Higgs Challenge

Jan Küchler

MLHEP 2016

26th June 2016

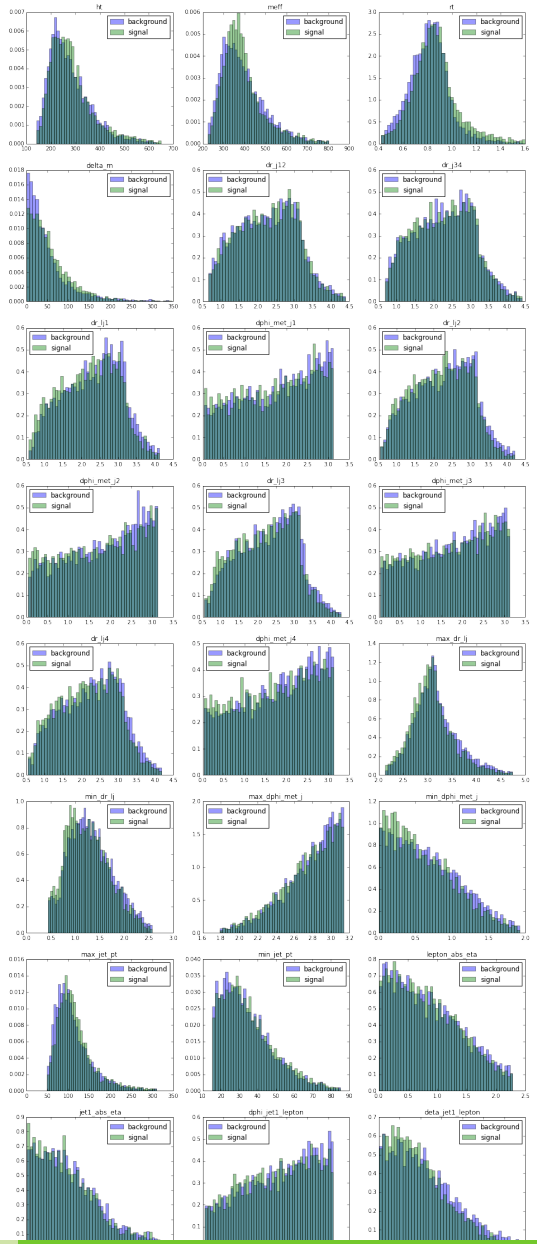
Classifier

BERGISCHE UNIVERSITÄT WUPPERTAL

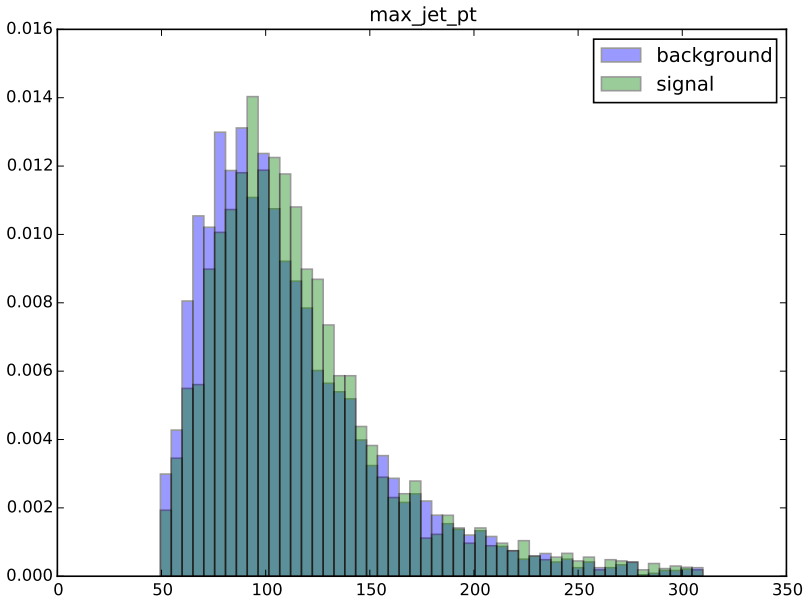
- main part: deep neural network
  - 5 dense layers, 700 neurons each
  - `batch_norm`
  - `dropout` ($p = 0.15$)
  - trained using `adadelta` on minibatches of 2000 events
  - all features (provided + custom ones)
- $\oplus$ `XGBoost` boosted decision tree
  - weighted mean, 10% from the BDT, 90% NN
  - 6-fold CV
  - hyperparameters optimised via grid search
  - 50 most important features
- $\oplus$ feature engineering

