Intro to machine learning

inga.strumke@ntnu.no

norwegian open ai lab

NTNU

The menu

Part 1: What is machine learning

Part 2: Example from particle physics, with challenges

Part 3: The black box

norwegian open ai lab

NTNU

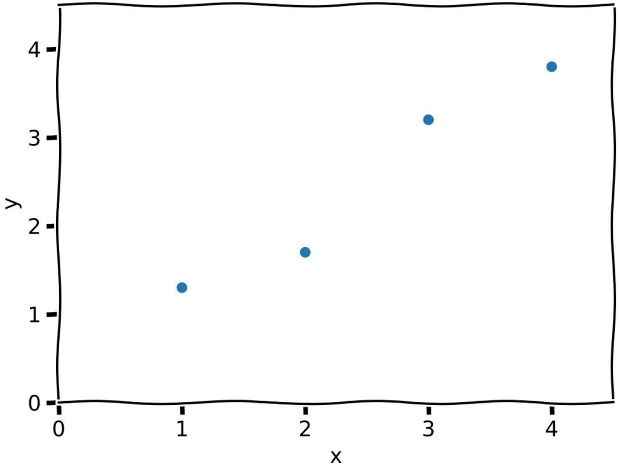
Part 1: What is machine learning

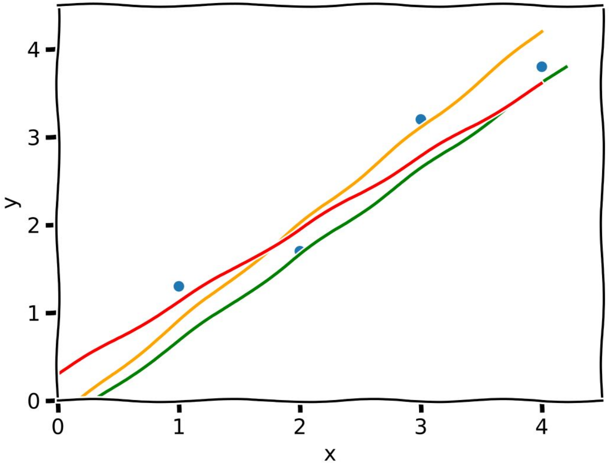
norwegian open ai lab

NTNU

Machine learning 101

Data is used to fit parameters such that a goal is reached, i.e. a loss function minimized

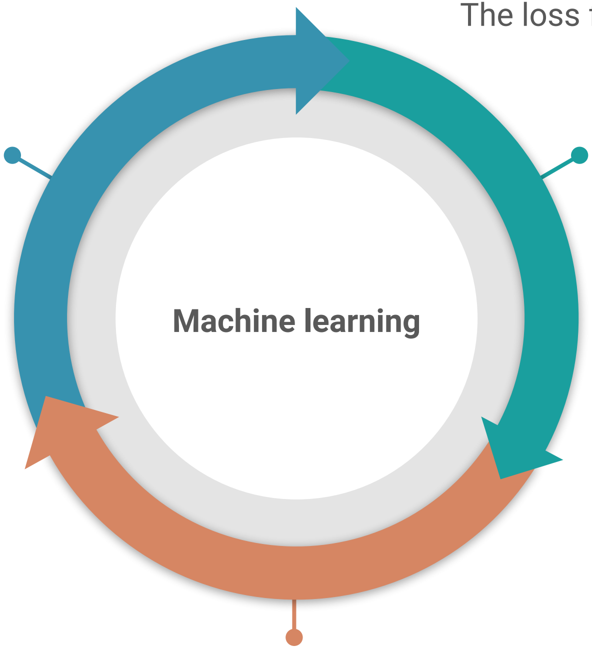




Machine learning 101

Feedback:
The loss function is evaluated on
the prediction

Prediction:
The model runs on data



Adjustment:
The model parameters are adjusted
using the gradient of the loss

norwegian
open ai lab

NTNU

data → parameter tuning → model

Machine learning 102: The three forms of learning

Supervised learning

Unsupervised learning

Reinforcement learning

Supervised learning

The data is **label**ed, i.e. contains the true value for each data point

The loss function expresses deviation of the prediction from the label

The **easiest** form, includes **regression** and **classification**

Supervised learning

Image data

Data	Label (Y)
	Cat
	Dog
	Bicycle

Tabular data

X_1	X_2	X_2	Y
12	15	0.1	5
15	23	0.5	6
13	45	0.3	8

Loss functions

Mean squared error
Mean average error
Cross-entropy
Huber loss
...

Always:

$f(Y, \text{prediction})$

Unsupervised learning

No known labels $\Rightarrow$ difficult to specify objective

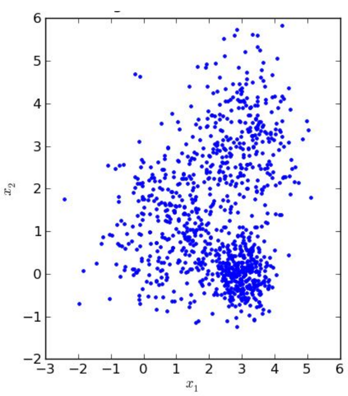
Used for **data analysis**, including dimensionality reduction, and **clustering**

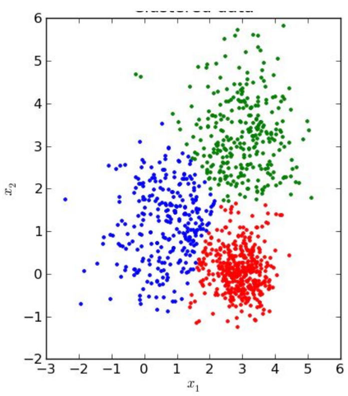
Unsupervised learning

Example:

Dimensionality reduction (e.g. **PCA, t-SNE**), followed by clustering (e.g. **k-means**)

X_1	X_2	X_2	X_3	...	X_{N-1}	X_N
12	15	0.1	500	...	1	5
15	23	0.5	200	...	3	6
13	45	0.3	600	...	2	8





Unsupervised learning

No known labels $\Rightarrow$ difficult to specify objective

Used for **data analysis**, including dimensionality reduction, and **clustering**

Unsupervised Clustering on Astrophysics Data: Asteroids Reflectance Spectra Surveys and Hyperspectral Images

