Some Recent Important Topics

Machine Learning

Fabio Vandin

January 13th, 2023

Understanding Deep Learning

Deep Learning seems to be of a *different nature* of other ML models

For example: the theory (VC-dimension and more advanced ones, e.g. Rademacher complexity) tells us that:

- The more parameters there are, the more data you need (more or less)
- They require *tons of samples*

In practice: they require much less data than the theory predicts! A new theory may be needed?

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107-115.

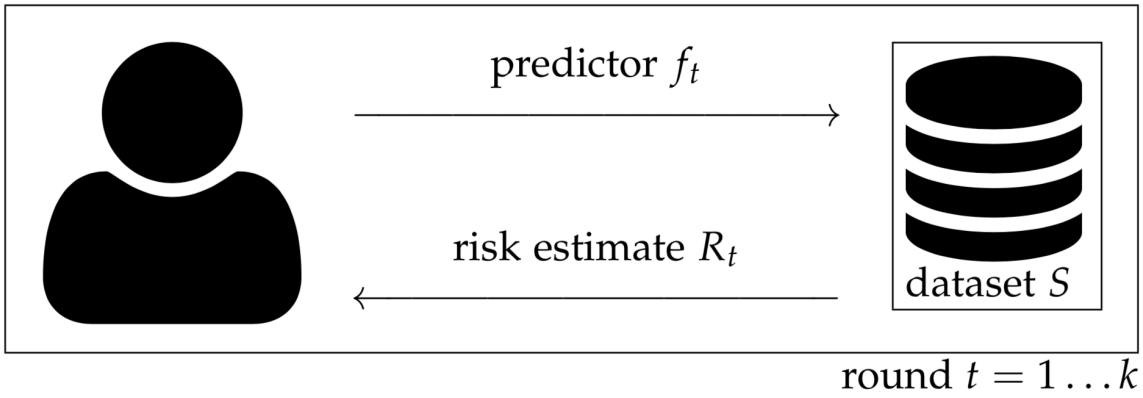
Understanding Deep Learning

Very often DL is used in online challenges

Framework:

- You are given a training set to learn a model
- Then you can post your model to the challenge and see its results on a (fixed) test set

You can repeat the steps above as many times as you want…



Does this lead to overfitting?

Theory: “yes!”

Understanding Deep Learning

Practice?

