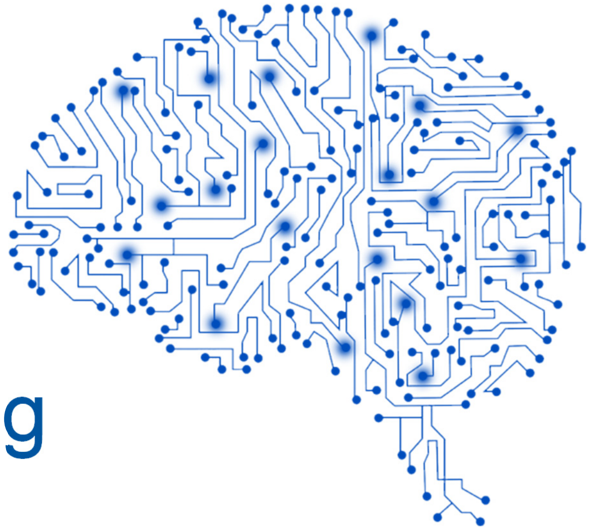
CERN



Application of Machine Learning to Beam Diagnostics

Elena Fol
CERN, Goethe-University Frankfurt

39th International Free Electron Laser Conference, 26 – 39 August 2019, Hamburg

Outline

I. Introduction to Machine Learning
   - General application fields
   - Relevant concepts and definitions

II. Machine Learning and Beam Diagnostics
   - Motivation and overview
   - Detection of Beam Position Monitors faults
   - Optics correction using regression models
   - Further examples

III. Conclusion and recommendations

Part I. Introduction to Machine Learning

Teaching machines to learn from experience

- Tasks that are extremely easy and obvious for us are difficult to program in traditional ways
- Impossible to learn every possible rule to perform a task
  - ➢ learn from examples instead

Teaching machines to learn from experience

- Tasks that are extremely easy and obvious for us are difficult to program in traditional ways
- Impossible to learn every possible rule to perform a task
  - ➢ learn from **examples** instead

Teaching machines to learn from experience

- Tasks that are extremely easy and obvious for us are difficult to program in traditional ways
- Impossible to learn every possible rule to perform a task
  - ➢ learn from **examples** instead

→ Cat?

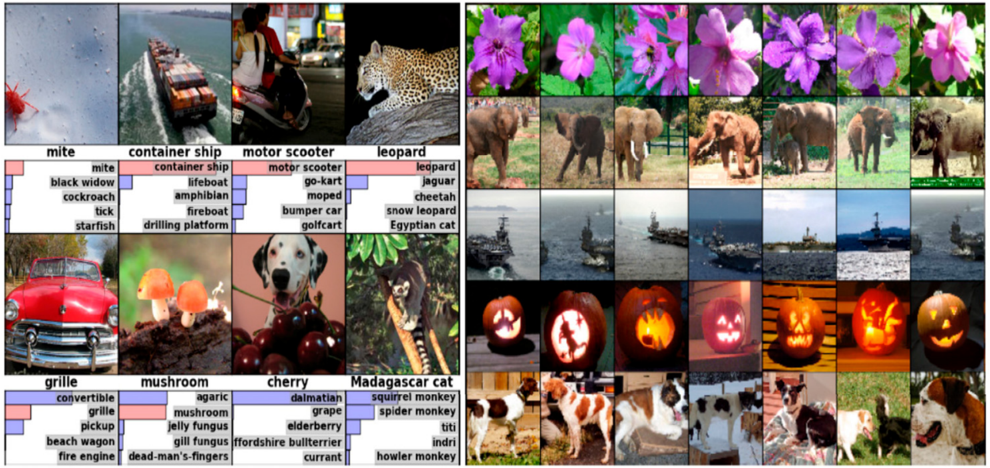
Machine Learning is extremely successful in many different fields:

- Computer vision
- Speech recognition
- Natural language and text processing
- Face recognition
- Financial market analysis, risk prediction
- Search engines
- Medical diagnostics
- Transactions fraud detection
- Recommendation engines, advertising
- Robotics, automation
- Video games
- Self-driving cars

MNIST handwritten digits dataset

http://yann.lecun.com/exdb/mnist/

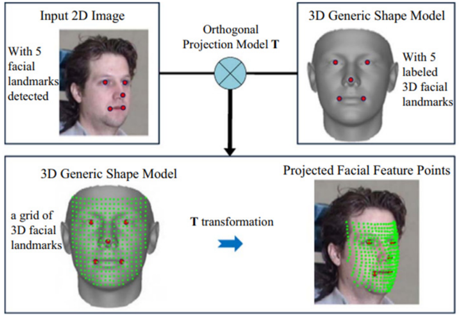
Y. LeCun, et.al, "Gradient-based learning applied to document recognition"



The ImageNet project

- Visual objects recognition (up to 78% accuracy on 1000 object classes)

G. E. Hinton et.al, "ImageNet Classification with Deep Convolutional Neural Networks"



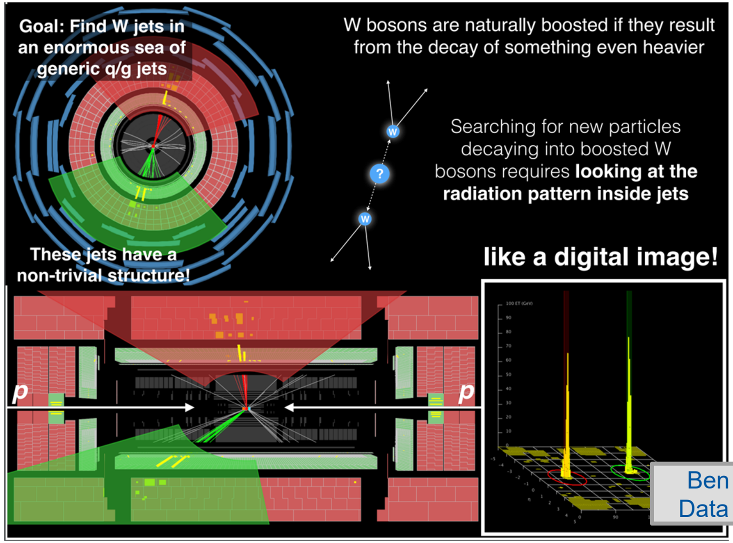
Face recognition and reconstruction

- Automatic detection of semantic regions
- Specific "layers" are sensitive to certain regions (e.g. eyes, nose, lips)

D. Changxing, T. Dasheg, "Pose-invariant face recognition with homography-based normalization"

AlphaGo from Google

- First match against Go European champion in 2015, 5:0 for AlphaGo
- In 2017 AlphaGo surpassed the performance of its previous versions and became the strongest Go player of all time *



Ben Nachman, CERN
Data Science Seminars

High Energy Physics

- ML is used in dark matter search, jets recognition, particle tracking, neutrino classification, shower simulations

* Silver, David et al. "Mastering the game of Go without human knowledge." *Nature* 550 (2017): 354-359.

Relevant ML concepts and definitions

*"… computer programs and algorithms that automatically improve with experience by **learning from examples** with respect to some class of task and performance measure, **without being explicitly programmed**." **

Supervised Learning

- Input/output pairs available
- Make prediction for unknown input based on experience from given examples

Automatic spam detection, object detection in computer vision, speech recognition, predictive control

Unsupervised Learning

- Only input data is given
- Learn structures and patterns

Anomaly detection, pattern recognition, clustering, dimensionality reduction

Reinforcement Learning

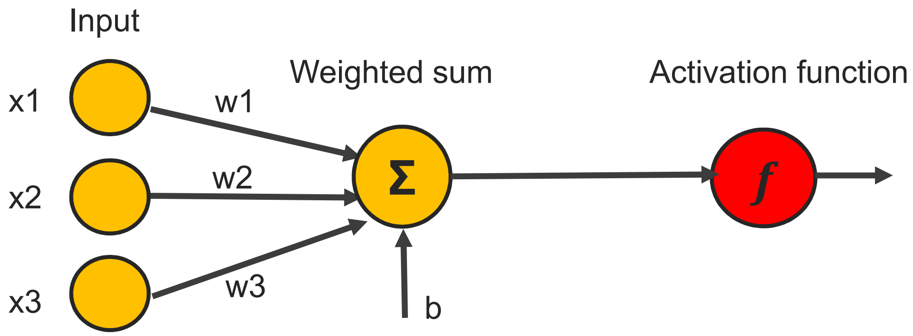
- No training data
- Interact with an environment
- Trying to learn optimal sequences of decisions

Robotics, industrial automation, dialog systems

* Thomas M. Mitchell. Machine Learning. McGraw-Hill, Inc., New York, 1997.

