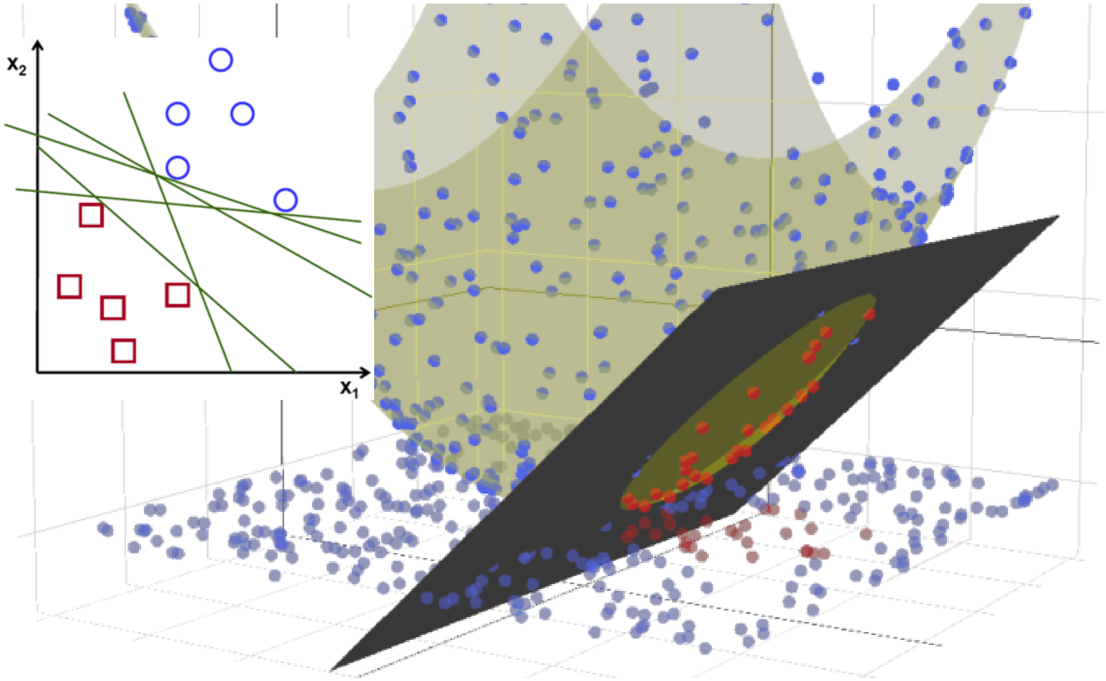
Machine learning
Lecture 6



Marcin Wolter

IFJ PAN

11 April 2018

- Practical exercise – simplified ATLAS analysis using root/TMVA

Measurement of Tau Polarisation in $Z/\gamma^* \to \tau\tau$ Decays in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector

STDM-2015-18

Version: 1.0

To be submitted to: Eur. Phys. J. C

Slides from: Paweł Malecki

Supporting internal notes

Analysis note: https://cds.cern.ch/record/2105517

Comments are due by: 11 June 2017

Analysis Team

[*email:* atlas-stdm-2015-18-editors@cern.ch]

Marzieh Bahmani, Philip Bechtle, Pawel Bruckman de Renstrom, Jane Cummings, William Davey, Sarah Demers, Klaus Desch, Jochen Dingfelder, Anna Kaczmarska, Christian Limbach, Pawel Malecki, Elzbieta Richter-Was, Lara Schildgen, Peter Wagner, Benedict Winter, Marcin Wolter

Editorial Board

[*email:* atlas-stdm-2015-18-editorial-board@cern.ch]

Eric Torrence (chair)
Quentin Buat
Stan Lai

Paper: https://cds.cern.ch/record/2283028

Introduction

- Polarization:

$$P_\tau = \frac{\sigma_{\text{right}} - \sigma_{\text{left}}}{\sigma_{\text{right}} + \sigma_{\text{left}}}$$

