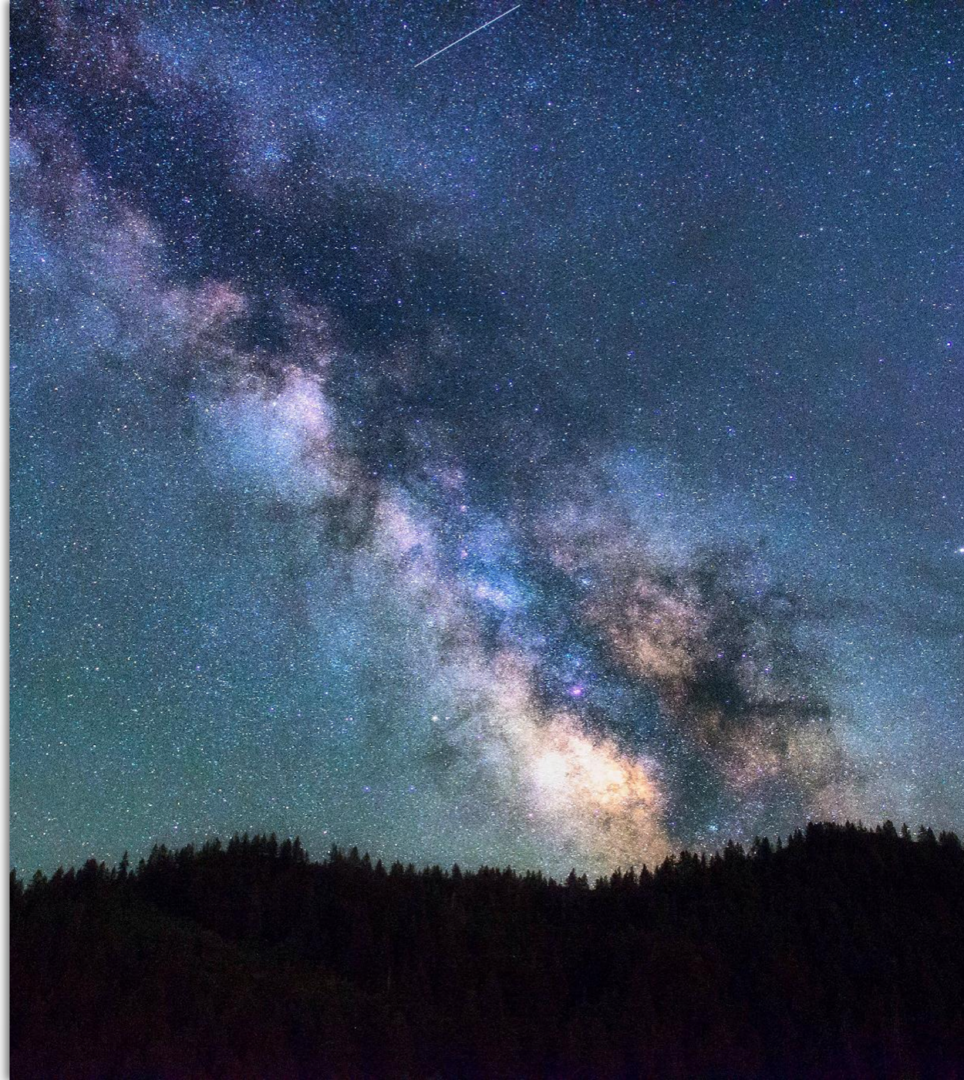
Dark Machines

http://darkmachines.org/ @dark_machines

Bob Stienen
ATLAS Machine Learning Workshop
November 11th, 2019



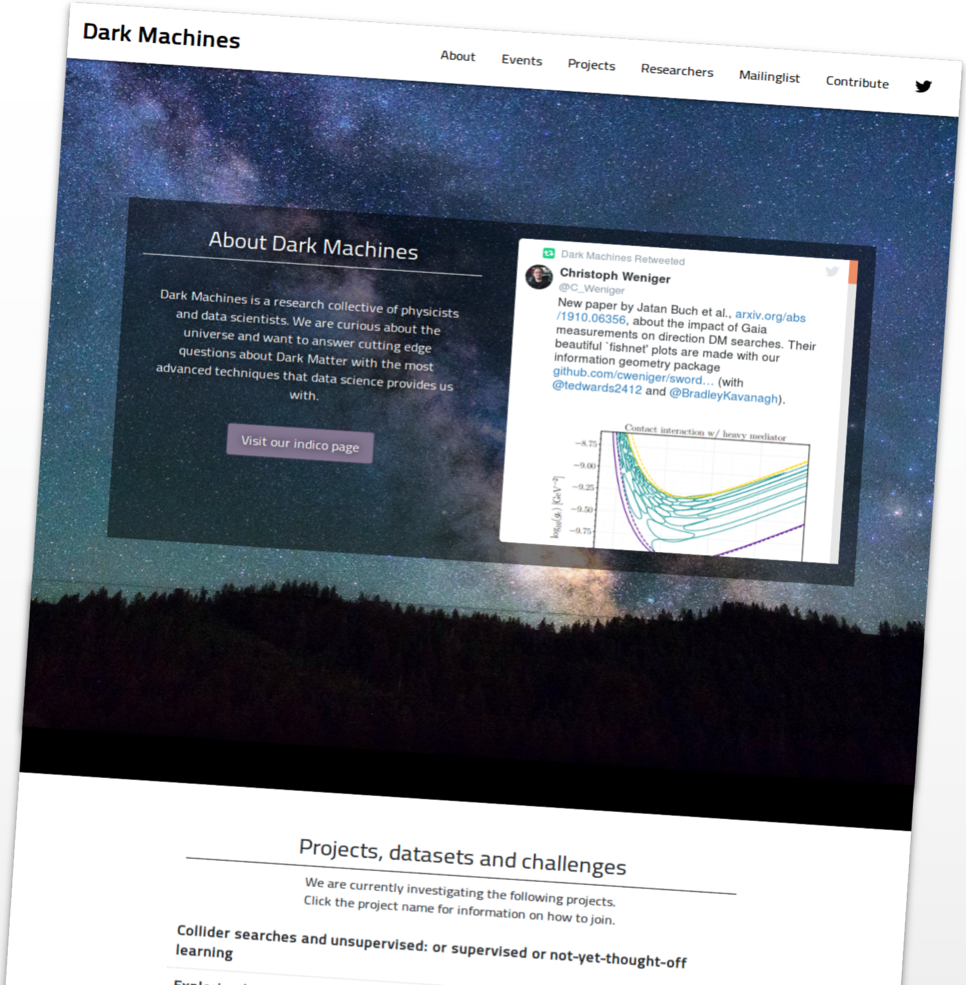
The goal of Dark Machines