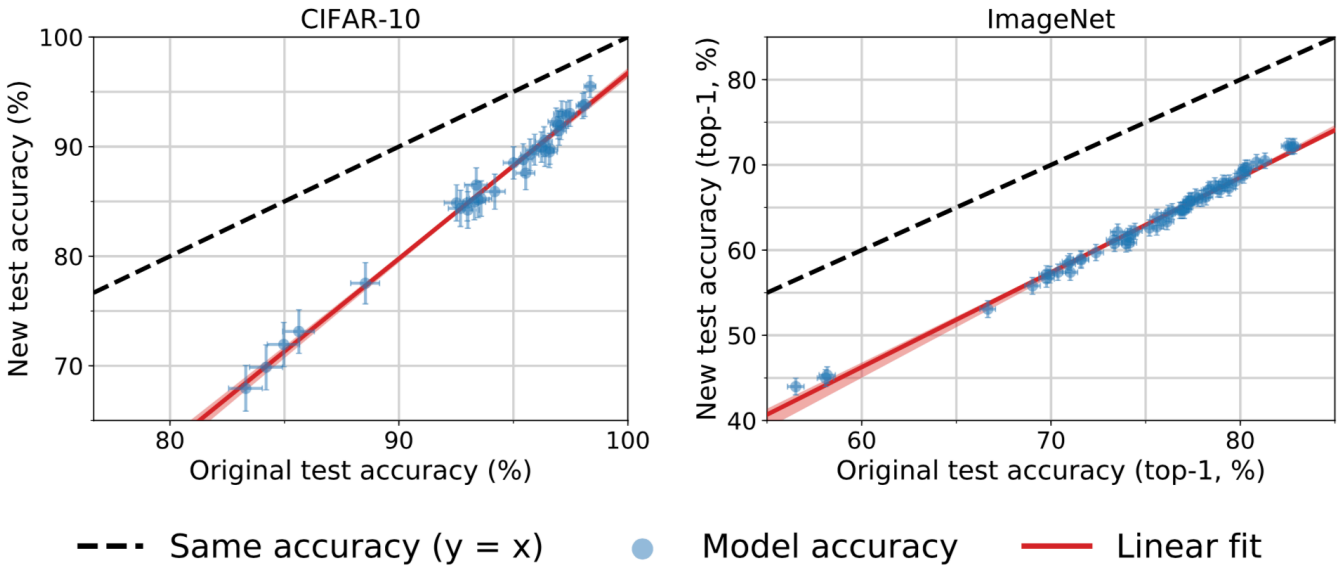


Figure 31: Model accuracy on the original test sets vs. new test sets for CIFAR-10 and ImageNet. Each data point corresponds to one model in a test bed of representative models.

Why does this happen? Still not well understood

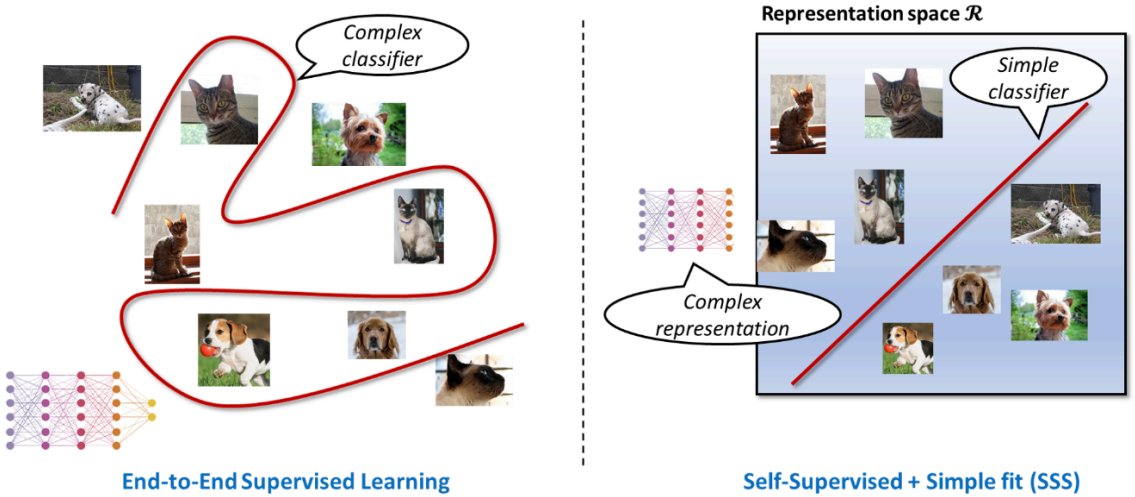
Understanding Machine Learning

Deep Learning learns useful representations of the data

Comparison of two supervised models:
1. traditional end-to-end supervised DL
2. Self-supervised learning followed by simple supervised linear model

Self-supervised learning: ignore labels of the data, try to learn representations of the input features (e.g., goal: reconstruct whole input given only part of it)

Result: the models have very similar performance!



ML Applications

We have mentioned several examples of the application of ML tasks

Classification:

- email spam prediction
- ads click prediction
- movie rating prediction
- …

Other ML Applications

There are other, more critical, applications of ML classification:

- granting loans to people
- recidivism prediction for criminals (whether to grant them bail)
- suitability of candidates for jobs

These are cases where the ML system can have severe consequences on people

Traditional Approach to Learn a Model

Task: classification

Our traditional goal: find a model that maximizes accuracy (i.e., minimize errors)

Is this enough in important applications?