December 2008

DOI: 10.1063/1.3059034

Laurent Galluccio · Olivier Michel · Philippe Bendjoya · Eric Slezak

arXiv > astro-ph > arXiv:1904.07248

Astrophysics > Instrumentation and Methods for Astrophysics

[Submitted on 15 Apr 2019]

Machine Learning in Astronomy: a practical overview

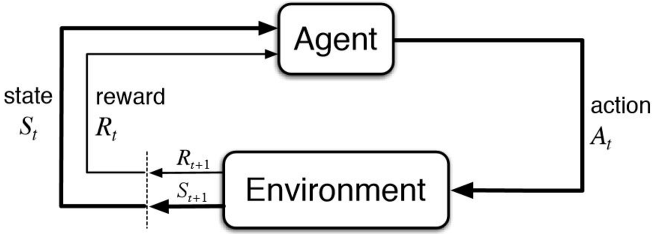
Dalya Baron

norwegian open ai lab

NTNU

Reinforcement learning

The model acts in an environment, i.e. collects its own data



Very difficult, many unsolved challenges, such as

- Goal hacking
- Exploration / exploitation tradeoff

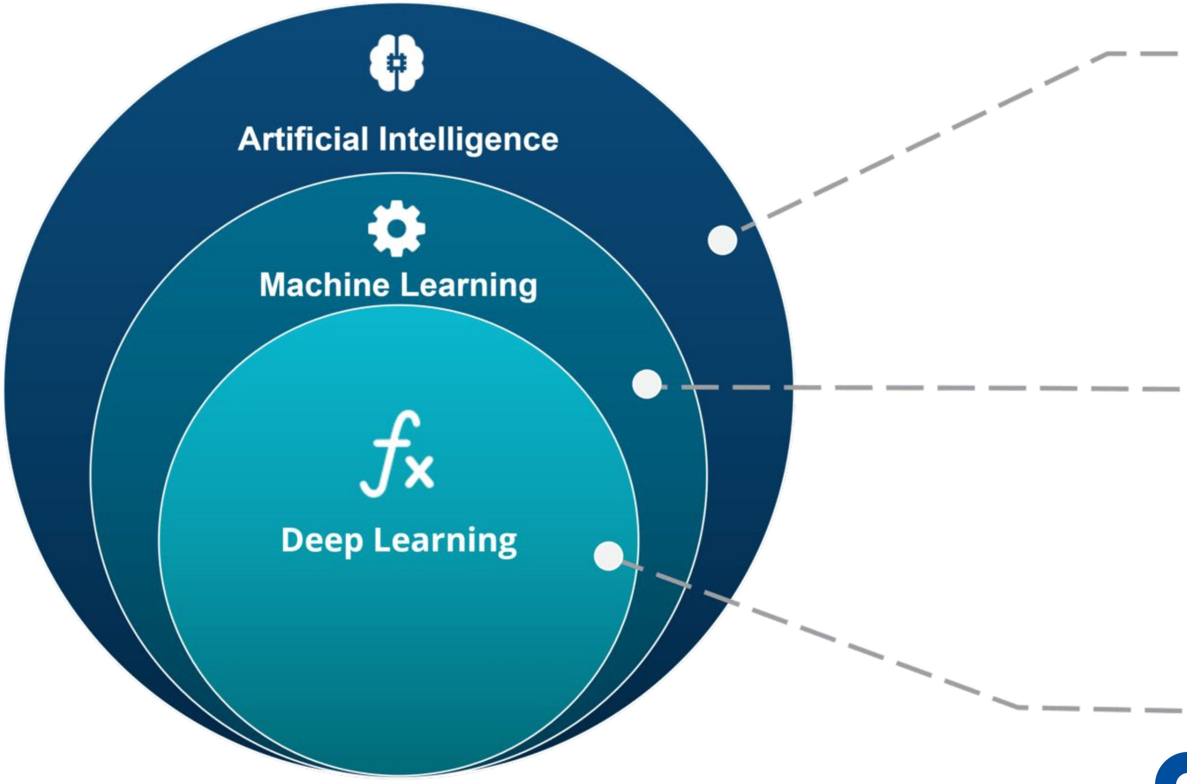
norwegian open ai lab

NTNU

norwegian open ai lab

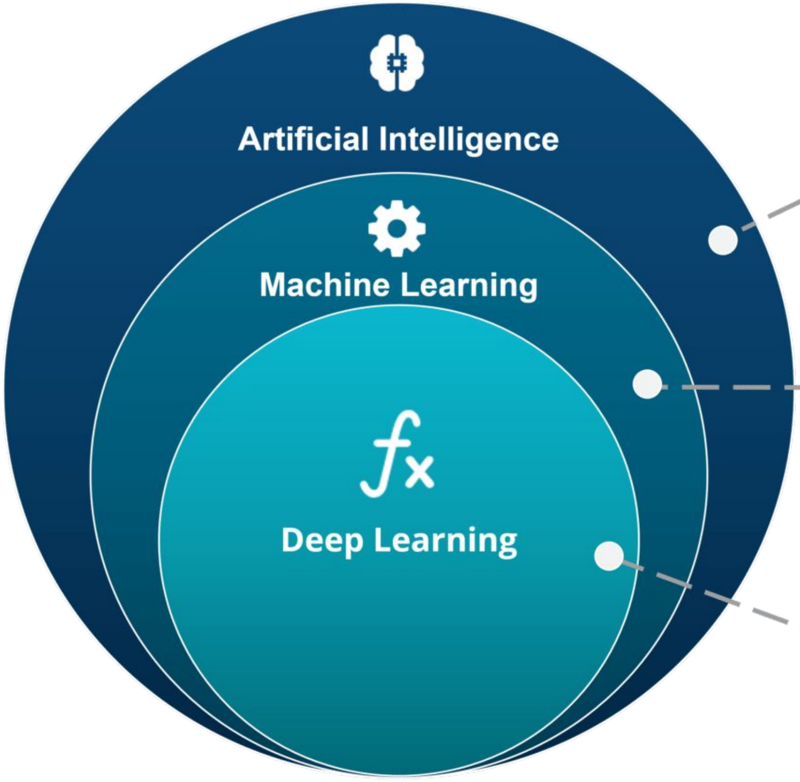
NTNU

Artificial intelligence



Academic discipline since 1950s
Collective term:
Reasoning, expert systems, robotics, formal logic, ..., machine learning

