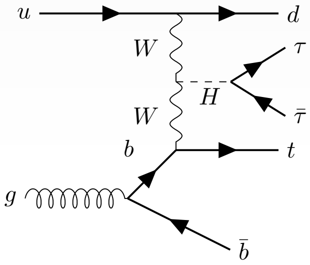
Categorical neural networks for background separation in the analysis of single top-quark production in association with a Higgs boson

Christian Kirfel

15th of September 2021

tHq
A binary classifier

Channel selection



- n-jets: at least 2 (b-jets: $>\mathbf{0}$)
- b-jet WP: 70 DL1r
- nLeptons & nTaus: $\mathbf{1e}/\mu\ \mathbf{2}\tau_{\text{had}}$
- $E_{\text{T,miss}}$: no cut (to 800 GeV)

- jets:
  - $p_T > 35$ GeV
  - $|\eta| < 4.5$
  - EMPFlow
- electrons:
  - $p_T > 20$ GeV leading 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: LooseAndBLayerLH ; isolation: no requirement
- muons:
  - $p_T > 20$ GeV leading 27 GeV
  - $0.01 < |\eta| < 2.5$
  - WP: Loose ; isolation: no requirement
- taus:
  - $p_T > 20$ GeV leading 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: RNNLoose
  - ASG recommended OLR (τ_{had} remove jets)

Features

eta_jf	forward jet eta
pt_jf	forward jet transverse momentum
mass_jf	forward jet mass
phi_jf	forward jet phi
eta_b	b-jet eta
pt_b	b-jet transverse momentum
phi_b	b-jet phi
mass_b	b-jet mass
m_met	Missing energy
Reco_w_mass_2	Reconstructed mass of the W case 1
Reco_w_mass_1	Reconstructed mass of the W case 2
fs_had_tau_1_pt	Hadronic tau 1 pt
fs_had_tau_1_eta	Hadronic tau 1 eta
fs_had_tau_2_pt	Hadronic tau 2 pt
fs_had_tau_2_eta	Hadronic tau 2 eta

deltaRTau	Delta R of the hadronic taus
deltaPhiTau	Delta phi of the hadronic taus
HvisMass	mass of LorentzV sum of hadronic taus
HvisPt	pt of LorentzV sum of hadronic taus
HvisEta	eta of LorentzV sum of hadronic taus
TvisMass	mass of reconstructed top
TvisPt	pt of visible top
TvisEta	eta of visible top
M_b_jf	Mass of LorentV sum of b and jf
HT	Sum of transverse energies
lep_Top_pt	Light lepton pt
lep_Top_eta	Light lepton eta

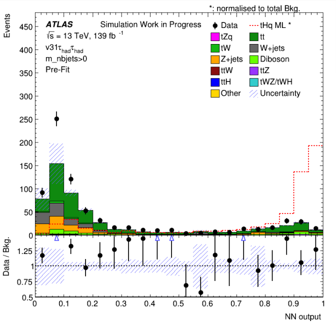
Training specifications for the lepditau channel

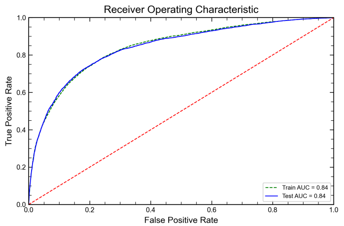
- Training a deep neural network using tensorflow.Keras
- Coarse optimisation using an evolutionary neural network
- Fine optimisation doing a grid search
- Optimised hyperparameters:
    - Number of nodes
    - Number of layers
    - Dropout percentage
    - Activation function
    - Weight initialisation
    - Optimiser
- Signal: tHq
- Background: Diboson, Z+jets, ttbar, W+jets, tW, tZq
- Using absolute weights