- Measured in 66 – 116 GeV Z mass range,

- Also measured in fiducial region that resembles signal region,

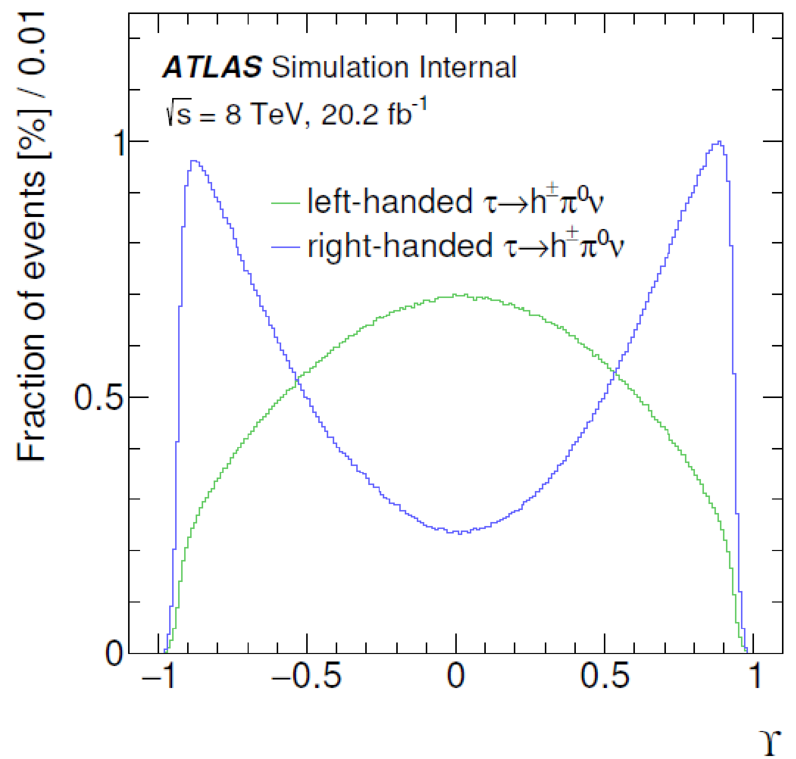
- Motivation:

- Sensitivity to $\sin^2\theta_W$,
- Complementary to LEP (pp->Z->ττ instead of ee->Z->ττ),
- Basis for future measurements and searches (e.g. Higgs CP, charged Higgs).

- Methodology:

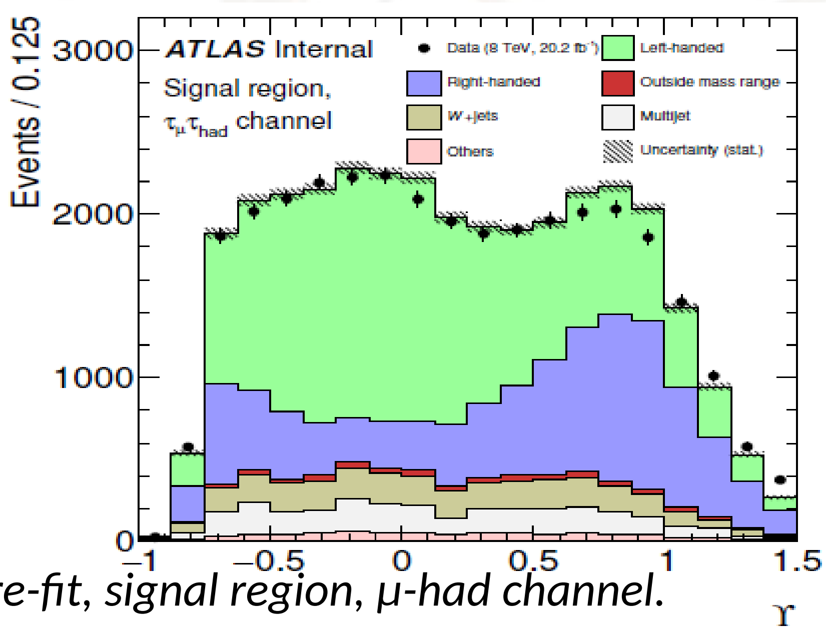
Template fit of a variable sensitive to tau helicity, *Y:*

$$\Upsilon_{\text{theory}} = \frac{E_{\pi^\pm} - E_{\pi^0}}{E_{\pi^\pm} + E_{\pi^0}} \approx \Upsilon = \frac{2 \cdot p_{\text{T,track}}}{p_{\text{T},\tau_{\text{had-vis}}}} - 1$$



Analysis flow

- Select Z→ττ (lep-had) events,
- Split into e-had and μ-had channels,
- Use lepton to trigger and tag event. Measure polarisation from hadronic decay,
- Use TauSpinner program to attribute helicity to tau leptons in MC,
- Estimate background:
  - W→ℓν and QCD dijets: data-driven,
  - Z→ℓℓ, ttbar: MC.