Exploit recent advances in Machine Learning to help in the search for Dark Matter

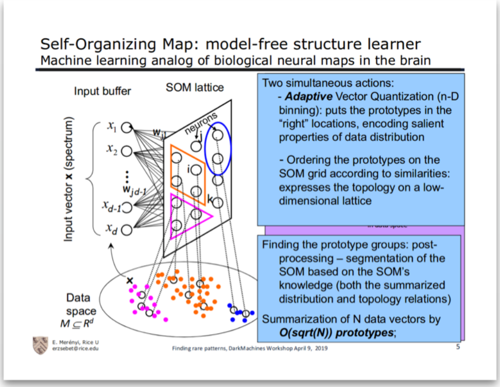
Who is Dark Machines

Online research collective with scientists from range of fields

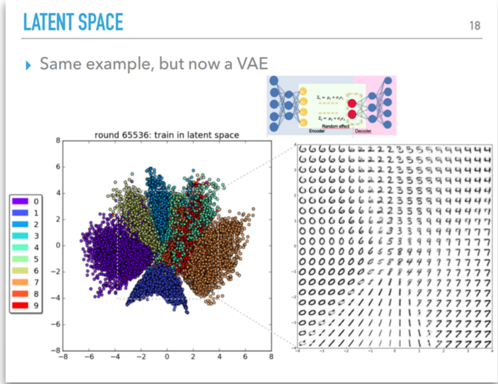
Many are ML experts, all are ML enthusiasts



