Politecnico di Bari

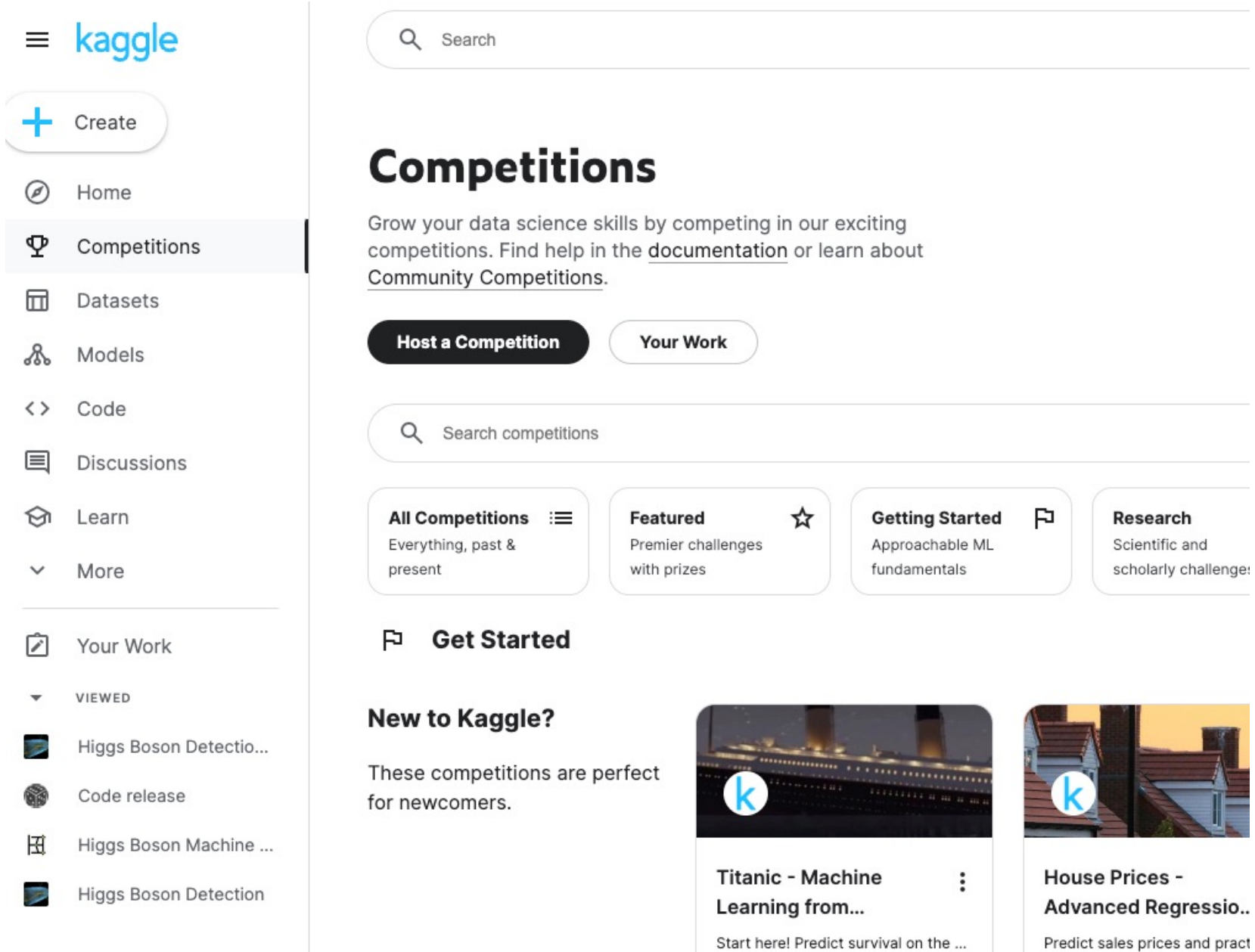
DIPARTIMENTO INTERATENEO DI FISICA

Dottorato in Fisica – XXXIX ciclo - 2024

Machine Learning techniques for particle physics

Federica Maria Simone - federica.simone@poliba.it

Machine Learning competition!



kaggle.com

Article | Published: 02 July 2014

Searching for exotic particles in high-energy physics with deep learning

P. Baldi ✉, P. Sadowski & D. Whiteson ✉

Nature Communications **5**, Article number: 4308 (2014) | Cite this article

Abstract

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on 'shallow' machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.

