Mitigation of systematic errors induced biases in ML-based selection

Francisco Matorras, Pablo Martínez
IFCA
Instituto de Física de Cantabria (Santander, Spain)
Luis Crespo Ruiz
Universidad de Cantabria (Santander, Spain)

Foreword

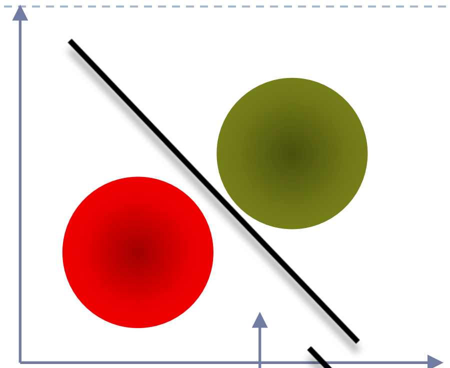
- This is an extract of the final project of Luis Crespo Ruiz to get his Degree in Physics
- **Application of multidimensional classification techniques to Particle Physics in the presence of systematic errors**
- https://repositorio.unican.es/xmlui/handle/10902/20598
- Not yet tested on real HEP analysis, but very promising results worth being shared

Motivation

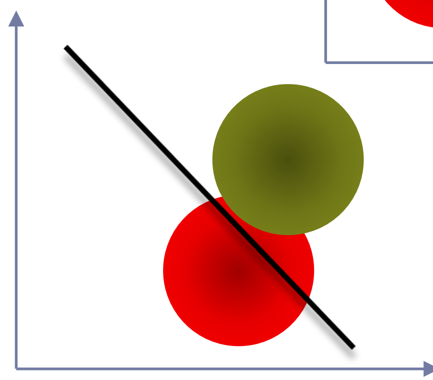
- No need to explain how spread are the different ML (or MVA) techniques in particle physics analyses
- When its application started (and still now) a repeated question was "what about systematics?"
- Systematic uncertainties are usually treated in an equivalent way as for a classical cut analysis:
  1. I do my analysis based on ideal samples
  2. I estimate the effect of the systematics in the variables
  3. I propagate the uncertainties through my selection (either cuts or MVA)
  4. I estimate the effect on my result: efficiency, NN output, cross section, fit...
- For ML based, this means **training in samples with ideal conditions**
- That's something, but
  - what if we rely on a variable that is poorly described and there are some other ideally less discriminant but better in real life?

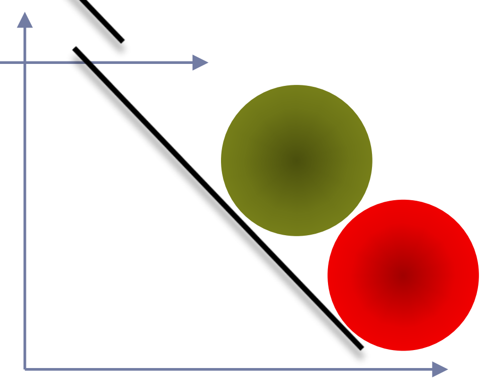
An example

- Your sample tells the algorithm to separate like that

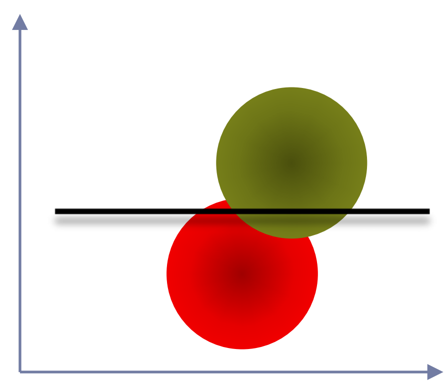


- But if your data is there or there…?





- It might be wiser to ignore the horizontal variable and cut



The method

- Can we make the ML algorithm learn the weaknesses of the variables in such cases?
- Propose to use the data augmentation technique
  - in this case, let’s feed the machine with replicas with the weakness incorporated
- Relatively simple to implement in HEP
  - Given our MC (or data) training samples, replicate each event several times according to a law driven by the systematics
  - Train on these altered samples (no need to perform the costly MC simulation)