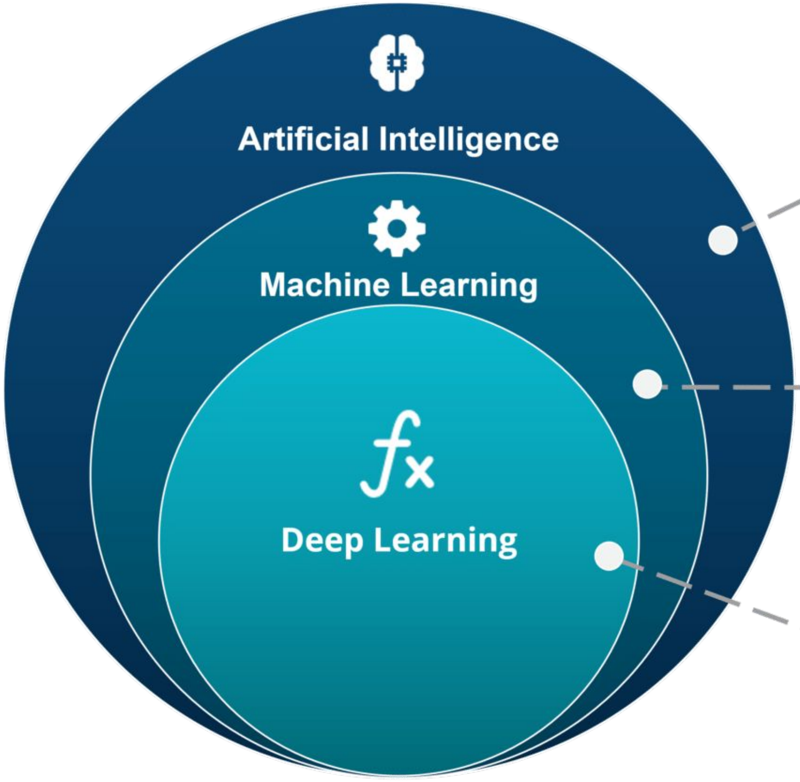
Artificial intelligence



Academic discipline since 1950s
Collective term:
Reasoning, expert systems, robotics, formal logic, ..., machine learning

Learning algorithms using **data** to achieve a goal, i.e. minimise a **loss**
Older roots than AI; basically statistics
Huge upswing during recent years

Artificial intelligence



Academic discipline since 1950s
Collective term:
Reasoning, expert systems, robotics, formal logic, …, machine learning

Learning algorithms using **data** to achieve a goal, i.e. minimise a **loss**
Older roots than AI; basically statistics
Huge upswing during recent years

Branch of machine learning

norwegian open ai lab NTNU

Deep learning 101

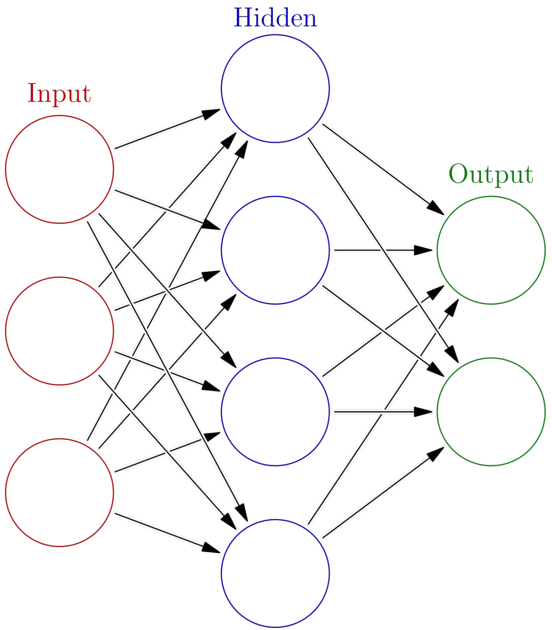
Still: data → parameters → model that makes predictions

Nodes organised in layers

Input layer: Data goes in. #nodes == #features

Output layer: Prediction.

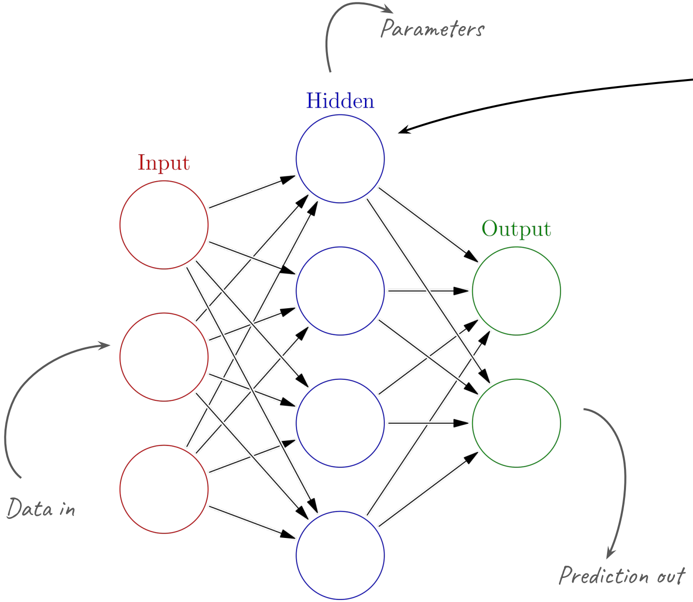
Hidden layers: Adjustable parameters.
>1 hidden layer ⇒ “deep”