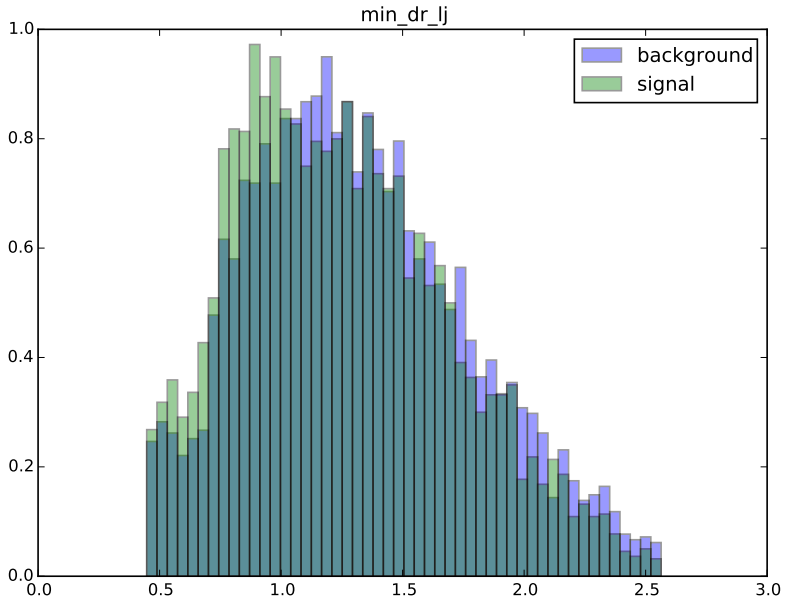
Feature importance

	Feature	XGBoost ranking
1	m_bb	2268
2	lepton_pt	2212
3	m_jjj	1986
4	m_jlv	1965
5	m_wwbb	1891
6	m_jj	1862
7	mem_pt	1808
8	m_wbb	1646
9	**max_jet_pt**	1558
10	**min_dr_lj**	1281
11	**delta_m**	1219
12	m_lv	1154
13	**rt**	1144
14	**ht**	1111
15	**max_dr_lj**	1063