https://doi.org/10.1038/ncomms5308

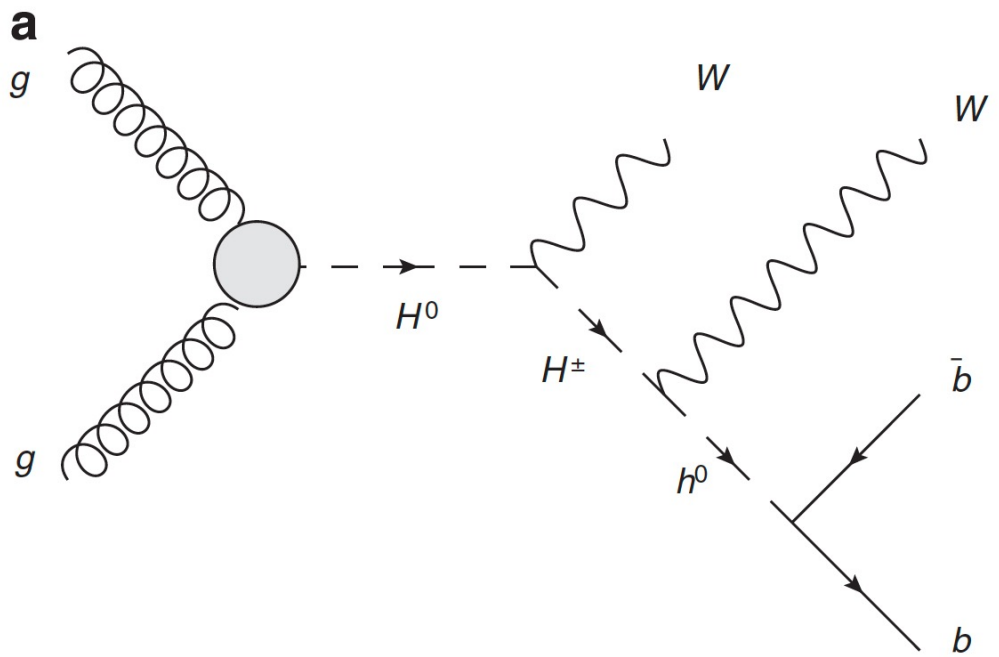
