LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Machine Learning in 45 Minutes

…and some Deep Learning applications in HEP

5th IDPASC/LIP Students Workshop
July 2019 - Braga

Miguel Crispim Romão
mcromao@lip.pt

Outline

- what is ML?
- taxonomy
- anatomy
- what is DL?
- applications

I won't teach you Machine Learning in 45 mins...
but here is a checklist of what we are covering today

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

What is Machine Learning?

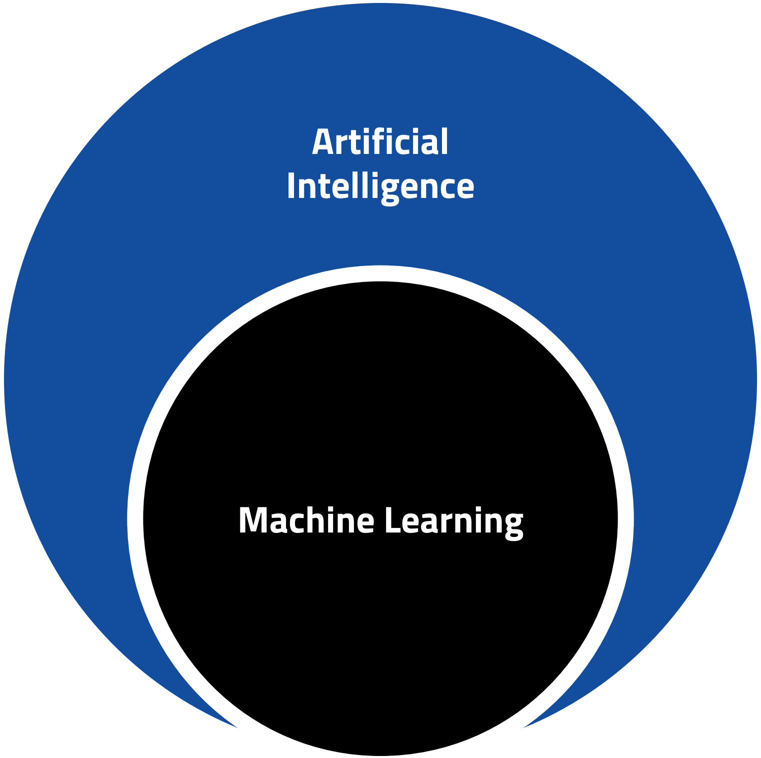
From an Artificial Intelligence Perspective

“ *Artificial Intelligence is the quest of creating machines that think and act intelligently*

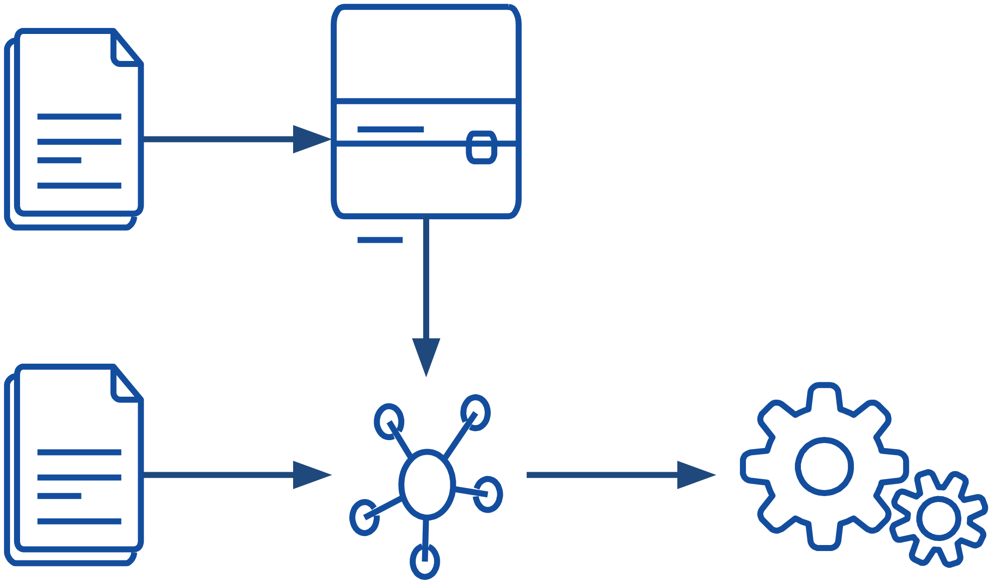
Artificial Intelligence is a big topic
and covers many problems

- Reasoning and Problem-solving
- Knowledge Representation
- Planning
- Learning
- Natural Language Processing
- Perception
- Motion and Manipulation
- Social Intelligence
- “General Intelligence”

Machine Learning is the subfield of AI that concerns how a machine can learn to perform tasks



A machine learns how to perform a task by creating a model that will act intelligently on new data