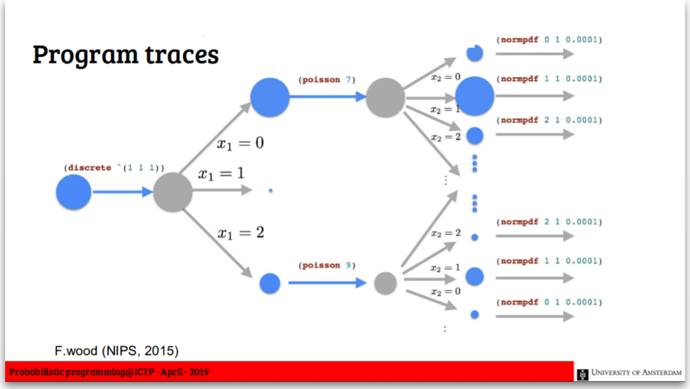
Flashes from our previous workshop



Self-organising maps



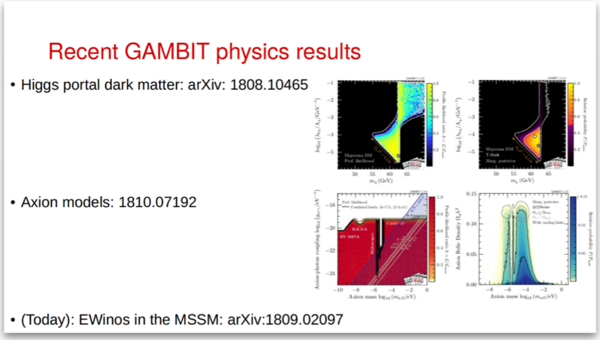
Variational Auto-encoders



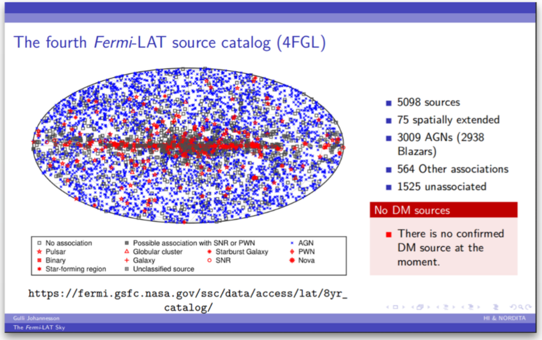
Probabilistic Programming

http://indico.ictp.it/event/8674/session/155/contribution/1120/material/slides/
http://indico.ictp.it/event/8674/session/155/contribution/1121/material/slides/
http://indico.ictp.it/event/8674/session/153/contribution/1109/material/slides/

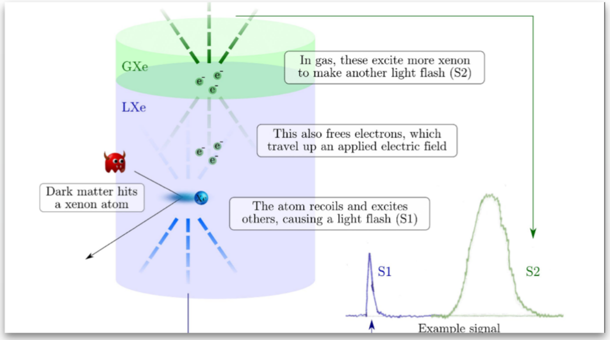
Flashes from our previous workshop



Gambit



Fermi-LAT



XENON

http://indico.ictp.it/event/8674/session/155/contribution/1123/material/slides/
http://indico.ictp.it/event/8674/session/153/contribution/1110/material/slides/
http://indico.ictp.it/event/8674/session/158/contribution/1141/material/slides/

What does Dark Machines do

Yearly workshop

2018 @ Lorentz Center, Leiden, The Netherlands

2019 @ ICTP, Trieste, Italy

2020 @ CERN (April 27th - May 1st)

Network of 216 members

Regular meetings

Currently 8 research lines (a.k.a. challenges)