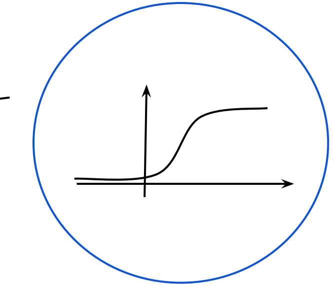


norwegian open ai lab

NTNU

Deep learning 101





Non linear function in each node

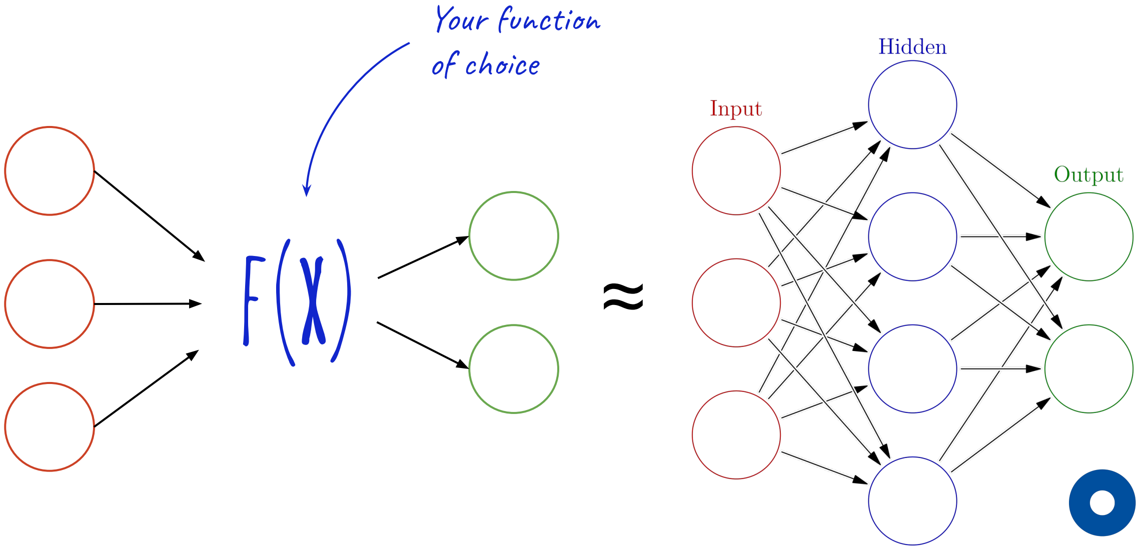
activation function

norwegian open ai lab

NTNU

Universal approximation theorems

A feed forward network with one hidden layer consisting of a finite number of nodes can approximate continuous functions on compact subsets of R, under certain (weak) assumptions about the activation function

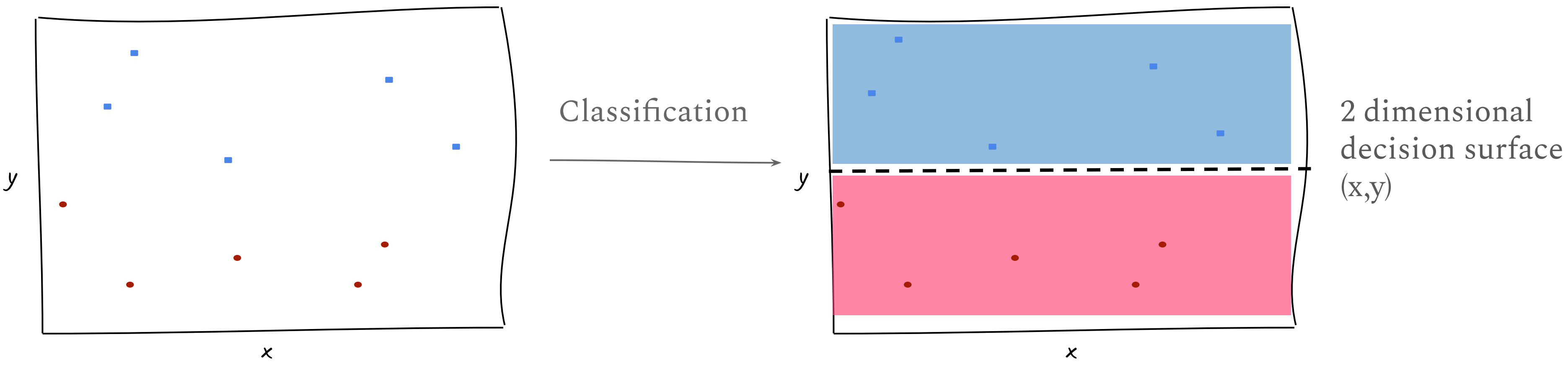


There always exists a neural network that can approximate your F

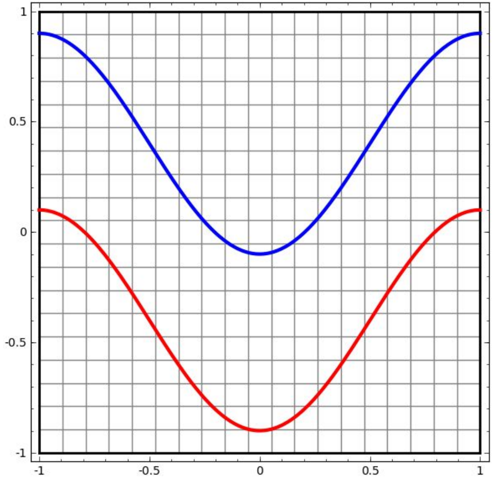
norwegian open ai lab

NTNU

Geometric intuition

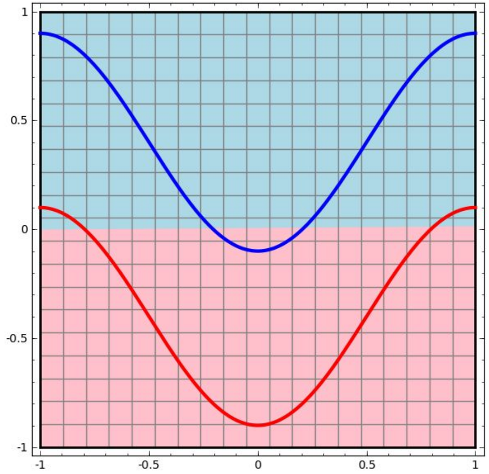


Neural network with no hidden layers or linear activation function



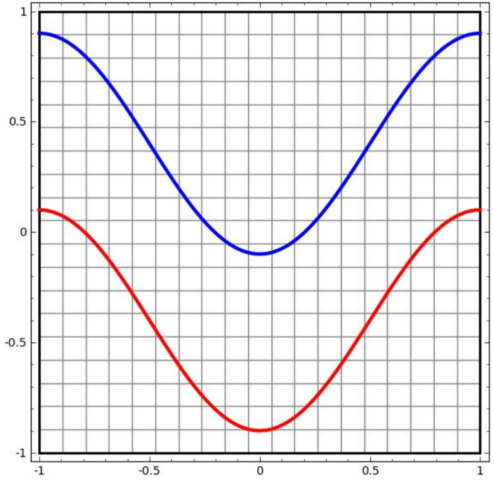
Classification?

