DOMAIN ADAPTATION FOR SYSTEMATICS

A case study

Victor Estrade (PhD student), with the help of Erwan Martin (internship)

July 20, 2017

University of Paris-Sud

Error sources

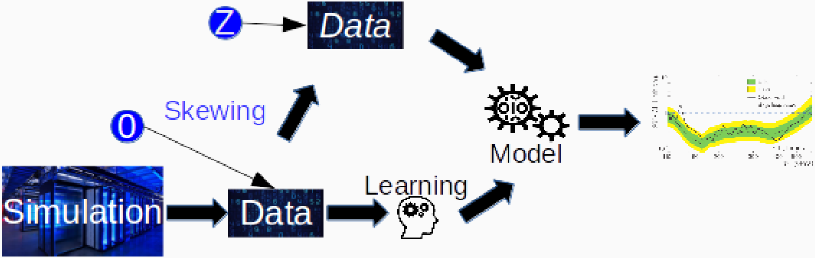
$$measure = value \pm err_{stat} \pm err_{syst}$$

Statistical error

- lack of training data
- wrong hyper-parameter settings
- wrong choice of algorithm
- etc

Systematic error

- measurement imperfections
- simulation limitations
- theoretical knowledge lacks



- $H \to \tau\tau$ simulation (HiggsML challenge data, Geant4)
- Skewing : rescaling τ energy $(+[1\%, 10\%])$
- [Adam-Bourdarios et al., 2014] & [Baldi et al., 2015]

Nail details

Physicist : We want to count higgs boson candidates and take into account systematics.

ML : Great ! I'd love to have an objective function ...

Physicist : minimize

$$\frac{\sigma_\mu}{\mu} = \sqrt{\left(\frac{\sqrt{s_0 + b_0}}{s_0}\right)^2 + \left(\frac{(s_z - s_0) + (b_z - b_0)}{s_0}\right)^2}$$

- s, selected signals (True posivites)
- b, selected backgrounds (False posivites)
- $*_z$, on the skewed test set
- $*_0$, on the nominal test set

Deep learning tools

- Data augmentation : train on both simulation and real data (or many simulations)
- Adversarial methods
  - *GAN* [Goodfellow et al., 2014] : Generate samples that are indistinguishable from real data
  - *DAN* [Ganin et al., 2015] : Find a common space for both simulation and real data
  - *PAN* [Louppe et al., 2016] : Correct the model to be pivotal
- *Tangent Propagation* [Simard et al., 1991] [Rifai et al., 2011] : Regularize training using known invariant

Available Hammers

What I tried

- Data augmentation : train on both simulation and real data (or many simulations)
- Adversarial methods
  - *GAN* [Goodfellow et al., 2014] : Generate samples that are indistinguishable from real data
  - *DAN* [Ganin et al., 2015] : Find a common space for both simulation and real data
  - *PAN* [Louppe et al., 2016] : Correct the model to be pivotal
- *Tangent Propagation* [Simard et al., 1991] [Rifai et al., 2011] : Regularize training using known invariant

Fast Presentation : Tangent Propagation

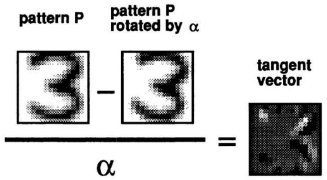
Key Ideas

- Regularize the derivative of the model according to the transformation.

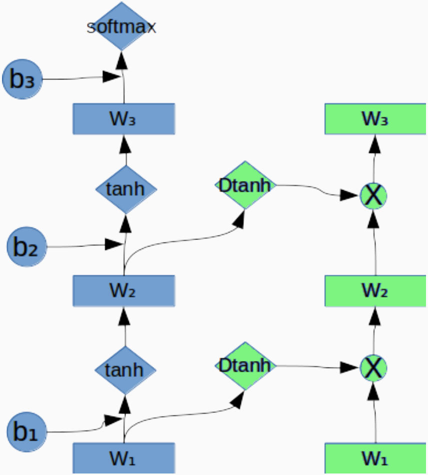
$$loss = E_{standard} + \lambda \sum_{x \in Data} \left| \frac{\partial G_w(s(x,\alpha))}{\partial \alpha} \right|^2_{\alpha=0}$$

- Compute this derivative with a forward propagation through a "linearized" network.

$$\left. \frac{\partial G_w(s(x,\alpha))}{\partial \alpha} \right|_{\alpha=0} = \nabla_x G_w(x). \left. \frac{\partial s(x,\alpha)}{\partial \alpha} \right|_{\alpha=0}$$