Artificial Neural Networks

- Basic processing unit = **neuron** (or perceptron) with following parameters:
  - **Weights** w from the inputs x_i
  - **Activation function** f
  - Output y of a single neuron: $y = f(\sum x_i w_i + b)$
- Neurons are stacked into **layers**
- Connected layers build a **network**



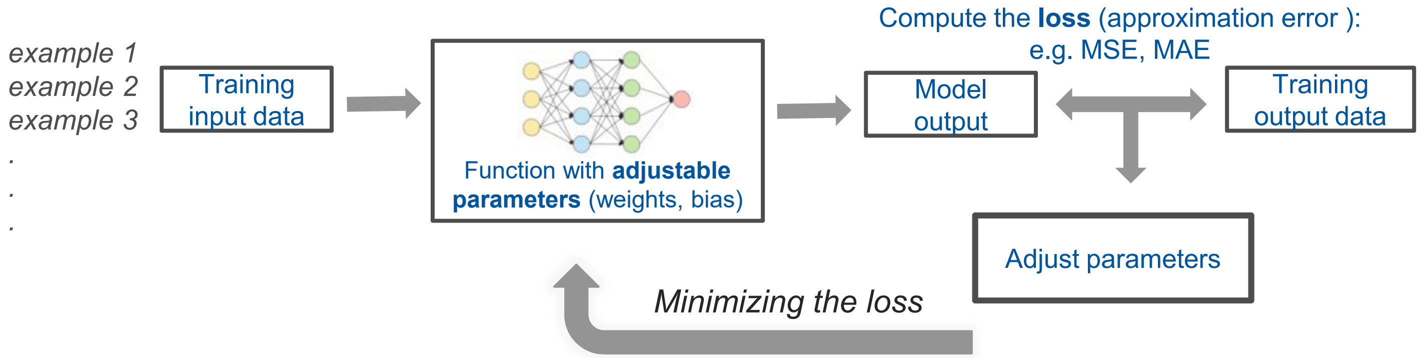
Supervised Learning

- Universal Approximation Theorem

A **simple neural network** including only a single hidden layer can approximate any bounded continuous target function with arbitrary small error.

(Cybenko, 1989, for sigmoid activation functions)

- **How does the learning work in practice?**



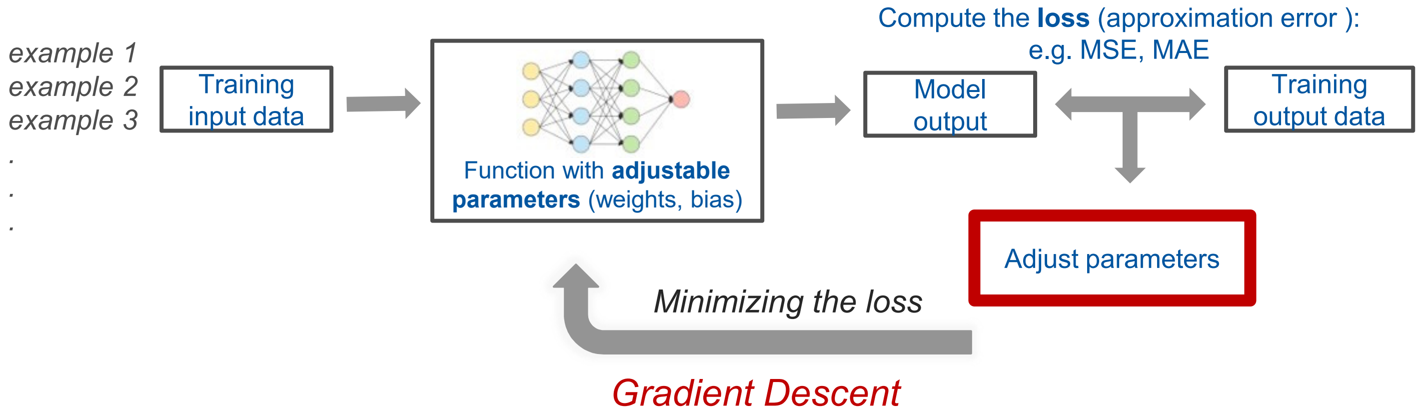
Supervised Learning

- Universal Approximation Theorem

A **simple neural network** including only a single hidden layer can approximate any bounded continuous target function with arbitrary small error.

(Cybenko, 1989, for sigmoid activation functions)

- **How does the learning work in practice?**



ML is more than Neural Networks…

- **Regression and Classification Models:** *resolve correlation between input variables and dependent target variables*
  - Simple Linear Regression, Multivariate Regression, Logistic regression, Support Vector Machine
- **Dimensionality reduction techniques:** *reduce the number of independent variables (features) without significant decrease on prediction accuracy*
  - Independent Component Analysis, Principle Component Analysis, Features Importance Analysis
- **Decision Trees**: *split the input data based on a sequence of variables (thresholds) to estimate the target output value or to separate data points into regions*
  - Ensemble methods: Train several slightly different models and take majority vote/ average of the prediction
- **Clustering:** *grouping or separating data objects into clusters*
  - Identify hidden patterns in the data, similarities and differences

ML is more than Neural Networks…

- **Regression and Classification Models:** *resolve correlation between input variables and dependent target variables*
  - Simple Linear Regression, Multivariate Regression, Logistic regression, Support Vector Machine
- **Dimensionality reduction techniques:** *reduce the number of independent variables (features) without significant decrease on prediction accuracy*
  - Independent Component Analysis, Principle Component Analysis, Features Importance Analysis
- **Decision Trees**: *split the input data based on a sequence of variables (thresholds) to estimate the target output value or to separate data points into regions*
  - Ensemble methods: Train several slightly different models and take majority vote/ average of the prediction
- **Clustering:** *grouping or separating data objects into clusters*
  - Identify hidden patterns in the data, similarities and differences

Machine Learning is about learning from the data, not about application of a particular "intelligent" technique.

Part II. Application to Beam Diagnostics

Motivation

Accelerators

Limitations of traditional optimization and modeling tools?

ML is a powerful tool for prediction and data analysis

Which limitations can be solved by ML with reasonable effort?

Some traditional optimization methods: Simplex, Random walk optimization

- Resolve linear correlations between input parameters and optimization objectives
- Relatively small amount of target parameters

Limitations:

- How to deal with non-linear behavior?
- Required computational resources for large amount of optimization targets
- Objective functions, specific rules and thresholds have to be known

Machine Learning methods can learn an arbitrary model from given examples without requiring explicit rules

Potential for Machine Learning in Beam Diagnostics

- Detection of instrumentation defects
- Beam control and lattice imperfection corrections
- Optimization and operation automation
- Virtual Diagnostics

Potential for Machine Learning in Beam Diagnostics

Detection of faulty Beam Position Monitors

Detection of instrumentation defects

Beam control and lattice imperfection corrections

Regression models for optics correction

Optimization and operation automation

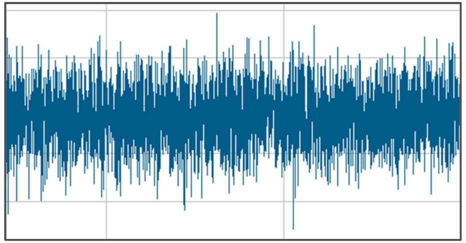
Virtual Diagnostics

Motivation: faulty BPMs in optics analysis

Optics measurements at LHC

BPMs record the turn-by-turn data measuring the oscillations of the excited beam

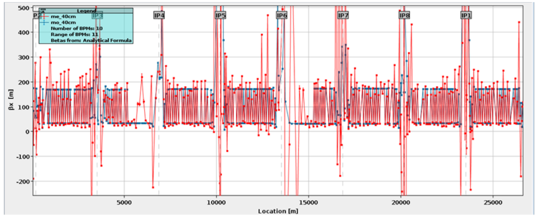
BPM turn-by-turn readings



Calculate optics functions (beta-beating, dispersion, etc.) based on **harmonic analysis of BPMs signal**

