

Advanced Methods in Data Analysis

Outline of the course:

1. Statistics and Data Analysis
2. Multivariate Techniques and Machine Learning
3. Physics Modeling, Simulation and Monte Carlo Method

First two parts I will focus on applications in physics, mostly in High Energy Physics.

The last part will cover as well few typical „Data Science” problems and solutions.

Acknowledgement: slides below „borrowed” from different courses on advanced analysis methods in HEP and Data Science.

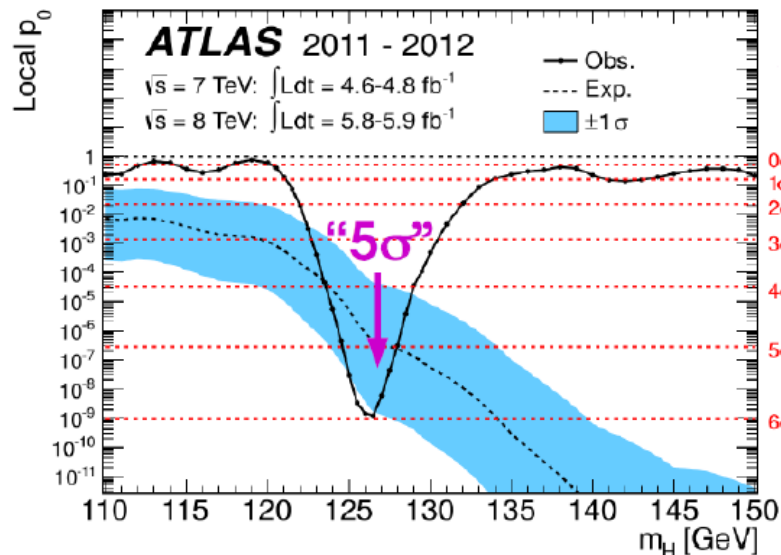
Part 1: Statistics and Data Analysis



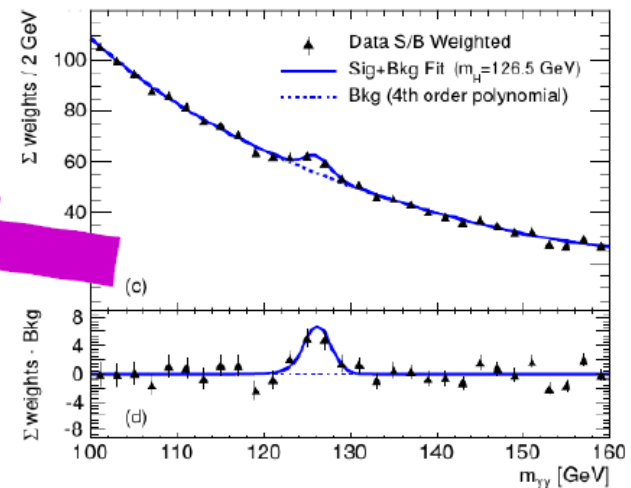
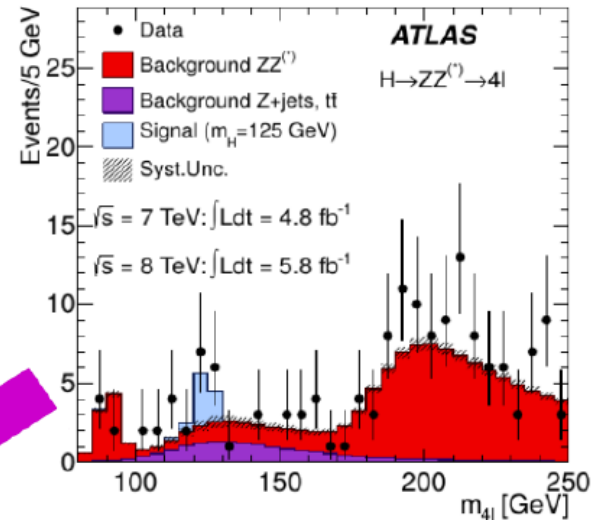
Part 1: Statistics and Data Analysis

Statistical methods play a critical role in many areas of physics

Higgs discovery : “We have 5σ ” !