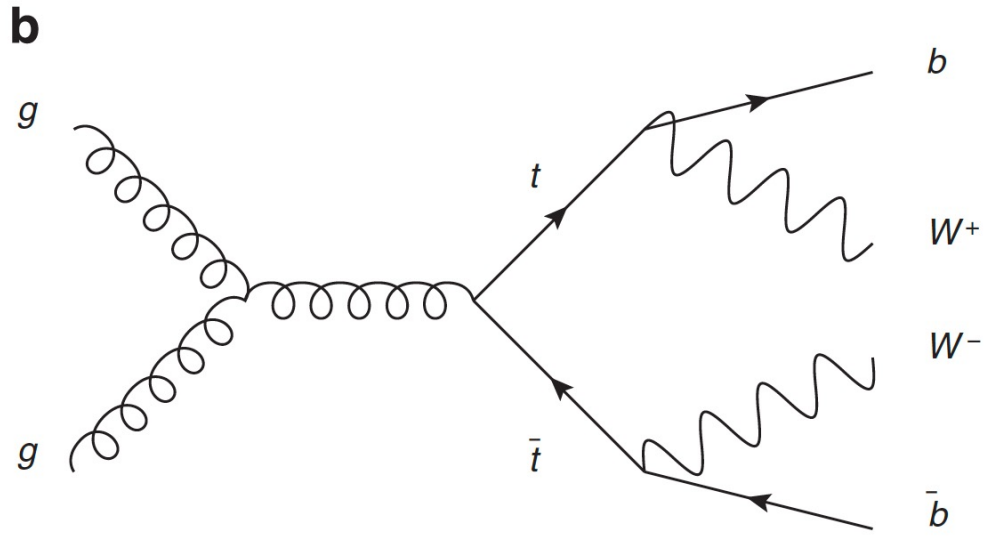
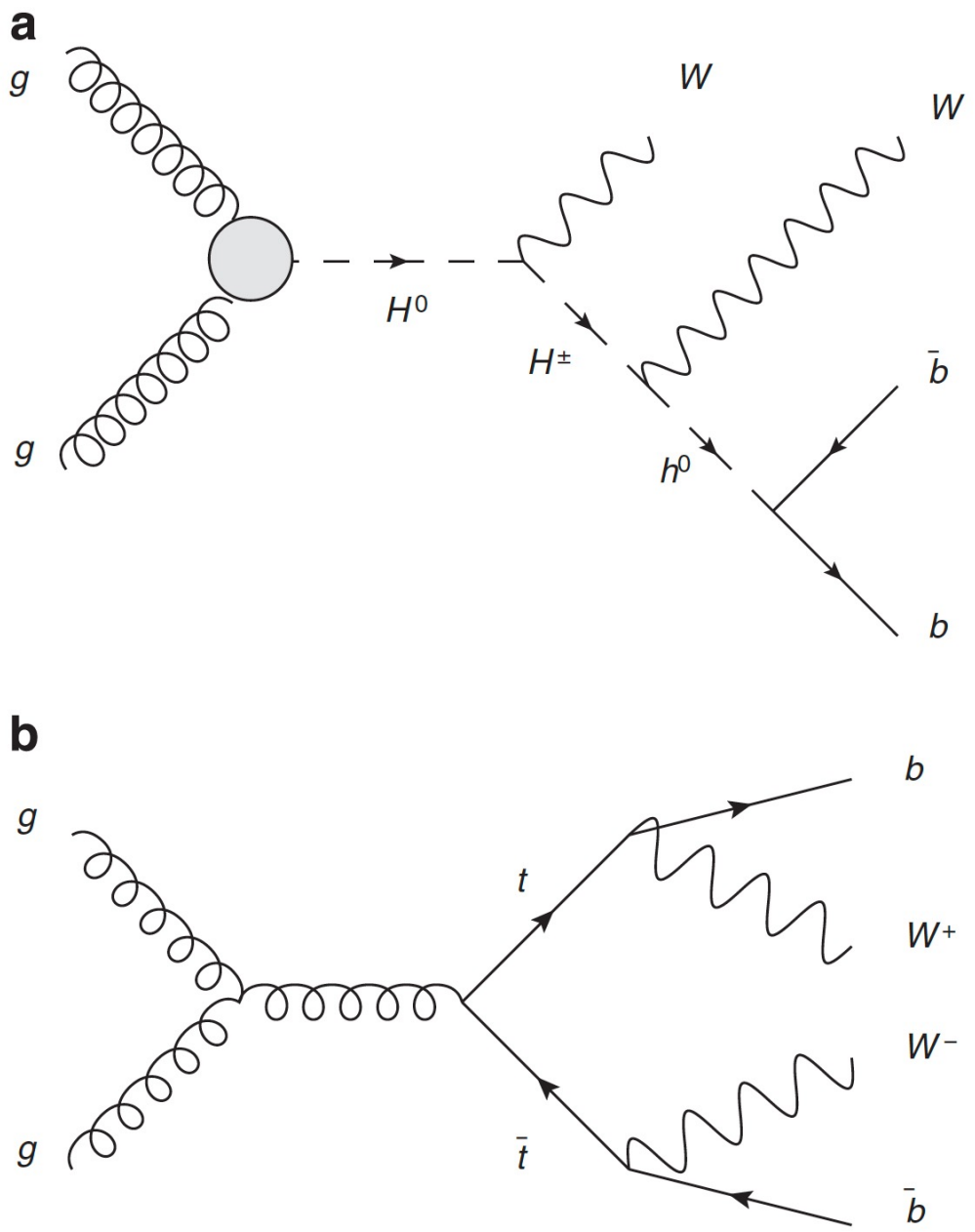