Linear :(



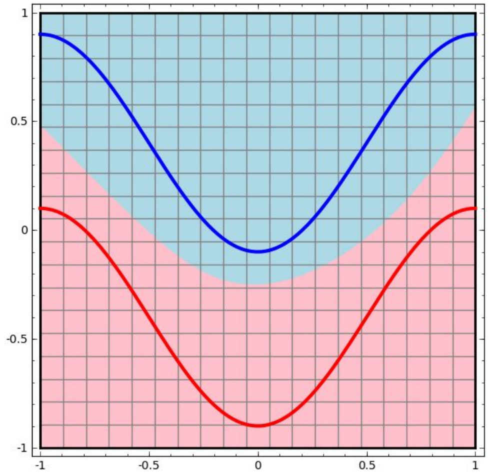
Figures from: https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

Neural network with at least one hidden layer and non-linear activation function



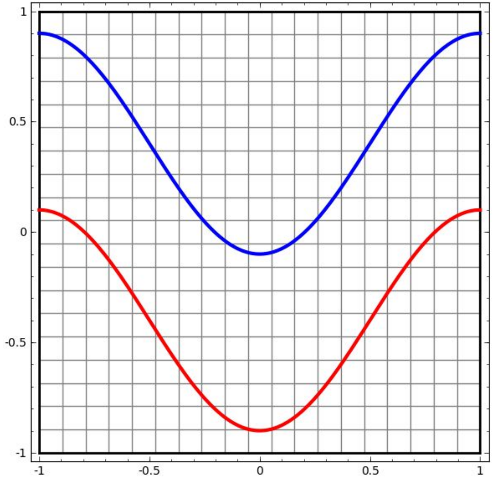
Classification

Non-linear!

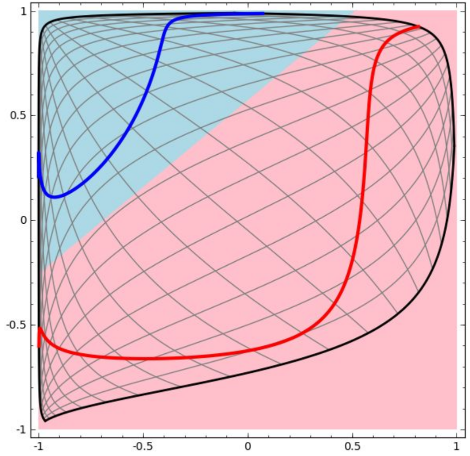


Figures from: https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

New representation of the data in the hidden layer

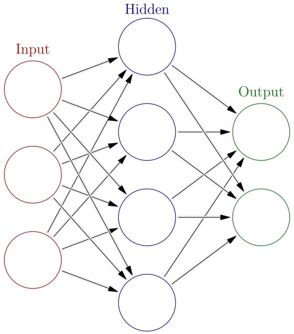


Linear separation in the space of the hidden layers



Figures from: https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

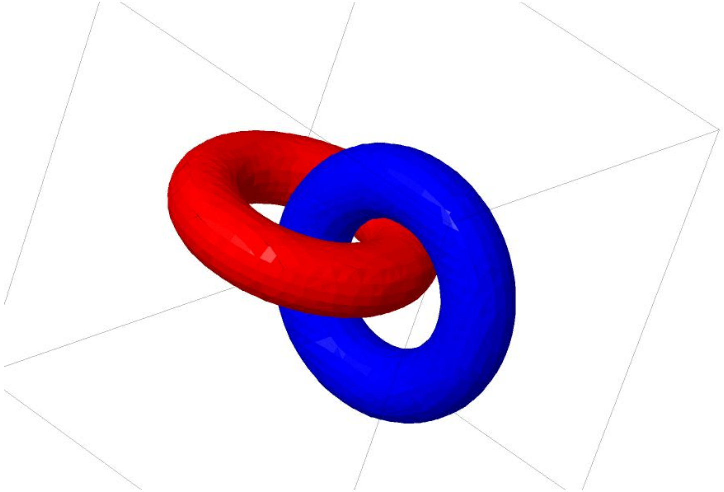
Geometric intuition



Separating the red from the blue regions requires a hidden layer with four nodes

Or: A four-dimensional representation of the data is necessary to create a decision surface

Neural networks do not change the **topology**; only the representation



norwegian open ai lab

NTNU

Figures from: https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

Part 2: Example from particle physics

The Higgs: The one and only <3

- In the Standard Model, we have one scalar SU(2) doublet Φ, and the Higgs potential

$$V_H = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$

with

- 3 d.o.f. to be absorbed by $W^{\pm}$ and Z^0

$$\Phi \sim \begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ v + \sigma(x) + i\eta_3(x) \end{pmatrix}$$

- 1 Higgs boson h and VEV v

The Higgs: The one and only <3

- Add an SU(2) doublet, call it a Higgs, and we have Φ_1 and Φ_2 and a much larger V_H.

$$\Phi = \Phi_1 + \Phi_2 \sim \begin{pmatrix} \phi_{11}(x) + i\phi_{12}(x) \\ \phi_{13}(x) + i\phi_{14}(x) \end{pmatrix} + \begin{pmatrix} \phi_{21}(x) + i\phi_{22}(x) \\ \phi_{23}(x) + i\phi_{24}(x) \end{pmatrix} \tag{3}$$

The universe has one Higgs, or five (or more, but not an arbitrary number. And not two.)

- 3 d.o.f. must still be absorbed by $W^{\pm}$ and Z^0
- 1 Higgs boson h and VEV v
- 4 d.o.f. left! $\Rightarrow$ 4 new Higgs bosons H, A, H^+ and H^-.

2HDM: Can we tell which one we found?