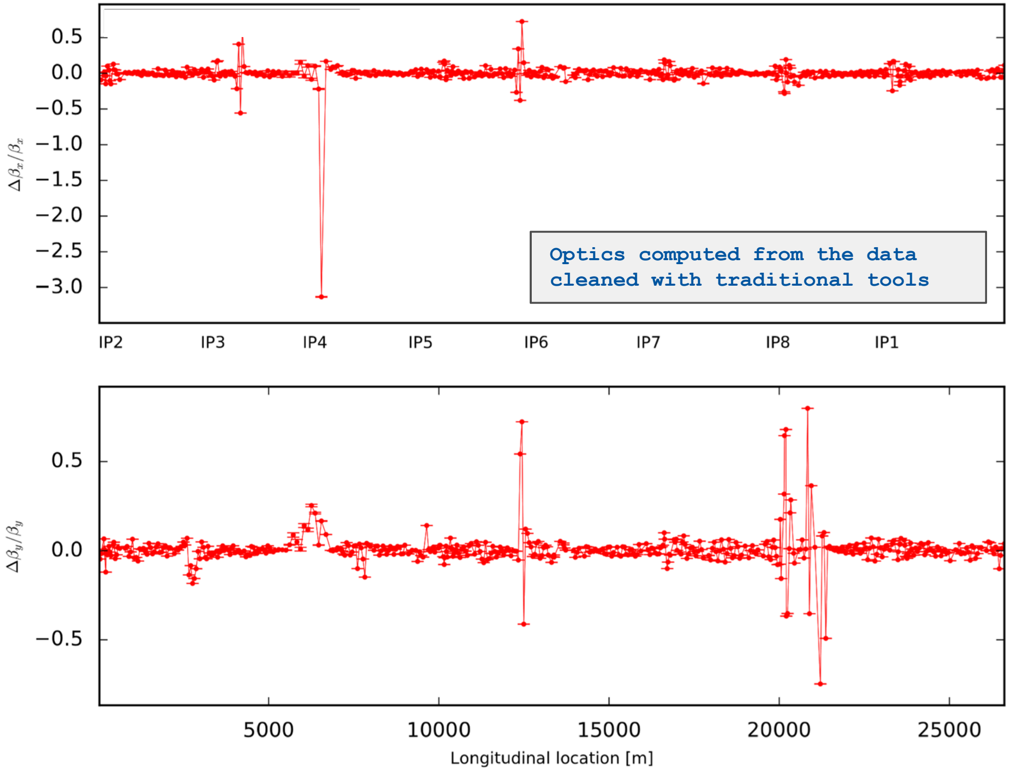
+ data cleaning

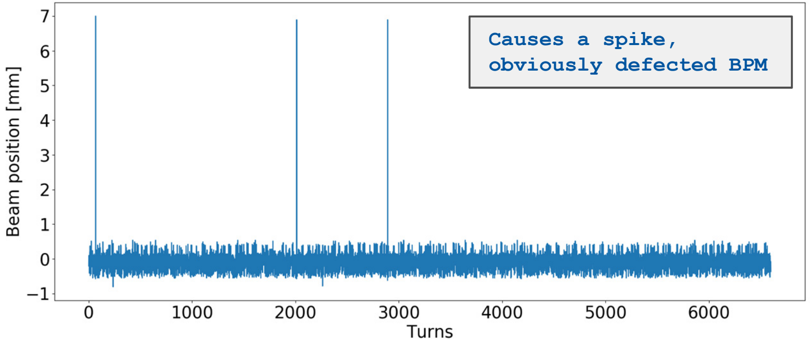
Computed optics



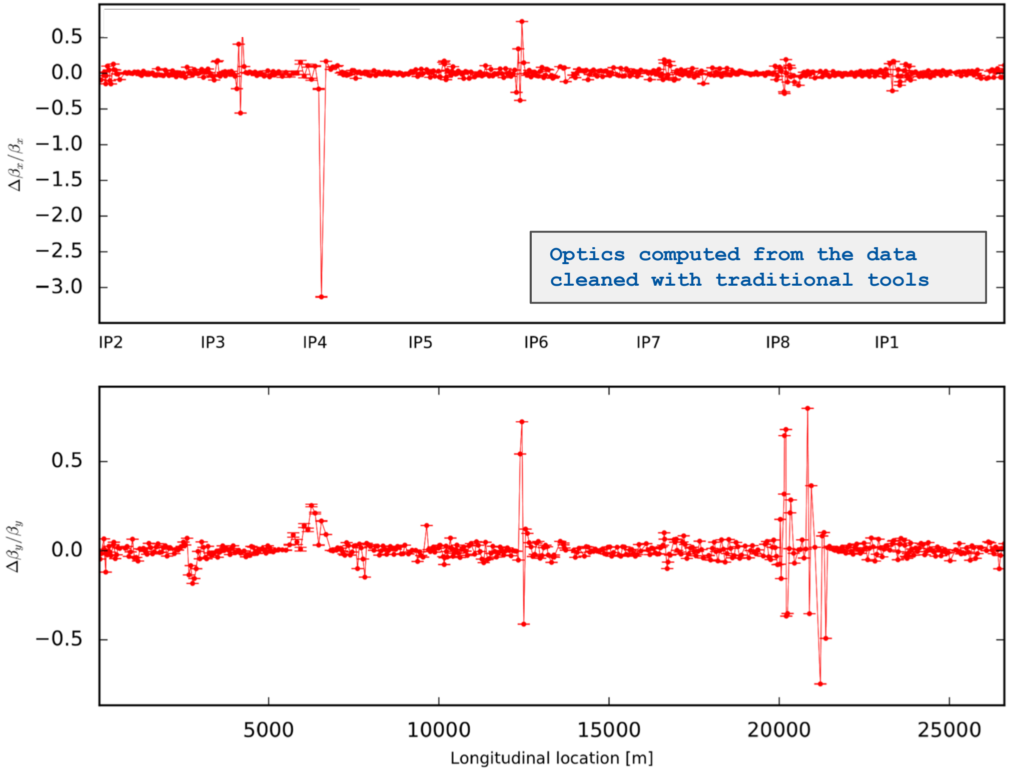
- Unphysical values coming from faulty BPMs signal still can be observed in reconstructed optics even after cleaning with available tools.
- Important to remove faulty BPMs since they affect the optics much more than missing good BPMs.
  - **ML as an alternative solution to improve the analysis**

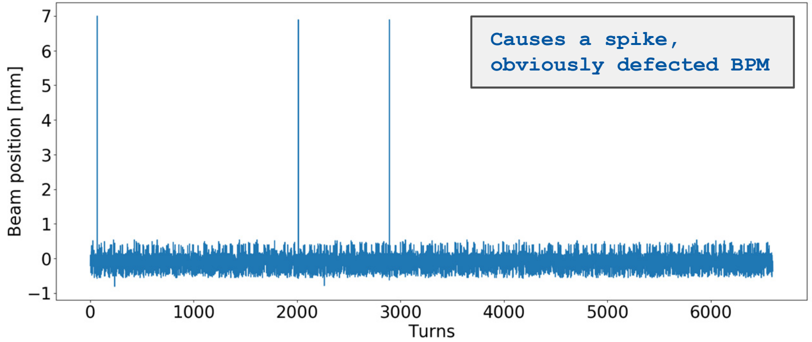
Outliers in the optics reconstructed from harmonic analysis of BPM turn-by-turn measurement