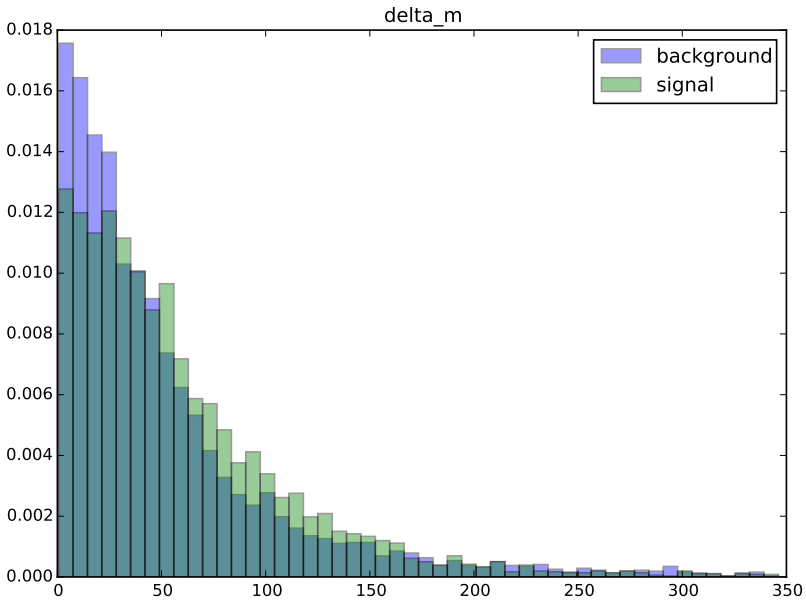
Feature Engineering

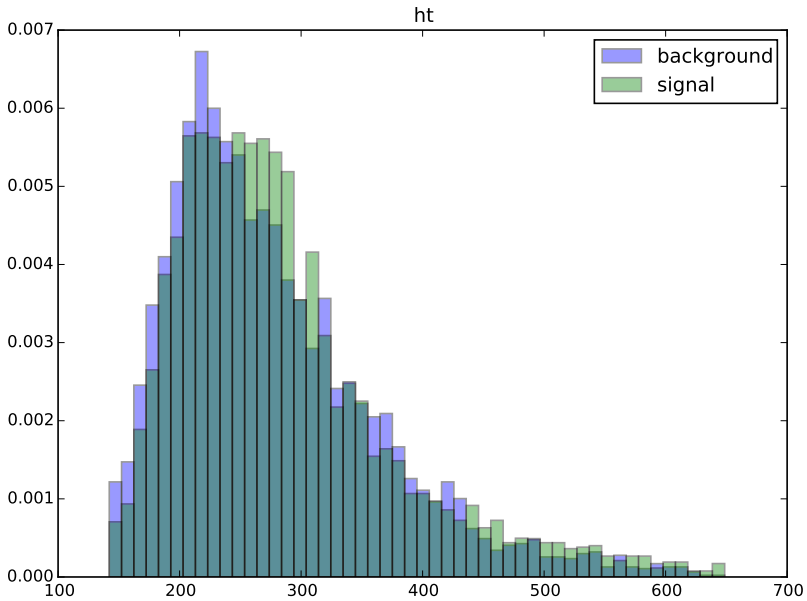




- $\max\left(p_{\mathrm{T,jet}}\right)$
- $\min\left(\Delta R\left(\ell, j\right)\right)$ $\qquad \Delta R(\ell, j) = \sqrt{(\eta_\ell - \eta_j)^2 + (\phi_\ell - \phi_j)^2}$

