Identification of Hadronically Decaying Tau Leptons at the ATLAS Detector Using Deep Neural Networks

Nico Madysa

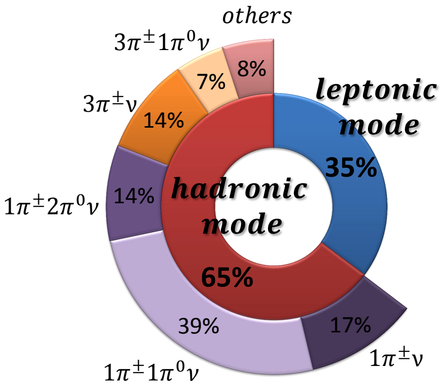
March 10, 2015

DPG-Frühjahrstagung Wuppertal

The Task

Distinction Between

- jets from hadronic tau decays (signal)
- jets from pure QCD processes (background)



Hadronic Tau Decay

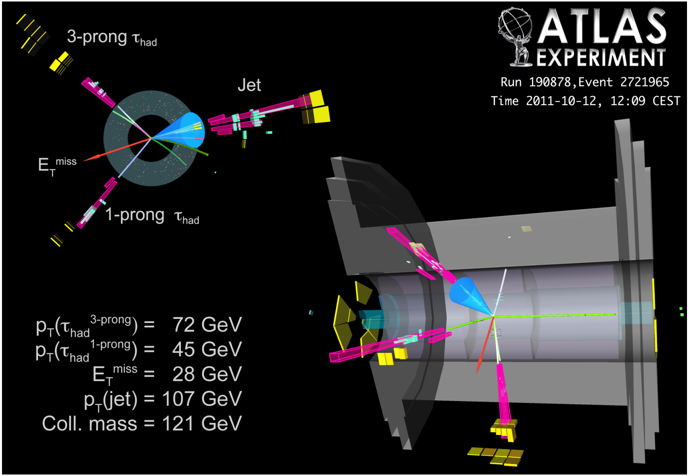


Figure: ATLAS Experiment © 2014 CERN

Binary Classfication

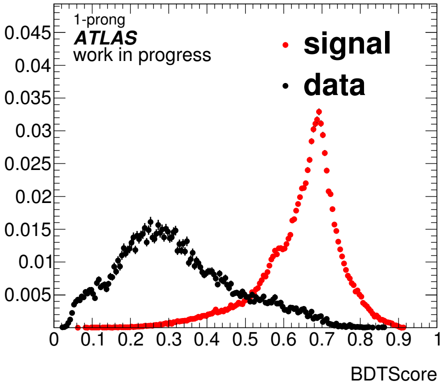
Binary Classifier

- input: data points $\vec{x} \in \mathbb{R}^n$
- output: discriminating score $y \in [0; 1]$

Training

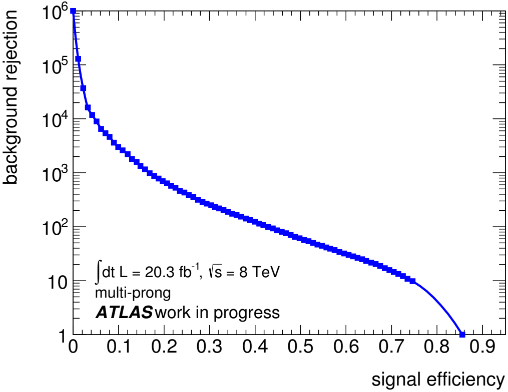
- training: iterative computation of the classifier's parameters
- training sample: data points with known results