Acting intelligently depends greatly on the model

- How suitable the chosen model is
- How well the model learned from training data
- How the model generalises to new data

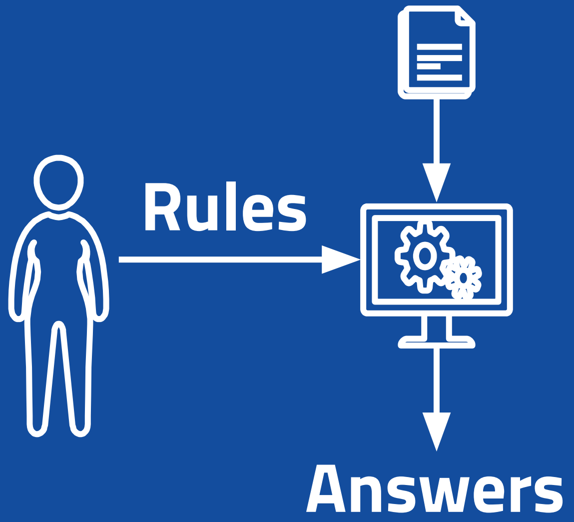
From a Statistics point of view: The model is an estimator. These issues are addressed by Statistical Learning.

From an AI point of view: the model is a function that encodes decision-making rules.

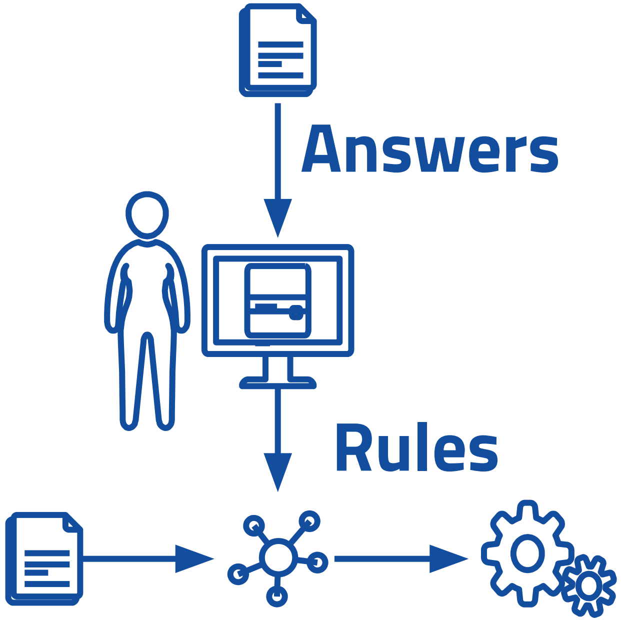
Self-Taught Code

Machine Learning is a different paradigm of computing: a program that learns what it has to do

Classical Programming



Machine Learning



Checkpoint

1

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

Machine Learning Taxonomy

What is out there and what tasks can we solve?

Machine Learning
Taxonomy: Types of Learning

The main differentiator is the type of learning, i.e. by **task**

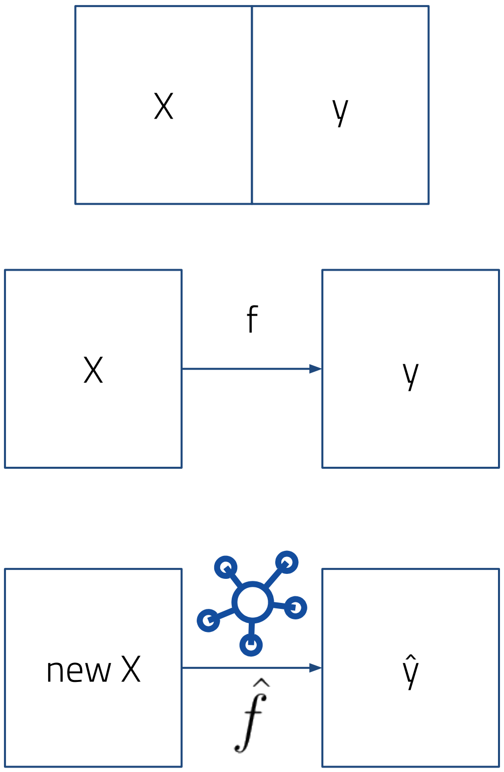
- Supervised
  - Data includes the answers
- Unsupervised
  - Algorithm embodies the answers
- Other types that we won't have time to talk about
  - Semi-supervised
  - Self-supervised
  - Reinforcement

Machine Learning
Taxonomy: Supervised Learning

- The training data includes the answer we want to reproduce
  - $\mathcal{D} = \{(X_i, y_i)\}$
  - X: Independent Variables/Features
  - y: Target Variables/Labels
- Assume (hope?) there exists a relation such that

