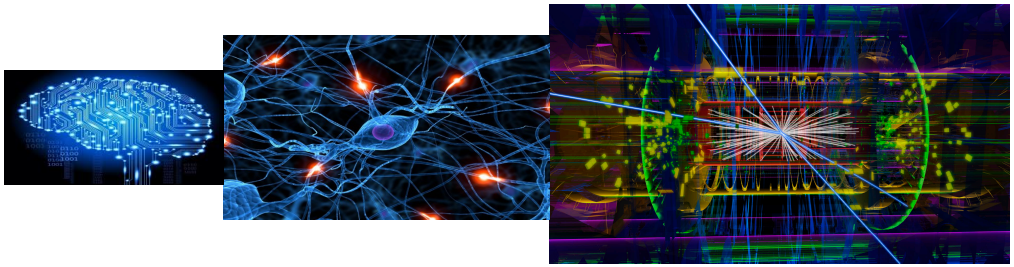
PHYSICS AT THE TERA SCALE
Helmholtz Alliance

DESY

MVA Tutorials

S. Gleyzer[1], H. Prosper[2], C. Rosemann[1]

[1]DESY, [2]Florida State University



DESY Statistics School 2014

April 3, 2014

Outline

Classification

- Exercise A: Simple Gaussians 30 min
  - Methods: eye, GA cut optimization
- Exercise B: Simple Gaussians 30 min
  - Methods: BDT and MLP
- Exercise C: Higgs to ZZ Example 90 min
  - Methods: Various

Regression

- Exercise D: Calorimeter example 90 min

Optional Exercises E, F: Advanced BDT/MLP/BNN

Tutorials

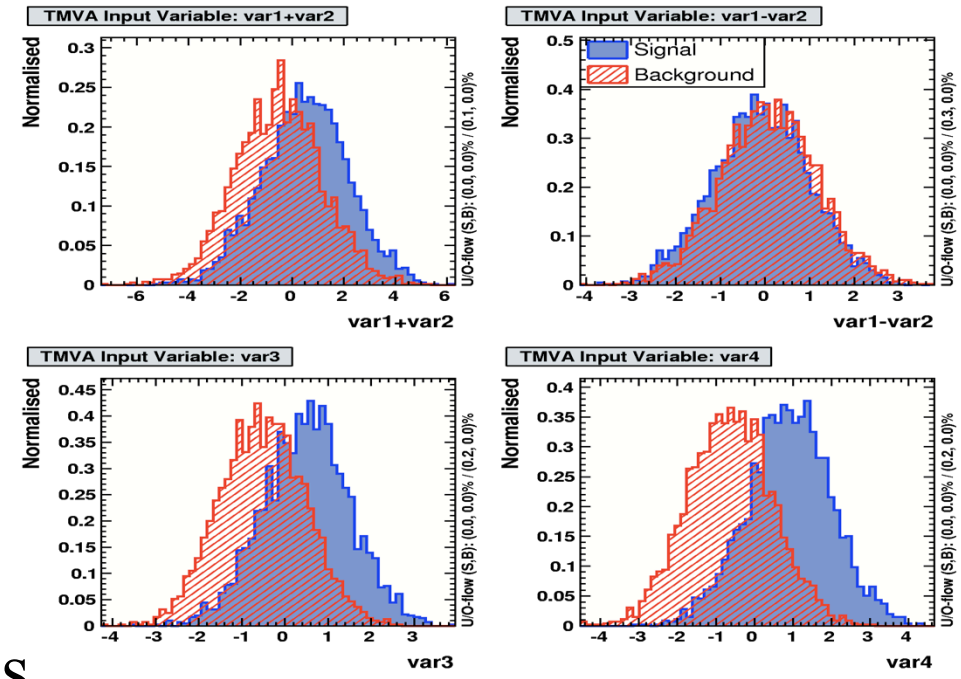
All the exercises are located in:

/afs/desy.de/group/school/statschool2014/

Exercise A: Simple cuts

1. Simple gaussians

- 4 variables
- Make "cuts" by eye



2. Cut optimization using TMVA

- Simple plus correlations
- Genetic Algorithm Cut Optimization

Exercise B: BDT+MLP

Simple gaussians + correlations

- Train BDT and MLP
  - compare with exercise A2
- Look at classifier performance
  - ROC curves
- Study tree and network architectures in detail
  - Do they make sense for this problem?
- Modify macro
  - try to optimize these two classifiers

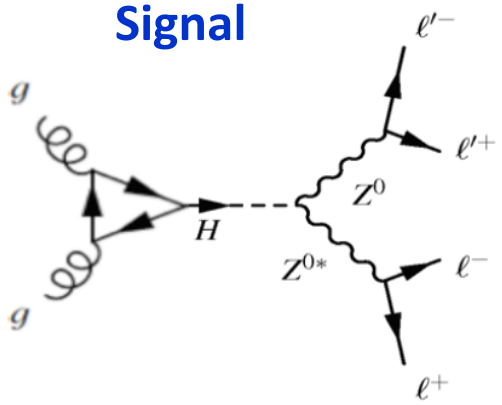
Exercise C: H to ZZ

H->ZZ-> 4leptons

- Train Random Grid Search (RGS) and Kernel Density Estimator (KDE)
- Train BDT and MLP classifiers using TMVA
- Plot Z_2 vs Z_1 mass for signal and background
- Look at classifier outputs and surface plots with plot.py
- What can you conclude from the RGS result?
- Compare different classifiers. Which would you use?