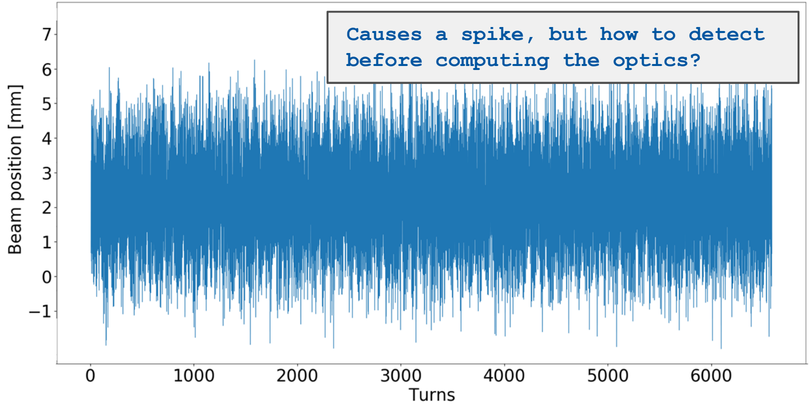




Outliers in the optics reconstructed from harmonic analysis of BPM turn-by-turn measurement







Anomaly detection: Detection of faulty BPMs

Statistical analysis of the past measurements shows that **~10%** of BPMs are faulty

General Idea:

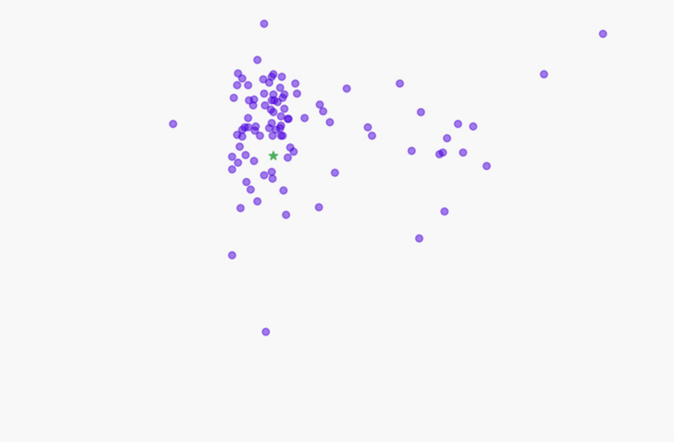
- Since it is unknown, which BPMs are really physically defected, we consider the appearance of non-physical spikes in reconstructed optics as artifact of bad BPMs
- We do not want to replicate current results, no training data set (input-output pairs) available: **Unsupervised learning approach**
- Assuming most of the BPMs measure correctly, the bad BPMs should appear as an **anomaly**
- Applied algorithms: K-means[1], DBSCAN[2], Local Outlier Factor[3], **Isolation Forest**[4] using *Scikit-Learn*

1. Stuart P. Lloyd. Least squares quantization in PCM
2. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" Ester, M., H. P. Kriegel, J. Sander
3. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May)., LOF: identifying density-based local outliers
4. Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. "Isolation forest." Data Mining, 2008. ICDM'08.

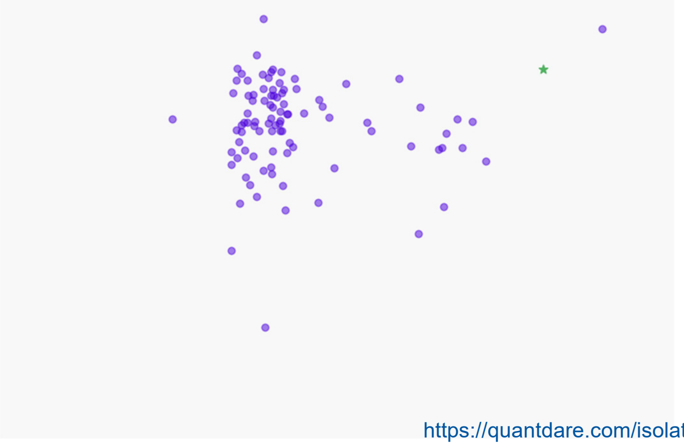
Isolation Forest (IF)

- Forest consists of several **decision trees**
- Selects a feature, randomly selects a split between minimum and maximum values
- Less splits are needed to **isolate an anomalous point –** anomaly score is based on number of splits
- **Contamination factor**: fraction of data points with the highest anomaly scores

Isolating a ”normal” point



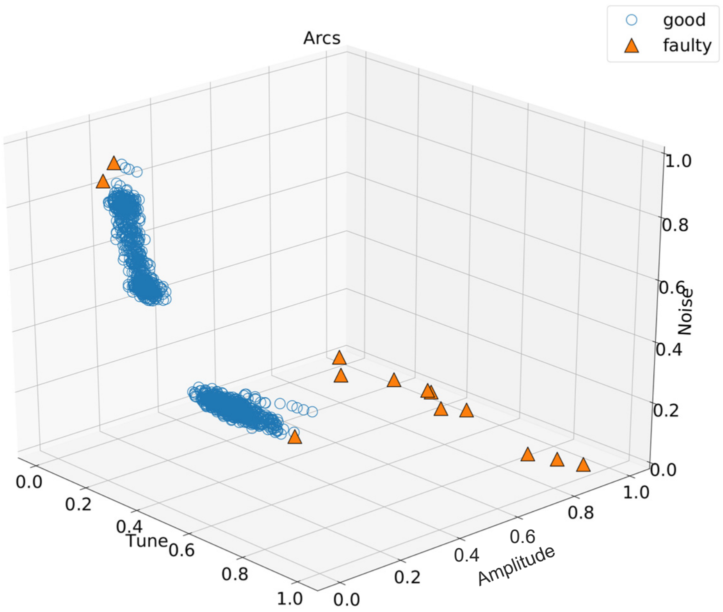
Isolating an outlier

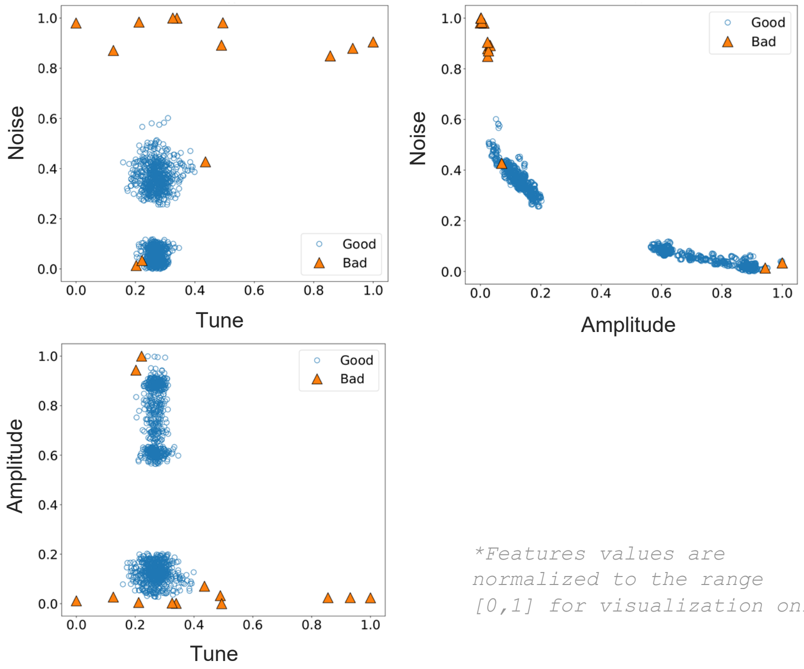


https://quantdare.com/isolation-forest-algorithm/

Faulty BPMs detection with IF: with harmonic analysis data as input features

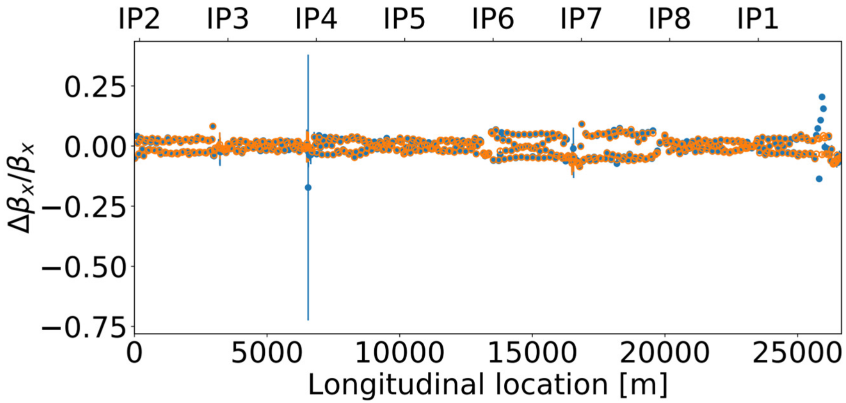
Applied during ions optics commissioning, Beam 2