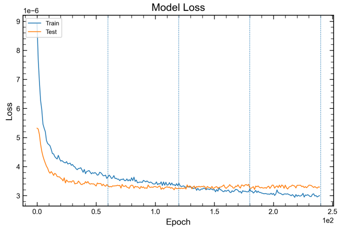
Hyperparameters

Hyperparameter	Setting
Model	Binary
Nodes	120
Layers	6
Dropout	0.65
Batchnormalisation	On
Activation	elu
Output activation	sigmoid
Batch size	1000
Optimisation	Adam
Weight Initialisation	Lecun Normalisation
K-folds	4

Results

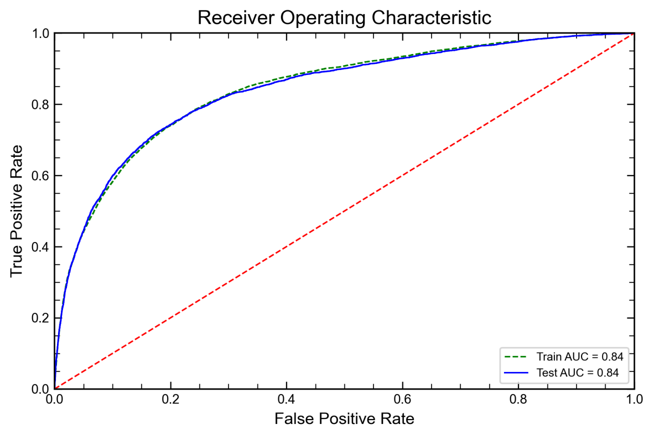






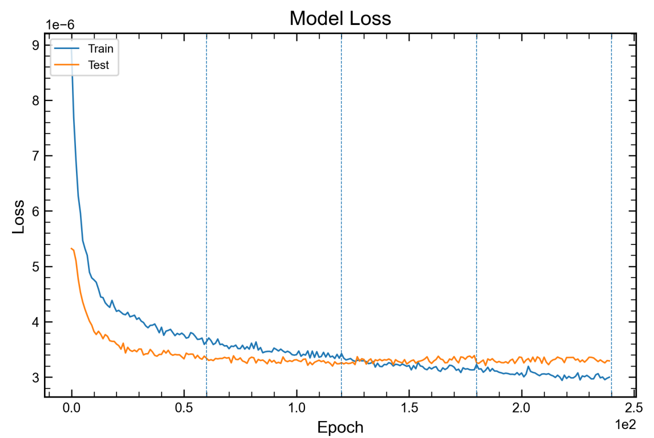
- Decent separation
- Stable agreement

The ROC curve



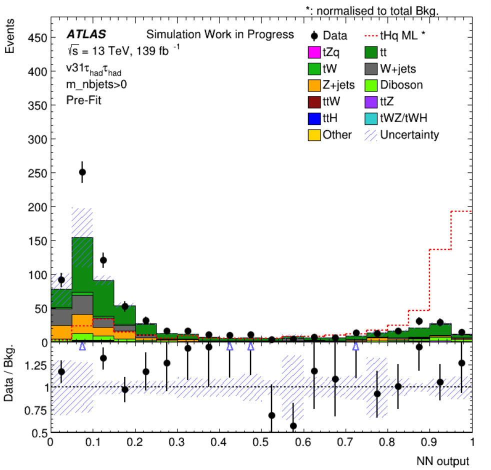
- 0.84 Area Under the Curve
- Good training and test agreement