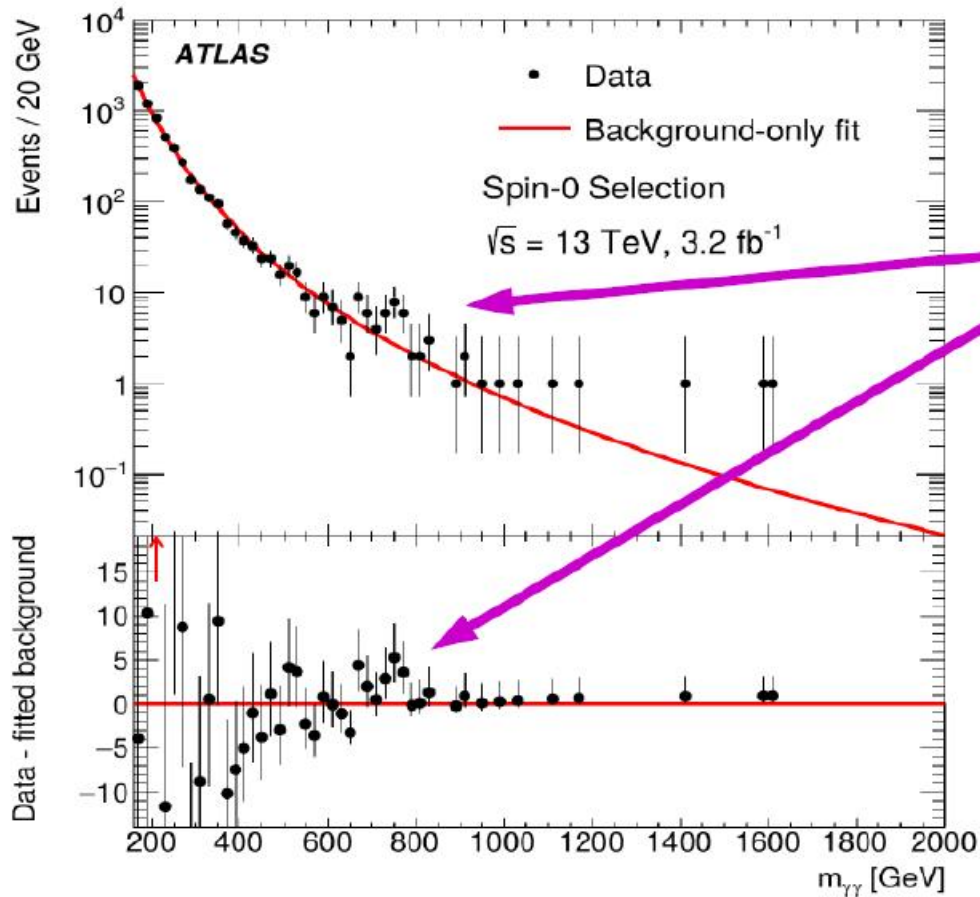
Phys. Lett. B 716 (2012) 1-29



From N. Berger, CERN Summer School, 2019

Part 1: Statistics and Data Analysis

Sometimes difficult to distinguish a bona fide discovery from a **background fluctuation**...



New Physics ?