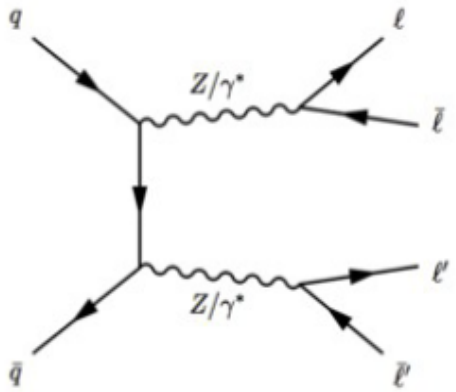
Exercise C: H to ZZ

Signal

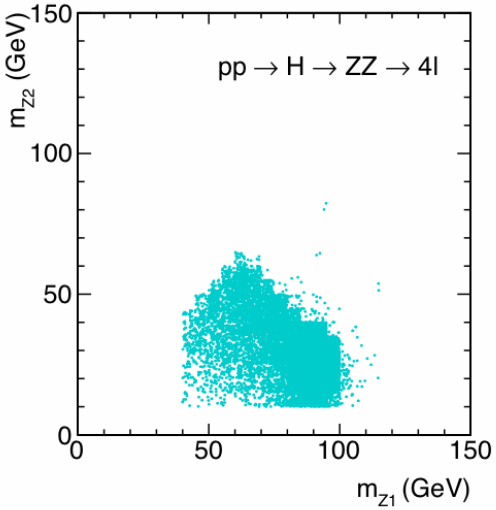


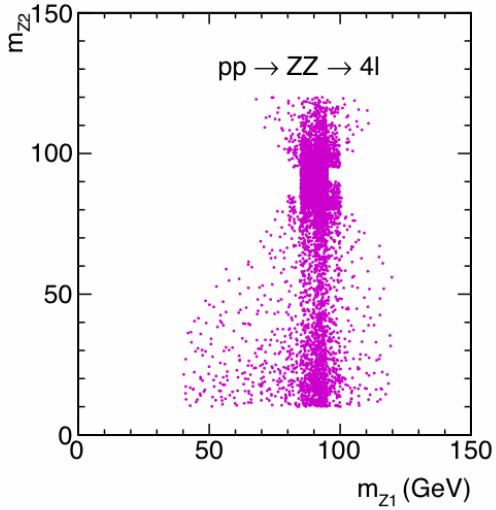
$$pp \to H \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$$

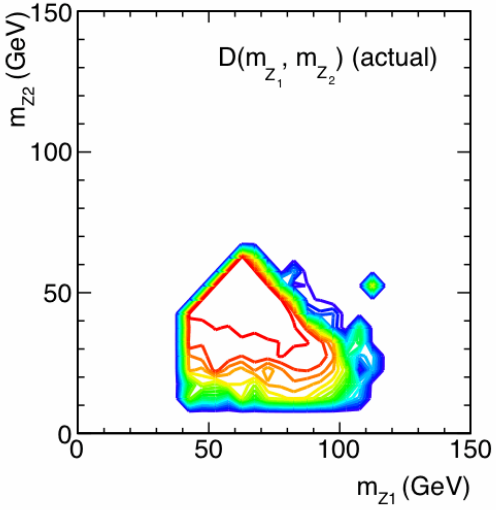
Background

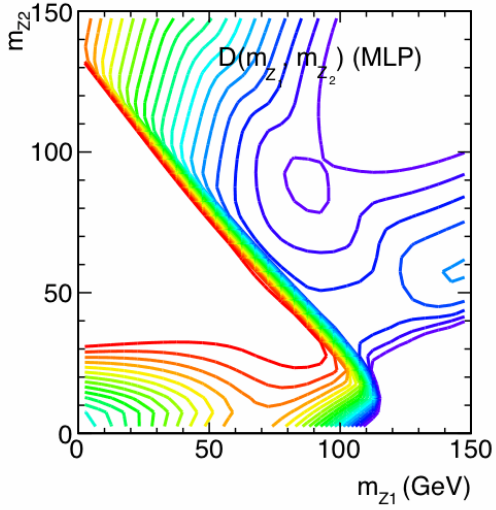


$$pp \to ZZ \to \ell^+\ell^-\ell'^+\ell'^-$$









Exercise D: Regression

Toy Calorimeter Regression

- 5 thin and 8 thicker layers, non-compensating
  - $E_0 \ldots E_{12}$
- **Leakage** and dead regions present
- One cluster per event from jets or single particles
- Energy sum over all layers $E_{SUM} = \Sigma e_i$
- True energy E_{TRUTH}
- Build a classifier to Estimate E_{TRUTH}/E_{SUM}

GOAL: smallest standard deviation of target vs. estimated

Exercise E: Classification

Optional Classification Exercise

- Complicated signal shape, non-linear correlations
- Train Likelihood, MLP or BDT
- Optimize classifiers as much as you can and compare results (in TEAMs)

Exercise F: Regression

Also Optional!

Bayesian Neural Network Regression

- f(x,y) = sin(x) • cos(y)
- Fit BNN to data
- Plot the result
  - Can you guess which is the true function

You are DONE!