What's new about Machine Learning?

Daniel Whiteson, UC Irvine
Jul 2022 / Snowmass in Seattle
It's everywhere!

"Machine Learning"

1984 - 2022

"Neural Network"

1977 - 2022
What's new about ML?

What society thinks I do

What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

What I actually do

from theano import

Graphic from D. Guest
Early days of HEP

ME CALL IT A "PARTICLE COLLIDER"

EARLY PHYSICISTS
ML in HEP is not new

1997:

Fig. 2. (a) The output $\eta$ of the neural network b tag for radiative returns to the $Z$ for 161 GeV q̅q Monte Carlo (histogram) compared to the data at 161 GeV (points). The shaded region shows the contribution from generated b-jets. (b) The performance of the neural network b tag (solid line) for Monte Carlo events, presented in terms of the efficiency for identifying b-jets versus the efficiency for rejecting light quark jets. The performance of the single most powerful b tagging input variable to the neural network is shown for comparison (dashed curve).
Is it something new?

Is modern ML something new, or just more of the same?
Is it something new?

Is modern ML something new, or just more of the same?

Daniel Whiteson
@DanielWhiteson

Is recent (> ~2013 deep learning moment) ML in particle physics "more of the same" or "qualitatively something new".
Is it something new?

Is modern ML something new, or just more of the same?

Daniel Whiteson
@DanielWhiteson

Is recent (> ~2013 deep learning moment) ML in particle physics "more of the same" or "qualitatively something new".

- More of the same: 39.7%
- More, not the same: 39.7%
- It's complicated (comment): 11%
- ML is nonsense: 9.6%

73 votes · Final results
Is it something new?

Is modern ML something new, or just more of the same?

My view: it’s both!
Outline

1. Much much more of the same

2. Something qualitatively new
Traditional role of ML

Why do we need machine learning?
Traditional role of ML

Why do we need machine learning?
Making a new particle

There are a lot of different reactions that can give you the Higgs. For example...

You can fuse two gluons...

Which gives you a Higgs...

...and the Higgs decays into bottom quarks.
Backgrounds

The problem is, there’s lots of other ways you can make two bottom quarks:

It’s one of the most common things to make.

All we can see are the decay products. And what you want to know is... did the Higgs exist?

The thing is, we can’t see inside these reactions...
Neyman-Pearson

NP lemma says that the best statistic is the likelihood ratio:

\[
\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha
\]

(Gives smallest missed discovery rate for fixed false discovery rate)
Functional space

All functions

Global Optimum
No problem

If you can calculate:

\[
\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha
\]

For which you need:

\[P(\text{data} \mid \text{theory})\]
In general

We have a good understanding of all of the pieces

Do we have

\[ P(\text{data} | \text{theory})? \]
In general

We have a good understanding of all the pieces.

Do we have

\[ P(\text{data} | \text{theory})? \]

We wouldn’t need ML if we could:

- Express the likelihood of seeing our data.
Darn

We can’t calculate 

\[ P(\text{data} | \text{theory}) \]

… but we can simulate it!
The nightmare

"data" is a 100M-d vector!
The nightmare

"data" is a 100M-d vector!
The nightmare

We wouldn’t need ML if we could:
- Express the likelihood of seeing our data
- Access infinite computing resources
- Develop infinitely-fast simulation
Summary statistics

<table>
<thead>
<tr>
<th>Raw</th>
<th>Sparsified</th>
<th>Reco</th>
<th>Select</th>
<th>Ana</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e7</td>
<td>1e3</td>
<td>100</td>
<td>50</td>
<td>1</td>
</tr>
</tbody>
</table>

We don’t need to analyze the raw data

...If we could summarize it perfectly
<table>
<thead>
<tr>
<th>Raw</th>
<th>Sparsified</th>
<th>Reco</th>
<th>Select</th>
<th>Ana</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e7</td>
<td>1e3</td>
<td>100</td>
<td>50</td>
<td>1</td>
</tr>
</tbody>
</table>