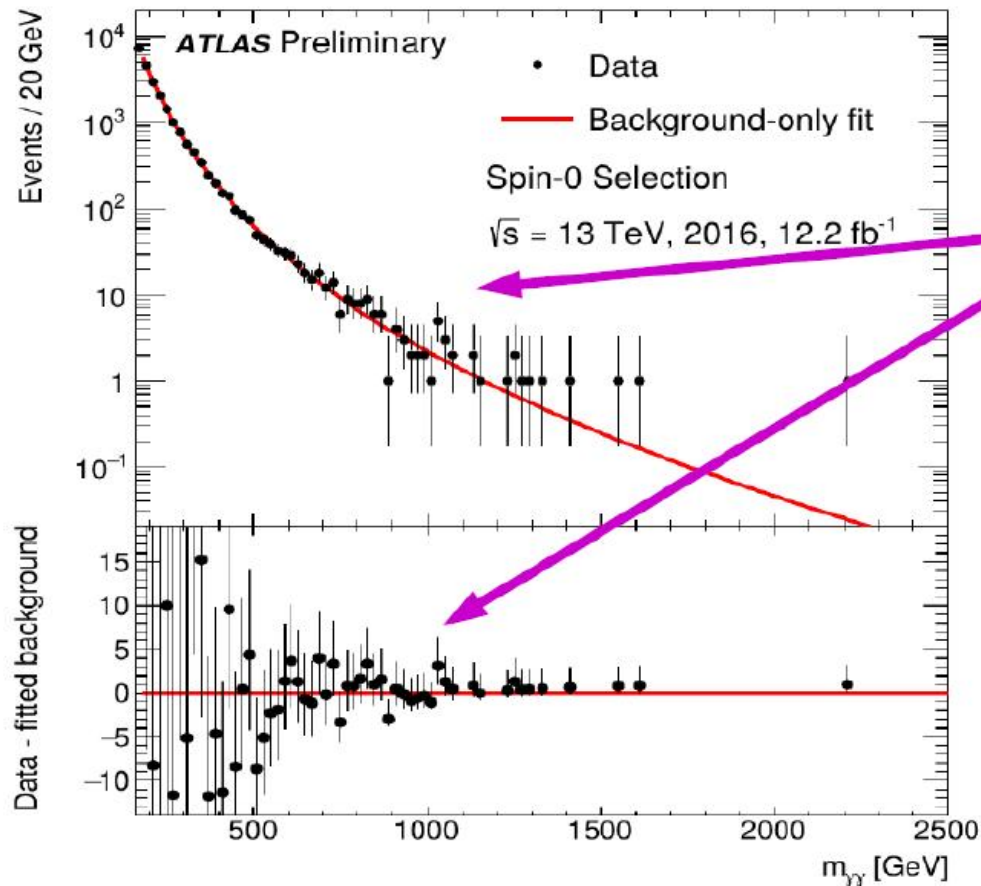
3.9 σ ? 2.1 σ ?



JHEP 09 (2016) 1

Part 1: Statistics and Data Analysis

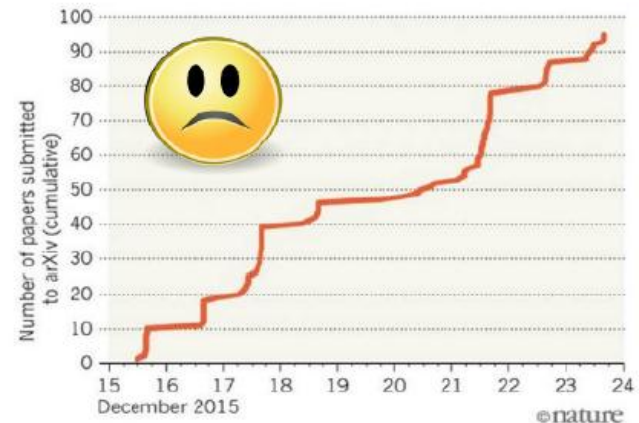
Sometimes difficult to distinguish a bona fide discovery from a **background fluctuation**...



A few months later...