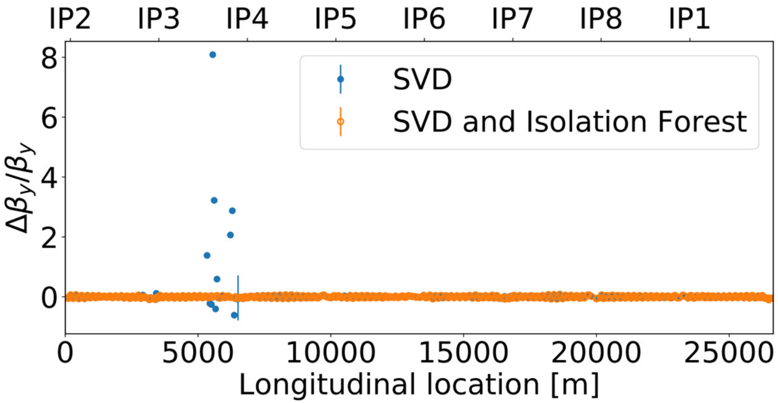




*Features values are normalized to the range [0,1] for visualization only

β-beating from the measurement cleaned with old techniques (SVD and threshold cuts) before and after applying IF:

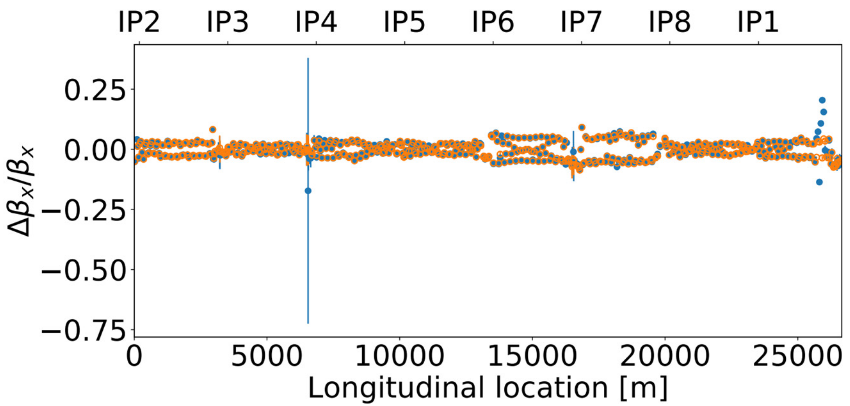


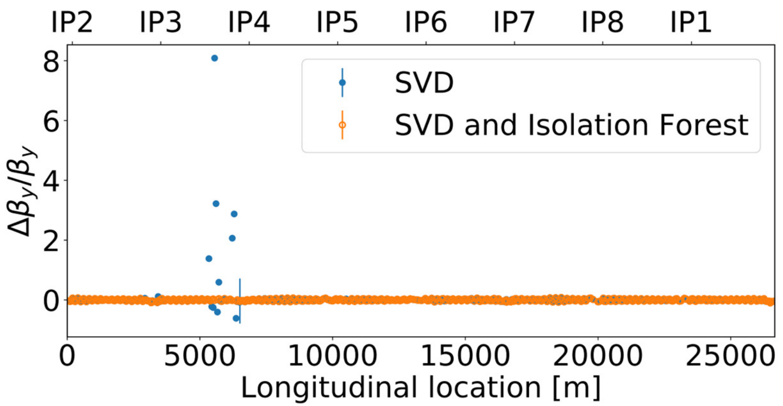


- ✓ This method is fully integrated into optics measurements at LHC
- ✓ Successfully used during beam commissioning and machine developments in 2018 under different optics settings

CERN

β-beating from the measurement cleaned with old techniques (SVD and threshold cuts) before and after applying IF:





- ✓ This method is fully integrated into optics measurements at LHC
- ✓ Successfully used during beam commissioning and machine developments in 2018 under different optics settings

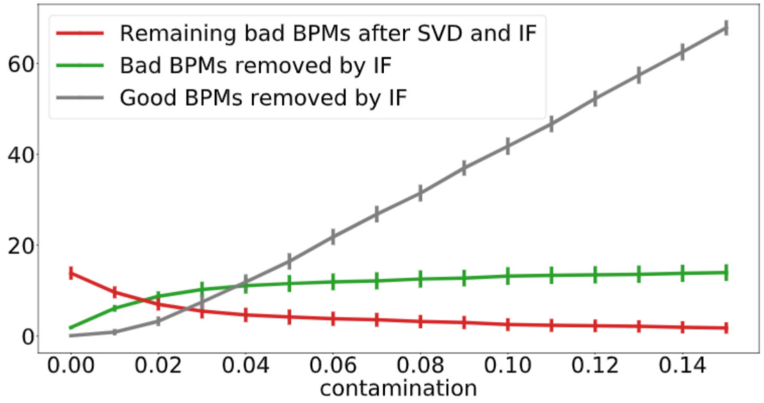
- IF detected most of BPMs which cause erroneous values in the optics
- Some of BPMs without obvious spikes are removed as well

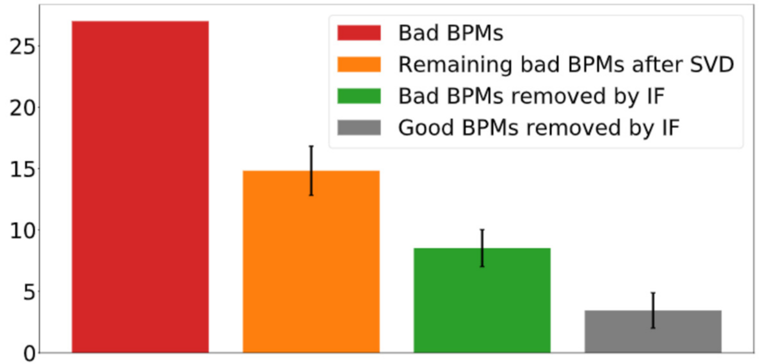
- ➢ Performance of IF depends on its input parameter: **contamination factor**
- ➢ **How does it affect the quality of optics measurements?**

Unsupervised learning: how to verify the results?

Simulations with known artificially introduced faulty BPMs:
Faults are known → **cleaning results can be verified.**

Attention! Simulated data is used only to verify the algorithm, there is no "training".



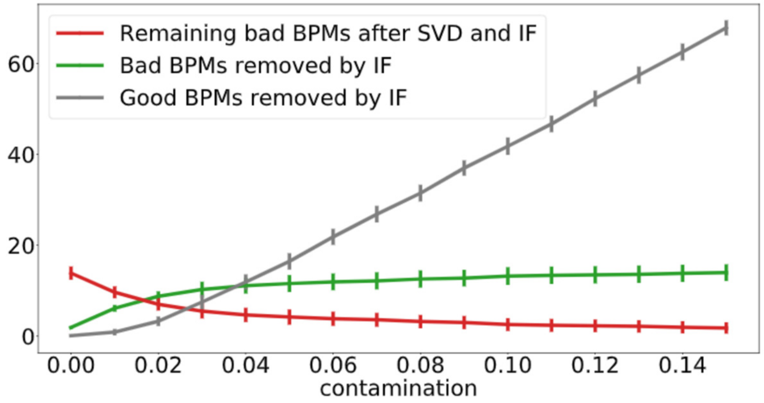


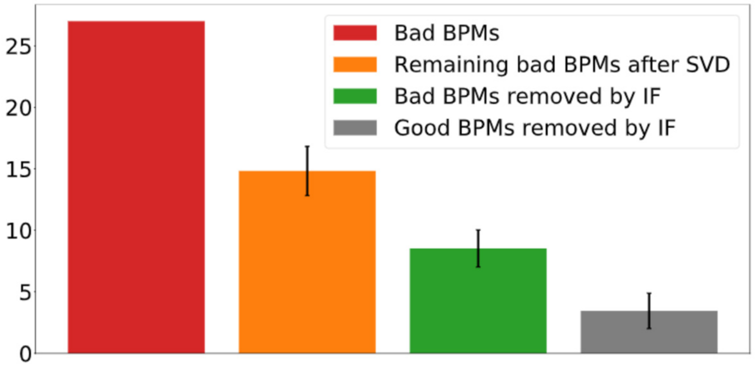
Averaged result over 20 simulated measurements, contamination = 0.02

Unsupervised learning: how to verify the results?