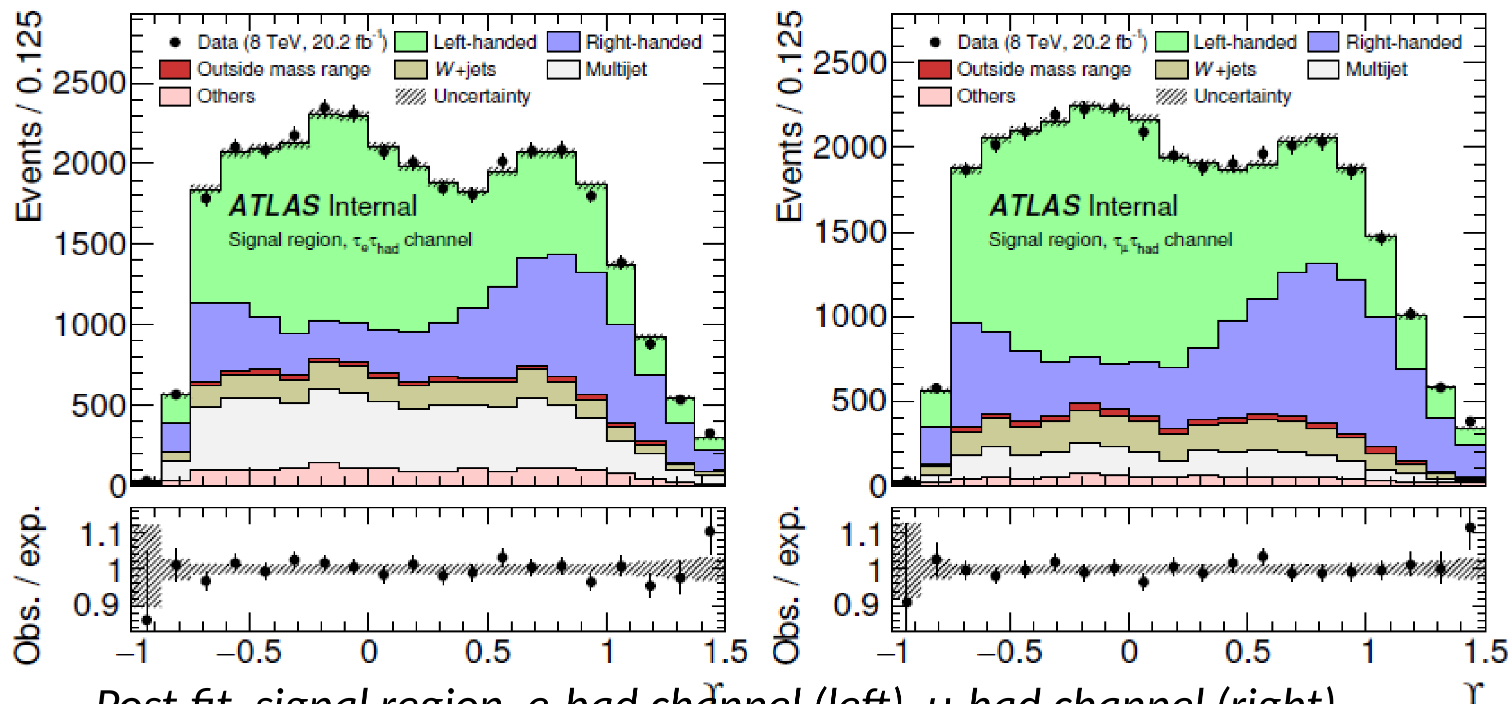


Pre-fit, signal region, μ-had channel.

- Perform ML fit with the obtained *Y* templates to asess the number of left- and right-handed taus.
- Systematic uncertainties:

Source of uncertainty	σ_{P_τ} in mass range	σ_{P_τ} in fiducial region
Modelling of signal process	$^{+0.026}_{-0.026}$	$^{+0.022}_{-0.022}$
τ_{had} identification	$^{+0.019}_{-0.020}$	$^{+0.023}_{-0.024}$
MC statistical	$^{+0.016}_{-0.016}$	$^{+0.018}_{-0.019}$
Signal sample splitting	$^{+0.015}_{-0.014}$	$^{+0.015}_{-0.015}$
TES and TER	$^{+0.013}_{-0.015}$	$^{+0.017}_{-0.019}$
Multijet estimate	$^{+0.013}_{-0.012}$	$^{+0.013}_{-0.012}$
PDF	$^{+0.006}_{-0.007}$	$^{+0.005}_{-0.005}$
W+jets shape	$^{+0.002}_{-0.002}$	$^{+0.003}_{-0.003}$
Other	$^{+0.004}_{-0.008}$	$^{+0.003}_{-0.003}$
Total systematic uncertainty	$^{+0.040}_{-0.039}$	$^{+0.039}_{-0.037}$
Statistical uncertainty	$^{+0.015}_{-0.015}$	$^{+0.016}_{-0.016}$

Results



Post-fit, signal region, e-had channel (left), μ-had channel (right).

