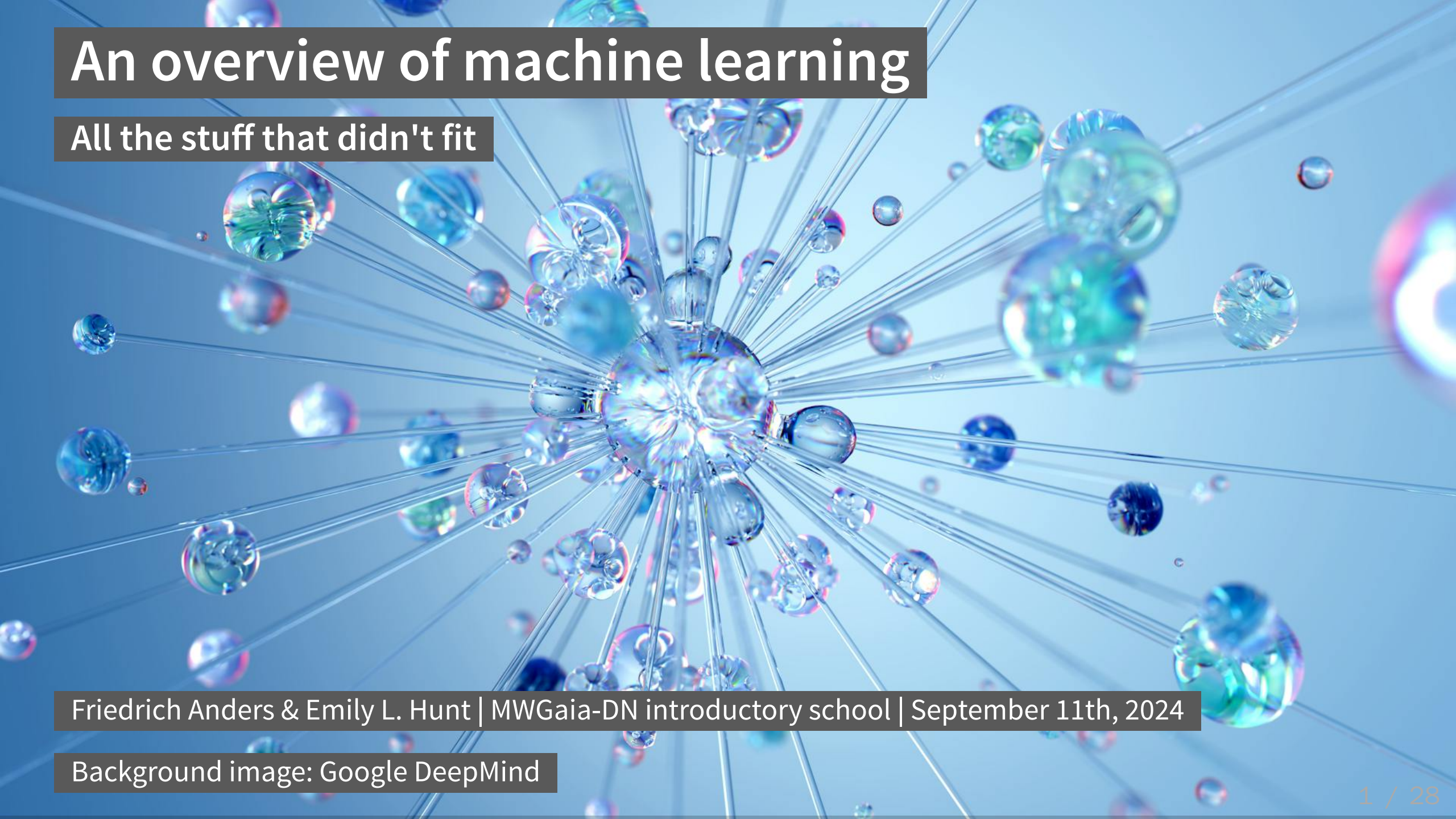
An overview of machine learning

All the stuff that didn't fit



Friedrich Anders & Emily L. Hunt | MWGaia-DN introductory school | September 11th, 2024

Background image: Google DeepMind

About this final session

We want to do three things:

1. Mention advanced things that we didn't have time for
2. Have group discussions about some concepts
3. Have some time for questions

Remaining problems with ML

Machine learning is not a perfect family of methods!

Can you think of some things that are missing?

Problem 1: Uncertainties

Most measurements in astronomy have uncertainties...

... but most methods in machine learning **cannot consider uncertainties**

This is a major open problem in ML

A refresher on uncertainties

A model fit to data has two types of uncertainty:

1. **Random uncertainties** due to the data/random chance
2. **Systematic uncertainties** due to model/data systematic mistakes

A refresher on uncertainties

A model fit to data has two types of uncertainty:

1. **Random uncertainties** due to the data/random chance
2. **Systematic uncertainties** due to model/data systematic mistakes

The fancy names for these are **aleatoric** and **epistemic** uncertainties.