~~New Physics ?~~

~~$3.9\sigma ? 2.1\sigma ?$~~

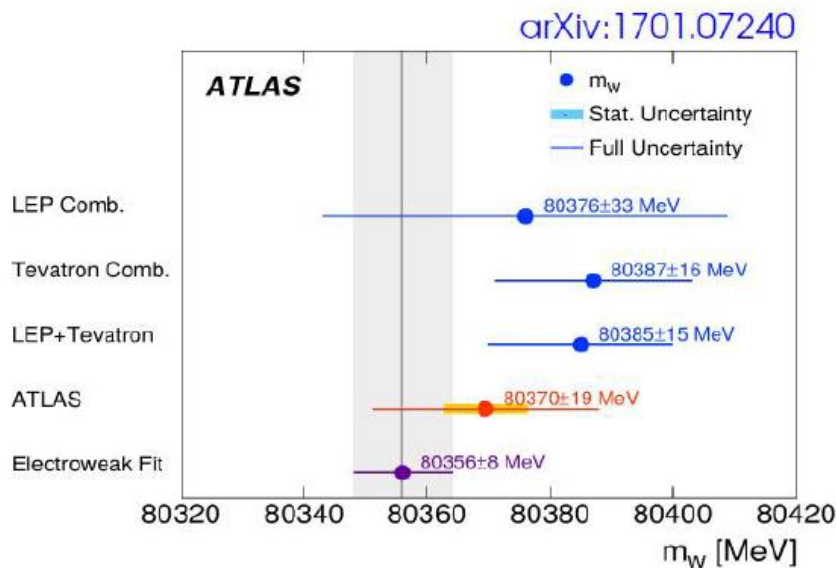


JHEP 09 (2016) 1

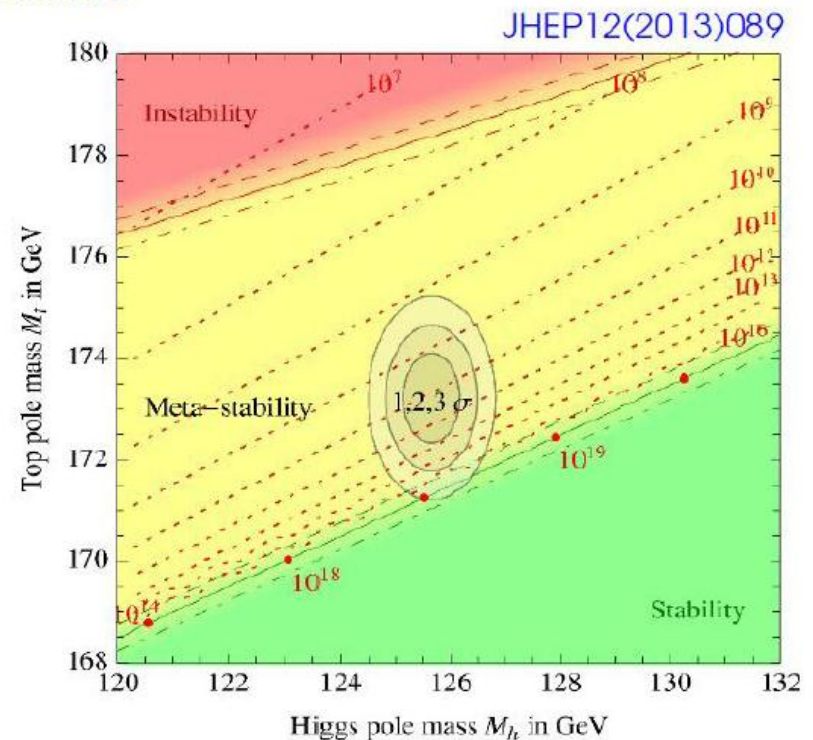
Part 1: Statistics and Data Analysis

Many important questions answered by **precision measurements**, especially if no new peaks found at high mass...

Key point = determination of **uncertainties**

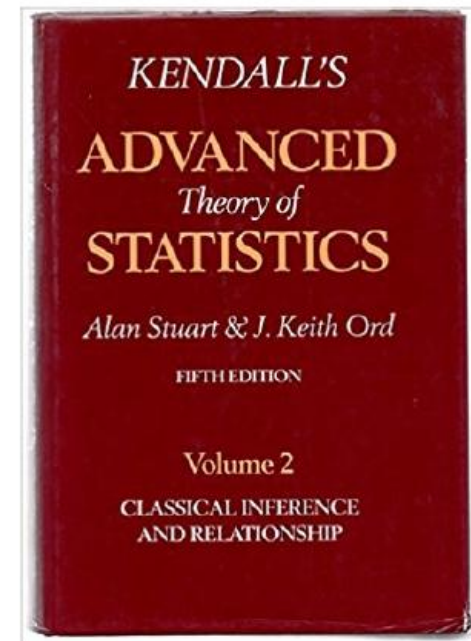
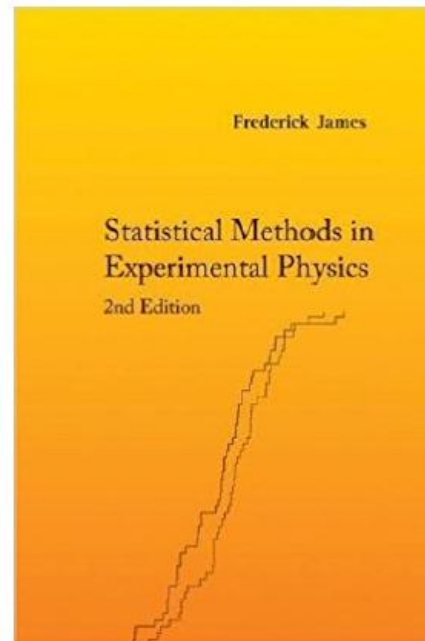
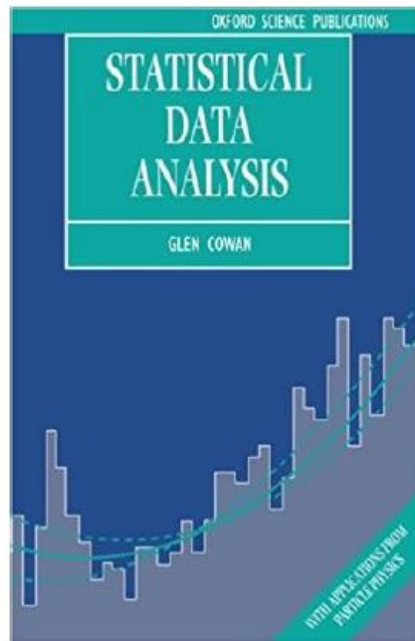