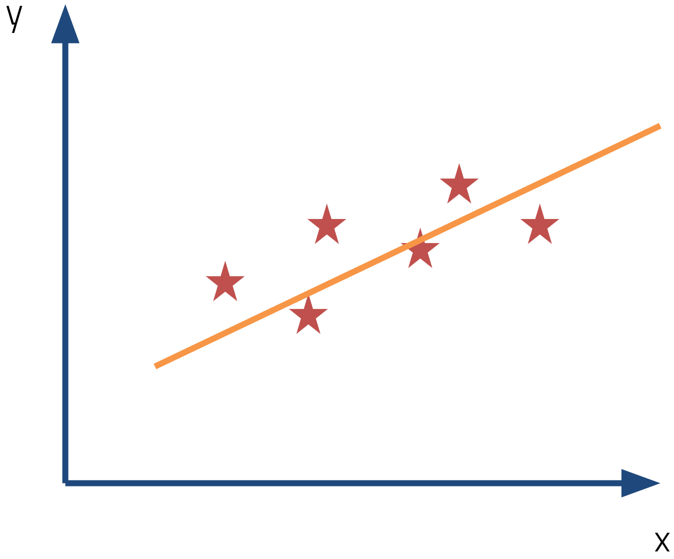
$$f : X_i \mapsto y_i$$

- The model will approximate f, $\hat{f}$
- The type of y defines two sub-classes
  - y is a real variable: **Regression**
  - y is categorical: **Classification**



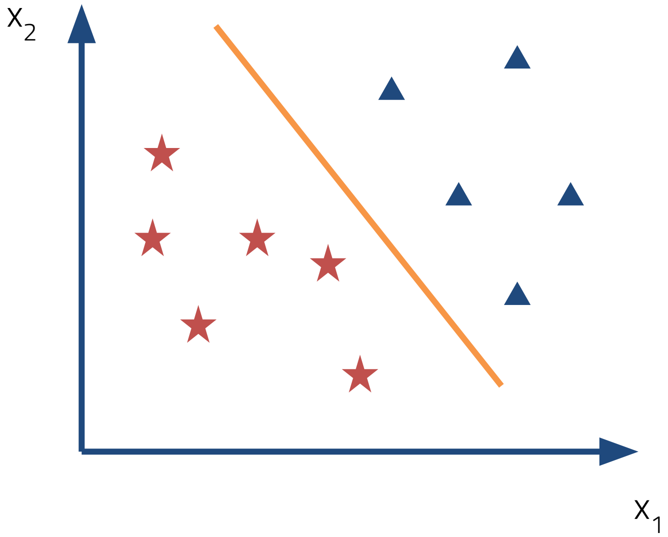
Regression Example

Linear Regression



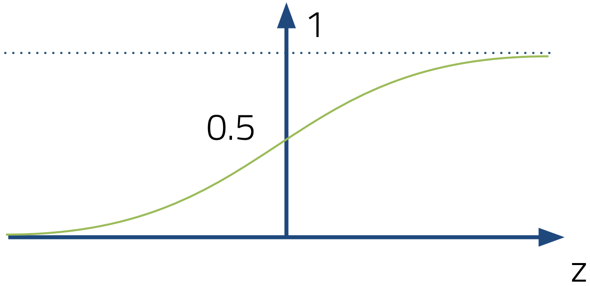
$$y = wx + b$$

Classification Example
Logistic Regression



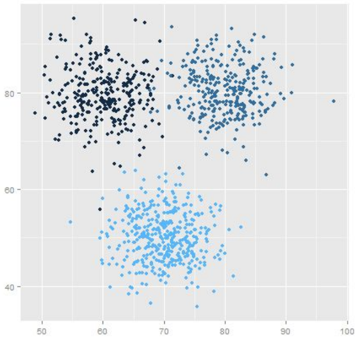
$$\sigma(x) = \frac{1}{1 + e^{-z}}$$

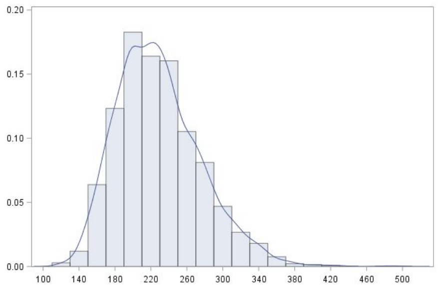
$$z = \vec{w} \cdot \vec{x} + b$$

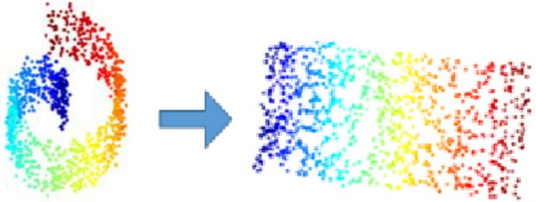


Machine Learning
Taxonomy: Unsupervised Learning

- The training data does not include the answer we want to reproduce
- $\mathcal{D} = \{X_i\}$
- The answer is embodied in the Learning Algorithm (i.e. provided by a human)
- The model will learn how to map the X to the answers, $\hat{f}$
- Answers define the type of model
  - Clustering
  - Density Estimation
  - Dimensional Reduction