Score Histograms



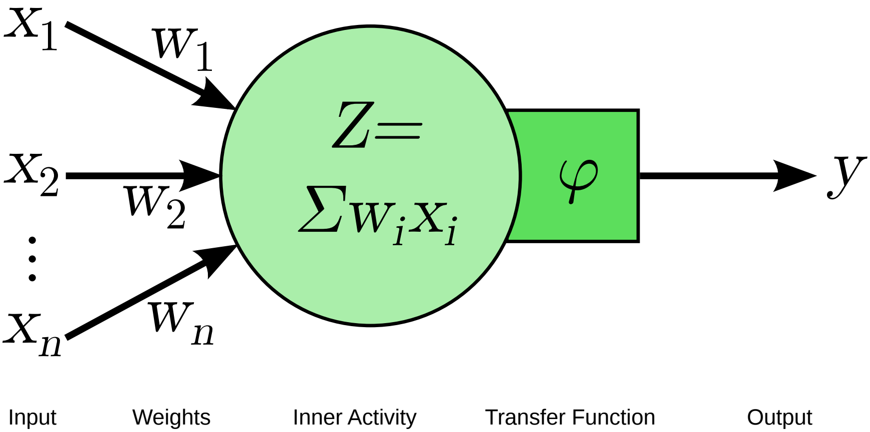
cut-off score: get discrete answer from continuous score spectrum

ROC Curves



figures of merit: *signal efficiency*, *background rejection*

Artificial Neurons



Single Layer Perceptrons

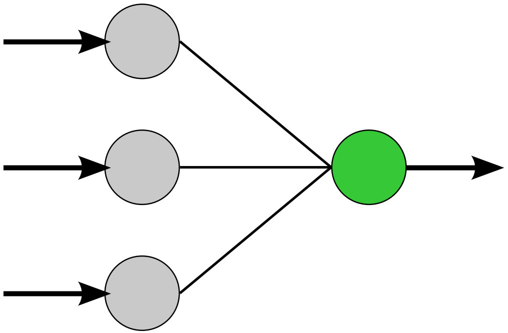
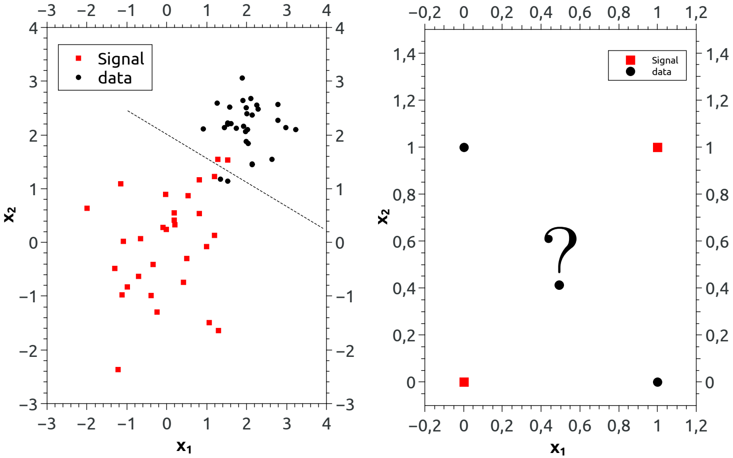


Figure: In binary classification, an SLP has only one active neuron.



(a) linearly separable problem (b) XOR problem

Multi Layer Perceptrons

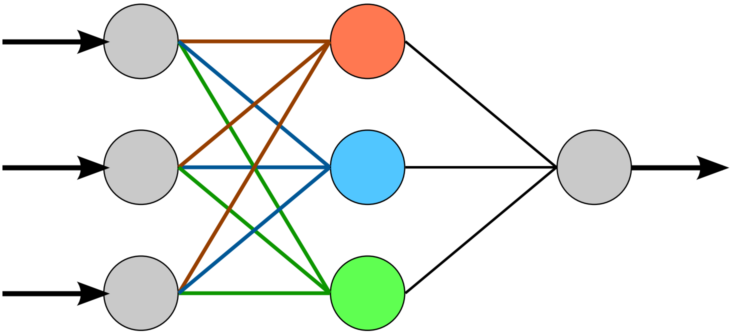


Figure: Each hidden neuron takes up one part of the problem.