Examples of challenges

Particle track reconstruction with ML

Sydney Otten, Michela Negro and Fabian Gieseke

Unsupervised collider searches

Andrea de Simone, Maurizio Pierini and Amir Farbin

Inclusive analysis of Fermi-LAT point sources

Luc Hendriks and Gabrijela Zaharijas

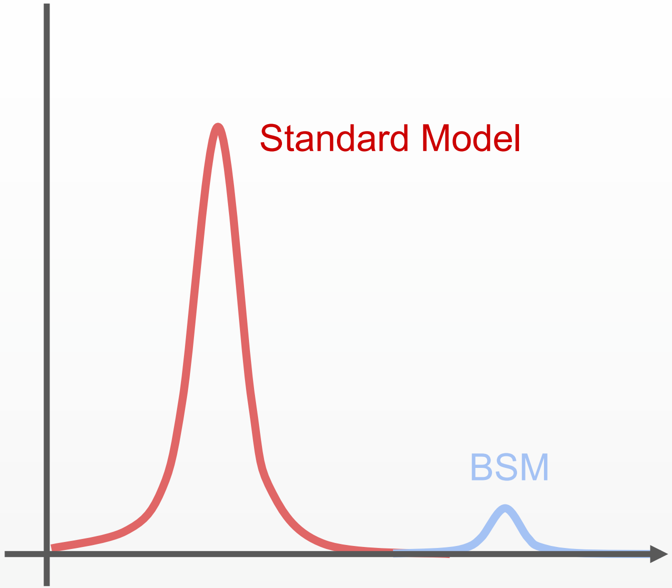
Exploring high dimensional parameter spaces

Martin White and Joaquin Vanschoren

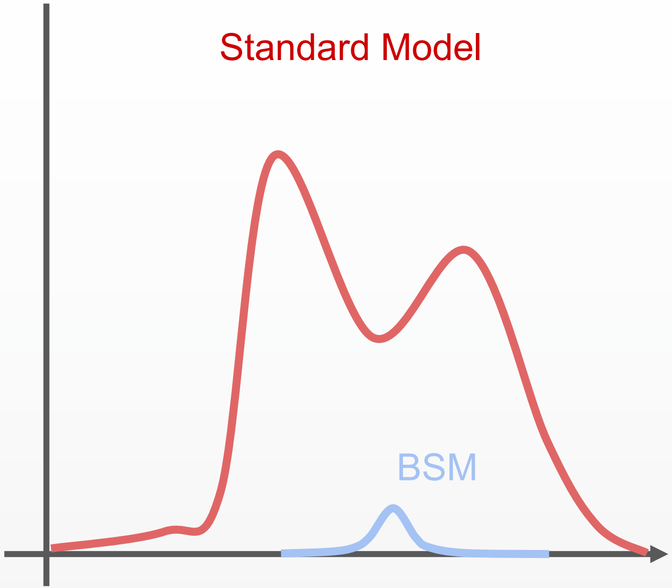
Library of trained models

Sanmay Ganguly and Bob Stienen

Unsupersived Collider Searches



Relatively easy to find



Relatively hard to find

But… both are important!

Unsupersived Collider Searches

Generated Standard Model events and events from BSM models

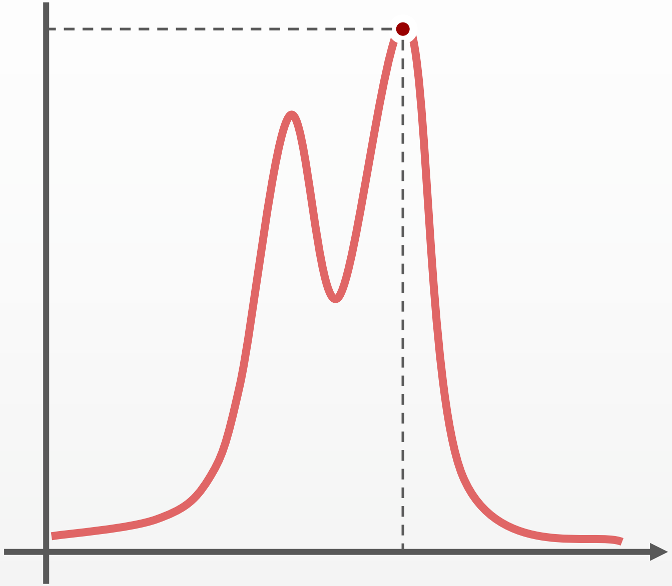
$O(10^9)$ events generated in order to archieve 10 fb^{-1}