Simulations with known artificially introduced faulty BPMs:
Faults are known → **cleaning results can be verified.**

Attention! Simulated data is used only to verify the algorithm, there is no "training".





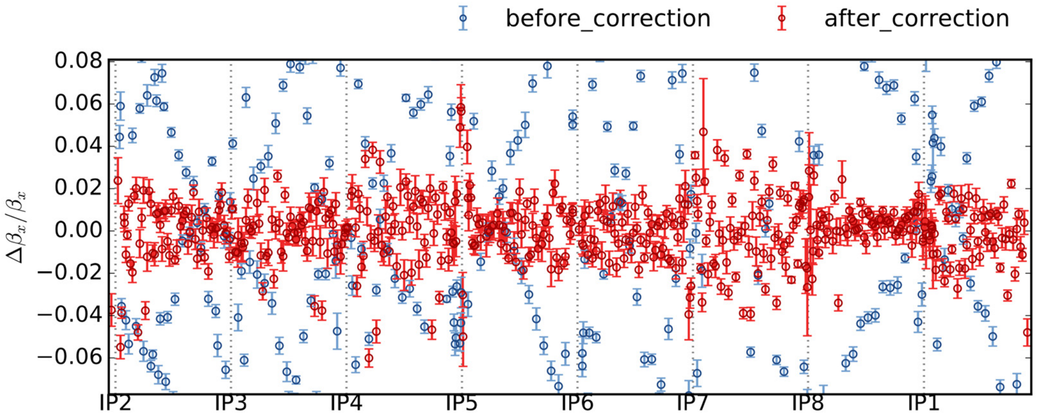
Averaged result over 20 simulated measurements, contamination = 0.02

Optimal contamination factor = 0.02

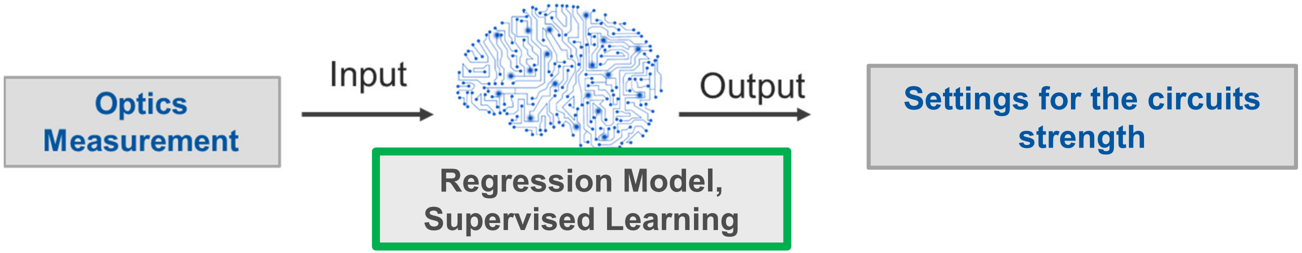
IF algorithm successfully identifies anomalies in measurement data caused by BPM faults without significant loss of good signal.

Optics measurements at the LHC

- One of the main parameters is **beta-beating**: ratio of the measured β-function with respect to the nominal designed function
- Corrections aim to **minimize the difference between the measured and design optics** by changing the strength of corrector magnets – single quadrupoles and quadrupoles powered in circuits.
- Optics corrections in the LHC are currently based on a response matrix between available correctors (single quadrupoles or quadrupoles **powered in circuits**) and observables.



Idea and challenges



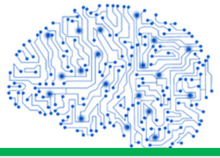
- Corrections are implemented by changing the strength of **circuits** – magnets powered in series
- Optics perturbations are caused by **single magnets** all around the ring

Training data has to consist of pairs:
"*input – <u>correlated to</u> – **known** target values*"

Idea and challenges

Optics Measurement

Input



Output

Settings for the circuits strength

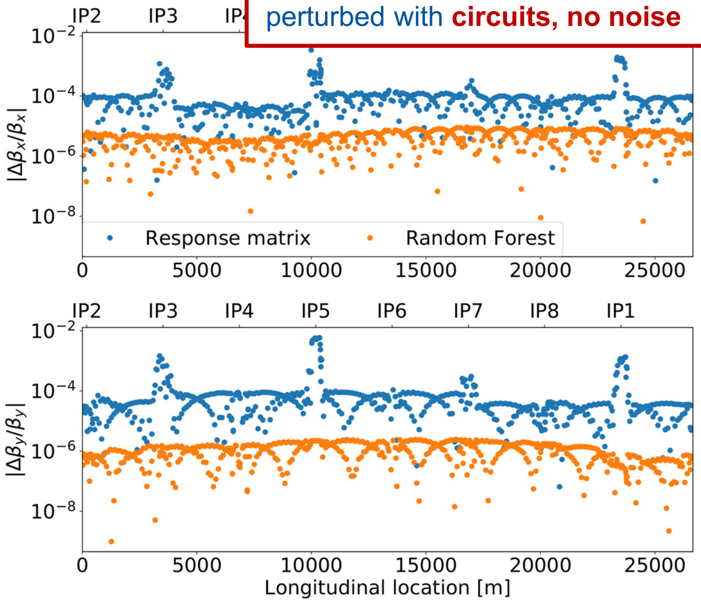
Regression Model,
Supervised Learning

- Corrections are implemented by changing the strength of **circuits** – magnets powered in series
- Optics perturbations are caused by **single magnets** all around the ring

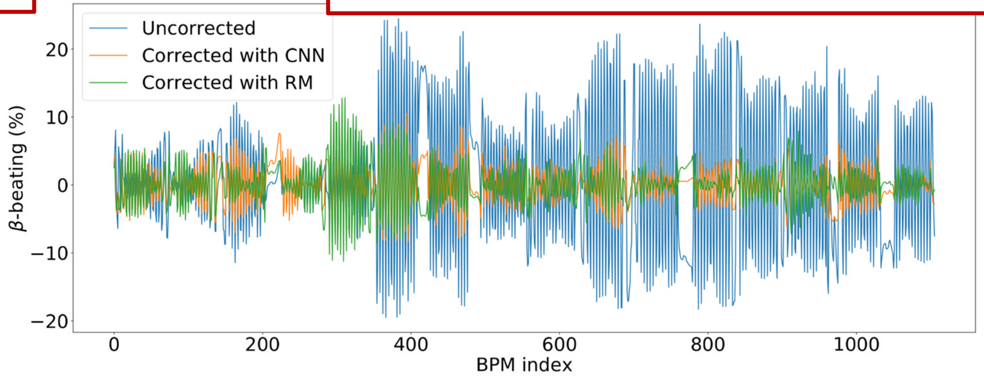
Training data has to consist of pairs:
"*input* – *correlated to* – ***known*** *target values*"

I. First results: training data using errors in the circuits

Corrections on simulated measurement perturbed with **circuits, no noise**



Corrections on simulated measurement perturbed with **single quads, noise added**



Convolutional Neural Network slightly overperformed Response Matrix on simulated measurement reducing ***rms β-beating*** *from 9.5% to 3.0% vs. 3.2% achieved by RM.*

