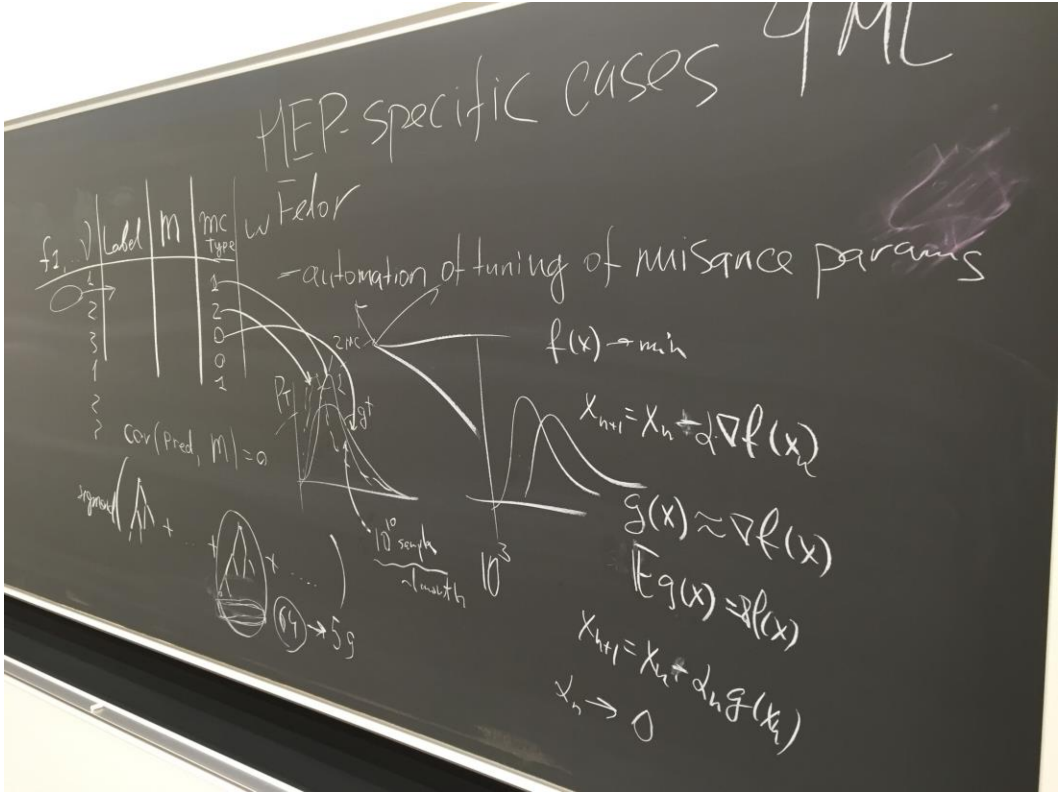
Open-Space discussions summary

Andrey Ustyuzhanin, David Rousseau, Alexander Baranov

HEP-specific cases of ML

- Nuisance parameter-based MC generative models
- How do you define quality criteria for generative models?
    - Make nuisance parameter additional label
    - Reduce mutual information between prediction and this label



Future ML & HEP challenges

We discussed mainly about the tracking machine learning challenge, where 3D points corresponding to high-lumi LHC simulation in ATLAS or CMS would be released.