Multi Layer Perceptrons

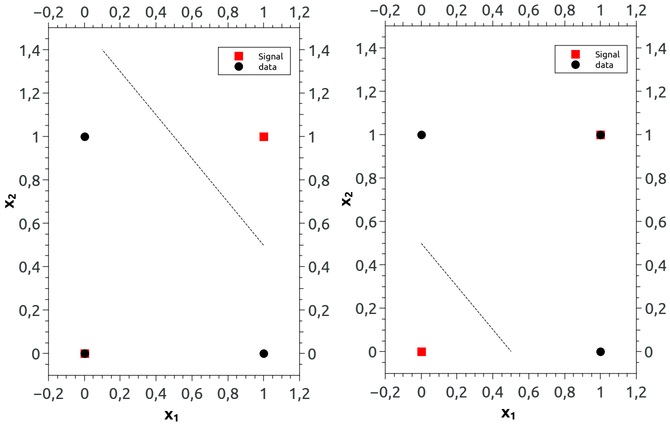


Figure: Solving the XOR problem with two hidden neurons.

Deep Neural Networks

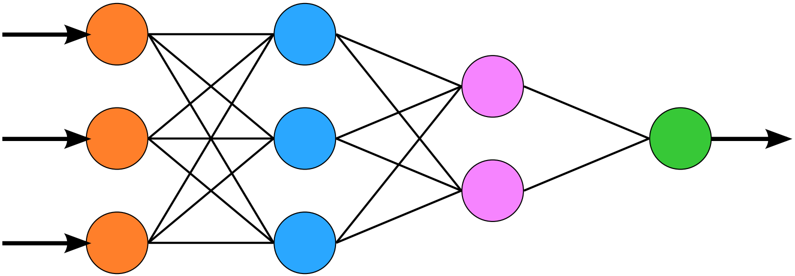


Figure: The first hidden layers transforms the input data.

Deep Neural Networks

Problem

- How many layers do we need?
- How many neurons per layer do we need?
- What training algorithm works best?

Solution

- scan over all possible options
- train and evaluate an ANN for each configuration
- compare with current tau identification algorithm

Implementation

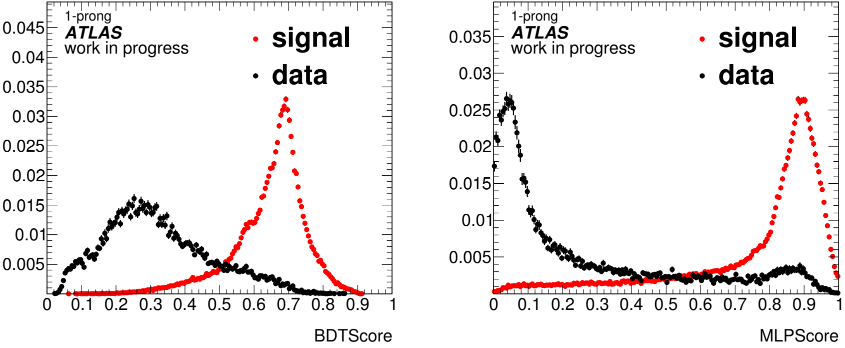
What we used

- TMVA's MLP algorithm
- training algorithm: Backpropagation & BFGS algorithm
- benchmark: Boosted Decision Tree (BDT)

Training sample

- signal: Monte-Carlo generated
  - contains processes $W \to \tau\nu_\tau$, $Z \to \tau\tau$, and $Z' \to \tau\tau$
- background: QCD jet data collected at ATLAS in the 2012 run