Challenge: "Find the BSM physics"

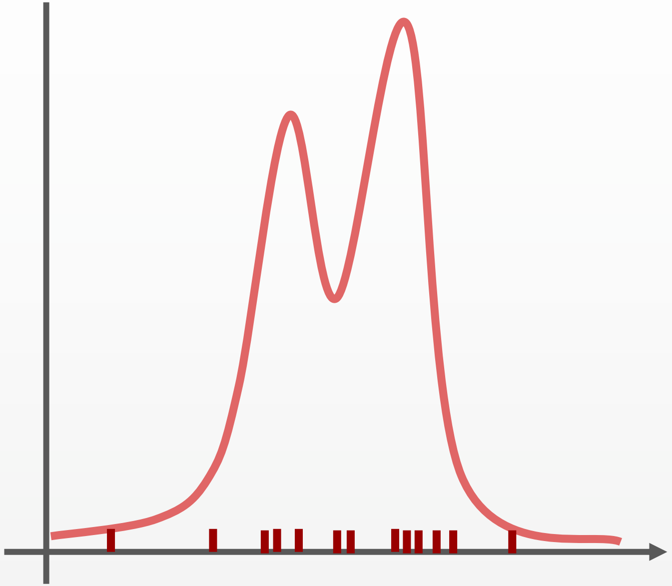
Open to community, everyone can participate. Join with your favourite algorithm!

Exploring high dimensional parameter spaces

Sampling for optimisation



Sampling for posterior estimation



Exploring high dimensional parameter spaces

Naive approach of random sampling is flawed:

- Concentrates points near the edge of the parameter space for high dimensional problems, leading to biased inference
- Easily misses high-likelihood regions in strongly peaked spaces

So more sophisticated algorithms are needed! But...

Exploring high dimensional parameter spaces

Sampling for optimisation	Sampling for posterior estimation
Random sampling	Rejection sampling
Stochastic gradient descent	MCMC
Genetic algorithms	PyBAMBI
Particle swarm optimisation	MultiNest

Which one should you choose?

Exploring high dimensional parameter spaces

Systematic investigation of sampling algorithms to find which algorithms work best in which situations

Python package written to do this (found on github, Travis CI and unit tests available, additions more than welcome)