The loss curve



- Test loss stable
- K-folds stable

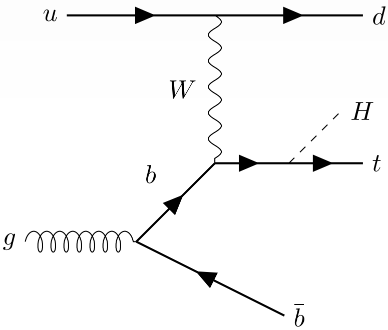
The response

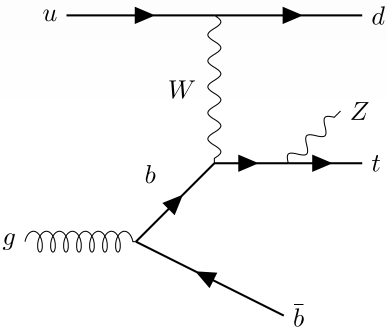


- Good separation shape

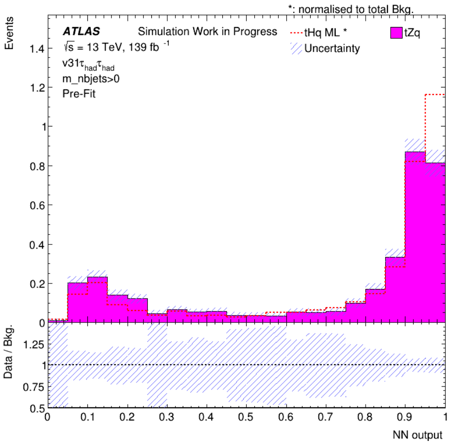
tZq
Weakness in a binary separation

tHq to tZq comparison





Response



- $t\bar{t}$ dominates the training
- tZq misclassified as signal

Categorical Neural Networks

The idea of categorical neural networks

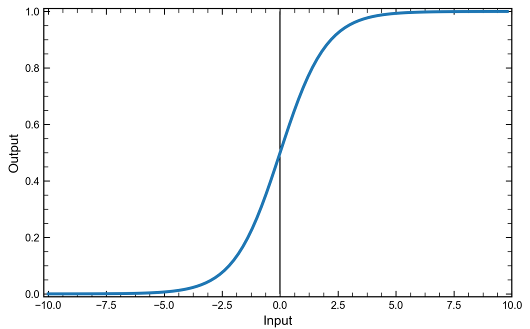
- Difficult backgrounds $\rightarrow$ additional targets

Target handling

- Using One-Hot-Encoding
- Vector instead of target
- Target values get translated to vector component

$$\mathrm{tH}q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mathrm{tZ}q = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} Background = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Binary final node - Sigmoid



- Bounded, differentiable
- Allows for backpropagation
- Assigning a single output value $\in (0, 1)$

$$\frac{1}{1 + e^{-s_i}}$$

- Expects only one input value

Categorical final node - Softmax

$$f(s) = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$$

- Takes an input vector equal to the number of targets
- Sum of vector components is 1
- normalises output to a probability distribution

Loss function - Crossentropy

Binary Crossentropy

$$Loss = -\sum_{i=1} y_i \log \hat{y}_i =$$

$$-y_1 \log \hat{y}_1 - (1 - y_1) \log(1 - \hat{y}_1)$$

- Measures model quality for two classes

Categorical Crossentropy

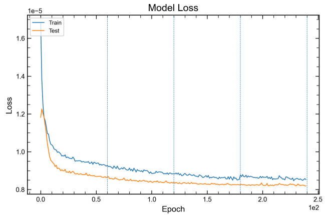
$$Loss = -\sum_{i=1} y_i \log \hat{y}_i$$

- Measures model quality for multiple classes

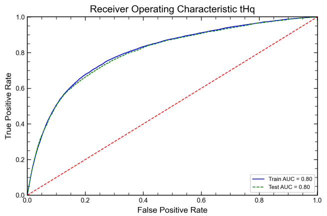
Hyperparameters

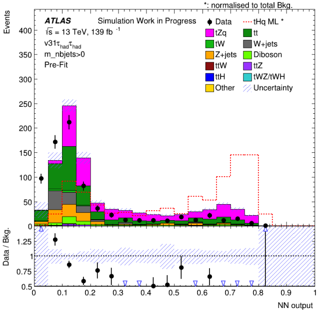
Hyperparameter	Setting
Model	Categorical
Nodes	120
Layers	6
Dropout	0.65
Batchnormalisation	On
Activation	elu
Output activation	Softmax
Batch size	1000
Optimisation	Adam
Weight Initialisation	Lecun Normalisation
K-folds	4