Diagram involving new exotic Higgs bosons



SM background with same final state!

https://doi.org/10.1038/ncomms5308



Assumptions and pre-selections:

Exotic Higgs boson mass:
$m(H^0) = 425$ GeV, $m(H^{\pm}) = 325$ GeV

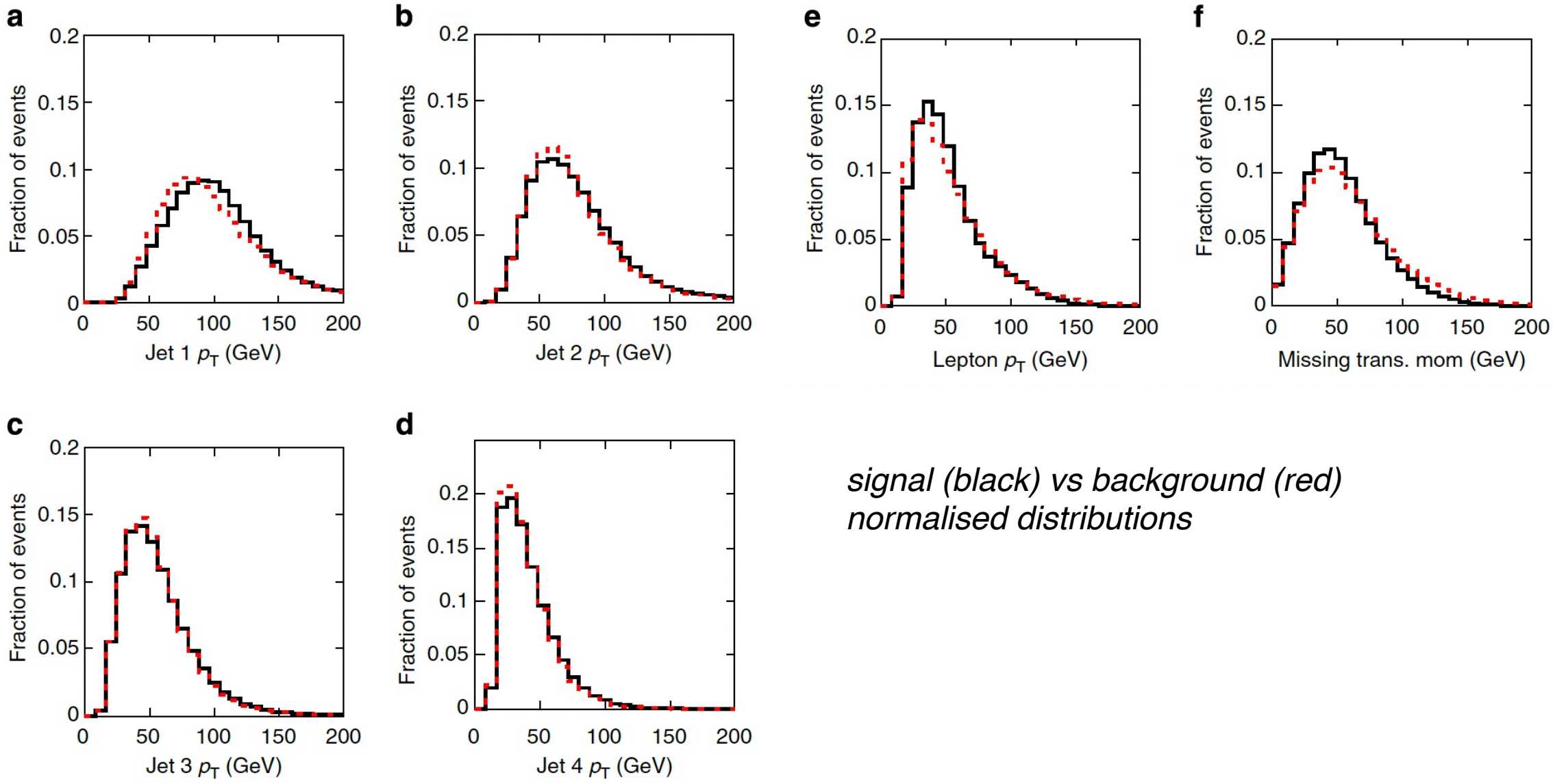
Focus on semi-leptonic decay mode, where one W decays into lepton + neutrino, the other W decays to two jets.
Therefore, the final state is: $W^+W^-b\bar{b} \rightarrow \ell\nu jjb\bar{b}$

Pre-selections:

- exactly **one electron or muon**, with $p_T > 20$ GeV and $|\eta| < 2.5$;
- at least **four jets**, each with $p_T > 20$ GeV and $|\eta| < 2.5$;
- b-tags on at least two of the jets.