Traditional Approach to Learn a Model (continue)

Toy example: assume we learn a model with the following performance for the problem of hiring candidates

PREDICTED LABEL	TRUE LABEL 0	1
(Do not hire) 0	44	6
(hire) 1	4	46

Overall accuracy: 90%!

Is this ok?

Traditional Approach to Learn a Model (continue)

Assume that:

- 69% of the candidates are men, 31% are women
- the performance for men is different from the performance for women

Traditional Approach to Learn a Model (continue)

MEN

PREDICTED LABEL	TRUE LABEL 0	TRUE LABEL 1
0	43	0
1	6	51

WOMEN

PREDICTED LABEL	TRUE LABEL 0	TRUE LABEL 1
0	45	20
1	0	35

The model should be *fair!*

A Real World Example

- COMPAS: Correctional Offender Management Profiling for Alternative Sanctions
- Measures the risk of a person to commit another crime(recidivism)
- In some states in the USA, judges use this system while deciding court cases, e.g., whether to release an offender on bail, or to keep him/her in prison.

PRO PUBLICA

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of $80.

- Black defendants were often predicted to be at a higher risk of recidivism than they actually were. Our analysis found that black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent).
- White defendants were often predicted to be less risky than they were. Our analysis found that white defendants who re-offended within the next two years were mistakenly labeled low risk almost twice as often as black re-offenders (48 percent vs. 28 percent).
- The analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 45 percent more likely to be assigned higher risk scores than white defendants.
- Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism. And white violent recidivists were 63 percent more likely to have been misclassified as a low risk of violent recidivism, compared with black violent recidivists.
- The violent recidivism analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 77 percent more likely to be assigned higher risk scores than white defendants.

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Unfair Machine Learning

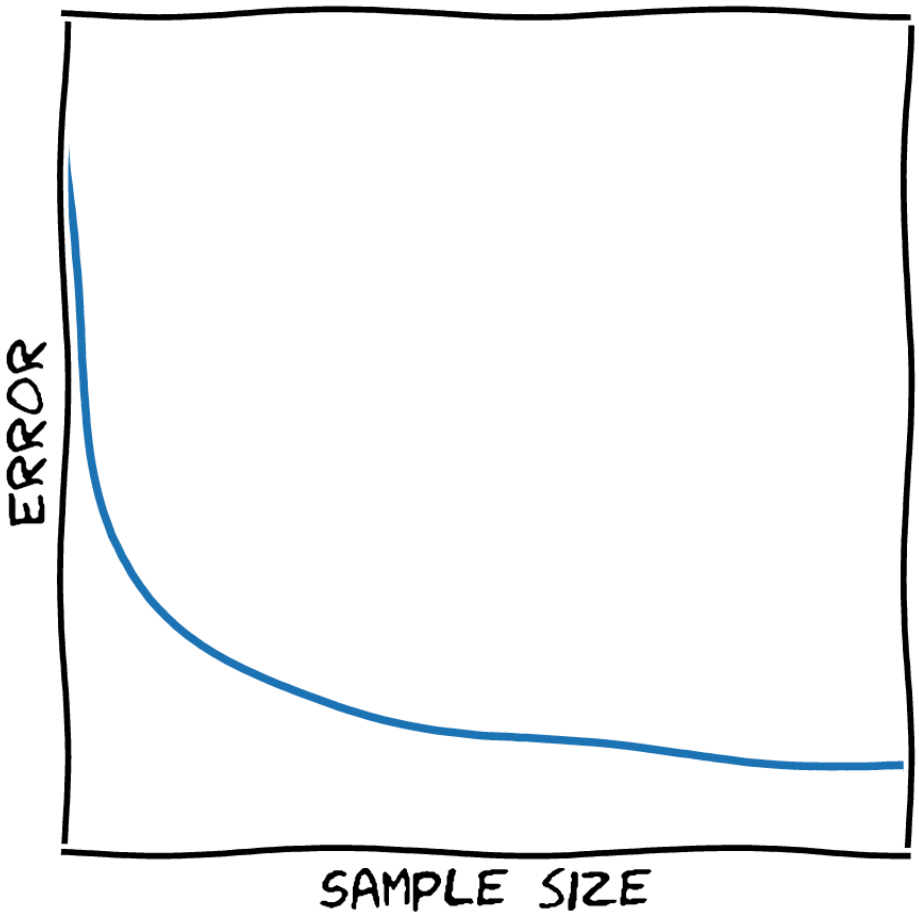
How can a machine learning model ending up being ***unfair*** without any explicit wrongdoing?