Five Higgses;

the old one H

two electrically charged ones H^+ H^-

two electrically charged **opposite CP states** H A

norwegian open ai lab

NTNU

2HDM: Can we tell which one we found?

$$A/H \to \tau\tau \to \pi^+\pi^0\nu\,\pi^-\pi^0\nu$$

When doing data analysis: Look at distributions of observable quantities

We only have real data from the 125 GeV Higgs

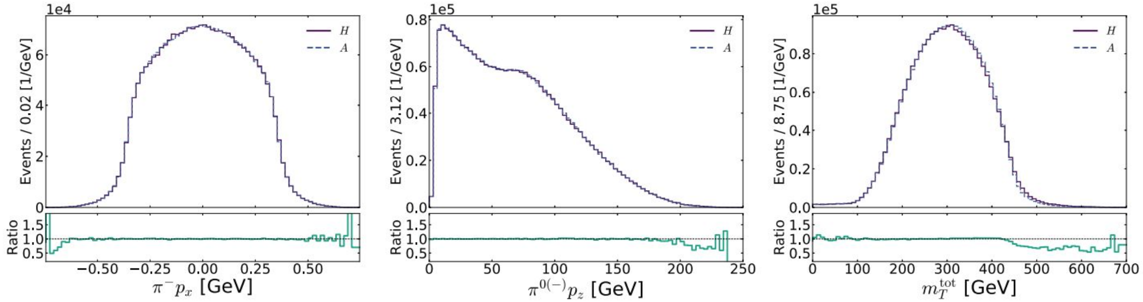
$\Rightarrow$ decide on 2HDM model and generate decays with Pythia.

norwegian open ai lab

NTNU

2HDM: Can we tell which one we found?

$$A/H \rightarrow \tau\tau \rightarrow \pi^{+}\pi^{0}\nu\,\pi^{-}\pi^{0}\nu$$

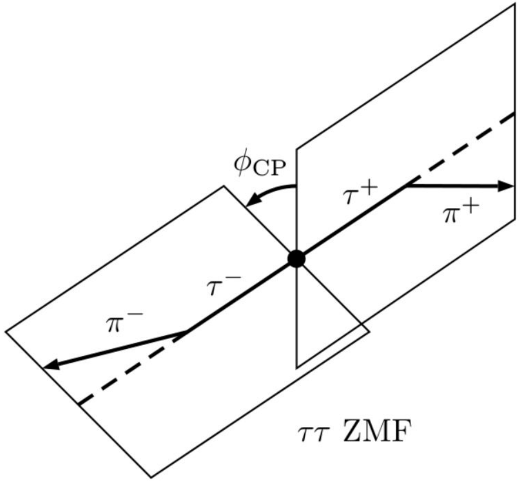


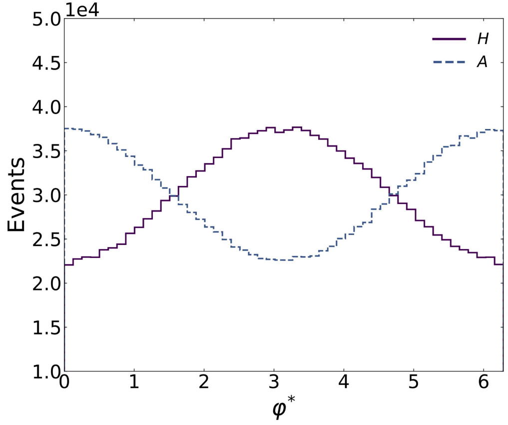
This is what bad news looks like

The competition

We <u>know</u> that there is hope.

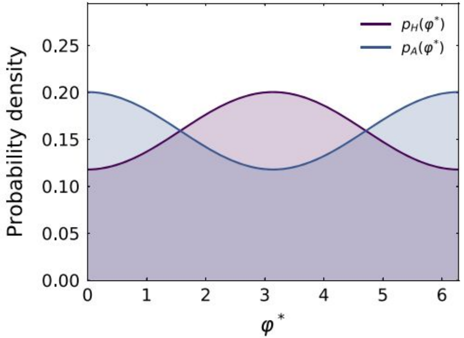
Traditional analyses look at the angle between the tau decay planes



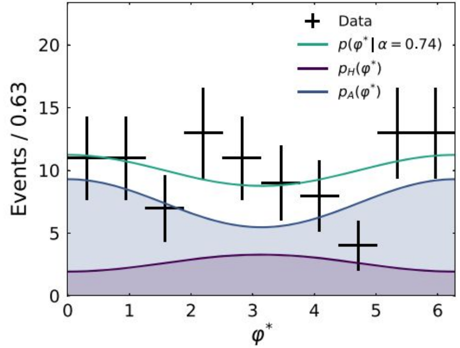


The competition

Template fit to the distribution of the φ*
(the angle between the τ decay planes)



(a)