Channel	P_τ in mass range	P_τ in fiducial region
$\tau_e - \tau_{had}$	-0.195 ± 0.024 (stat) $^{+0.048}_{-0.050}$ (syst)	-0.331 ± 0.026 (stat) $^{+0.049}_{-0.052}$ (syst)
$\tau_\mu - \tau_{had}$	-0.129 ± 0.020 (stat) $^{+0.045}_{-0.046}$ (syst)	-0.259 ± 0.021 (stat) $^{+0.046}_{-0.046}$ (syst)
Combination	-0.141 ± 0.015 (stat) $^{+0.041}_{-0.041}$ (syst)	-0.268 ± 0.016 (stat) $^{+0.040}_{-0.041}$ (syst)

Theory predictions: P_τ = -0.152 ± 0.001 P_τ^{fid} = -0.270 ± 0.006
LEP value (Z-pole): P_τ = -0.1439 ± 0.0043

Exercise
ATLAS Z→tau tau selection

- Monte-Carlo:
  - mc12/Ztautau.root - signal
  - Powheg_ttbar.root - background
  - Wenu.root - background
  - Wmunu.root - background
  - Wtaunu.root - background
  - Zee.root - background
  - Zmumu.root - background
- Data:
  - data12/Egamma.PhysCont.grp14.root - electron sample
  - **data12/Muons.PhysCont.grp14.root - muon sample**

  (at the beginning start working with the muon sample)
- The cross section/luminosity is included in the event weight *WeightLumi,* which is used to weight the MC events.

Tau Tau → hadronically (jet) + muon (electron) + neutrino

Preselection

```
	if(!(	evtsel_is_dilepVeto > 0 && evtsel_is_tau > 0 &&
fabs(evtsel_tau_eta) < 2.47 && evtsel_is_conf_lep_veto == 1 &&
evtsel_tau_numTrack == 1 && evtsel_lep_pt > 26 &&
fabs(evtsel_lep_eta) < 2.4 && evtsel_transverseMass < 70))
continue;

	if (!(  evtsel_is_oppositeSign>0 &&  evtsel_is_mu>0 &&
evtsel_is_isoLep>0 )) continue;
```

Selecting muon sample

ATLAS Z→tau tau selection

- Variables used for training:
  - *evtsel_tau_et*
  - *evtsel_dPhiSum*
  - *evtsel_tau_pi0_n*
  - *evtsel_transverseMass*
  - *sum_cos_dphi*

- Spectator
  - *vis_mass*

- Program:
  - TMVAClassificationMW.py and TMVAClassificationApplication.py

    First makes a basic training, the second applies the trained methods to the data.

ATLAS Z→tau tau selection

- Install root and TMVA
- Copy the programs and data:
  - Inside IFJ PAN:
    - http://nz14-46.4.ifj.edu.pl/cwiczenieATLAS/exerciseATLAS.tgz
  - From outside IFJ PAN:
    - http://nz14-46.ifj.edu.pl/cwiczenieATLAS/exerciseATLAS.tgz
- Run the code:
  - python TMVAClassificationMW.py
  - python TMVAClassificationApplicationMW.py
- Modify it and play with it:
  - **Optimize parameters of a selected method**
  - **Maybe we can skip some variables (or add)?**
  - **Try to use individual variables used to calculate *sum_cos_dphi.* Maybe Deep Neural Network and more variables might help?**