Machine Learning
Taxonomy: Parametric vs Non-parametric

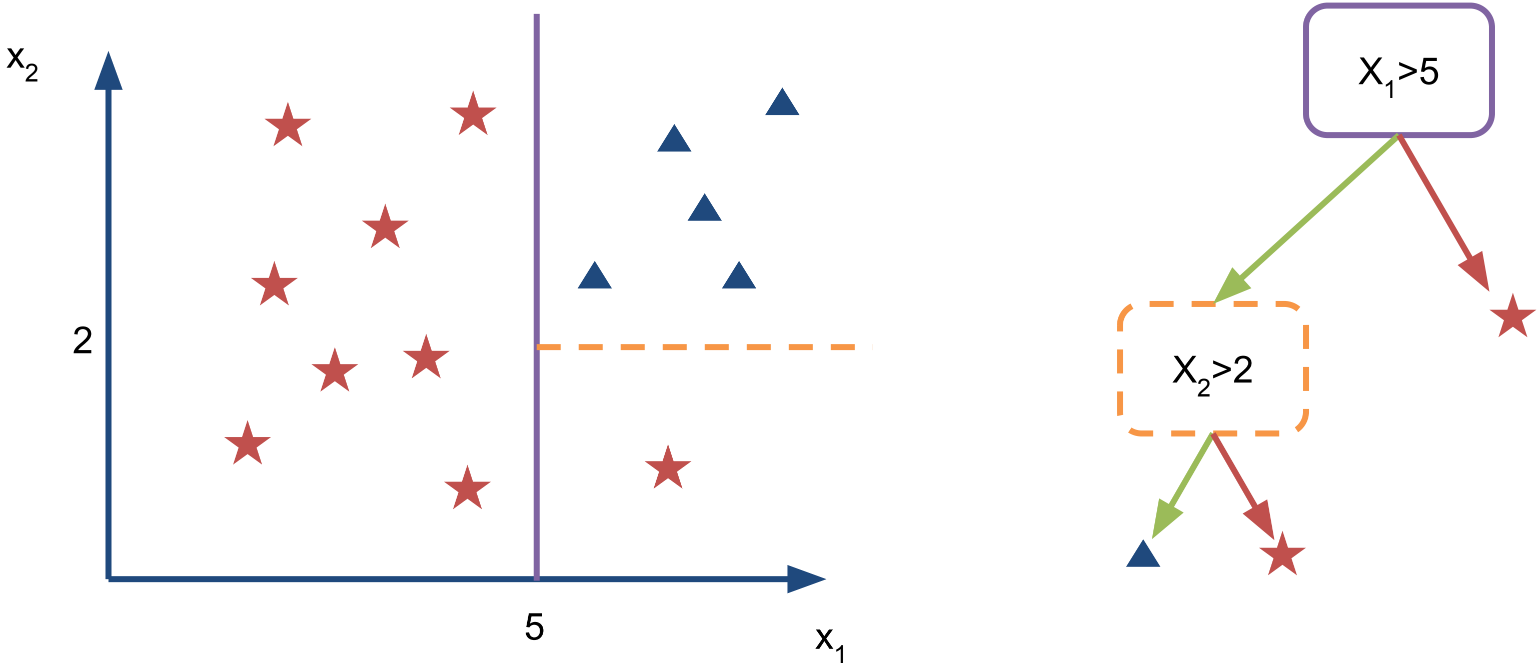
If $\hat{f}$ has parameters:

- Parametric
  - Functional form set before training
    - Linear and Logistic Regressions
    - Neural Networks
  - Training = Adapting Parameters

If it allows for “infinite” parameters:

- Non-parametric
  - Functional form depends on the training set
    - Trees
    - Histograms and K-Nearest Neighbours
  - Training = “Memorising” Training Set

Machine Learning
Taxonomy: Explicit Non-parametric example



Checkpoint
2

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

Machine Learning Anatomy

How a Machine Learning Algorithm works and produces a model

Machine Learning
Anatomy: The components of an Algorithm

- A family of models:
  - A function that performs the task
- A Loss function:
  - Measures how badly the model performs the task
- An optimisation goal:
  - To minimise the Loss function
- An Optimiser:
  - A process to modify the model to achieve the optimisation goal

