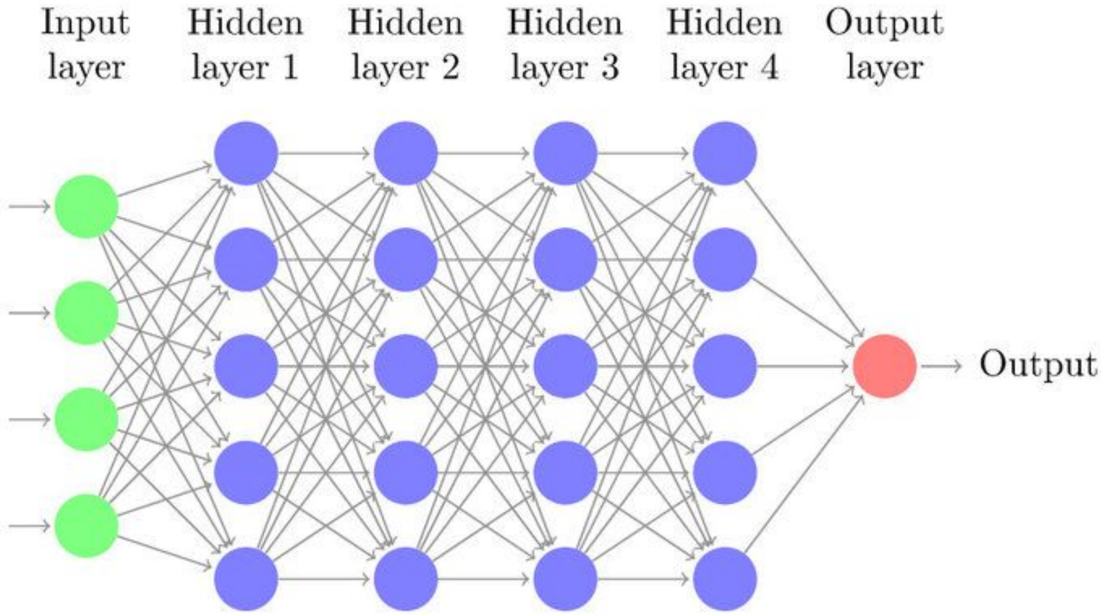
Unsupervised learning

Cross validation

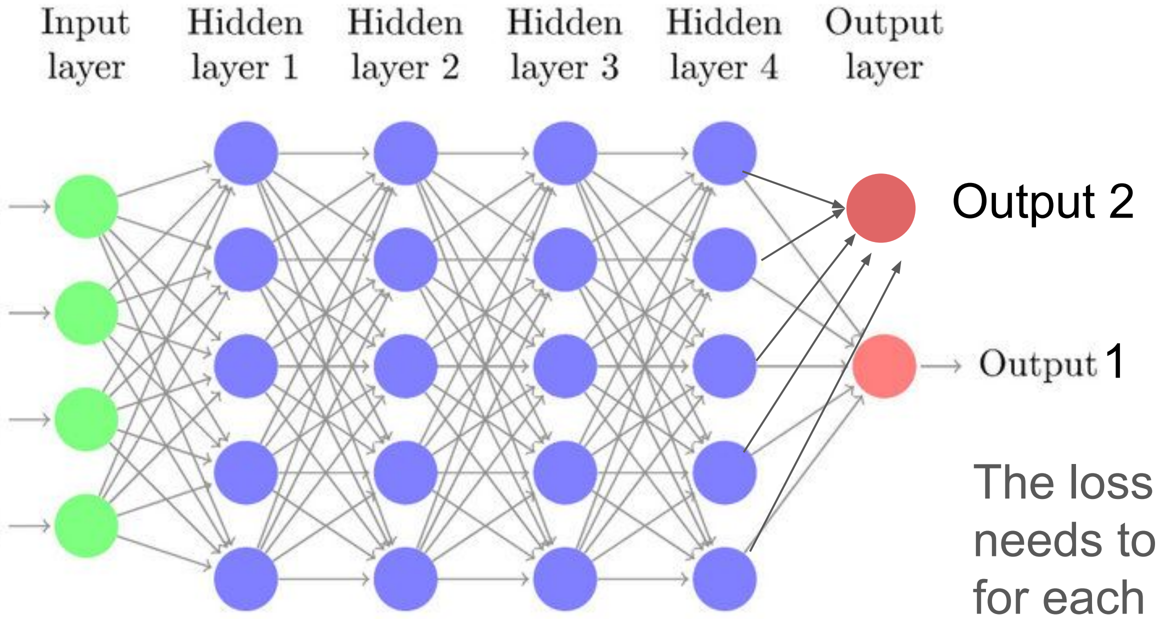
Hyper parameter tuning

New topic: multi-task learning

Multi-task learning: making a single model do multiple things



Multi-task learning: making a single model do multiple things



The loss function needs to include terms for each output