I. First results

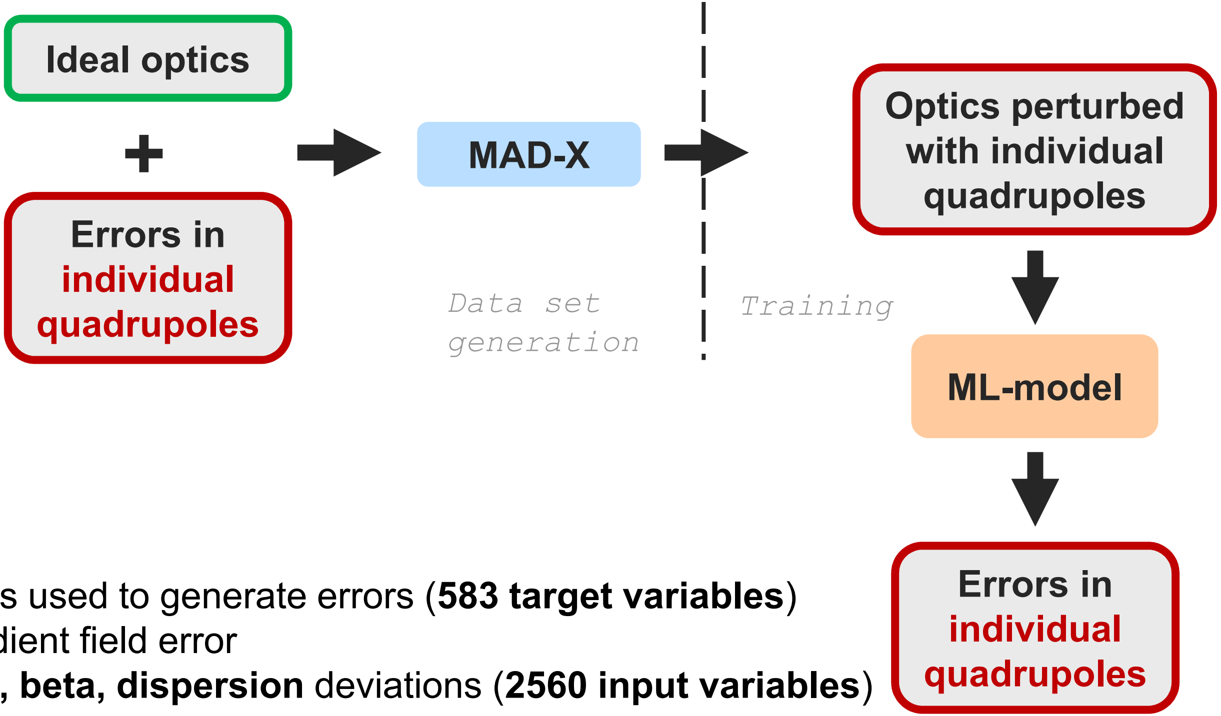
- Several regression models have been trained on the circuits-perturbed data *(Random Forest, CNN, Orthogonal Matching Pursuit, Ridge Regression, Linear Regression)*

➢ All methods demonstrate similar performance
➢ **Linear Regression** is significantly faster to train and easier to interpret.

Linear Regression using ML vs. Response Matrix approach:

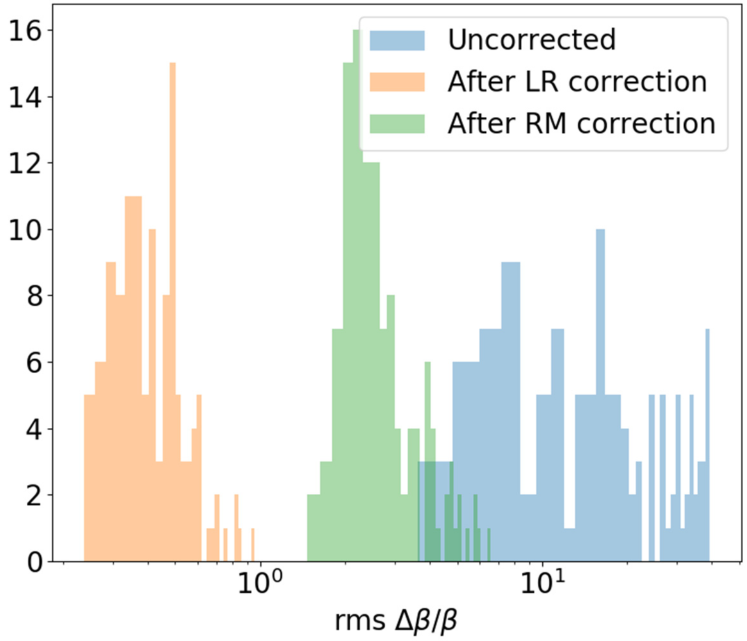
- Both methods are very similar
- However, **the way how we build the model is different**:
    - ML models compute a kind of *average response* that is good for all training samples which contain perturbations introduced by several magnets
    - Response Matrix computes optics response of a single change in a single magnet, weights assigned to the observables have to be obtained through experience / simulations

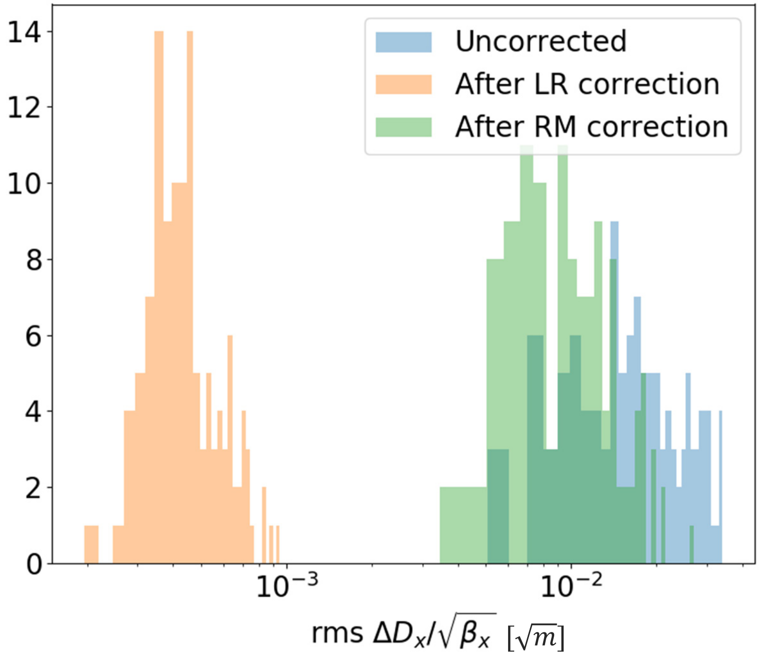
II. Prediction of individual quadrupole errors



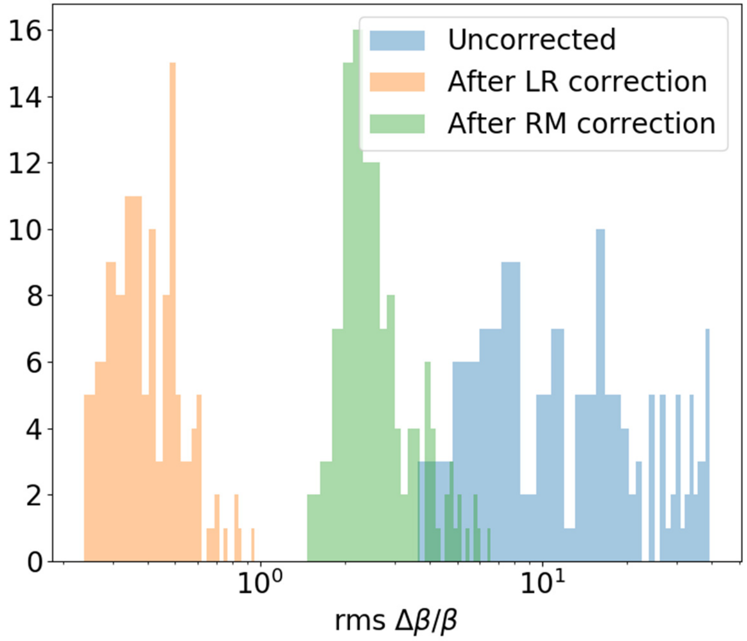
- All quadrupoles used to generate errors (**583 target variables**)
- Optimistic gradient field error
- Input is **phase, beta, dispersion** deviations (**2560 input variables**)