Testing the method

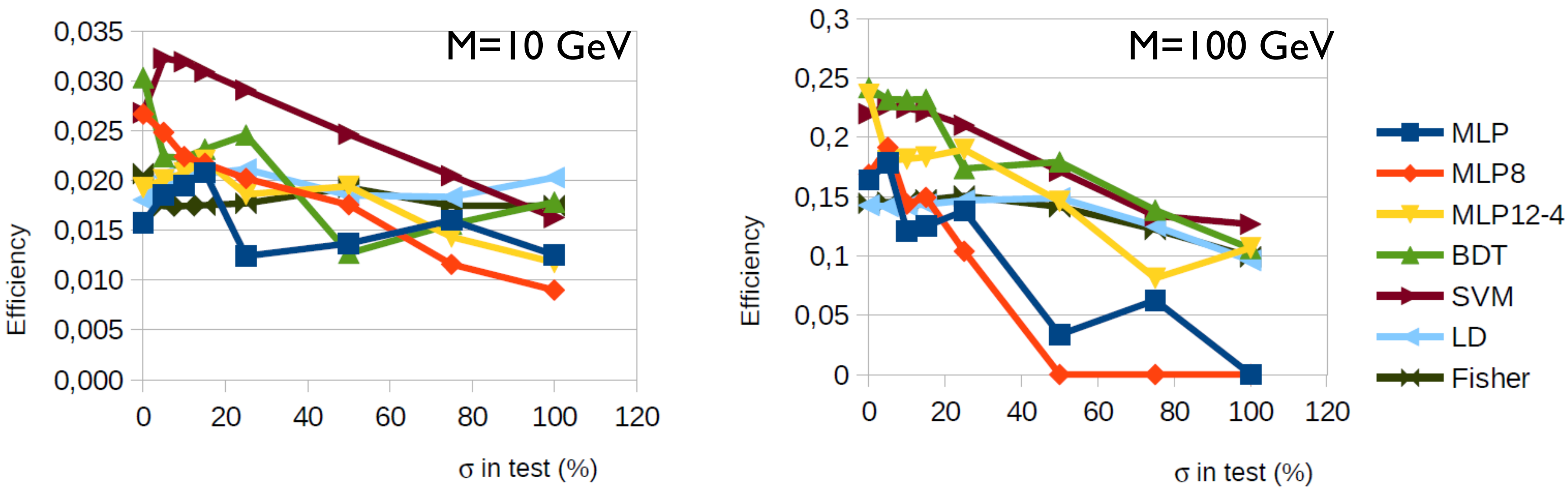
- As often happens, in ML difficult to demonstrate the general validity
- Run instead on an example
  - GEN + smearing-based example (true physics, simplified detector)
- Classification of the production of a dark matter candidate in association with a top pair (ttDM) versus the SM production of ttbar, for different masses of DM
  - Check the extremes in mass, for low mass they are basically indistinguishable while for high mass there is a clear separation
  - Few variables: 3-momentum for two jets and two leptons and MET
- Train different algorithms (several MLP, BDT, LD, SVM, Fisher)
- Compare performance on systematic samples after training under different conditions

Systematic on resolution

As an example, assume a systematic that implies worsen the resolution of some of the variables, with a random gaussian noise