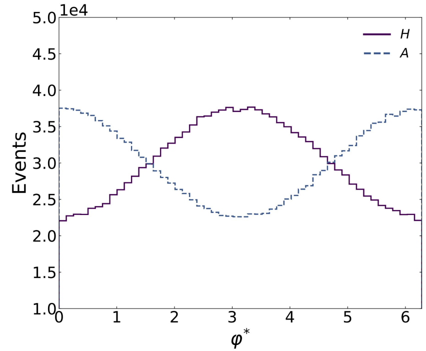
norwegian open ai lab

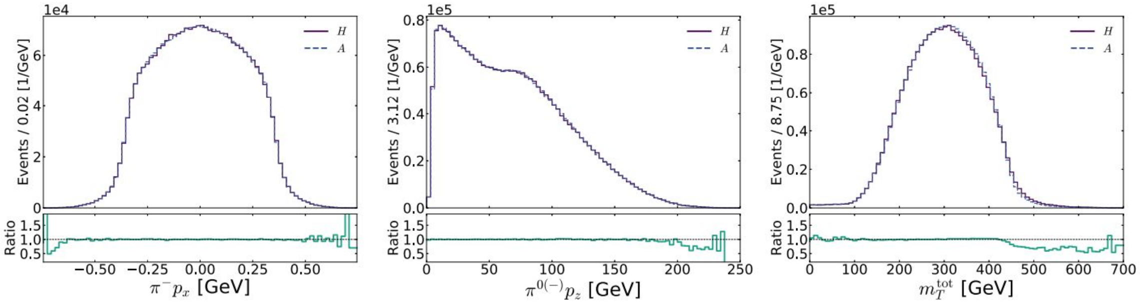
NTNU

Machine learning

The task of the neural network is to find more of these kinds of relations

In ML speak: **extracted features**





Traditional method vs machine learning

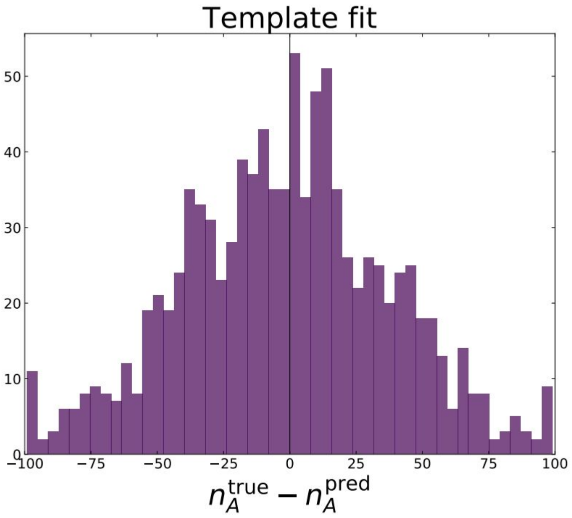
Trained deep neural network on 2 million H and A decay events

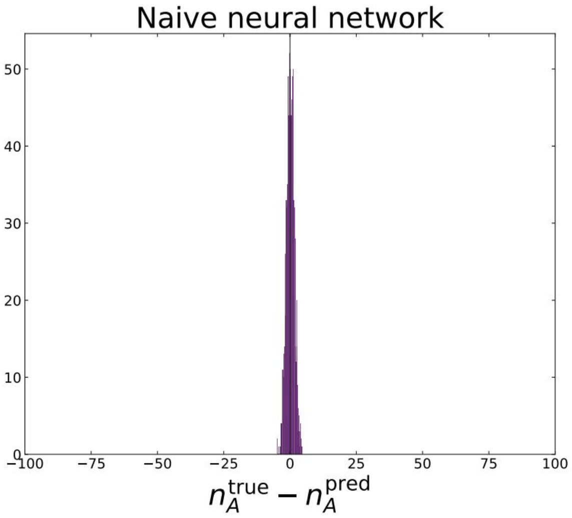
Tested both traditional method and neural network on 100 datasets of 200 events each...

Traditional method vs machine learning

Trained deep neural network on 2 million H and A decay events

Tested both traditional method and neural network on 100 datasets of 200 events each...





Not sure whether to publish in Science or Nature

norwegian open ai lab

NTNU

Too good to be true?

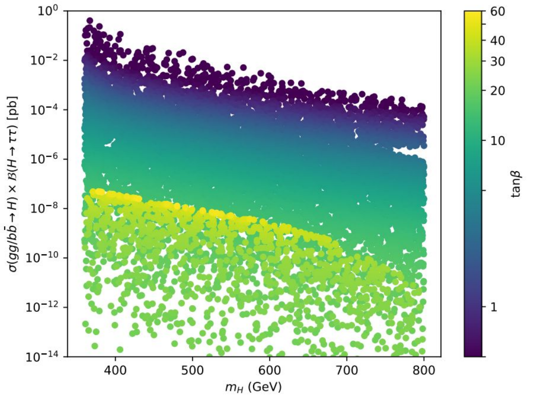
Too good to be true?

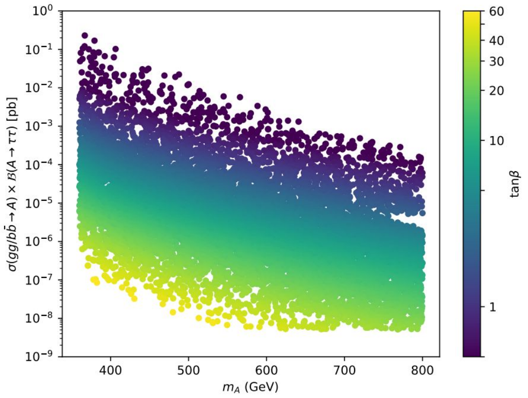
Yes. Totally.

Generated data ⇒ distributions depend on theory parameters

⇒ cannot generate data without making assumptions

⇒ cannot train machine learning model without **building in** assumptions!





norwegian open ai lab

NTNU

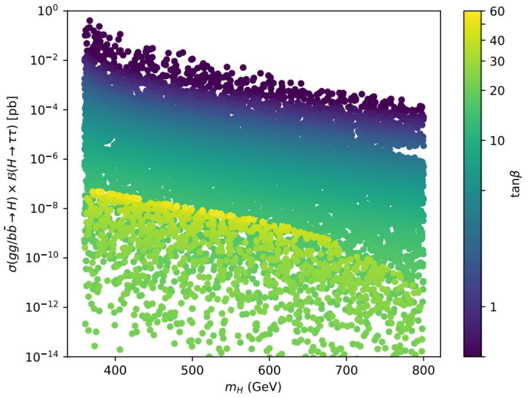
Too good to be true?

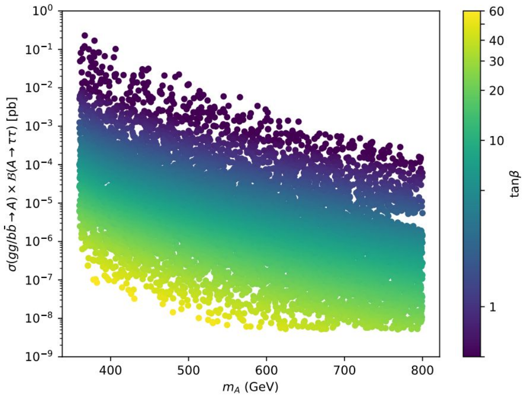
Yes. Totally.

Generated data $\Rightarrow$ distributions depend on theory parameters

$\Rightarrow$ cannot generate data without making assumptions

$\Rightarrow$ cannot train machine learning model without **building in** assumptions!





norwegian open ai lab

NTNU

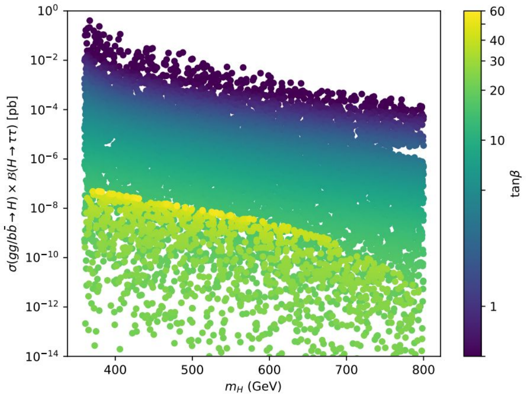
Too good to be true?