Consistency of the SM...



... or the fate of the universe

Part 1: Statistics and Data Analysis



Some other courses available online:

Glen Cowan's [Cours d'Hiver](#) and [2010 CERN Academic Training](#) lectures

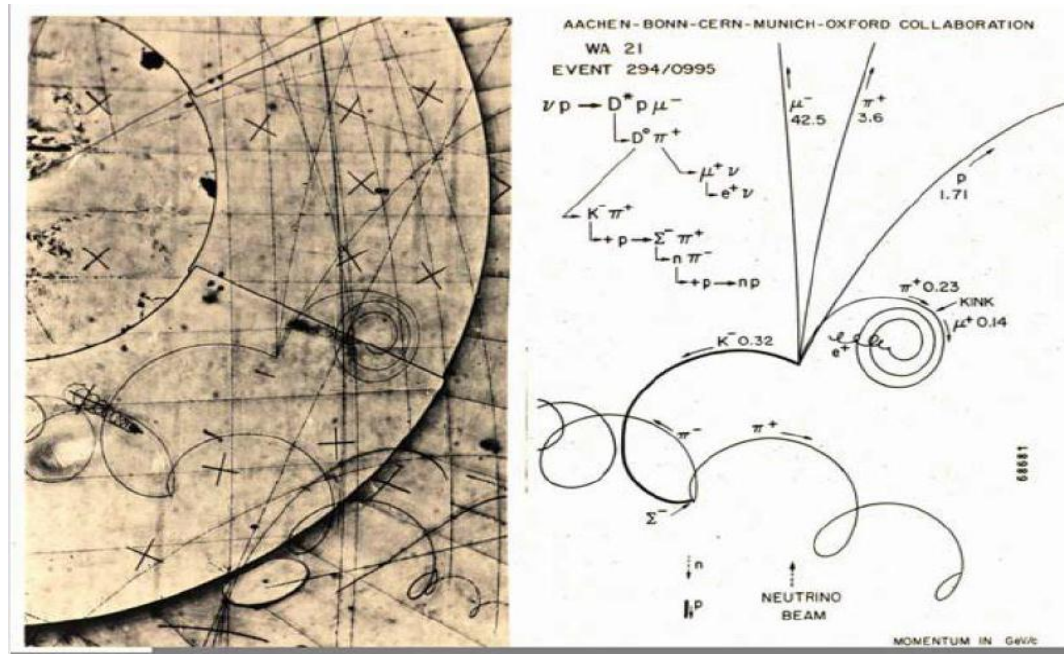
Kyle Cranmer's [CERN Academic Training](#) lectures

Louis Lyons' and Lorenzo Moneta's [CERN Academic Training Lectures](#)

Part 2: Multivariate Analysis and Machine Learning

In HEP everything started multivariate.