Results

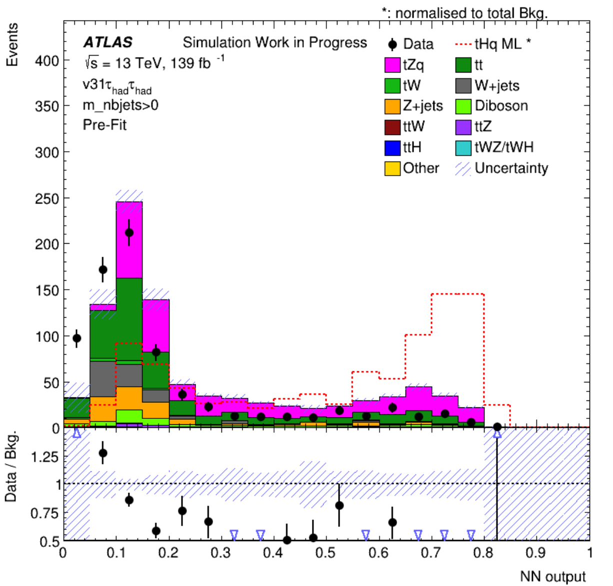


- Stable training
- Good AUC

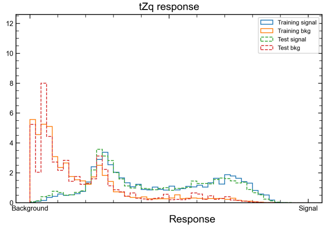


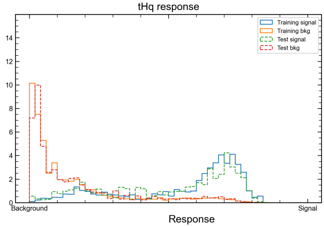


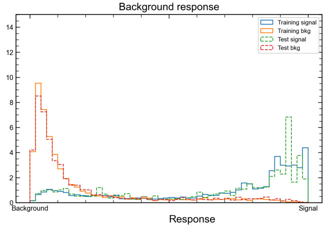
Response

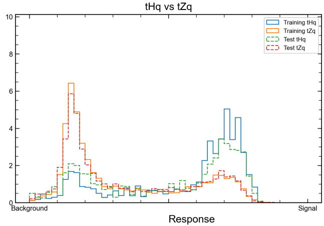

*: normalised to total Bkg.
ATLAS Simulation Work in Progress
√s = 13 TeV, 139 fb -1
v31τhadτhad
m_nbjets>0
Pre-Fit
Data
tHq ML *
tZq
tt
tW
W+jets
Z+jets
Diboson
ttW
ttZ
ttH
tWZ/tWH
Other
Uncertainty
Events
Data / Bkg.
NN output

Additional responses

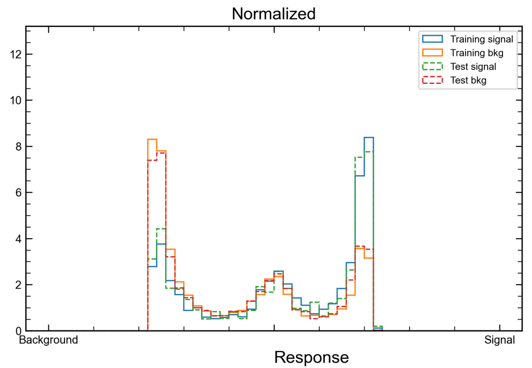








Stacking two networks



- Stacking two networks for additional tZq separation
- Recently tested
- Not obviously worth the effort

Conclusion

- Signal-like backgrounds are a weakness of binary classifiers
- Categorical neural networks are a promising approach and easy to implement
- The pollution of the signal by similar backgrounds is visibly decreased
- Combining the likelihoods could produce a more potent classifier
- Better Higgs mass reconstruction promises even more improvement