ATLAS Z→tau tau selection

Try to look, how the visible mass (spectator) changes after selection

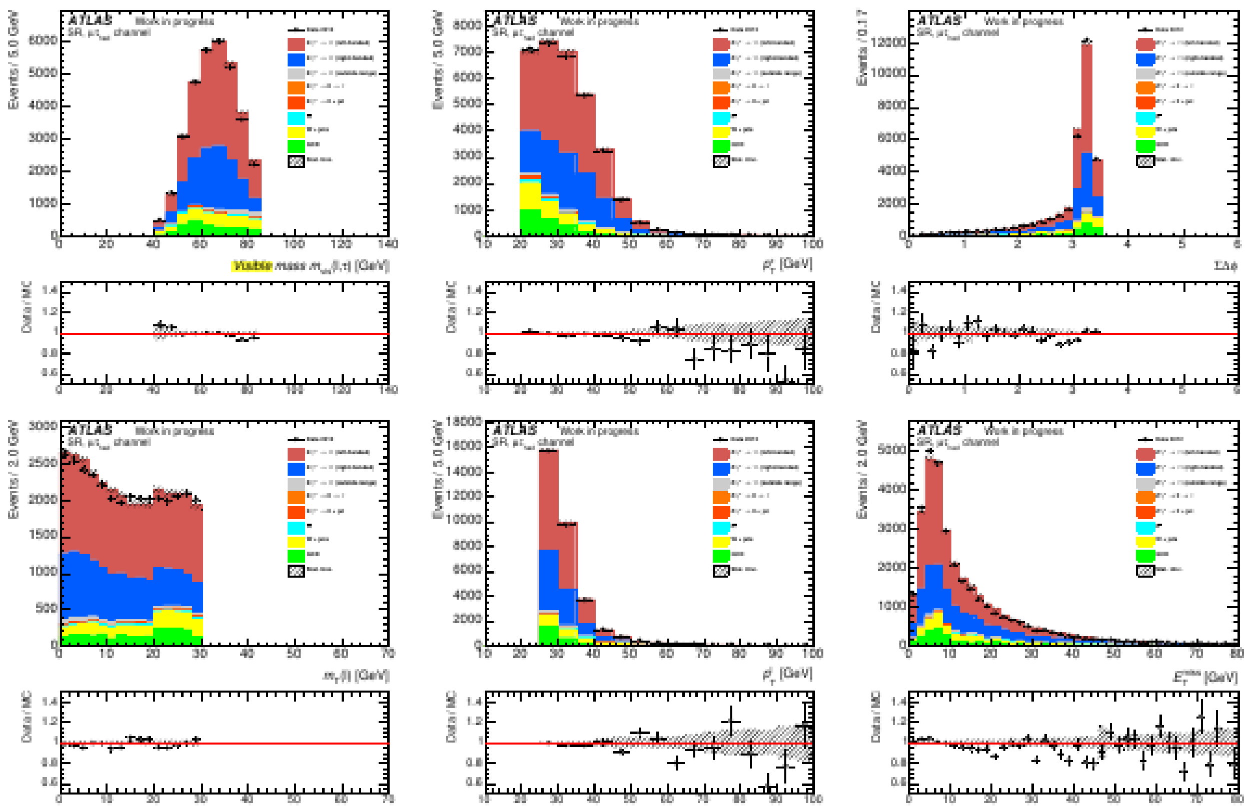


Figure 41: Distributions of variables observed in $Z \to \tau\tau$ (μ-had channel). From top-left: visible mass of τ-lepton system, τ transverse momentum, sum of polar angles between τ and missing-E_T and between lepton and missing-E_T, transverse mass of the lepton-missing-E_T system, lepton transverse momentum and missing-E_T.

ATLAS Z→tau tau selection

Event Selection and Background Estimate

Region / Cut	Signal Region	Same Sign	W Control Region	QCD Control Region
Single lepton trigger + offline lepton pT	evtsel_is_mu / evtsel_is_el			
Isolated Lepton	evtsel_is_isoLep			!evtsel_is_isoLep
Medium Tau ID	evtsel_is_tau			
Veto dileptons	evtsel_is_dilepVeto			
Muon Veto + medium Electron Veto	evtsel_is_conf_lep_veto_medium			
Single Prong tau	evtsel_tau_numTrack == 1			
Transverse Mass	evtsel_transverseMass < 30		evtsel_transverseMass > 70	
Sum Delta Phi	evtsel_dPhiSum < 3.5		evtsel_dPhiSum > 3.5	
Opposite Sign	evtsel_is_oppositeSign	!evtsel_is_oppositeSign	evtsel_is_oppositeSign / !evtsel_is_oppositeSign	evtsel_is_oppositeSign / !evtsel_is_oppositeSign

Cuts used in the analysis to select *Z → taut tau*
How do they compare to our Machine Learning result?

Machine Learning and HEP

- 90'ies - Neural Nets used by LEP experiments
- BDT (Adaboost) invented in 97
- Machine Learning used extensively at D0/CDF (mostly BDT, also Neural Nets) in the 00'ies
- Last years – mostly BDT built in TMVA ROOT package (popular among physicists). Neural Nets and other techniques treated as obsolete.

- **Not much work within LHC experiments on studying possible better MVA techniques.**
- **Enormous development of Machine Learning in the outside world in the last 10 years ("Big Data", "Data Science", even "Artificial Intelligence" is back).**

- **We have to catch up and learn from computer scientists:**

Make an open Higgs challenge!

- **Task: identify H->tau tau signal out of background in the simulated data.**

How did it work ?

- People register to Kaggle web site hosted https://www.kaggle.com/c/higgs-boson . (additional info on https://higgsml.lal.in2p3.fr).
- …download training dataset (with label) with 250k events
- …train their own algorithm to optimize the significance (à la s/sqrt(b))
- …download test dataset (without labels) with 550k events
- …upload their own classification
- The site automatically calculates significance. Public (100k events) and private (450k events) leader boards update instantly. (Only the public is visible)
- 1785 teams (1942 people) have participated
- most popular challenge on the Kaggle platform (until a few weeks ago)
- 35772 solutions uploaded