Estimating aleatoric (random) uncertainties

In principle, we can do techniques like:

- Ask our model to estimate means **and** uncertainties (e.g. mixture density networks)
- Sample our model many times with different samples of our data (MCMC analogue)

These work! But this is **the easy one...**

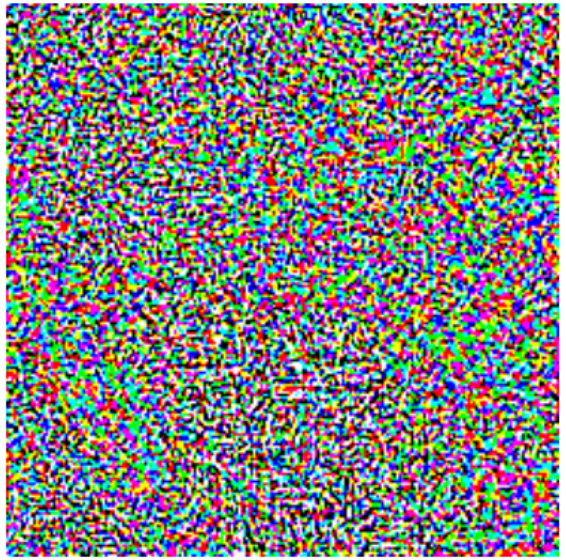
Estimating epistemic uncertainties

Unfortunately, machine learning models **absolutely suck** at estimating the **uncertainty of the model itself**

Machine learning models are often **confidently wrong**

Classic FGSM paper example

$+ .007 \times$



$=$

$\boldsymbol{x}$

"panda"
57.7% confidence

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$\boldsymbol{x} +$
$\epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

Why does this happen?

ML models make unreliable predictions on data **unlike** what they were trained on

"If you don't train me on it, then I don't know what I'm doing"

This is often referred to as the poor **out of distribution (OOD) performance** of ML models

What you can try

What you can try

- **Ensemble learning** (training an ensemble of many models) seems to work relatively well

What you can try

- **Ensemble learning** (training an ensemble of many models) seems to work relatively well
- Some models like **conditional invertible neural networks** seem to work relatively well

What you can try

- **Ensemble learning** (training an ensemble of many models) seems to work relatively well
- Some models like **conditional invertible neural networks** seem to work relatively well
- **Bayesian neural networks** are a **scam** (they require too many approximations to typically be useful)

Problem 2: one-shot learning

Machine learning can be much better than humans at some things (e.g. speed), but **hilariously bad** at other things

Maybe the best example is so-called "one-shot learning."

What problems does this give for scientists?

- We generally need **large, labelled training datasets**
- Typical ML approaches can be bad at **outlier detection**
- General-purpose "foundation" models are a long way away

Problem 3: interpretability

It's hard to know **how** an ML model derived its answer.

Sometimes, it can help to use tools (like **SHAP**), but ML models are generally hard to interpret

Some related methods (like **symbolic regression**) can provide good alternatives

Problem 4: logic

Another thing ML cannot do is **basic reasoning**

Models like ChatGPT are more like a **giant memory bank** than an 'artifical brain'

Memorising things only gets you so far...

Will these problems get fixed?

That's a **multi-trillion dollar** question that nobody can answer

... but we can speculate!

Transformers

New types of architecture are being invented regularly

... the most promising of which is the **transformer**

They are especially good at processing **sequential data** (like language) and power the 'AI revolution' (a.k.a. the AI hype bubble)

How do they work?

Transformers use 'multi-head attention blocks': basically a combination of big matrix multiplications and neural networks

I can highly recommend the **3blue1brown** YouTube videos about them

Are they good for astronomy?

We don't have many problems that need language processing (so: no)

However, transformers can do oddly well at other tasks (like classifying time-series data, even vision problems with vision transformers)

So, I expect they will gradually increase in popularity - unlikely to obsolete most existing methods, though

What will be the final architecture?

If I could tell you, I would have a permanent position in academia already...

I predict the next 10-20 years will see current models like neural networks be surpassed

How? Exactly when? Who knows...

Chance for questions

Before we go onto the next bit - does anyone have questions/thoughts to add?

Discussion time!

Finally, to finish, let's talk about some points and see what we think.

Some ground rules

1. **There are no wrong answers to these topics**
2. Respect each other's opinions, even if you disagree

Machine learning methods are better than traditional ones.

Do you agree/disagree? Why?

Machine learning methods are 'black boxes' - it's hard to understand how they actually work.

Do you agree/disagree? Why?

Supervised ML methods are better than unsupervised ones.

Do you agree/disagree? Why?

What are the important tricks for getting machine learning to work?

Do you have ideas for how you could use ML methods in your own work?

That's all, folks!

Thank you so much for your time and for listening over the past two days!

We'll be available for the rest of the week for help/questions =)