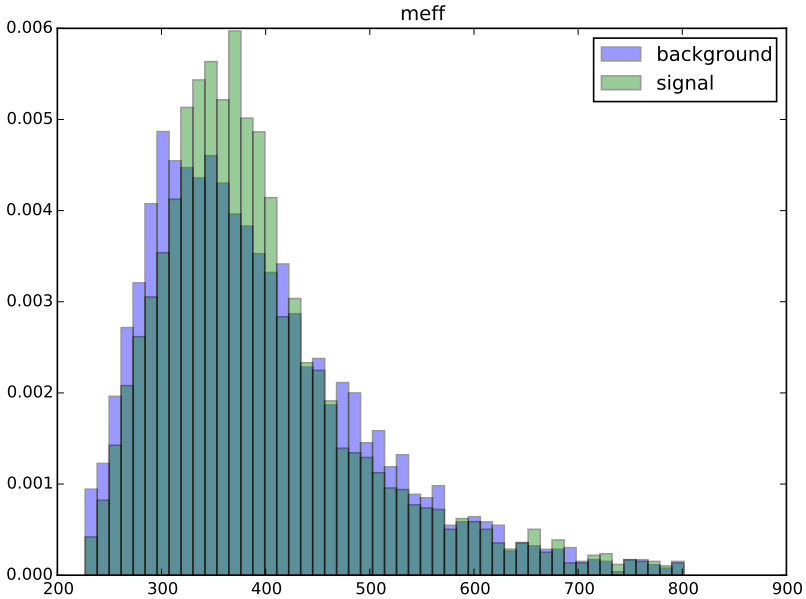
Feature Engineering

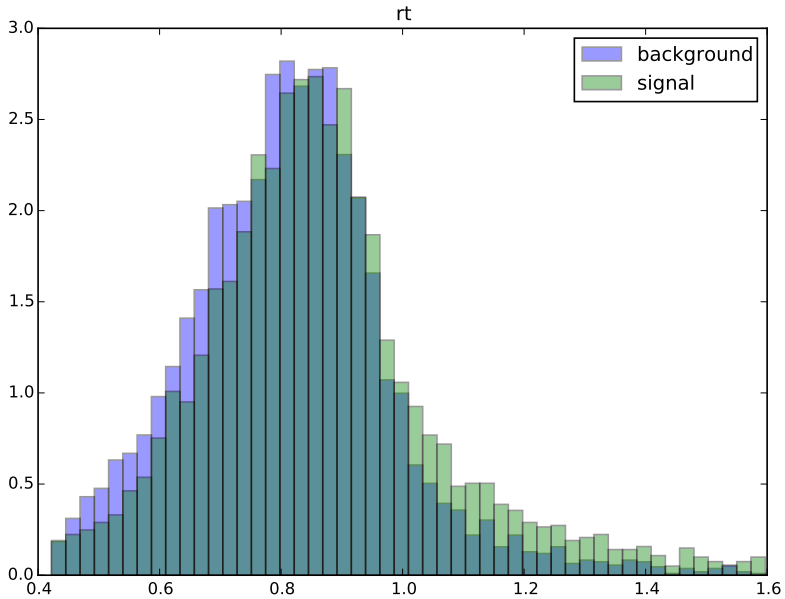




- $\Delta m = |m_{\mathrm{jjj}} - m_{j\ell\nu}|$
- $H_{\mathrm{T}} = \sum p_{\mathrm{T,jet}}$

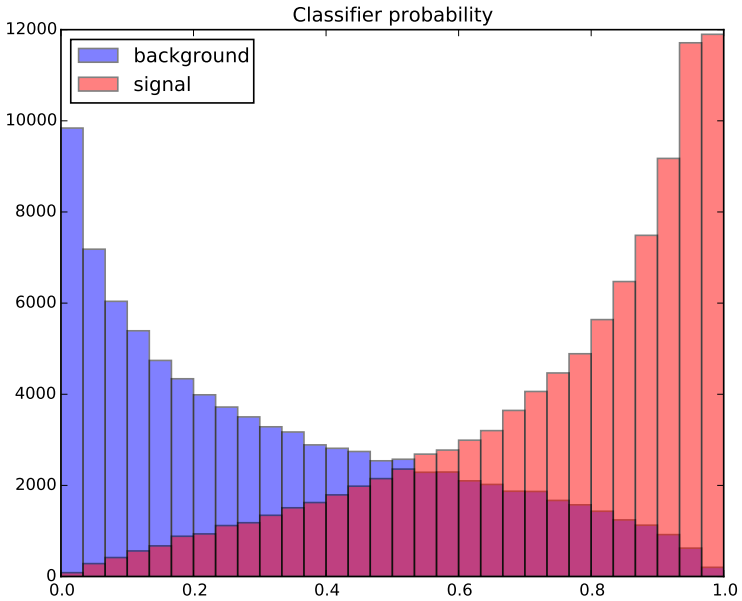
Feature Engineering

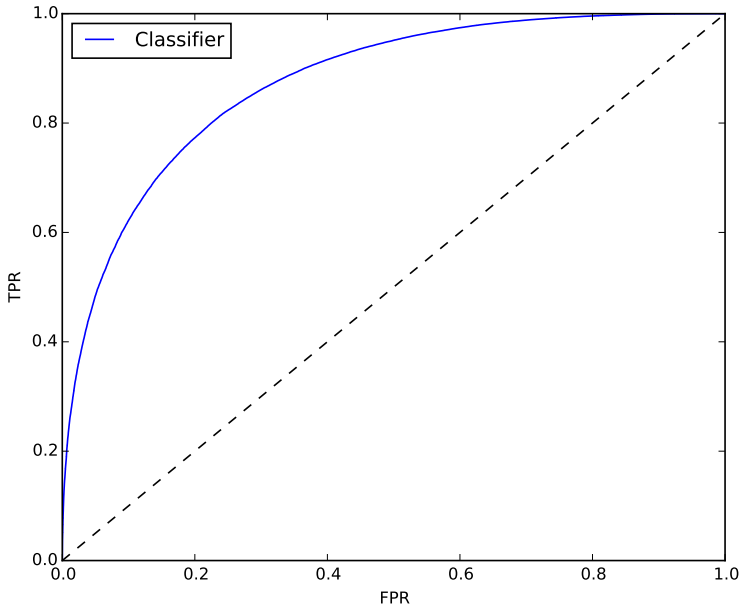




- $m_{\mathrm{eff}} = \sum p_{\mathrm{T,jet}} + p_{\mathrm{T},\ell} + E_{\mathrm{T}}^{\mathrm{miss}}$
- $R_{\mathrm{T}} = m_{\mathrm{eff}} / m_{WWbb}$

Classifier and ROC curve





- Jupyter notebooks:
  - [Features and BDT]
  - [Neural Network]
  - [Combination and final evaluation]

  ⇒ `https://github.com/JanKuechler/mlhep2016/tree/master/higgs_kaggle`