Funded by: Paris Saclay Center for Data Science, Google, INRIA

Final leaderboard

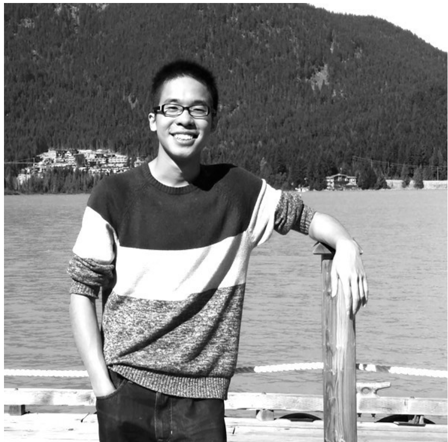
#	Δrank	Team Name ‡ model uploaded * in the money		Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑1	Gábor Melis ‡ *	7000$	3.80581	110	Sun, 14 Sep 2014 09:10:04 (-0h)
2	↑1	Tim Salimans ‡ *	4000$	3.78913	57	Mon, 15 Sep 2014 23:49:02 (-40.6d)
3	↑1	nhlx5haze ‡ *	2000$	3.78682	254	Mon, 15 Sep 2014 16:50:01 (-76.3d)
4	↑38	ChoKo Team		3.77526	216	Mon, 15 Sep 2014 15:21:36 (-42.1h)
5	↑35	cheng chen		3.77384	21	Mon, 15 Sep 2014 23:29:29 (-0h)
6	↑16	quantify		3.77086	8	Mon, 15 Sep 2014 16:12:48 (-7.3h)
7	↑1	Stanislav Semenov & Co (HSE Yandex)		3.76211	68	Mon, 15 Sep 2014 20:19:03
8	↓7	Luboš Motl's team	Best physicist	3.76050	589	Mon, 15 Sep 2014 08:38:49 (-1.6h)
9	↑8	Roberto-UCIIIM		3.75864	292	Mon, 15 Sep 2014 23:44:42 (-44d)
10	↑2	Davut & Josef		3.75838	161	Mon, 15 Sep 2014 23:24:32 (-4.5d)
45	↑5	crowwork ‡	HEP meets ML award XGBoost authors Free trip to CERN	3.71885	94	Mon, 15 Sep 2014 23:45:00 (-5.1d)
782	↓149	Eckhard		3.49945	29	Mon, 15 Sep 2014 07:26:13 (-46.1h)
991	↑4	Rem.		3.20423	2	Mon, 16 Jun 2014 21:53:43 (-30.4h)

The winners

- See http://atlas.ch/news/2014/machine-learning-wins-the-higgs-challenge.html
- 1 : **Gabor Melis** (Hungary) software developer and consultant : wins 7000$.
- 2 : **Tim Salimans** (Neitherland) data science consultant: wins 4000$
- 3 : **Pierre Courtiol** (nhlx5haze) (France) ? : wins 2000$
- HEP meets ML award: (team crowwork), **Tianqi Chen** (U of Washington PhD student in Data Science) and **Tong He** (graduate student Data Science SFU). Provided **XGBoost public software** used by many participants.

https://github.com/dmlc/xgboost

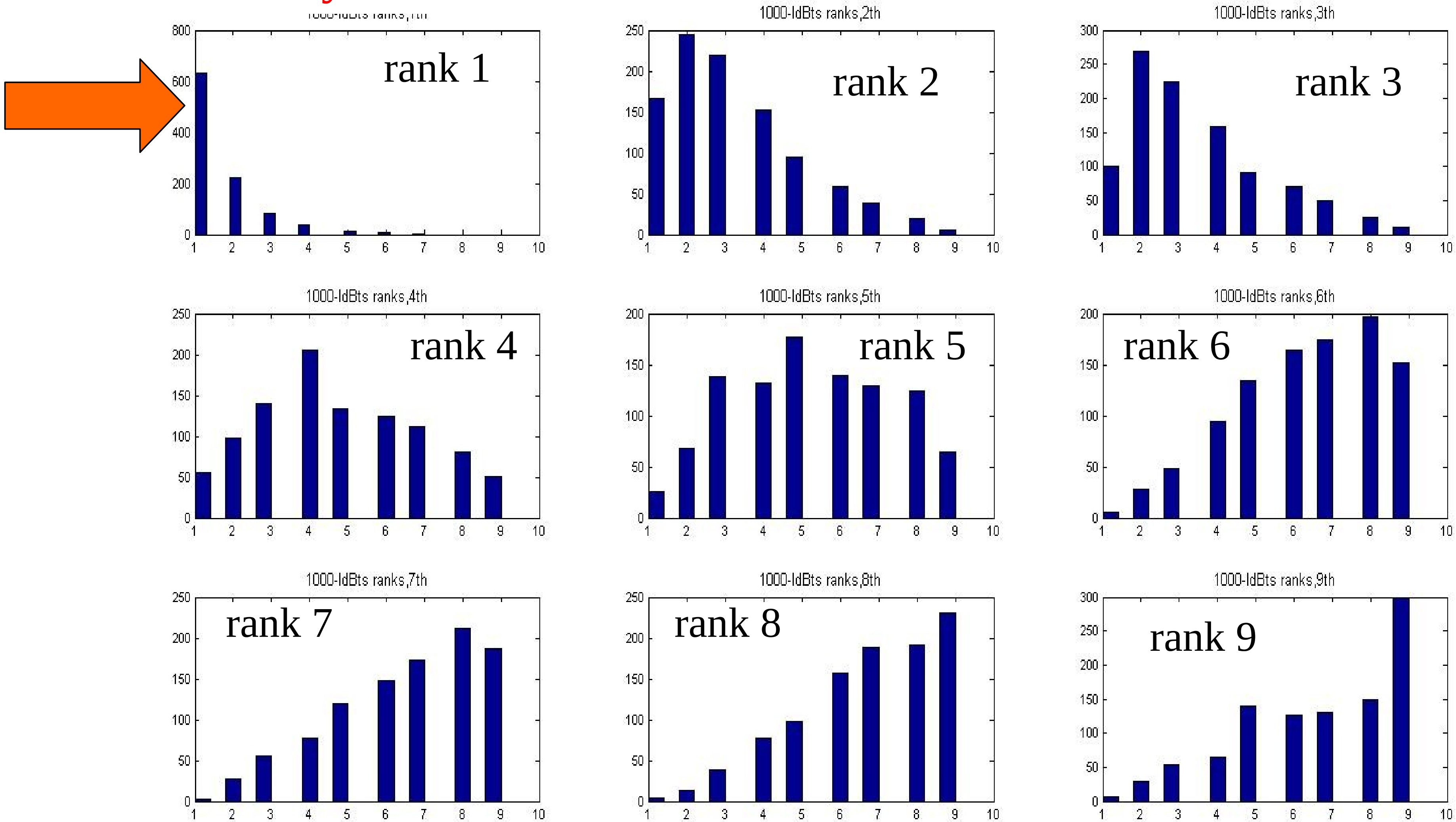
?