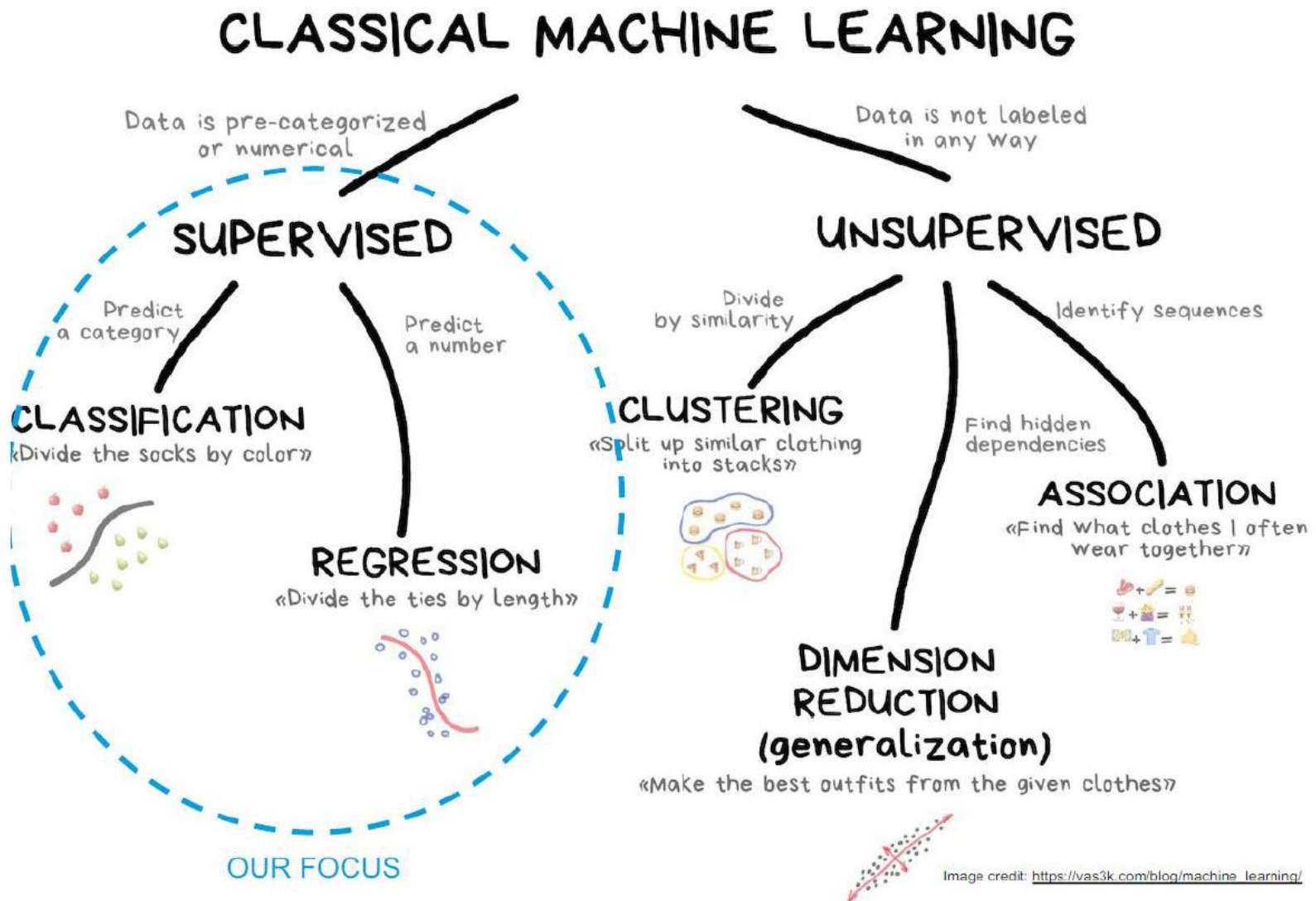
Below: intelligent „Multivariate Pattern Recognition” used to identify particles



https://www.fehcom.net/WA21/wa21_01.html

Nowdays: let computer help you.