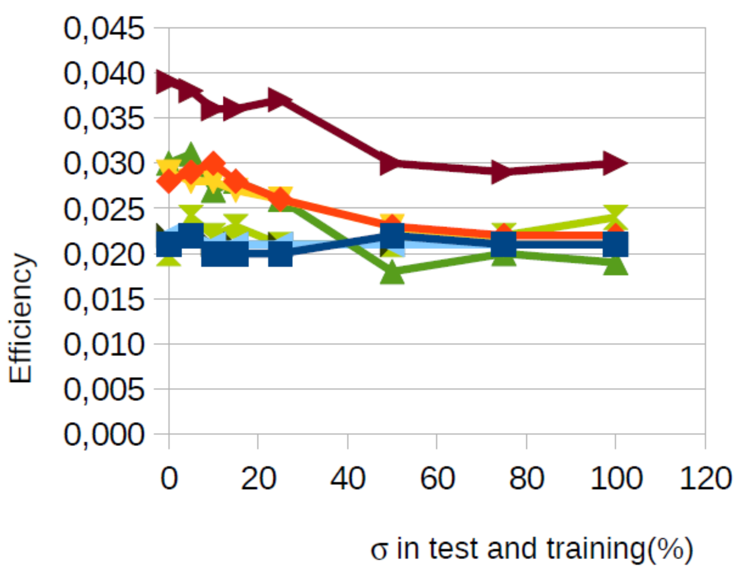
Jet energy resolution

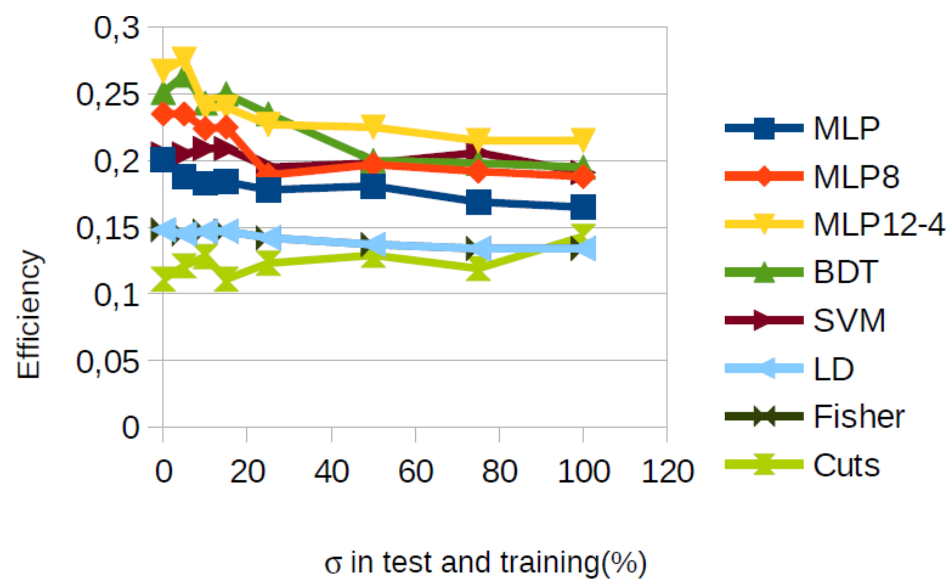
- Check the effect of jet energy resolution
- Add an additional gaussian smearing to the jet energy (and propagate to MET)
- Evaluate the signal efficiency at a different working points (1% efficiency for background in the plots shown)
- When trained on zero-systematic samples and tested on systematic samples important degradation for "some" of the methods. Usual estimation of systematics



Training on data augmented

- Not surprisingly if we train with a smeared sample most of the effect is corrected
  - Here training on samples smeared with equal σ as test sample
- Systematic nearly cancelled even for very large effects
- You might argue that this is obvious but still not always done…



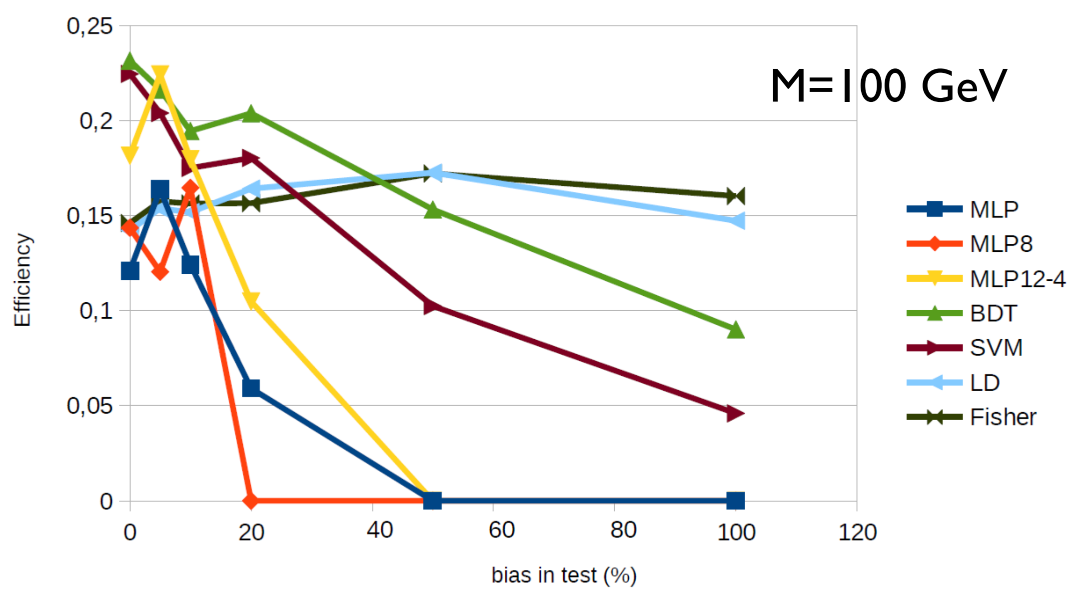


Systematic on scale/calibration

As another example, assume a systematic that implies a correlated bias in some of the variables