- The model could learn biases in the data!
- Even if you do not include a feature in the input, there may be correlated features (e.g., height)
- *Sample size disparity*

Reasons for Unfair Models?

Sample size disparity:

- usually, more data means smaller error
- by definition, there is less data for minority groups



Fair Machine Learning Models

For many real-life applications, ML models not only need good performace (e.g., high accuracy) but also need to be **fair**

Need to measure/define fairness

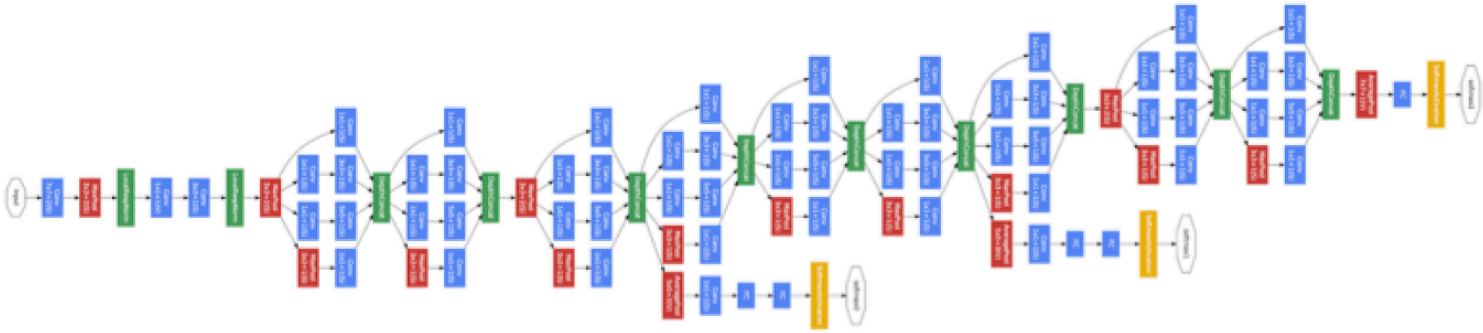
- several definitions have been proposed

A lot of research on designing fair machine learning models!

How Can We Understand if a Model is Unfair?

We need to understand *why* a model is making the predictions we see

Easy?