Rank distribution after bootstrap

Distribution of rank of participant of rank i after 1000 bootstraps of the test sample.

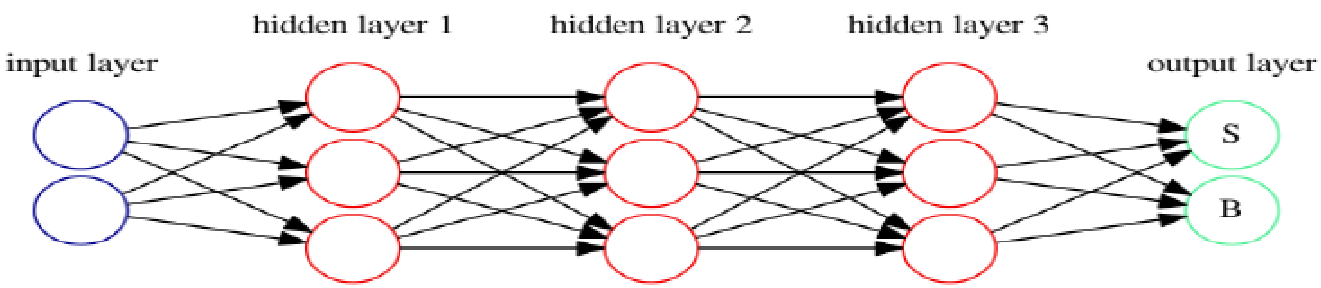
! Gabor clearly better



XGBoost)

Deep neural network

- Hierarchical feature extraction – first build abstract objects, than find dependencies between them.
- Deep neural network (DNN)- an artificial neural network with multiple hidden layers of units between the input and output layers.
- Extra layers - composition of features from lower layers, potential of modeling complex data with fewer units than a similarly performing shallow network.



- inputs: normalized features (˜30), some log transformed
- 3 hidden layers of 600 neurons each
- output layer: 2 softmax units (one for signal, one for background)
- activation function: “max channel” in groups of 3
- trained to minimize cross entropy
- regularization: dropout on hidden layers, $L_1 + L_2$ penalty and a mild sparsity constraint input weights