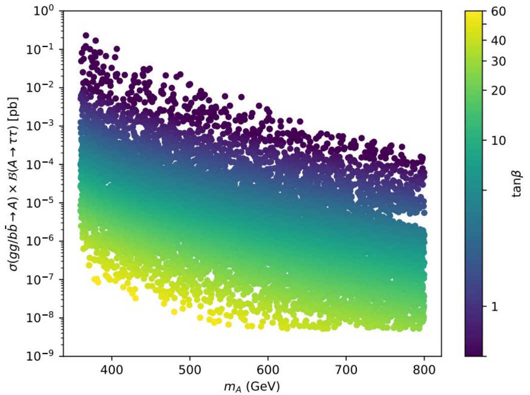
Yes. Totally.

Generated data ⇒ distributions depend on theory parameters

⇒ cannot generate data without making assumptions

⇒ cannot train machine learning model without **building in** assumptions!

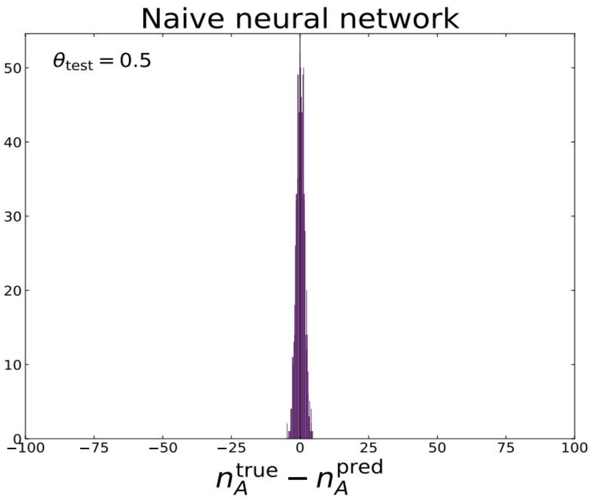


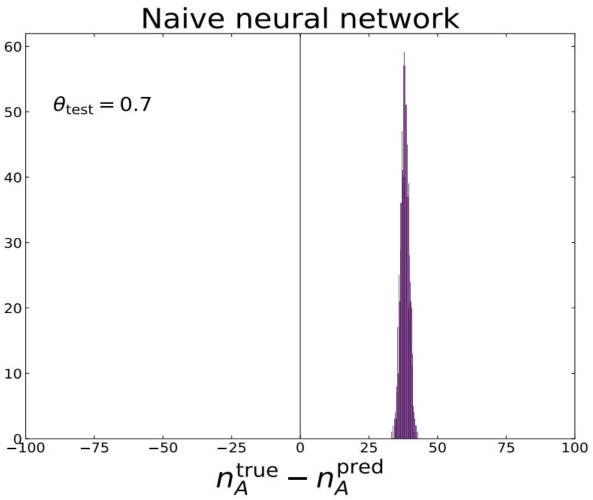


Too good to be true.

Define $\theta = \dfrac{n_A}{n_A + n_H}$, $\theta \in [0, 1]$.

Evenly distributed train set has $\theta_{\mathrm{train}} = 0.5$





Very overlapping features $\Rightarrow$ all points lie close to decision surface $\Rightarrow$ very strong prior dependence.

REUTERS

World Business Markets Breakingviews Video More

Amazon scraps secret AI recruiting tool that showed bias against women

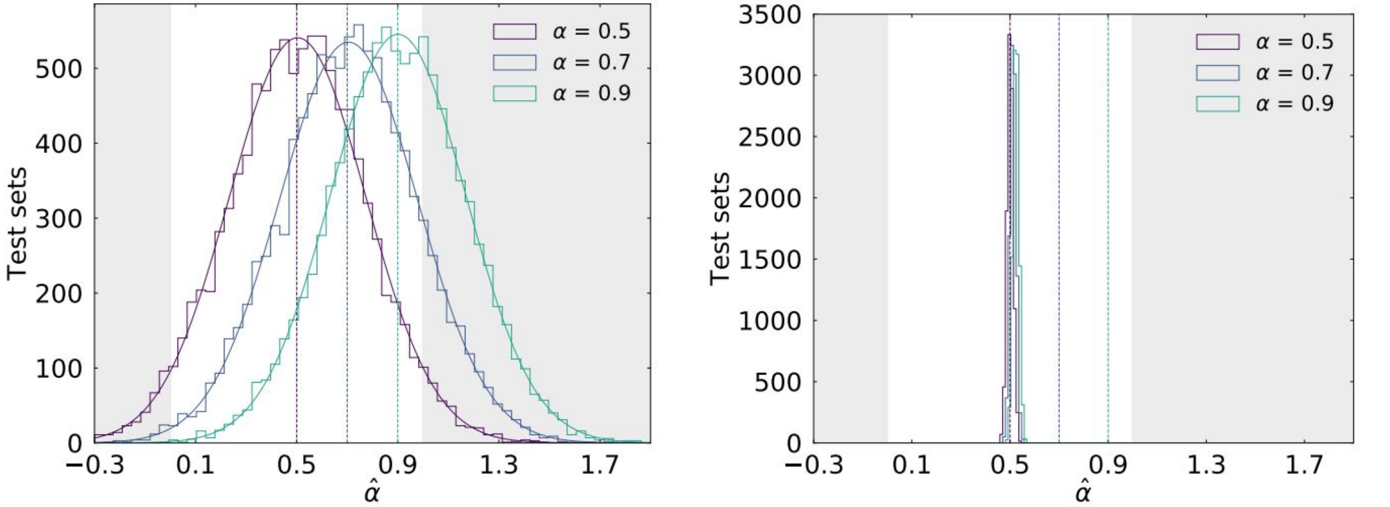
By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Traditional method still wins.

Actually, the ML method is biased to the point of being completely useless



To be continued ...

norwegian
open ai lab

NTNU

Get in touch if you want to do machine learning on physics data

inga.strumke@ntnu.no

norwegian open ai lab

NTNU