Convolution
Pooling
Softmax
Other

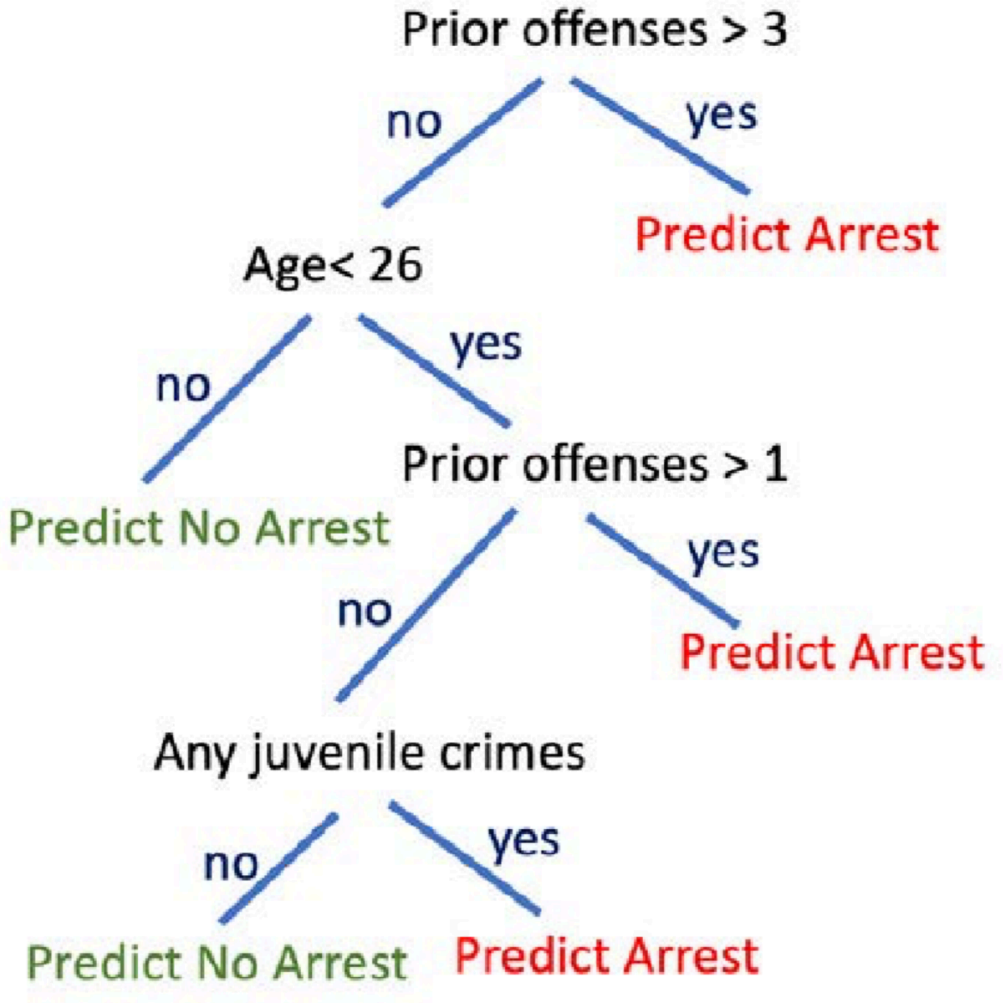
Interpretability vs Explainability

Possibilities:

- learn an *interpretable* model
- develop methods to explain the predictions of a (black-box) model