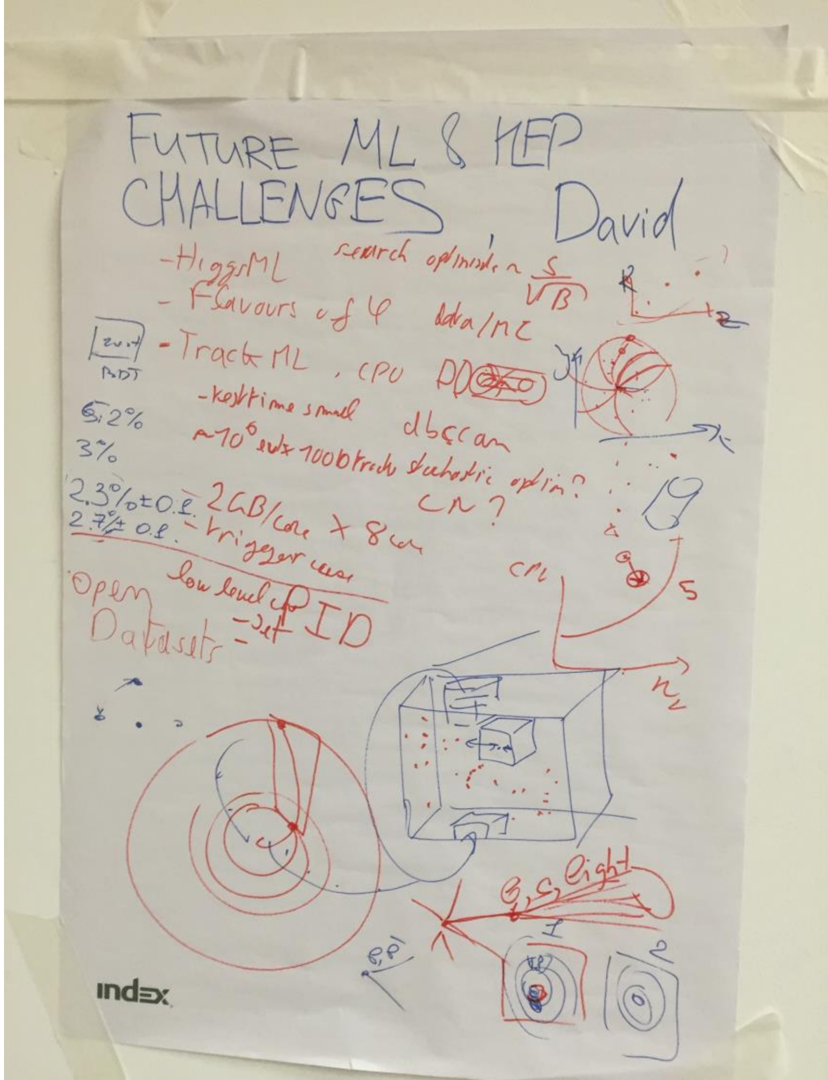
Possibly a **3D CNN** could do it. Once unfold in direction space, the points are nearby.

GPU ? probably not since our commodity are intel CPU (except possibly for trigger).

Other possible challenges:

- access to low level data, for example 3D calo info for particle ID,
- tracks for jet tagging, etc….

permanent accessibility to public dataset



Advanced classification

systematics aware training: how to train a classifier minimizing the systematics error (in addition to the statistical one ?).

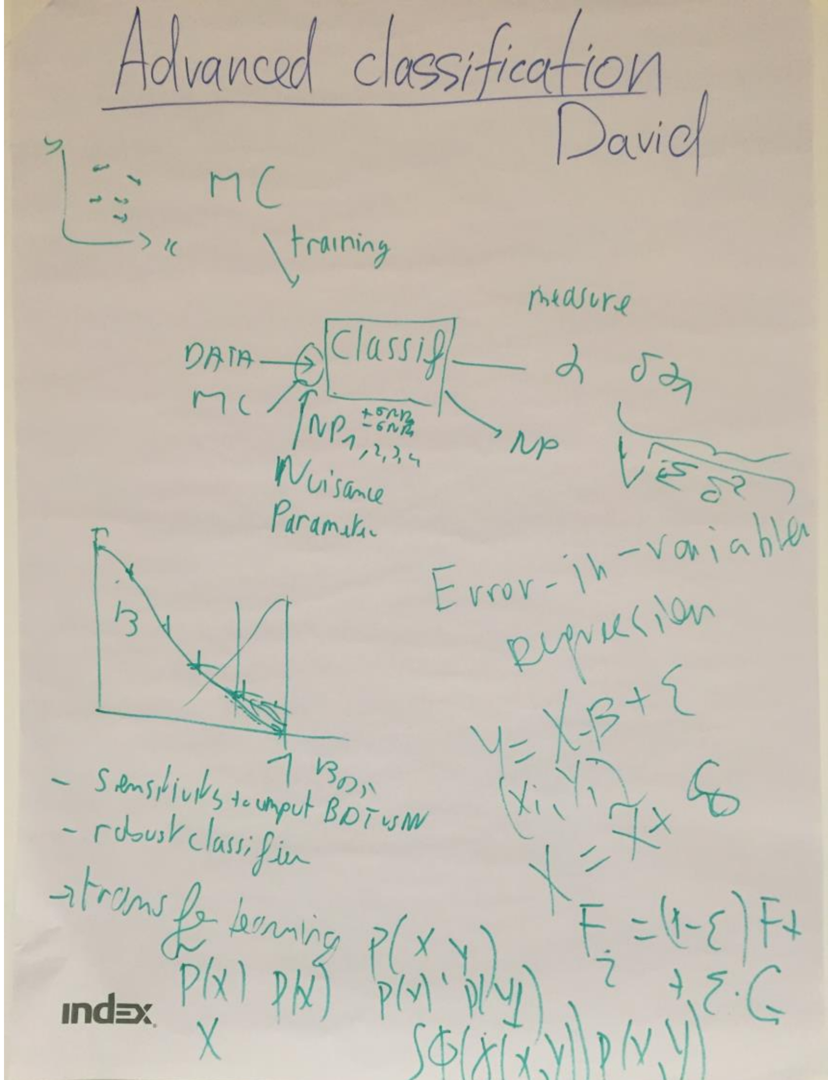
Typically systematics are evaluated by redoing the measurement changing one nuisance parameter NP1, and observing the change on the measurement delta 1. All the NPi are changed one by one, the total systematic is sqrt (Sigma delta_i^2) (this is assuming Gaussian, uncorrelated, distribution) (not always true but let's assume this). Let's assume also the range of change of the NPi is perfectly known.