Classical Learning



Machine Learning

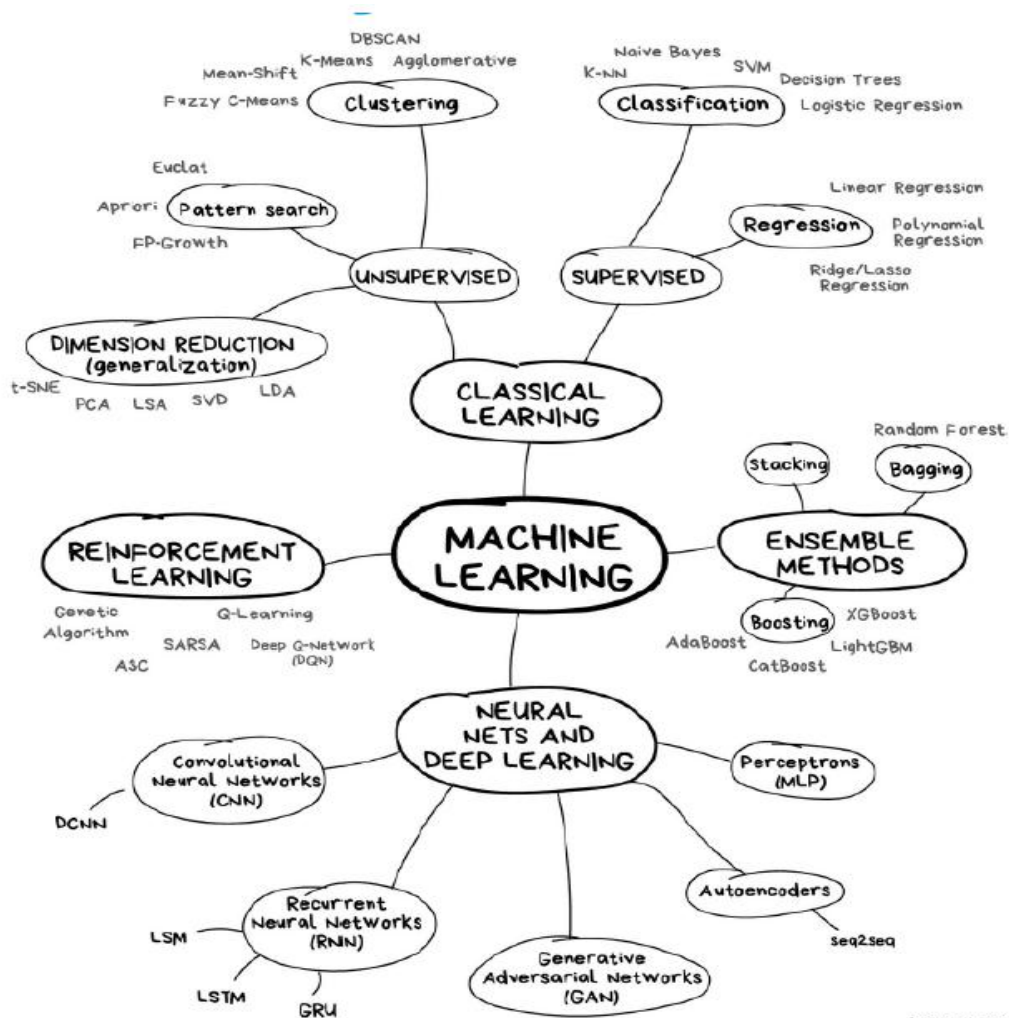


Image credit: https://vas3k.com/blog/machine_learning/

Classifiers and their properties

H. Voss, Multivariate Data Analysis and Machine Learning in High Energy Physics
<http://tmva.sourceforge.net/talks.shtml>

Criteria		Classifiers								
		Cuts	Likelihood	PDERS / k-NN	H-Matrix	Fisher	MLP	BDT	RuleFit	SVM
Performance	no / linear correlations	☹	😊	😊	☹	😊	😊	☹	😊	😊
	nonlinear correlations	☹	☹	😊	☹	☹	😊	😊	☹	😊
Speed	Training	☹	😊	😊	😊	😊	☹	☹	☹	☹
	Response	😊	😊	☹/☹	😊	😊	😊	☹	☹	☹
Robustness	Overtraining	😊	☹	☹	😊	😊	☹	☹	☹	☹
	Weak input variables	😊	😊	☹	😊	😊	☹	☹	☹	☹
Curse of dimensionality		☹	😊	☹	😊	😊	☹	😊	☹	☹
Transparency		😊	😊	☹	😊	😊	☹	☹	☹	☹

What is the model?