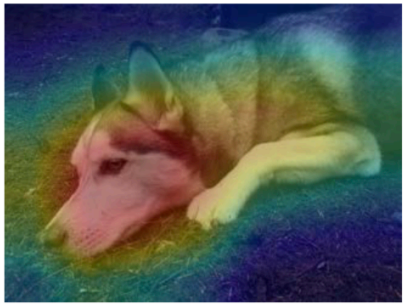
Interpretable Models

Simple models that we can interpret



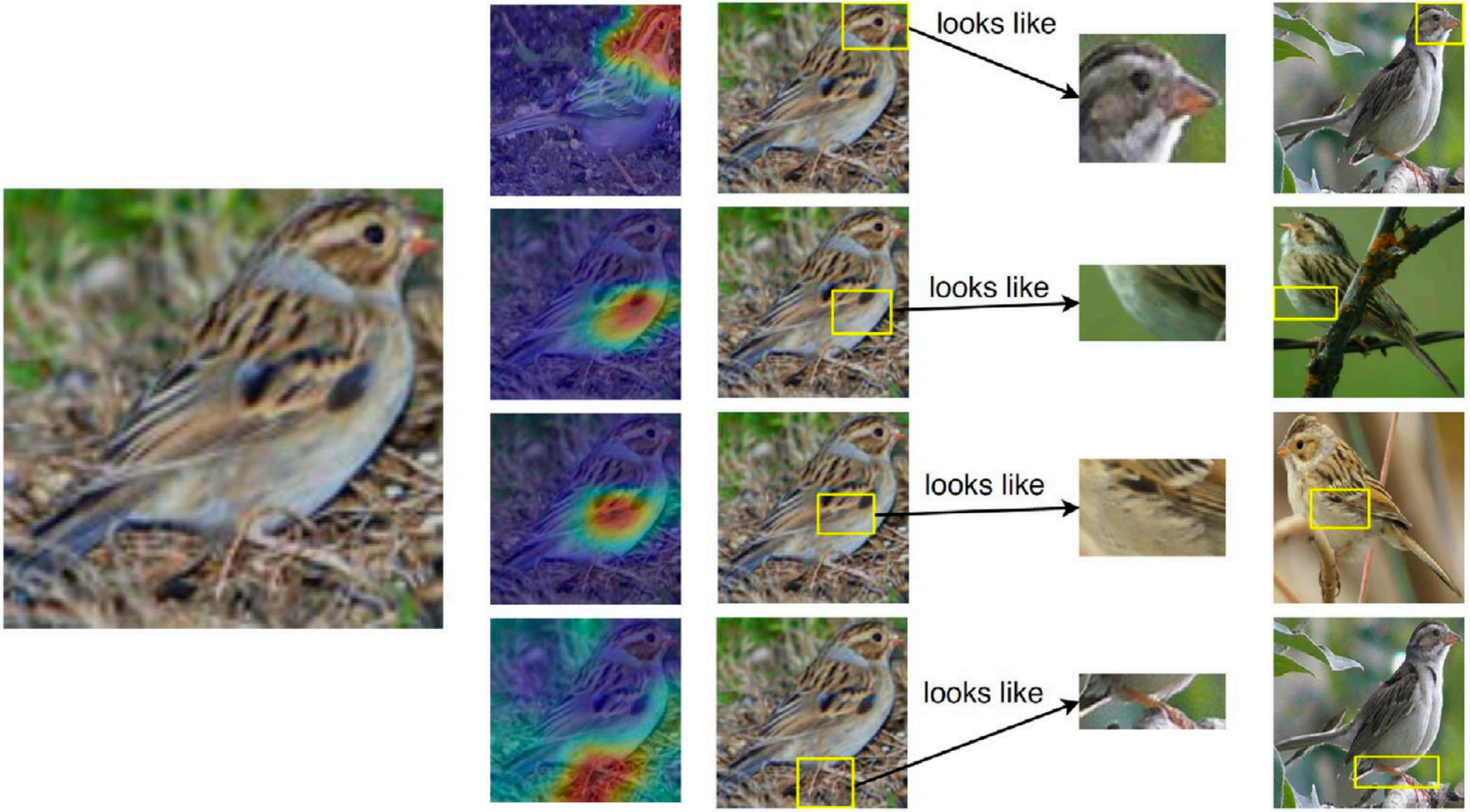
Explainability

Several attempts to develop methods to explain the predictions of black-box models (e.g., neural networks)

	Test Image	Evidence for Animal Being a Siberian Husky	Evidence for Animal Being a Transverse Flute
Explanations Using Attention Maps		 "Explanation"	

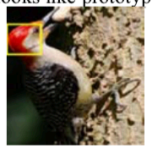
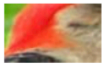
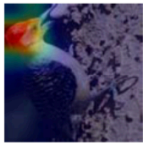
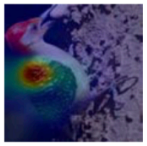
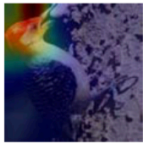
Do you trust the model?

A Different Approach



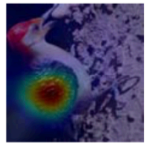
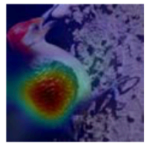
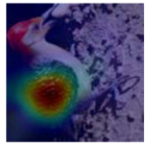
Why is this bird classfied as a red-bellied woodpecker?