So the question is how to make the training aware of the systematics, to minimise them.

- Typically the NPi quantify the data vs MC differences, so there is a connection with transfer learning. Need to evaluate the uncertainty on the transfer learning.

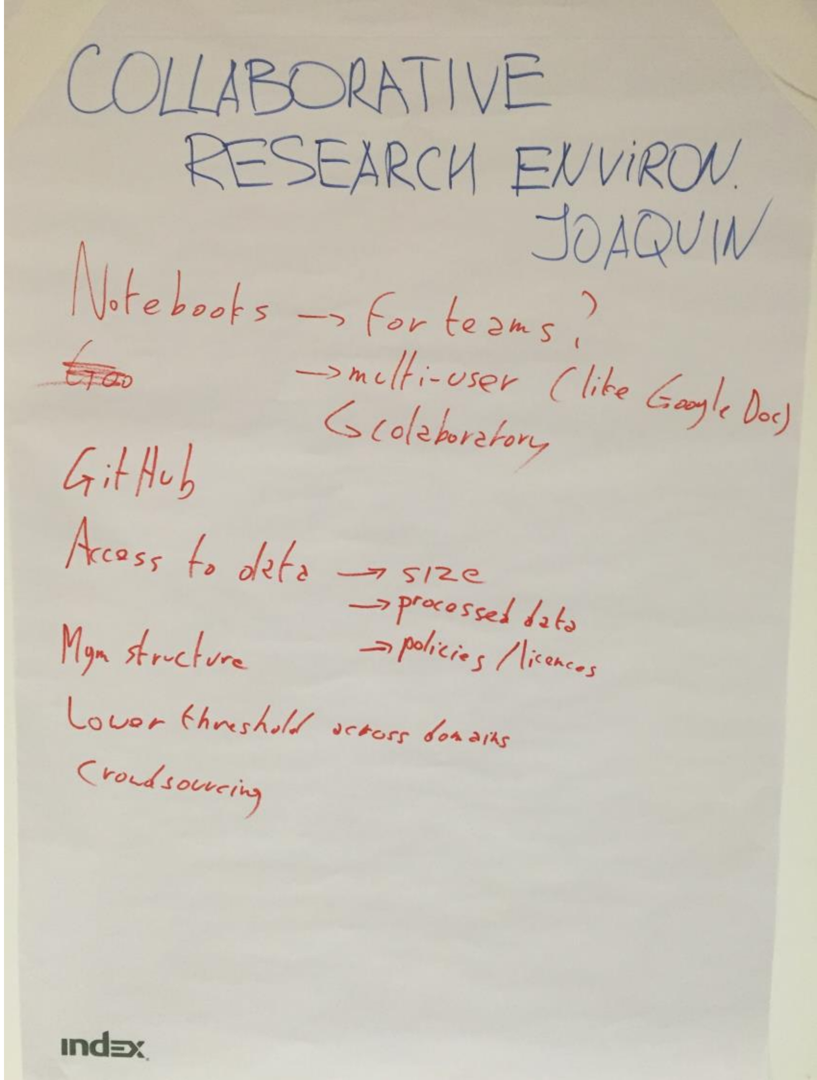
 - one can have the NPi as input to the training (similar to parametrised learning, see recent Baldi/Cranmer paper), which would be equivalent to retrain with the NPi changes. But what to do then ? Can be better when NPi is actually constrained by the analysis.