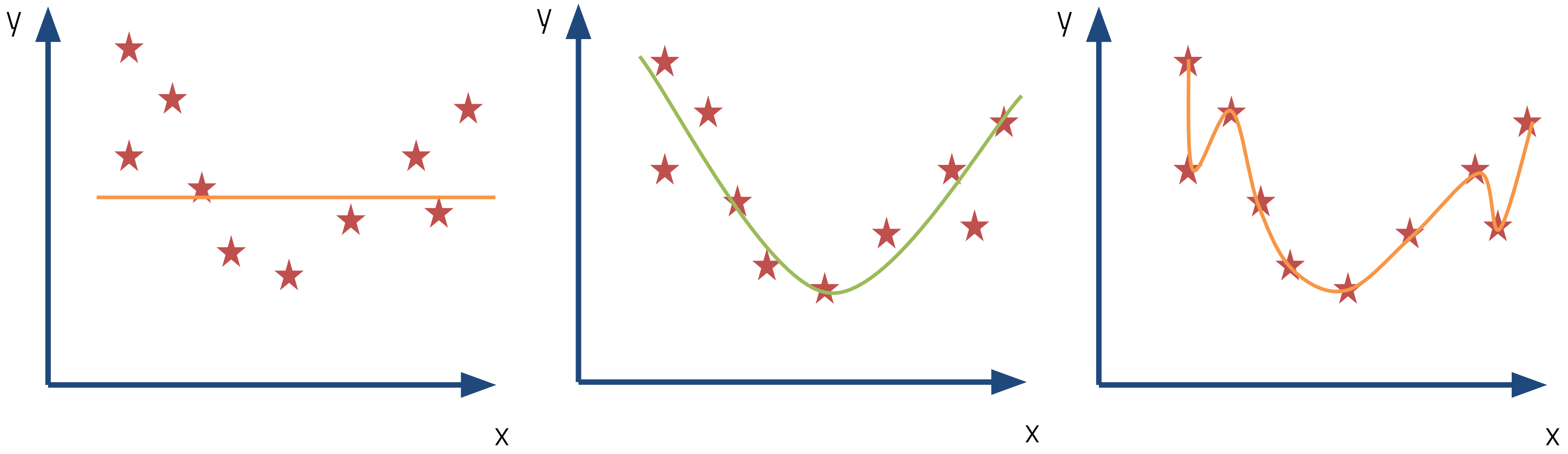
Anatomy
Model

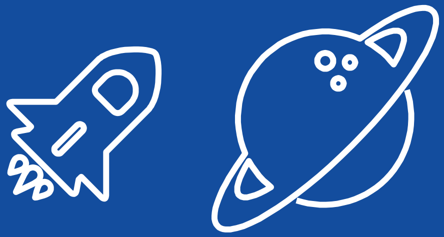
- No single rule to pick the best model: **No Free Lunch Theorem**
- Many attributes will impact decision:
  - Capacity
    - Flexibility to fit f on training data
  - Generalisation
    - How well it performs on new data
  - Training Time
  - Inference Time

Anatomy
Model Capacity

A model with insufficient capacity will fail to fit f: **underfitting.**

A model with too much capacity will fit the noise: **overfitting.**





Regularisation

In practice, one usually overestimates the capacity needed and then applies regularisation to prevent overfitting

Anatomy
Loss Function

Loss function measures how badly the model is performing the task ⇒ Depends on the task

Common Loss functions:

- Regression: Mean Squared Error

$$L = \frac{1}{N}\sum_{i}^{N}(y_i - \hat{f}(x_i))^2$$

- Classification: Cross-Entropy

$$L = -\frac{1}{N}\sum_{i}^{N}\sum_{c}^{K} y_{i,k}\log p_{i,k}$$

Anatomy
Optimisation Goal

Taken that L measures how badly the model is performing: Minimise L

- For parametric models this becomes the statement:

$$\hat{f}(x) = \hat{f}(x, \hat{w}) \ s.t. \ \nabla_w L = 0$$

- For non-parametric models we have:

$$\hat{f}(x) = \mathrm{argmin}_f L(f)$$

Anatomy
Optimiser

In general, an optimiser is an iterative process of updating the model such that the Loss function becomes minimal

This is often Loss- and model-dependent and specific to each algorithm

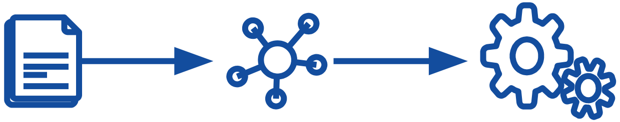
For parametric models, a very common broad class of optimisers is that of **Gradient Descent**

$$w^{t+i} = w^t - \eta \nabla L$$

Anatomy
Training and deploying

The optimisation process is often called **Training** or **Fitting**

Once the optimiser converges, the model is **trained** and can be used to perform the task on **new data**



Checkpoint

3

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

Deep Learning

In 5 minutes or less

Deep Learning
Neural Networks

Neural Networks are simply a **class of parametric Machine Learning algorithms**

They provide **a lot of capacity** and are easy to regularise

Neural Networks have striking **representational power** and are capable of **feature abstraction**

All state-of-the-art Machine Learning applications are based on Deep Learning and implement Neural Networks