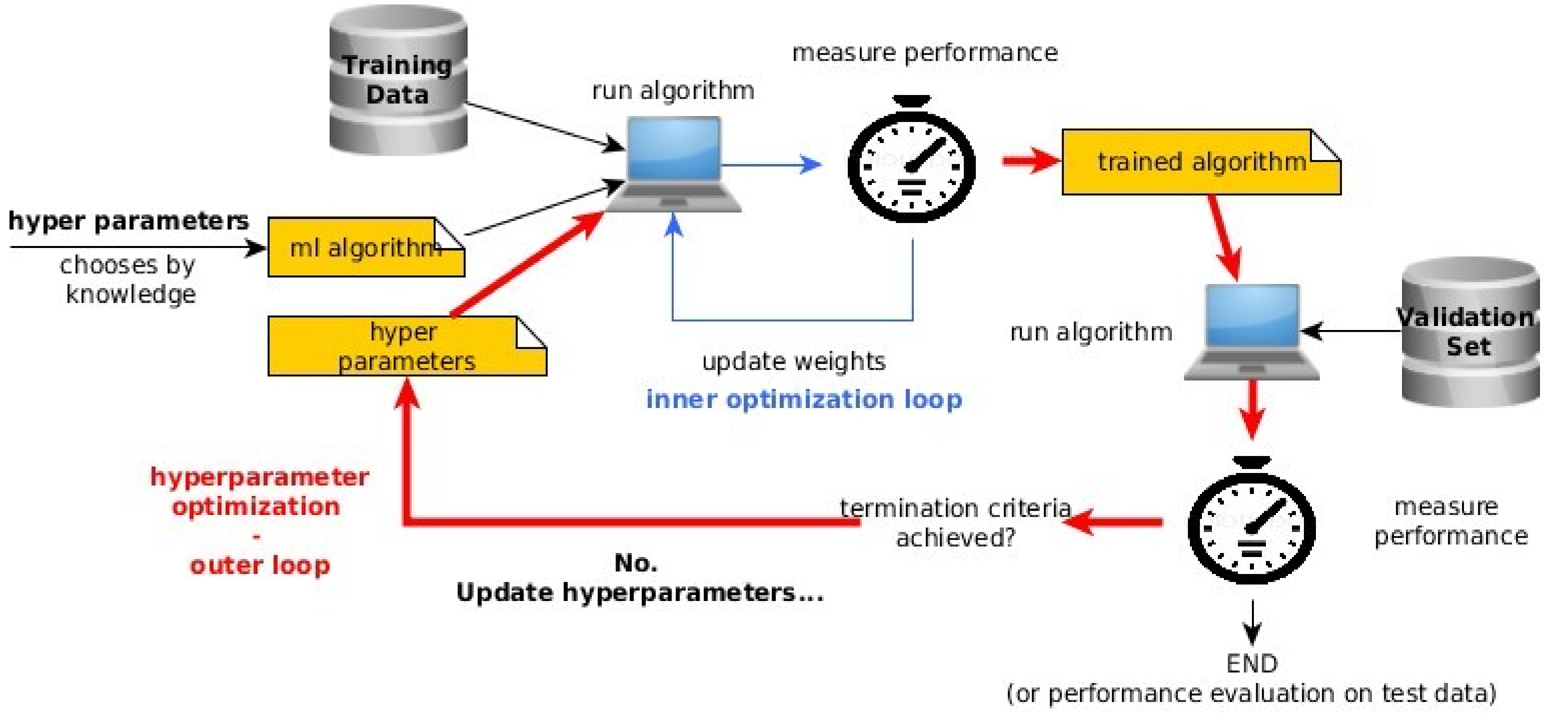
Challenge winning
Gabor's deep neural network
(from Gabor's presentation)

- CV bagged NNs: 3.83
- CV bagged xgboost: 3.79

Remark:
Few years ago some experts claimed neural networks are an obsolete tool :)

Automatic optimization of hyperparameters

- Manual optimization of NN (or any other method) is time consuming.
- Fortunately the Bayesian optimization methods can rival and surpass human domain experts in finding good hyperparameter settings.
- SMAC, SPEARMINT, TPE (and others) are doing that with great success: http://www.cs.ubc.ca/~hutter/papers/13-BayesOpt_EmpiricalFoundation.pdf



Analiza podczas praktyk studenckich

- Próbowaliśmy powtórzyć HiggsChallenge podczas praktyk studenckich.
- Udało się za pomocą TMVA (konwersja danych do formatu root) oraz pakietu XGBoost
- Optymalizacja parametrów XGBoost za pomocą programu hyperopt

A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss (2009)
TMVA 4 Package Documentation
https://tmva.sf.net

Tianqi Chen, Tong He, Bing Xu and Michael Benesty (2014)
XGBoost Package Documentation
https://github.com/dmlc/xgboost

James Bergstra, Dan Yamins, and David D. Cox (2013)
Hyperopt Package Documentation
https://github.com/hyperopt

Rozwiązania Kaggle-Higgs vs Hyperopt

Porównianie wyników uzyskanych przez nas automatycznie z wynikami z najlepszymi znalezionymi parametrami dla XGBoost.

Kto	9. K-H	M. Wolter	Nasze obliczenia
Maks. głębokość	9	10	9
Wsp. uczenia	0.01	0.089	0.059
Liczba drzew	3000	150/250/500	300
Liczba testów	-	300	100
Sub_sample	0.9	1	0.9
Maks. ROC	0.987	0.933/0.934/0.933	0.934

Sub_sample - jaka cześć danych brana jest do procesu uczenia - wprowadza pewną losowość i zapobiega przeuczaniu

Jak widać wyniki przez nas osiągnięte są znacznie słabsze. Prowadziliśmy poszukiwania w innym regionie parametrów.