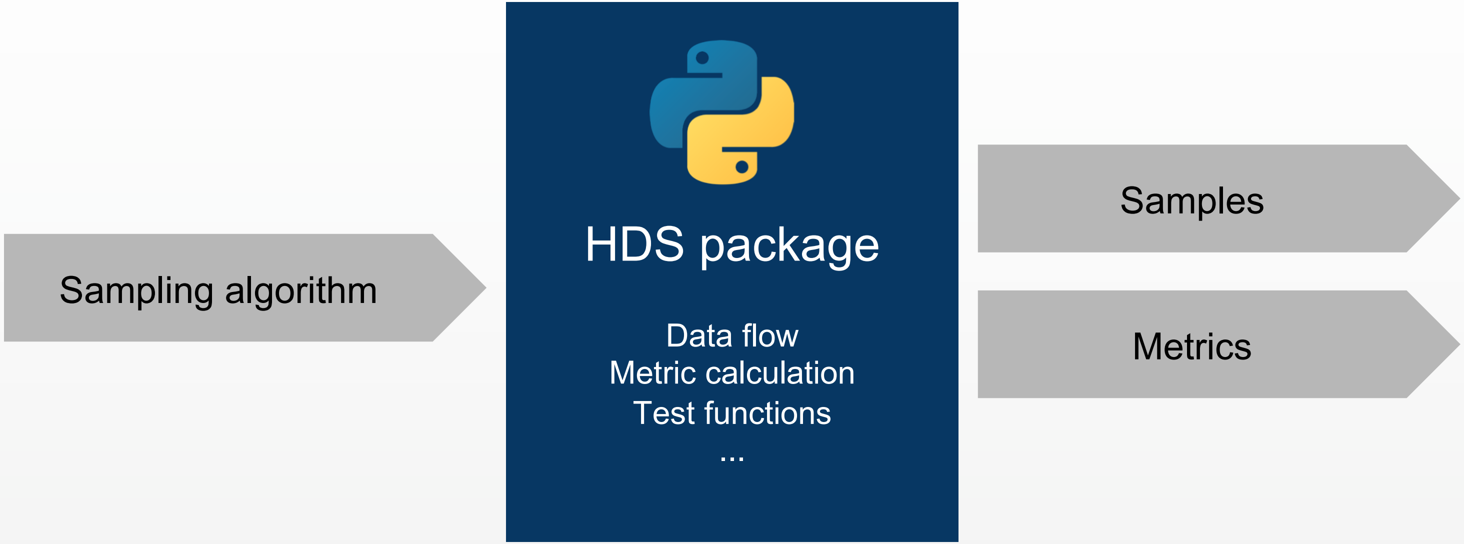
Structure of package

Package is finished developing, algorithms are being added

Results will be available hopefully soon

Exploring high dimensional parameter spaces

Systematic investigation of sampling algorithms to find which algorithms work best in which situations



Exploring high dimensional parameter spaces

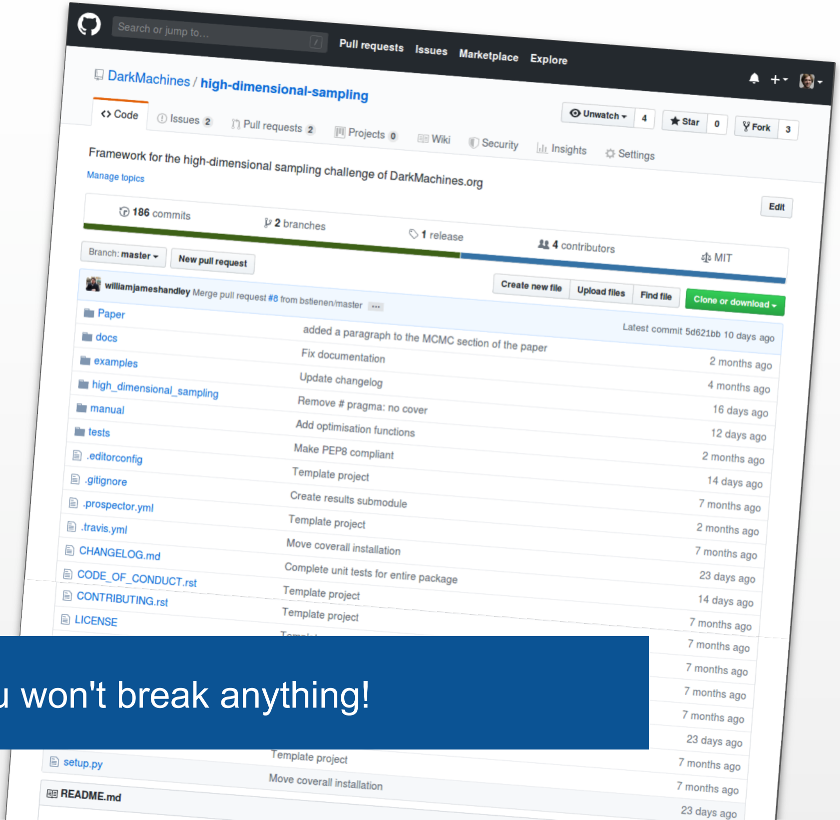
Systematic investigation of sampling algorithms to find which algorithms work best in which situations

Open source code

Development through GitHub and pull requests

Travis CI integration

Unit tests with almost 100% coverage



In other words: Join us! You won't break anything!