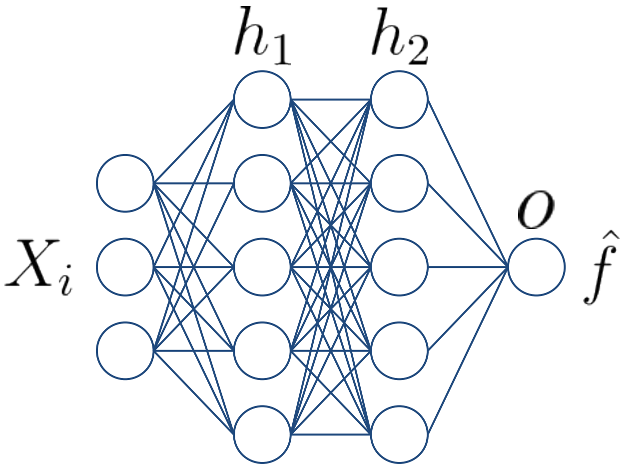
Deep Learning
Neural Networks

A Neural Net is composed of layers formed of units (neurons)

Each layer is a function of the outputs of the previous layer

The output, $\hat{f}$, is a composite function of all the inner layers and the inputs

Evaluating $\hat{f}(x)$ is called **feed-forward**



$$\vec{h}_1 = a_1(\mathbf{w_1} \cdot \vec{x} + \vec{b}_1)$$

$$\vec{h}_2 = a_2(\mathbf{w_2} \cdot \vec{h}_1 + \vec{b}_2)$$

$$o = a_o(\vec{w}_o \cdot \vec{h}_2 + b_o)$$

$$a_i = \{\tanh, \sigma, \mathrm{ReLu}, \ldots\}$$

Deep Learning
Neural Networks

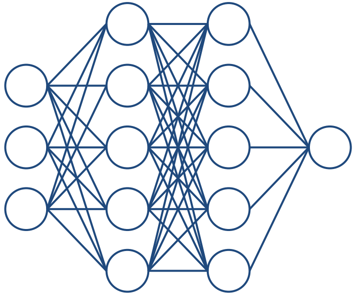
Optimisation is done using Gradient Descent implemented using **Backpropagation**:

- The parameters of each layer are updated in isolation
- Updated parameters only depend on how the error propagates from the output to that layer

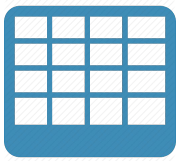
There are many optimisers that perform this operation
- Adam, Nadam, RMSprop, Adagrad, Adadelta, …

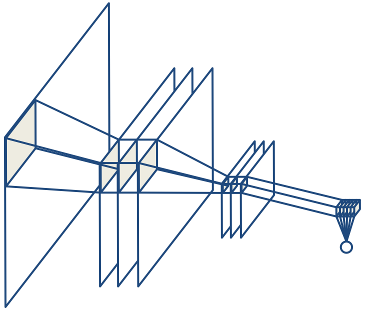
Deep Learning
Types of Neural Networks

Different architectures exist that excel at different tasks

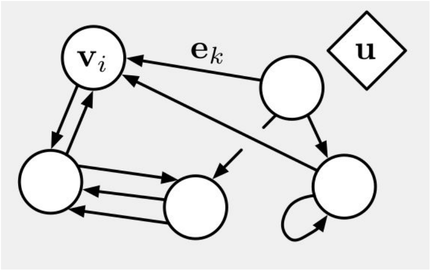


Dense

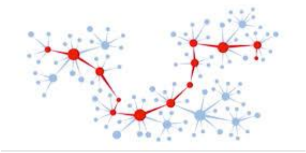




Convolutional



Graph



Checkpoint

4

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

Applications in HEP

Just a few highlights

Applications
Machine Learning in HEP

Machine Learning has been used in HEP for a long time

However, new developments that have led to the current Deep Learning era are still being explored and there is a growing research effort in bringing Deep Learning to HEP

For reviews on applications to Particle Physics:

- 1807.02876: Machine Learning in High Energy Physics Community White Paper
- 1905.06047: Supervised deep learning in high energy phenomenology: a mini review

Applications
Jet Discrimination

Jets are the cornerstone observables at colliders, from which we infer the physical processes that originated them

While traditional Machine Learning has been used to perform particle identification, Deep Learning has the potential of providing new approaches

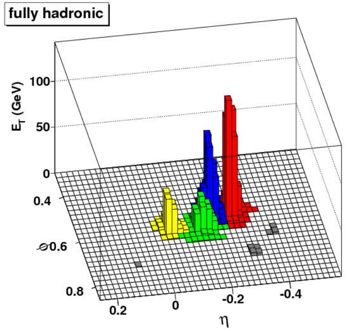
- Can we train models to detect *generic* new physics signals?
- Can deep models provide better signal efficiency and lead to better bounds on new physics?
- Can we use calorimeter images for this task, as opposed to observables from reconstructed objects?
- Can we learn new physics from what the machine has learned?