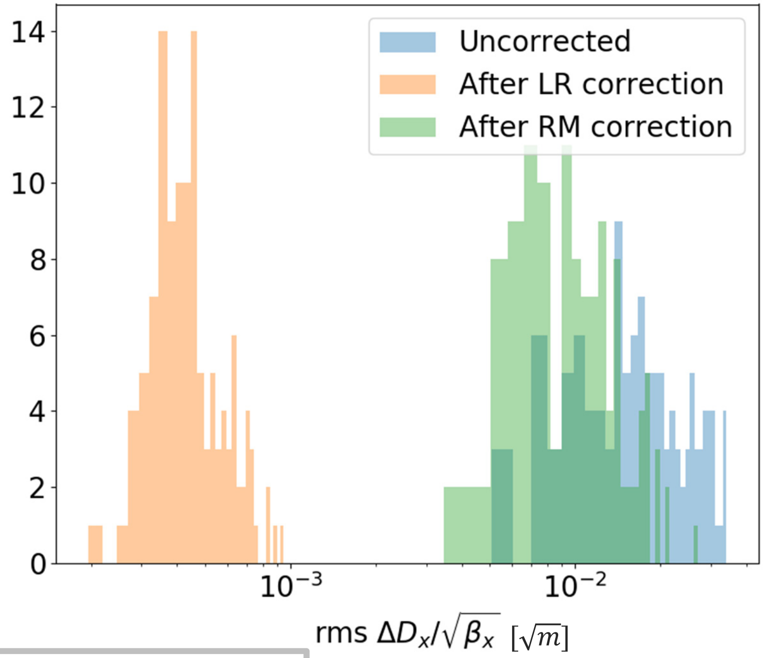
Expected β-beating and normalized dispersion after correction





Expected β-beating and normalized dispersion after correction





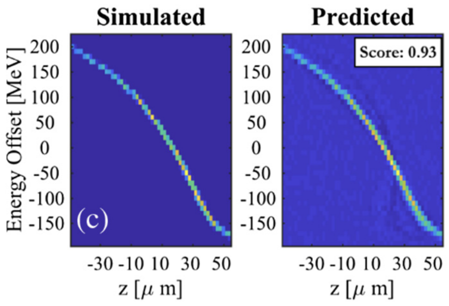
ML-based regression model using individual quadrupoles achieves much better correction than Response Matrix using circuits, for both β-beating and dispersion

Some other recent applications

Virtual Diagnostics

- **Longitudinal phase space (LPS) prediction using ANN, based on the correlation between LPS distribution and various accelerator parameters.** *[C. Emma et.al "Machine learning-based longitudinal phase space prediction of particle accelerators", Phys. Rev. Accel. Beams (21), 112802 (2018)]*

- **Prediction of multiple beam parameters based on cathode laser images.** *["First steps toward incorporating image based diagnostics into particle accelerator control systems using Convolutional Neural Network", A.L. Edelen et al. NAPAC16 (TUPOA51)]*
- **Estimation of oscillation amplitude and synchrotron damping time based on LPS measurements** *[“Machine Learning Application in Bunch Longitudinal Phase Measurement”, X.Y. Xu, Y.B. Leng, and Y.M. Zhoul, IPAC’19 (WEPGW064)]*
- **Instability detection for the LHC transverse feedback system using Isolation Forest** *[L. Coyle et al., https://wiki.epfl.ch/fcc-epfl-lpap/machinelearning]*



Simulated and predicted 2D LPS distributions for FACET-II.
C. Emma

Linear Optics Correction

- **Bayesian approach for linear optics correction** *[Y. Li, R. Rainer, W. Cheng, "Bayesian approach for linear optics correction", Phys. Rev. Accel. Beams (22), 012804 (2019)]*

Some other recent applications

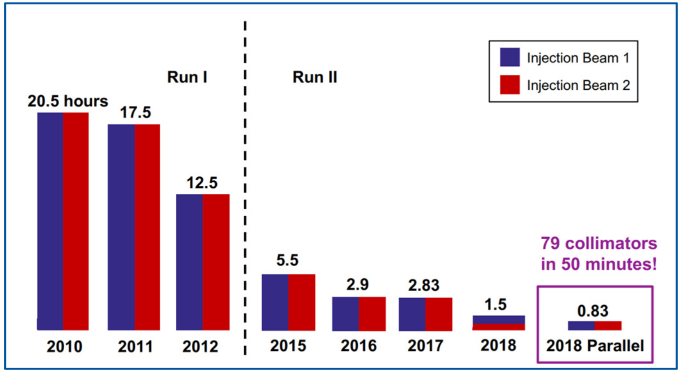
Optimization and Tuning

- **Fast and adaptive feedback system to stabilize the dynamics and control instabilities.** *[T. Boltz et al. “Feedback Design for Control of the Micro-Bunching Instability based on Reinforcement Learning”, IPAC’19 (MOPGW017)]*

- **Automatic performance optimization and first steps towards reinforcement learning at the CERN Low Energy Ion Ring** *[S. Hirlaender, 2nd ICFA Mini-Workshop on Machine Learning for Charged Particle Accelerators, PSI, Villingen, February 2019]*

- **Automatic alignment of LHC collimators based on beam loss spike recognition using an ensemble of several ML models.** *[G. Azzopardi et al. “Operational Results of LHC Collimator Alignment using Machine Learning”, IPAC19 (TUZZPLM1)]*

- **Prediction of complex diagnostics where modeling of every experimental aspect is not possible.** *[A. Sanchez-Gonzalez, et al . “Machine learning applied to single-shot x-ray diagnostics in an XFEL”, https://arxiv.org/abs/1610.03378]*



The time required to align LHC collimators using ML compared to previous operational results.
G. Azzopardi

- **Maximization of the average pulse energy in FELs by tune up to 105 components simultaneously based on average bunch energy.** *[A. Scheinker et al., "Model-independent tuning for maximizing free electron laser pulse energy", Phys. Rev. Accel. Beams (22), 082802 (2019)]*

Part III. Recommendations and conclusion

Where can we use ML in accelerators?

- Simultaneous optimization of several beam parameters
- Prediction of beam behavior
- Automation of diagnostics and operation
- Lattice imperfection correction
- Detection of instrumentation defects
- Maintenance of accelerator systems

… more great ideas are welcome during discussion!

Practical advice

- Often data preprocessing is needed before any model can be applied: rescaling, feature engineering, denoising, outlier elimination
  - ➢ data vizualisation can help
- Start with simple models - increase complexity only if needed
- Estimate model generalization (split into training, test and validation sets)

Frameworks to use:

- Prototyping, fast and easy implementation (very good documentation): http://scikit-learn.org/
- High-level package for Neural Networks: – https://keras.io/
- Deep Learning, specific complex model architectures: https://www.tensorflow.org/ http://deeplearning.net/software/theano/
- Reinforcement Learning: OpenAI Gym https://gym.openai.com/

CERN

Conclusions

Important to identify where ML can surpass traditional methods

- How much effort is needed to implement a ML solution? Is appropriate infrastructure for data acquisition available? Enough resources to perform the training?

> Universal Approximation Theorem
> A simple neural network including only a single hidden layer can approximate any bounded continuous target function with arbitrary small error.

Does not say how big the effort could be…

Successful examples

- Automation of particular accelerator components e.g. collimation system
- Modeling is not possible, fast diagnostics needed, e.g. single-shot x-ray diagnostics
- Reasons of perturbations of beam behavior are known, e.g. optics correction

  → ***Supervised Learning, e.g. Neural Networks, but also Multivariate Regression***
- Discover anomalies or important correlations, e.g. detection of instrumentation failures

  → ***Unsupervised Learning, e.g. clustering, decision trees***
- Performance optimization for operation, adaptive feedback systems

  → ***Reinforcement Learning***

Cat!

Thank you for your attention!

www.cern.ch