https://doi.org/10.1038/ncomms5308

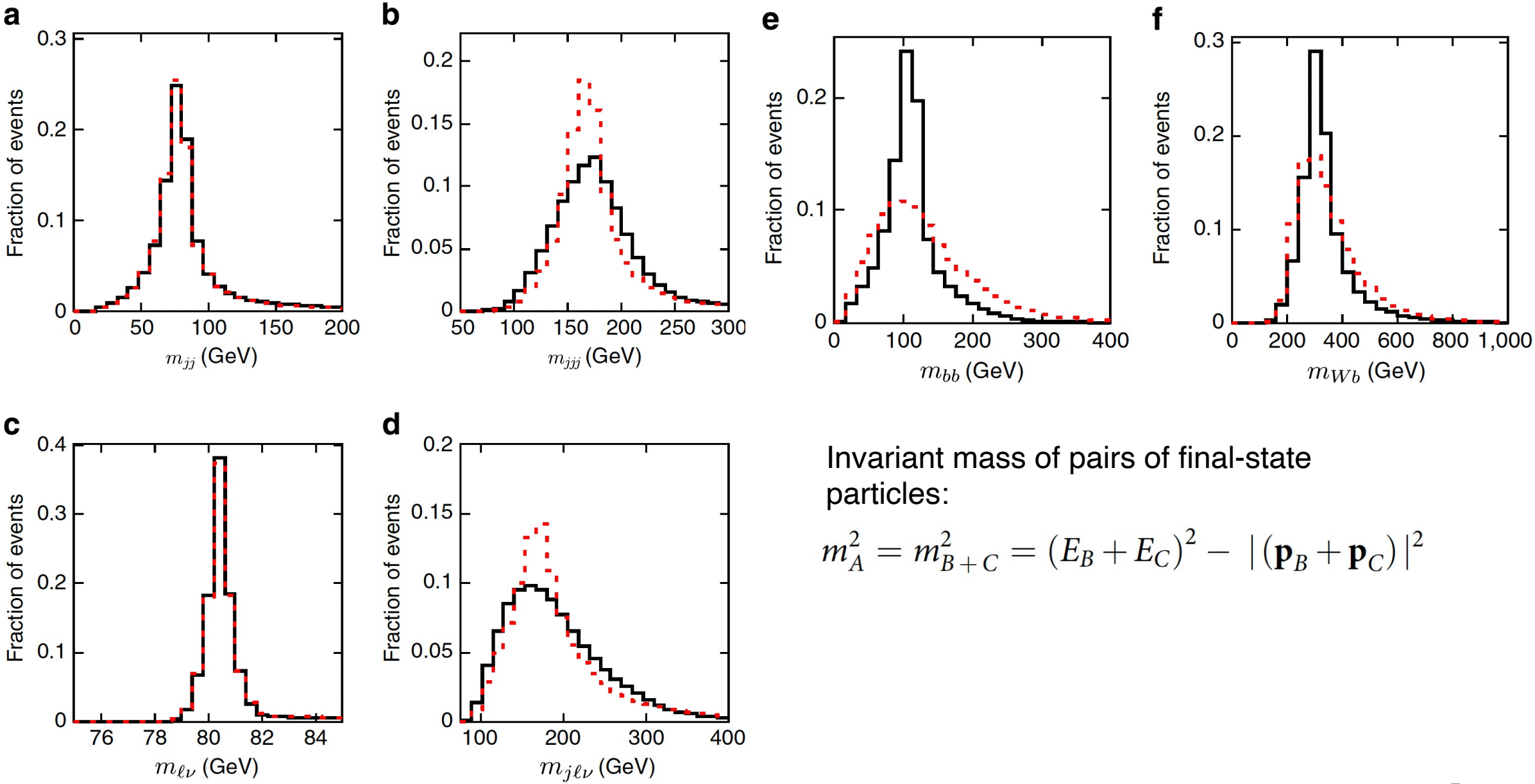
Low-level features:



signal (black) vs background (red) normalised distributions

https://doi.org/10.1038/ncomms5308

High-level features:



Invariant mass of pairs of final-state particles:

$$m_A^2 = m_{B+C}^2 = (E_B + E_C)^2 - |(\mathbf{p}_B + \mathbf{p}_C)|^2$$

https://doi.org/10.1038/ncomms5308

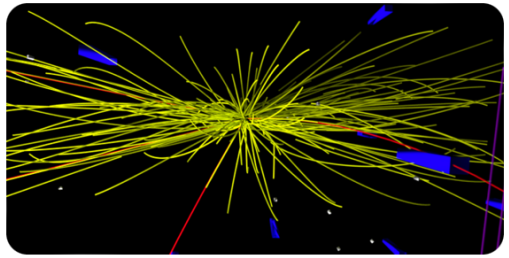
Hands-on (for real, this time)

SLDSRT · COMMUNITY PREDICTION COMPETITION · 3 MONTHS AGO

Late Submission

Higgs Boson Detection 2024

Searching for exotic particles in high-energy physics with machine learning



Overview | Data | Code | Models | Discussion | Leaderboard | Rules

Dataset Description

You will need the following files:

File descriptions

- **train.csv** - the training set. A 50k by 29 matrix, with a header. The first column is the label, while the other columns are the 28 features.
- **test.csv** - the test set. A 50k by 28 matrix, with a header.
- **sampleSubmission.csv** - a sample submission file in the correct format. First column is id, followed by prediction in range [0,1].

Files
3 files

Size
74.3 MB

Type
csv

License
MIT

Hands-on (for real, this time)

Starting-point notebook:
https://github.com/fsimone91/course_ml4hep/blob/2024/notebooks/2024/7_ML_competition.ipynb

Data: https://archive.ics.uci.edu/dataset/280/higgs

Rules:
- Use 200k examples
- Don’t use external data
- The evaluation metric is the Area Under the ROC curve