 - there are techniques to map one multidimensional P(X) onto another one Q(X) (like data and MC) (without reweighting) to be studied



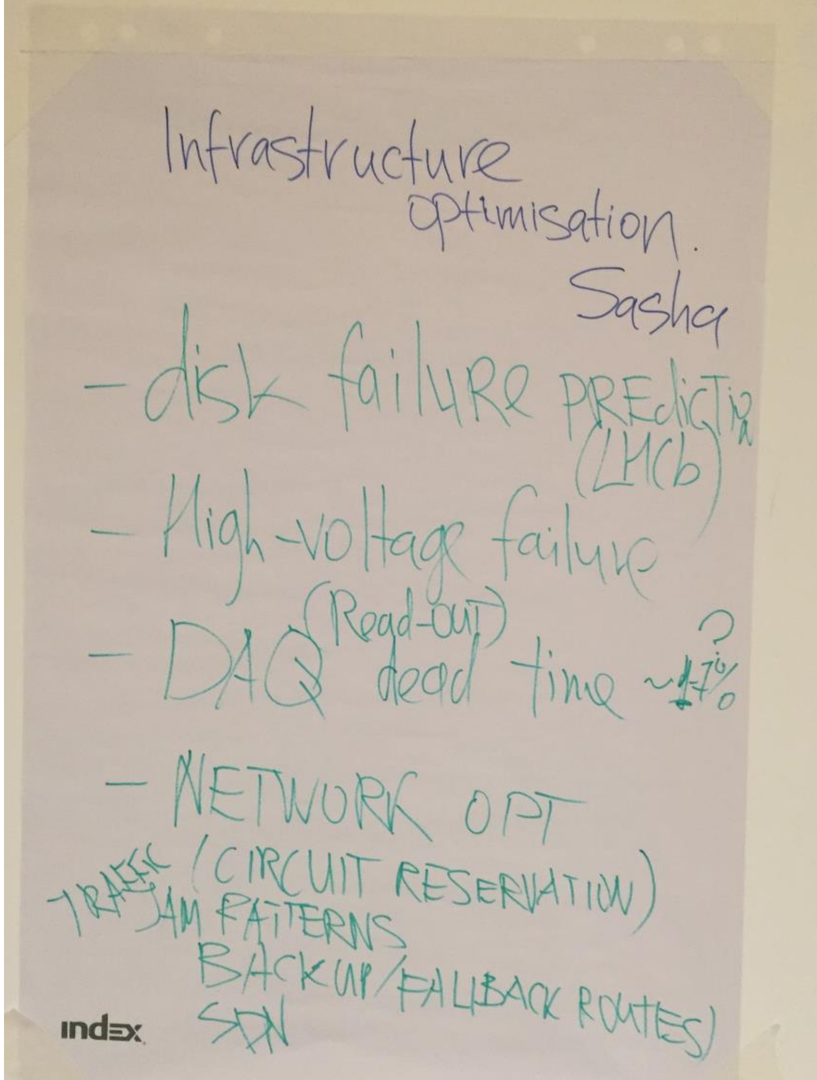
Collaborative research environment

- Access to data!
  - Size
  - Processing
  - Policies / licenses
- Environment for collaboration
  - Jupyter
  - Github
  - OpenML
  - Crowdsourcing