Applications
Jet Discrimination from images

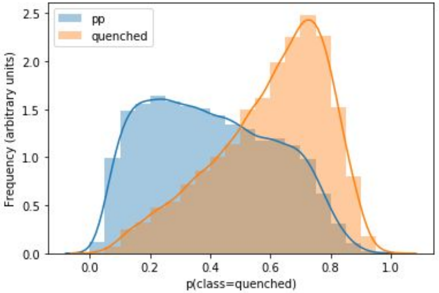
Convolutional Neural Nets are being studied to discriminate between Quark and Gluon jets (1612.0155) using (η-φ) planes

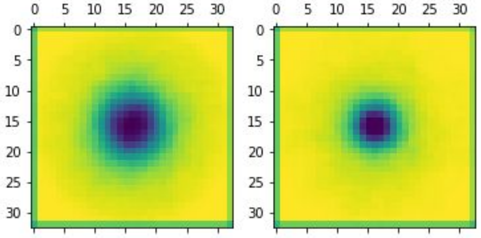
-> Study if we can use a similar approach to differentiate between vacuum and quenched jets (WIP)

=> We want to "reverse engineer" what the Net uses to learn the Physics of quenched jets



1204.0525



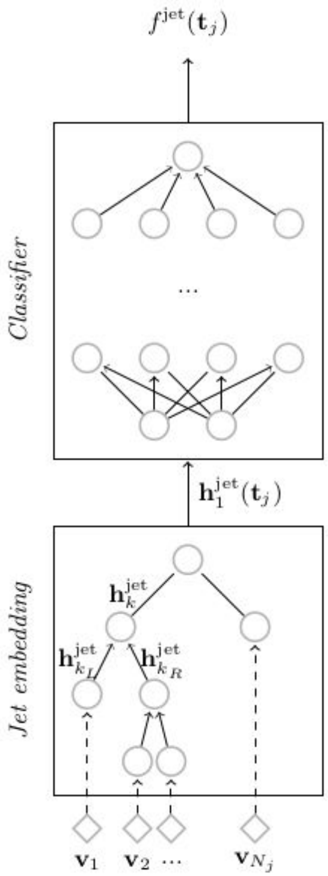


Applications

Jet Discrimination from “jet-trees”

Recursive Nets have been used to for jet discrimination using the jet reconstruction algorithm tree

Input	Architecture	ROC AUC	$R_{\epsilon=50\%}$
particles	k_t	0.9195 ± 0.0009	74.3 ± 2.4
particles	C/A	$\mathbf{0.9222} \pm \mathbf{0.0007}$	81.8 ± 3.1
particles	anti-k_t	0.9156 ± 0.0012	68.3 ± 3.2
particles	asc-p_T	0.9137 ± 0.0046	54.8 ± 11.7
particles	desc-p_T	0.9212 ± 0.0005	$\mathbf{83.3} \pm \mathbf{3.1}$
particles	random	0.9106 ± 0.0035	50.7 ± 6.7



1702.00748

Applications

One model to discriminate all

Exclusion bounds often rely on specific model variations and a selection of parameter-space benchmark points

But the background, i.e. the Standard Model, is the same for all new physics models and parameter combinations

In the same vein as other studies (1808.08979, 1811.10276, 1906.02694) we are exploring the potential of **Anomaly Detection** in Deep Learning to create a discriminator to find **any** VLQ signals

Applications
Beyond experiment... BSM model building

Experimental HEP is not the only data-intensive endeavour in HEP

When model building, one is often faced with vast parameter spaces and a myriad variations of a model

Recent developments have shown that Deep Learning can be used to perform highly efficient parameter-space scans (1708.06615, 1906.03277)

$$\mathcal{L} \sim \exp[(o_{\mathrm{model}}(\lambda_j) - o_{\mathrm{exp}})^2/\sigma]$$

Can we completely (or partially) automate model building? (I think so)

Applications
But it's not only Particle Physics!

Some recent developments:

- 1704.04650: Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy
- 1701.00008: Deep Neural Networks to Enable Real-time Multimessenger Astrophysics
- 1711.03121: Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data

Strangely, I haven't found many (or at all) references to the usage of Deep Learning to Supernovae detection, etc. **Any astronomers in the audience who want to collaborate?**

Checkpoint

5

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

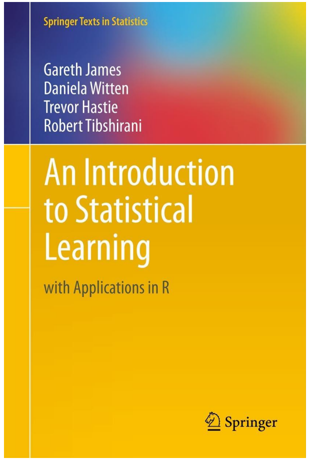
Current state of applications in HEP
Some ideas and general directions of research