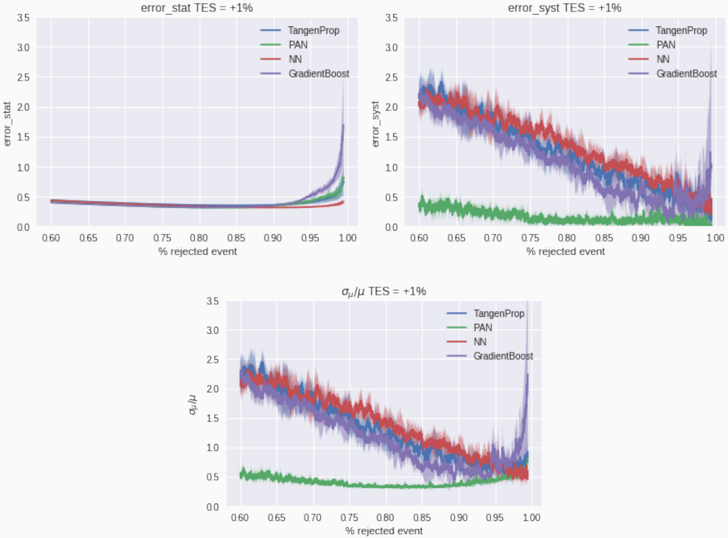


Jacobian Net

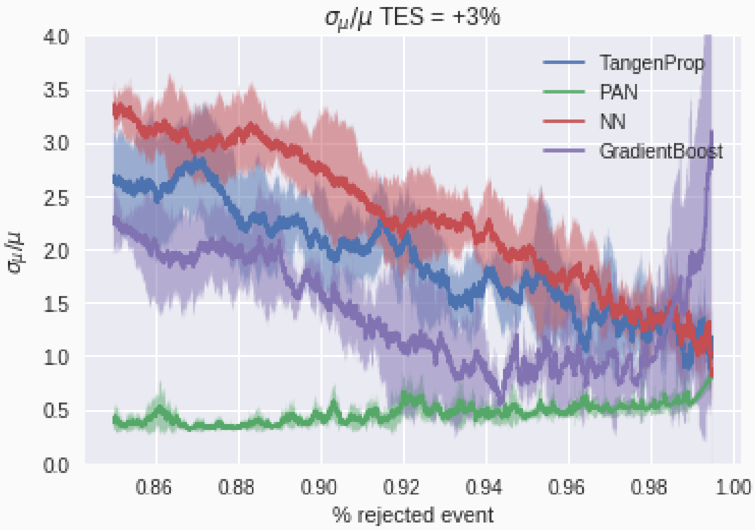


Tangent Propagation : standard network + ”linearized” model insertion

Systematics dominate



Adversarial training wins this nail



Discussion

Results

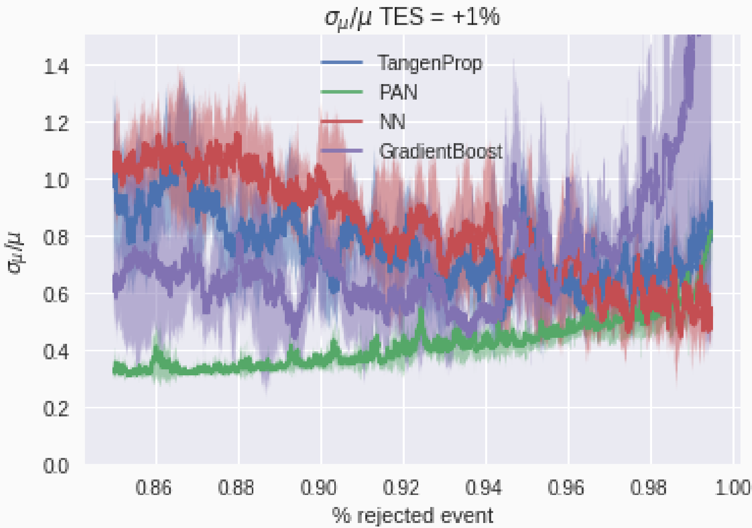
- Adversarial is hard to train but worth it.
- Boosting is still is the game (if high level features).
- Not giving up on Tangent Propagation.