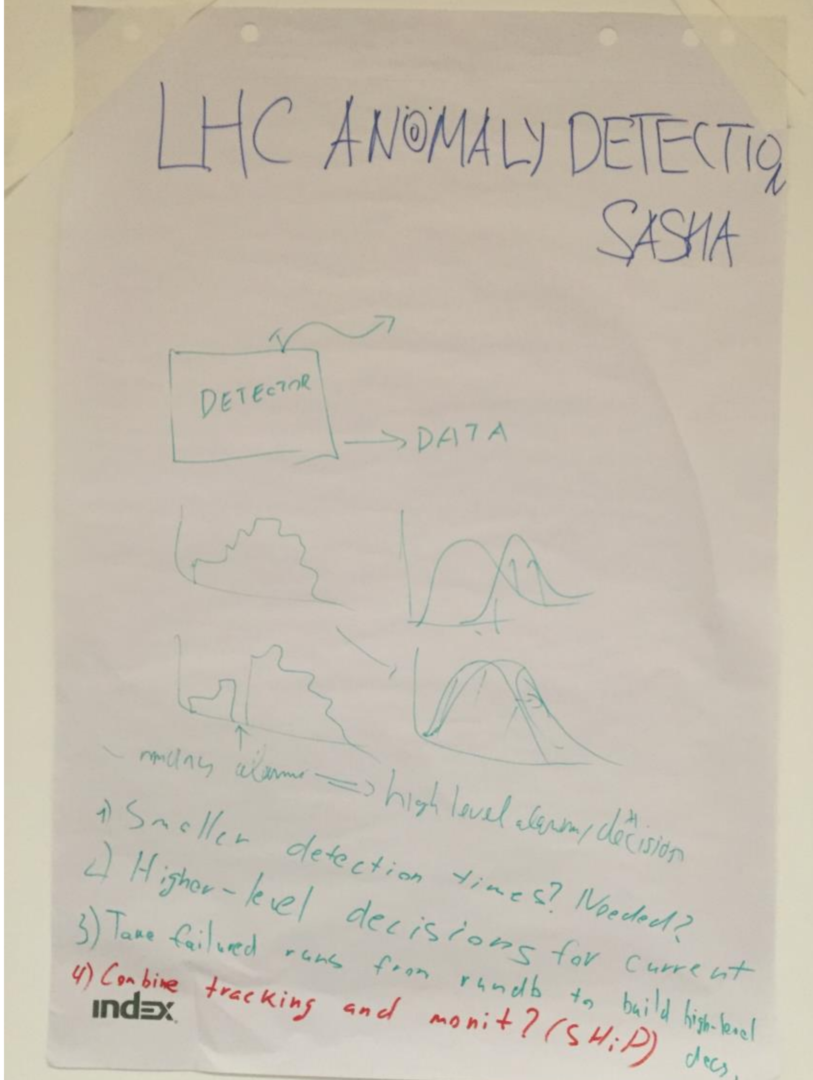
Infrastructure optimization

- Run-time optimization
  - Traffic optimization
- Offline optimization
  - GRID disk storage
- Anomaly detection & prediction
  - Disk failure
  - High-voltage
  - DAQ dead-time
  - …
- Steps
  - identify quality metrics so humans could be unhooked
  - Data collection?
  - Actions?
  - Models for the prediction? Simulation
  - Prototype & fail early



Anomaly Detection

- Current anomaly detection is time-fixed (runs, lumi-sections,…)
- Do we need to have finer-grained anomaly detection?
- What are higher-level user-friendly indicators based on current histogram-comparison rules?
- Analyze histogram database for ‘bad’ runs.
- Tracking and failure detection synthesis was proposed. E.g. if you see some channel is always [un-]triggered, then it probably failed.



Overall

- New conversation dynamics
  - Dense
  - Beer-chatting but with clear focus (no chatting) without beer
  - https://en.wikipedia.org/wiki/Open_Space_Technology
- Questionable value until you take part in it
- Ensemble of weak learners is stronger than single strong learner

  (motivation for RandomForest algorithm)
- Screenshots to remember on the following slides

REAL-TIME DATA PROCESSING. ALISON

How to parallelize workloads
CPU → GPU.