Score Histograms

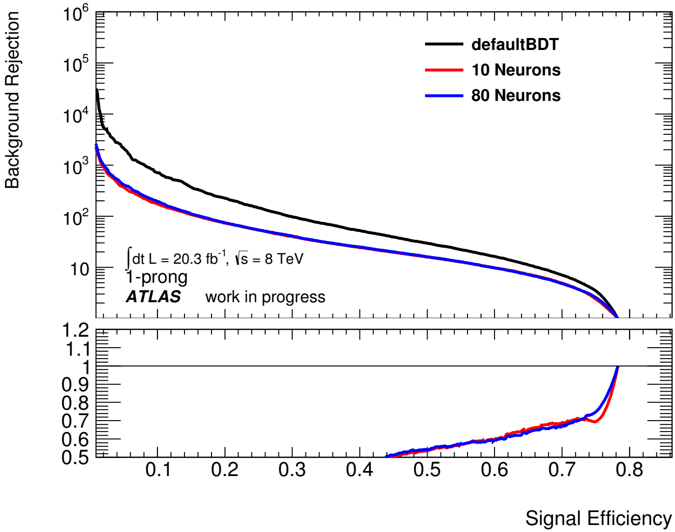


(a) Boosted Decision Tree

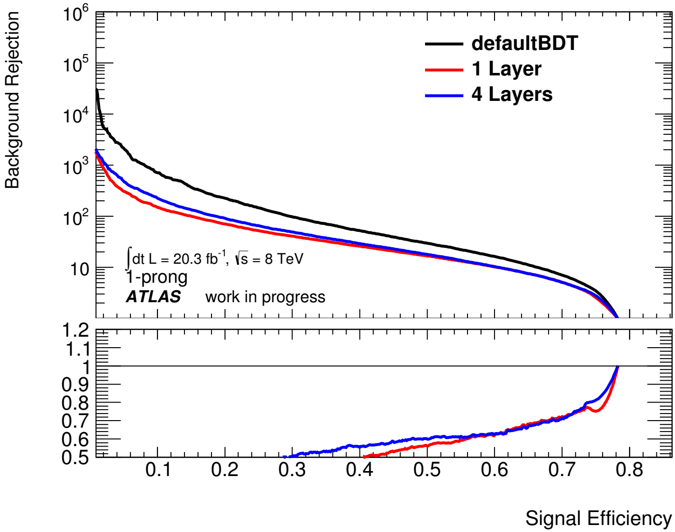
(b) BFGS, 1 hidden layer, 40 neurons

Figure: BDT/MLP Score Histograms for 1-Prong Jets

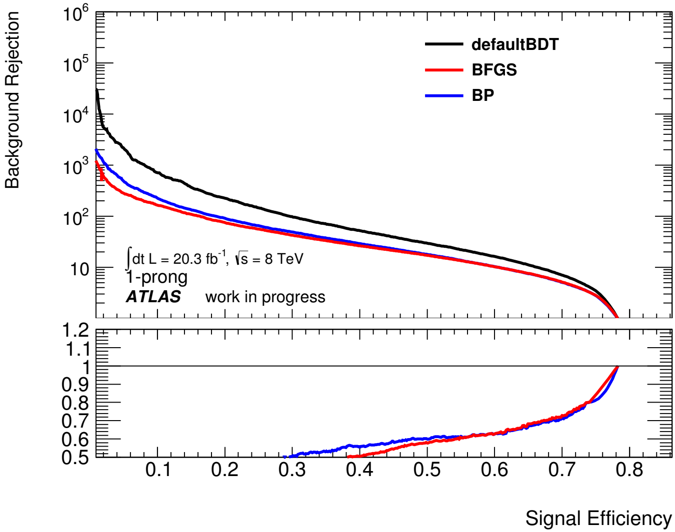
Number of Neurons (Backprop, 1 Hidden Layer)



Network Depth (Backprop, 20 Neurons/Layer)



Training Methods (4 Hidden Layers @ 20 Neurons)



Summary

Summary

- first attempt of NN-based tau identification
- comparison with current tau ID algorithm (BDT)
  - difference by a factor of ~ 2
- evaluation of performance under variation of
  - training method
  - neurons per layer
  - number of layers
- greatest influence: number of layers

Outlook

Outlook

- Is there a better choice of ID variables?
- Will even more layers improve the performance further?
- How do other training methods perform? (Quickprop, RProp, ...)
- Where does overtraining begin?
- If we reach overtraining, can more sophisticated algorithms help?