- ▶ This is not an apple just its graphical representation

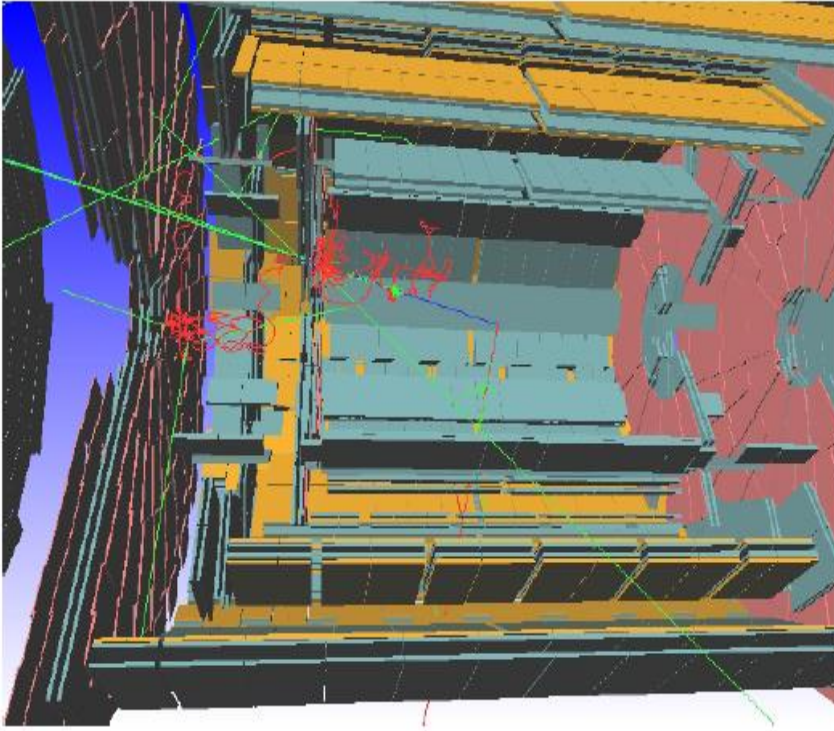
Many skills are needed to build a new model, to run it and analyze its results.

- ▶ Computational Science is an emerging, multidisciplinary domain, based on the idea of “**computational thinking**”.
- ▶ A computer-based description offers a new language, a new methodology to address scientific challenges, far beyond the scope of traditional numerical methods, and in fields where these classical approaches hardly apply.

Part 3: Physics modeling, simulation and Monte Carlo methods

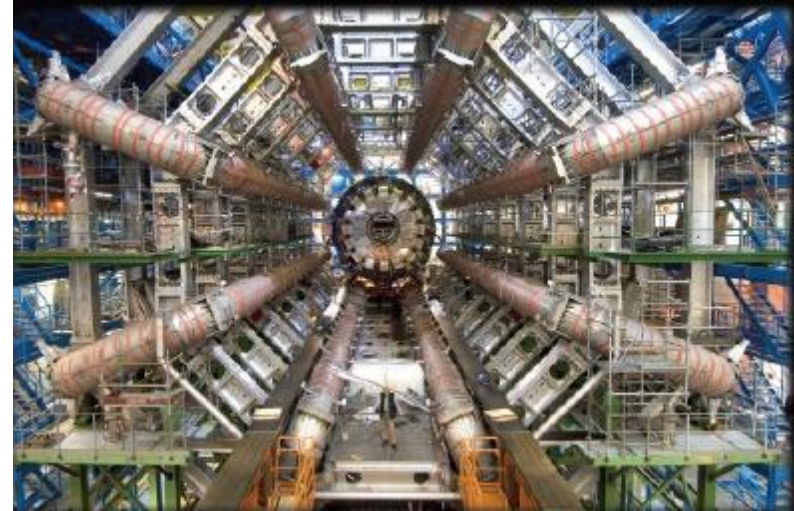
GEANT4

Visualised model of the detector used for simulation



GEANT4 is also used to determine the performance of X-ray and gamma-ray detectors for astrophysics

Detector



Getting your ETCs for the lectures

- **A written exam on the theory part.**
- **List of topical questions available after each module**
- **You will be asked to answer 5 questions out of 25-30 on the list.**