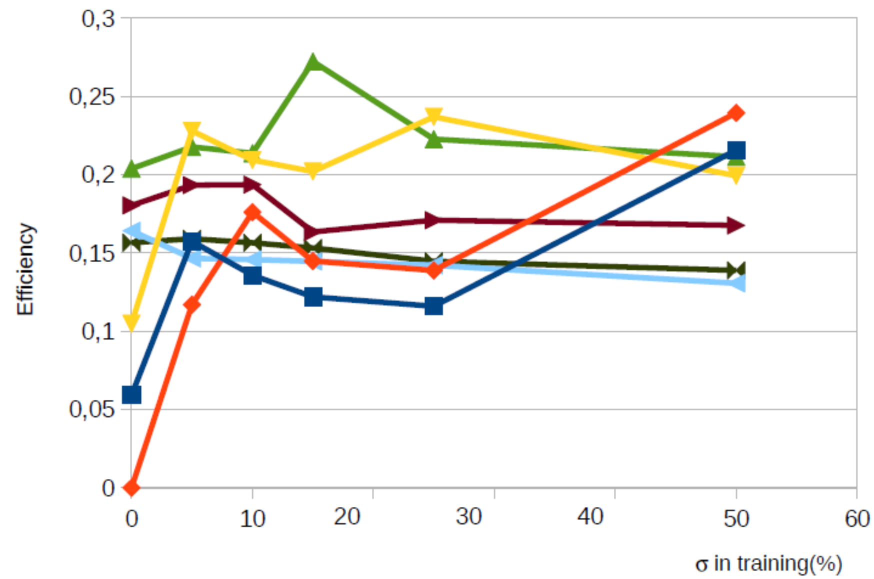
Jet energy scale

- Imagine instead a scale/calibration effect
  - All jets energy is wrong by a given fixed amount
- Jet energy on test samples scaled by a constant term
- Jet energy in training samples is smeared
- In all cases MET is recalculated
- When trained on zero-systematic samples and tested on systematic samples, catastrophic degradation for "some" of the methods (NN).



Training on data augmented

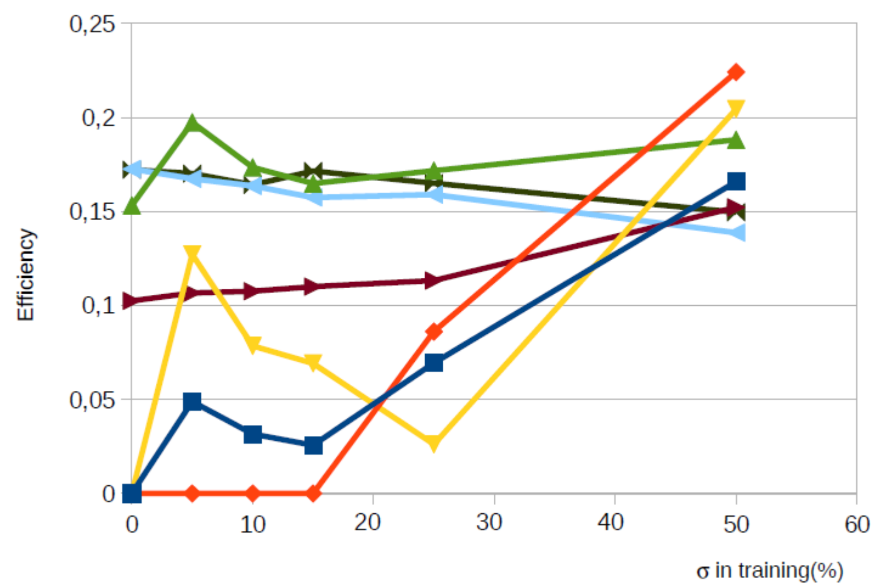
- Let's try to train with smeared samples
- Try for different size of the smearing
- Most of the effect for a 20% bias is cancelled when training with smeared sample with 5-20% sigma
- Similar result for a wide range of smearing



Test with a 20% bias, M=100 GeV

Training on data augmented

- Check with a larger (huge) bias of 50% (all energies scaled by 1.5)
- Again, response is mostly recovered when training with a smearing of similar size as the bias



Test with a 50% bias, M=100 GeV

Results

- Cannot draw general conclusions from this simplistic example but:
  - As it is very well known, the effect of the systematics is very strongly dependent on the type of algorithm, the working point and even the particular training.
  - Not so difficult to find examples where systematic uncertainties totally destroy the performance of ML algorithms.
  - Training on smeared samples cures most of the effect of resolution systematics, when the smear is comparable to the systematic error
  - Training on smeared samples cures most of the effect of scale systematics, when the smear is comparable to the systematic error

Conclusions and outlook

- A very simple method based on data augmentation is proposed to mitigate the effect of systematic errors in ML-based analyses
  - Based on training on samples augmented from the original samples, which include in some way the effect of systematics
  - Don't need to resimulate events
- Easily implemented for most systematics, in a similar way as we usually calculate them
- It is implemented at the level of the variables, so it is valid for any ML algorithm.
- So far, only tested on simplified examples, but results promising
  - Can recover performance even for very large systematic uncertainties
  - Very promising but need to check on real physic examples