Backup

Artificial Neurons

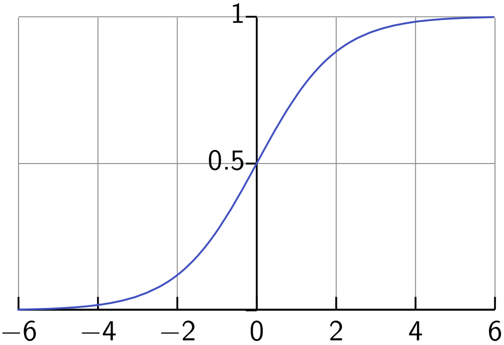


Figure: Transfer Function $\varphi(x)$

TMVA Options

BDT Options

- NTrees=100
- DoBoostMonitor=True
- H=False
- UseYesNoLeaf=False
- nCuts=200
- PruneMethod=NoPruning
- BoostType=AdaBoost
- AdaBoostBeta=0.2
- MaxDepth=8
- V=False
- MinNodeSize=0.1
- SeparationType=GiniIndex

MLP Options

- RandomSeed=42
- NeuronType=sigmoid
- NCycles=200
- BPMode=sequential
- TrainingMethod=[...]
- HiddenLayers=[...]
- VarTransform=N
- ConvergenceTests=5
- ConvergenceImprove=1e-08

ID Variables

Variables for 1-prong

- tau_calcVars_corrCentFrac
- tau_calcVars_corrFTrk
- tau_pi0_vistau_m
- tau_ptRatio
- tau_seedCalo_trkAvgDist
- tau_pi0_n
- tau_ipSigLeadTrk
- tau_seedCalo_wideTrk_n

Variables for 3-prong

- tau_calcVars_corrCentFrac
- tau_calcVars_corrFTrk
- tau_pi0_vistau_m
- tau_ptRatio
- tau_seedCalo_trkAvgDist
- tau_pi0_n
- tau_massTrkSys
- tau_seedCalo_dRmax
- tau_trFlightPathSig

Training Samples

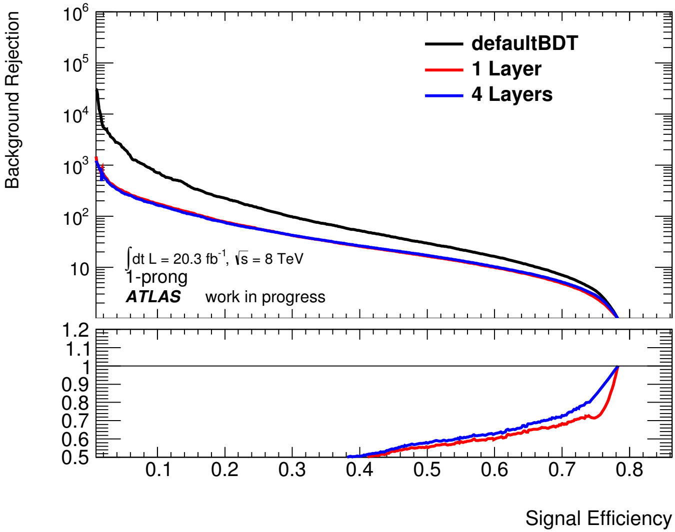
Signal Monte Carlo Samples

- $W \to \tau\nu_\tau$ (64k events)
- $Z \to \tau\tau$ (68k events)
- $Z' \to \tau\tau$, $m_{Z'} = 250\,\text{GeV}$ (62k events)
- $Z' \to \tau\tau$, $m_{Z'} = 500\,\text{GeV}$ (62k events)
- $Z' \to \tau\tau$, $m_{Z'} = 1000\,\text{GeV}$ (61k events)

Background Samples

- period D QCD jet data collected at ATLAS in the 2012 run (326k events)

Network Depth (BFGS, 20 Neurons/Layer)



Training Methods (1 Hidden Layer, 20 Neurons)