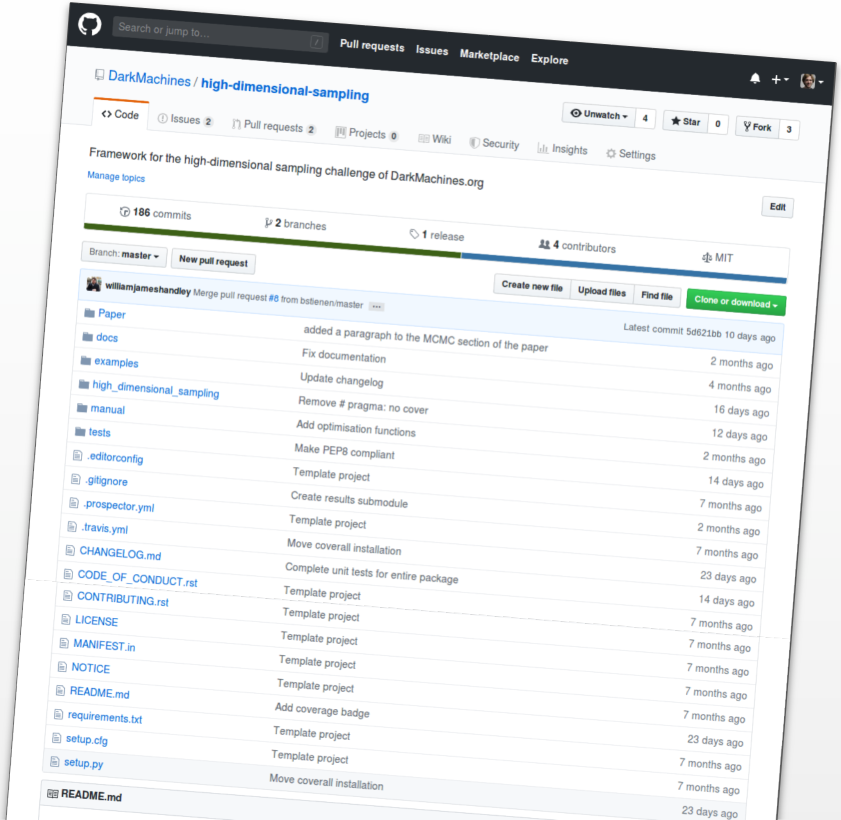
Exploring high dimensional parameter spaces

Systematic investigation of sampling algorithms to find which algorithms work best in which situations

Core of the package is finished

Algorithms are being added to the package

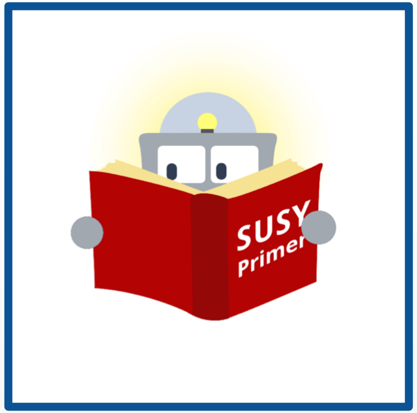
Stay tuned for results!



https://github.com/DarkMachines/high-dimensional-sampling

Library of trained models

Example of a trained model made public

19-dimensional pMSSM model point →  → ATLAS exclusion

DOI: 10.1140/epjc/s10052-017-4814-9

Library of trained models

Communicating trained models is not standard yet, even though they become more and more integral part of analyses in experiment and theory

Reasons

Using the models requires understanding machine learning (or ML libraries)

Model might be library dependent

Each model needs its own manual

Easy mistakes easy to make
e.g. querying a prediction outside of training box

Library of trained models

Facilitate easy communication of models, so that trained models can be used by external researchers

Easy to use for both implementer and end-user

Consistent interface for all trained models, reducing learning curve

Collection of trained models in a searchable library

Library framework

Handles post- and preprocessing of data following rules of implementer

Server-client structure to allow easy implementation in HEP workflows

Understandable warnings and errors

Library of trained models

Facilitate easy communication of models, so that trained models can be used by external researchers

Currently in the pipeline

Electroweak cross-section in pMSSM19

W' and Z' cross-sections

Doublet Higgs model

Loop integrals (input momenta, masses)

Global fits Gambit Zenodo data

Higgs model with many nuisance parameters

Example relic density trained on observables

Reconstruction efficiencies of LLSps

+ your model?

Dark Machines

Exploit recent advances in Machine Learning
to help in the search for Dark Matter

http://darkmachines.org/ @dark_machines

Workshops

Network

Exciting challenges

Infrastructure for collaboration

Subscribe to the newsletter