Future : What kind of hammer for what kind of nail

- Build a zoo with data stained with systematics
- Try GANs, DANs

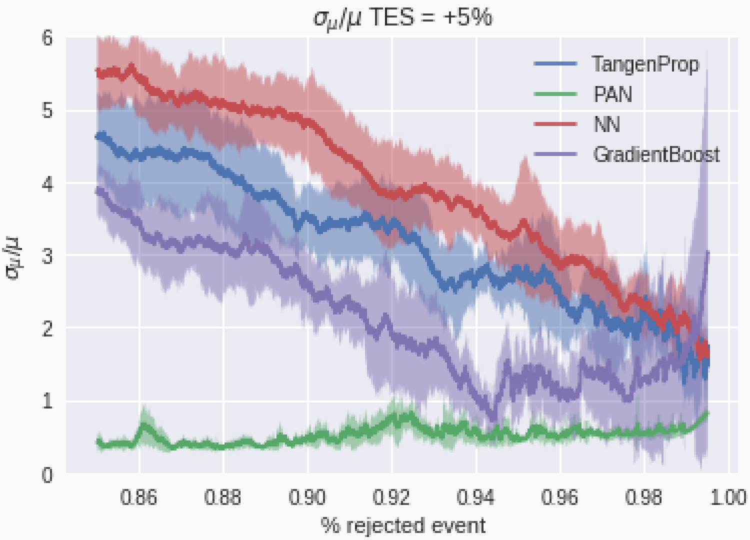
Questions ?

Backup 1%



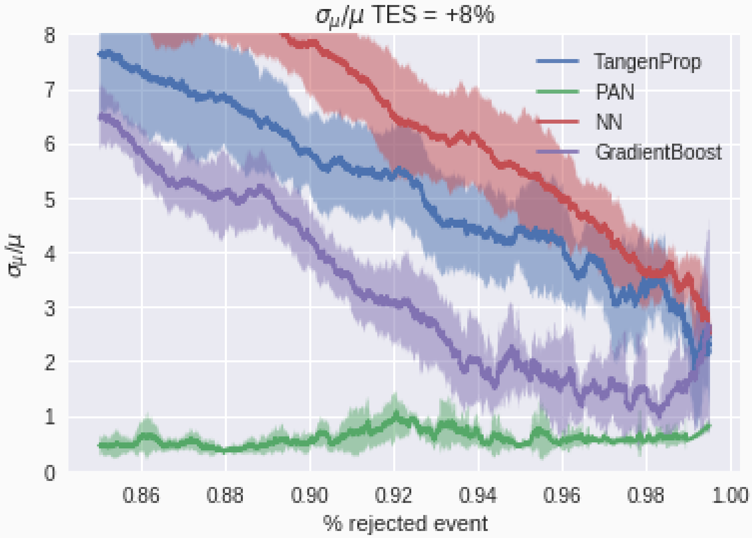
σ_μ/μ TES = +1%
TangenProp
PAN
NN
GradientBoost
σ_μ/μ
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.86
0.88
0.90
0.92
0.94
0.96
0.98
1.00
% rejected event

Backup 5%



σ_μ/μ TES = +5%
6
5
4
3
2
1
0
σ_μ/μ
0.86
0.88
0.90
0.92
0.94
0.96
0.98
1.00
% rejected event
TangenProp
PAN
NN
GradientBoost

Backup 8%



σ_μ/μ TES = +8%
σ_μ/μ
% rejected event
TangenProp
PAN
NN
GradientBoost
0
1
2
3
4
5
6
7
8
0.86
0.88
0.90
0.92
0.94
0.96
0.98
1.00

ML resistance

For the ML community not focusing on optimizing classification/regression error is counter intuitive.

Magic words :

”Non additive (function) error”

Adam-Bourdarios, C., Cowan, G., Germain, C., Guyon, I., Kégl, B., and Rousseau, D. (2014).
The Higgs boson machine learning challenge.
In *HEPML@ NIPS*, pages 19–55.

Baldi, P., Sadowski, P., and Whiteson, D. (2015).
Enhanced higgs boson to $\tau^+\tau^-$ search with deep learning.
Phys. Rev. Lett., 114:111801.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2015).
Domain-Adversarial Training of Neural Networks.
arXiv:1505.07818 [cs, stat].
arXiv: 1505.07818.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative Adversarial Networks.
arXiv:1406.2661 [cs, stat].
arXiv: 1406.2661.

Louppe, G., Kagan, M., and Cranmer, K. (2016).
Learning to Pivot with Adversarial Networks.
arXiv:1611.01046 [physics, stat].
arXiv: 1611.01046.

Rifai, S., Dauphin, Y., Vincent, P., Bengio, Y., and Muller, X. (2011).
The Manifold Tangent Classifier.
In *NIPS*, volume 271, page 523.

Simard, P. Y., Victorri, B., LeCun, Y., and Denker, J. S. (1991).
Tangent Prop - A Formalism for Specifying Selected Invariances in an Adaptive Network.
In Moody, J. E., Hanson, S. J., and Lippmann, R., editors, *NIPS*, pages 895–903. Morgan Kaufmann.