We wouldn’t need ML if we could:

- Express the likelihood of seeing our data
- Access infinite computing resources
- Develop infinitely fast simulation
- Derive perfect summary statistics

...If we could summarize it perfectly
### Summary statistics

<table>
<thead>
<tr>
<th>Raw</th>
<th>Sparsified</th>
<th>Reco</th>
<th>Select</th>
<th>Ana</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e7</td>
<td>1e3</td>
<td>100</td>
<td>50</td>
<td>1</td>
</tr>
</tbody>
</table>

We don’t need to analyze the raw data

…If we could summarize it perfectly

**Traditional ML (< 2012)**

Mostly allowed for >1 analysis feature
Combine a few features

We don’t need to analyze

...If we could summarize
Functional space

All functions

Global Optimum
How complex?

Essentially a functional fit with many parameters

Single hidden layer
In theory any function can be learned with a single hidden layer.
How complex?

Essentially a functional fit with many parameters

- Single hidden layer
  - In theory any function can be learned with a single hidden layer.
  - But might require very large hidden layer
Shallow space

Shallow networks

Shallow Optimum

Global Optimum

All functions
Neural Networks

Essentially a functional fit with many parameters

**Consequence:**
Networks are not good at learning non-linear functions. (like invariant masses!)

**In short:**
Couldn’t just throw data at NN.
Search for Input

No low-level inputs

Limited input size

Painstaking search through input space.
Deep networks

New tools let us train deep networks.

How well do they work?
Expanding space

- All functions
- Deeper networks
- Shallower networks
- Shallow Optimum
- Deep Optimum
- Global Optimum
Real world applications

**Head turn:** DeepFace uses a 3-D model to rotate faces, virtually, so that they face the camera. Image (a) shows the original image, and (g) shows the final, corrected version.
Low level data

Calorimeter pixels

Lists of tracks

Networks beat experts
**Summary statistics**

<table>
<thead>
<tr>
<th>Raw</th>
<th>Sparsified</th>
<th>Reco</th>
<th>Select</th>
<th>Ana</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e7</td>
<td>1e3</td>
<td>100</td>
<td>50</td>
<td>1</td>
</tr>
</tbody>
</table>

Networks can handle higher dimensionality and lower-level data.
The new frontier

Expertise is not obsolete!

If you know something about the problem, don’t use a completely general solution.

Engineer your network structure!

e.g., network structures which respect symmetries
Constraining space

All functions

Shallower networks

Deeper networks

Shallow Optimum

Constrained Deeper networks

Deep Optimum

Global Optimum
1. Much much more of the same

2. Something qualitatively new
Graph networks

Represent structured data
Generative models

Do more than classify

Generate data from noise

Optimal transport: new ways to compare distributions

2101.08944: Learn the detector from data!
Away from supervision
Background fitting

Away from ad-hoc background shapes:
ML for design

Optimize everything

Automatic Differentiation

Numerical gradients $\Delta L/\Delta \phi$ hopeless in trillion-D, need exact gradients $\partial L/\partial \phi$

Automatic Differentiation: careful application of chain rule to computer programs

```python
import jax
import jax.numpy as jnp

def func(x):
    y = x
    for i in range(4):
        y += x[i]**2 + jnp.sin(x[1]) + jnp.exp(-x[2])
    y = y.sum()
    return y

gfun = jax.value_and_grad(func)
gfun(jnp.array([2.0, 1.0, -4.0]))
```

TensorFlow  

... but also C++, Fortran, ...

L. Heinrich

See also: 1806.04743
ML for Theory!

How do we search large spaces?

Parameter Value (TeV)

String theory applications: 1707.00655, 1903.11616
Summary

**Modern ML**
Much more flexible and capable
Tackling previously intractable problems

**Many creative new ideas**
Widening in scope
Attacking new problems