Evidence for this bird being a red-bellied woodpecker:

Original image (box showing part that looks like prototype)	Prototype	Training image where prototype comes from	Activation map	Similarity score	Class connection	Points contributed
				6.499	× 1.180	= 7.669
				4.392	× 1.127	= 4.950
				3.890	× 1.108	= 4.310
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Total points to red-bellied woodpecker: 32.736

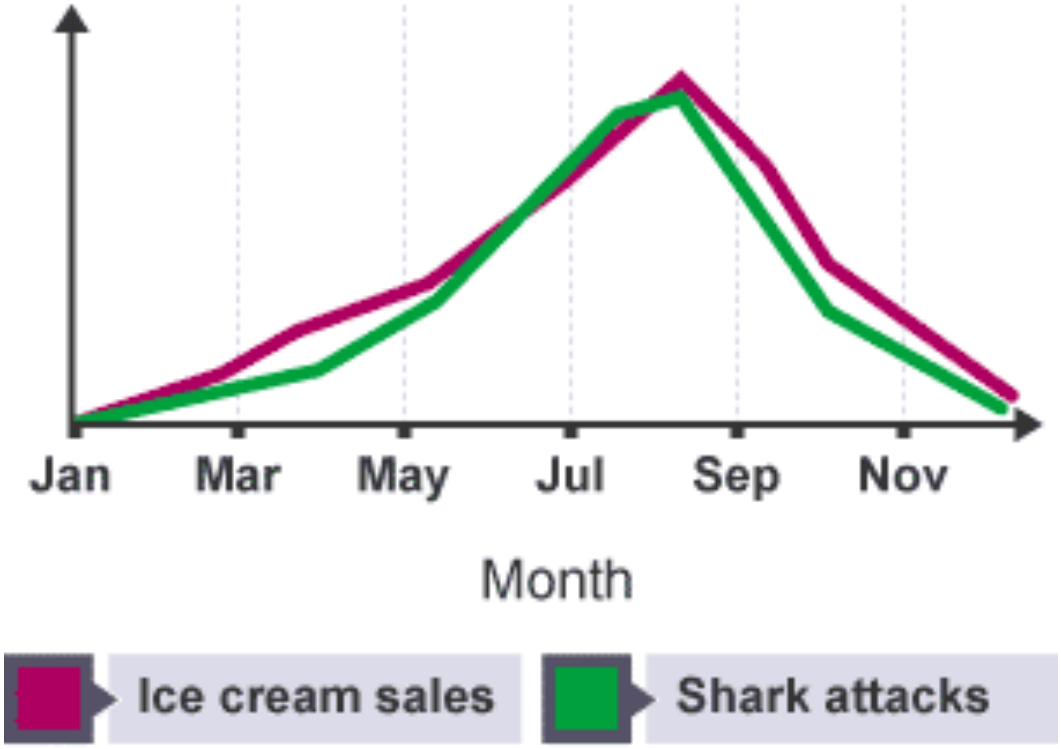
Evidence for this bird being a red-cockaded woodpecker:

Original image (box showing part that looks like prototype)	Prototype	Training image where prototype comes from	Activation map	Similarity score	Class connection	Points contributed
				2.452	× 1.046	= 2.565
				2.125	× 1.091	= 2.318
				1.945	× 1.069	= 2.079
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Total points to red-cockaded woodpecker: 16.886

A Last Topics: Causality

Machine learning models are based on correlations: *do not provide guarantees on finding causal relations*



Causal inference: AI and ML methods to detect causal relations

Some References

- *Understanding Deep Learning*:
  - book from Moritz Hardt and Benjamin Recht https://mlstory.org/
  - work and blog from Boaz Barak https://www.boazbarak.org/
- *Fairness*: see the work and material from Moritz Hardt https://mrtz.org/
- *Interpretability and Explainability*: see the work and material from Cynthia Rudin https://users.cs.duke.edu/~cynthia/
- *Causality*: see material from Judea Pearl http://bayes.cs.ucla.edu/jp_home.html