Compression
Reducing complexity

AUTO ML TOOLS
Alison

- accessibility of tools
- definition | language model to share deep learning models

"NVIDIA Deep Learning Course"

HEP-specific cases 4 ML
w Fedor

- automation of tuning of nuisance params

f1 ...	Label	m	mc Type
	1		1
	2		2
	3		0
	1		0
	?		1
	?		

Cov(pred, m) = 0

$f(x) \to \min$

$x_{n+1} = x_n + \alpha \nabla f(x_n)$

$g(x) \approx \nabla f(x)$

$E\, g(x) = \nabla f(x)$

$x_{n+1} = x_n + \alpha_n g(x_n)$

$x_n \to 0$

10^6 samples, smooth

10^3

Improving Classifiers

- CV (Nested CV)
- regularization
- test vs control set
- training set

LHC ANOMALY DETECTION

SASHA

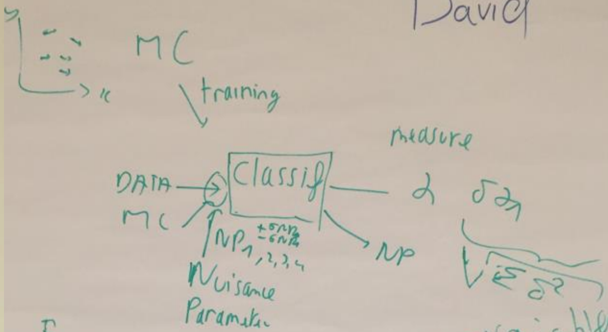
~ ... ⟹ high level alarm/decision

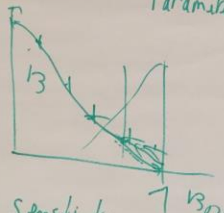
1) Smaller detection times? Needed?
2) Higher-level decisions for current
3) Take failured runs from rundb to build high-level decs.
4) Combine tracking and monit? (SHiP)

index

Advanced classification

David





- Sensitivity to input BDT vs NN
- robust classifier

→ trans for learning

$P(X)$ $P(N)$

X

$p(x, y)$

$p(v)$ $p(y)$

$\int \Phi(\gamma(x, y))\, p(v, y)$

Error-in-variables regression

$y = X \cdot \beta + \varepsilon$

(x_i, y_i)

$X = X^*$

$F_i = (1 - \varepsilon) F + \varepsilon \cdot G$

index

COLLABORATIVE RESEARCH ENVIRON.

JOAQUIN

Notebooks → for teams?

~~ETGO~~ → multi-user (like Google Doc)

↳ colaboratory

GitHub

Access to data → size

→ processed data

→ policies / licences

Mgm structure

Lower threshold across domains

Crowdsourcing

index

UNSUPERVISED LEARNING

- Do we need pattern mining for HEP or other physics related tasks?
  - $l_1\ l_2\ l_3$: + $l_1 - l_2 - l_3$
  - $l_1\ l_2\ l_4$: + frequent rare Dendrue Moore
  - $l_2\ l_4$: −
  - l_2 : 3 $l_1 l_2$: 2 A → B bread → milk 100%
- SOM, RBF RBF = exp(−γ‖x − x_l‖²)
- Representation learning
  - k-means
  - word2vec
  - auto-encoders

FUTURE ML & HEP CHALLENGES, David

- HiggsML search optimisation $\frac{S}{\sqrt{B}}$
- Flavours of φ data/MC
- TrackML, CPU
  - keep time small dbscan
  - ~10^6 evts 1000 tracks stochastic optim? CN?
- 2GB/core × 8 cores
- trigger ...
- low level PID
- Open Datasets

BDT 6.2% 3% 2.3% ± 0.8 2.7% ± 0.8

b, c, light

Infrastructure optimisation

Sasha

- disk failure prediction (LHCb)
- High-voltage failure
- DAQ (Read-out) dead time ~1%?
- NETWORK OPT (CIRCUIT RESERVATION)
  - TRAFFIC JAM PATTERNS
  - BACKUP/FALLBACK ROUTES)
  - SDN

index