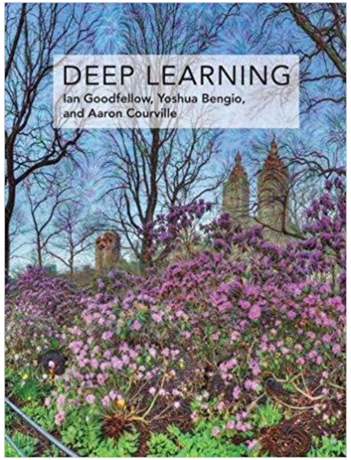
Where to go next
Some resources to help you go forward in this adventure

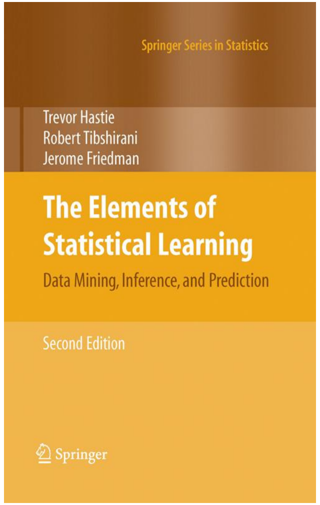
Further resources

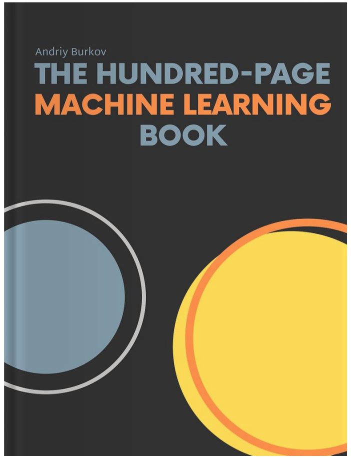
Some of them are free

These are free

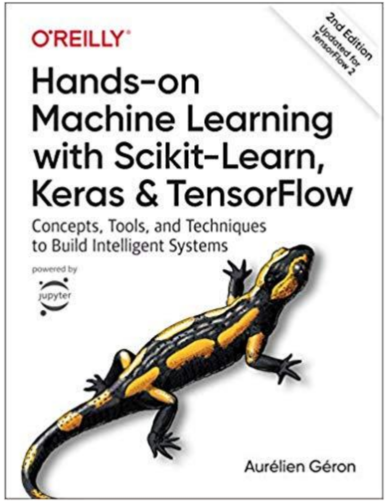


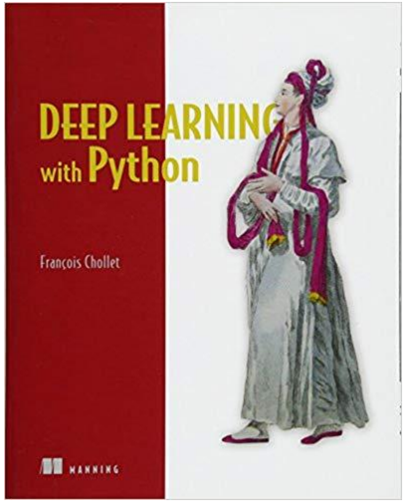


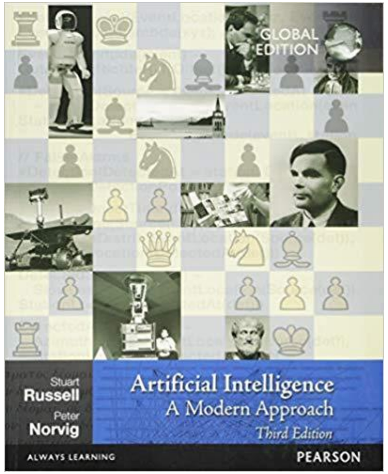


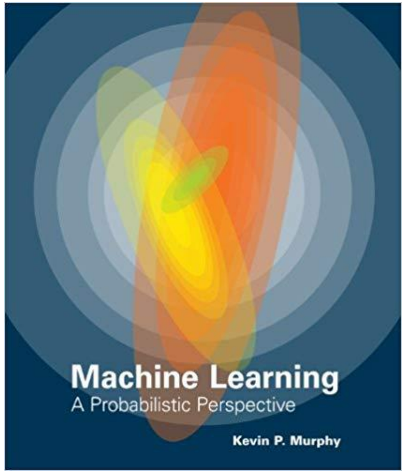


Not free, but very good









Checkpoint
6

What Machine Learning is
Context within AI and Statistics. Algorithm vs Model

Types of Machine Learning algorithms
Supervised. Classification. Regression. Unsupervised. Parametric vs Non-parametric

How a Machine Learning algorithm works
Training. Loss. Optimiser. Generalisation. Capacity. Regularisation

What Deep Learning is and why it matters
Neural Networks. Feed Forward. Backpropagation. Dense and Convolutional Nets

Current state of
applications in HEP
Some ideas and general directions of research

Where to go next
Some resources to help you go forward in this adventure

Thanks!

Any questions?