Categorical Neural Networks

Christian Kirfel

1st of September 2021

UNIVERSITÄT BONN

Neural Networks
An introduction

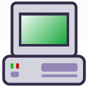
Neural Networks - Processing information

Humam senses

- Extraction of relevant info
- Impossible for machines

Human brain

- Web of neuron cells
- Input from surrounding cells
- Single combination $\rightarrow$ action



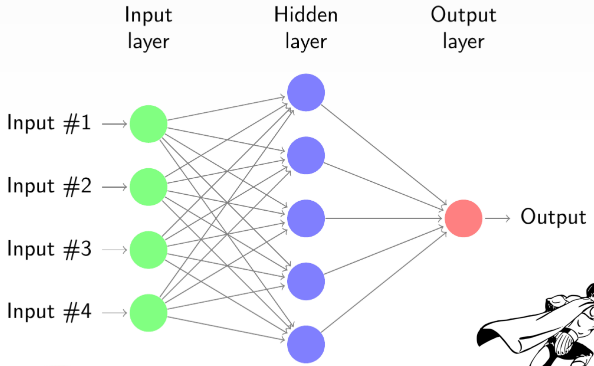
Input variables

- Preprocessed by user
- e.g. kinematic variables

Net of nodes

- Nodes = simple processors
- Connected by linear function
- Combination forms non-linear model

Neural network structure



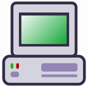
Neural Networks - Choosing the next step

Evaluation of an action

- Simple perceptions: pain, satisfaction
- Expectation

Decision for a next step

- Trial and error
- Learning from experience



Loss function

- Supervised learning: compare to the desired outcome
- Loss = estimator for quality

Optimisation

- Back-propagation impact of parameters' on the loss
- Adjust parameters to minimise plot

Hyperparameter optimisation

What is a hyperparameter?

- During the learning process the neural network optimises its internal parameters
- Some parameters are still set by the user according to the task of the network
- These are called hyperparameters

How does one optimise the choice?

- Neural networks provide several metrics to estimate result and performance
- To optimise the hyperparameters one usually runs several configurations to find a good set of parameters

Hyperparameters to remember

The core

Nodes	number of computational units
Layers	depth of computational units
Epochs	duration of training
Batchsize	batch of data we look at per time

Stuff you will encounter

Dropout	regularises, creates flexibility
Optimiser	calculates the next step
Activation function	creates non-linearity

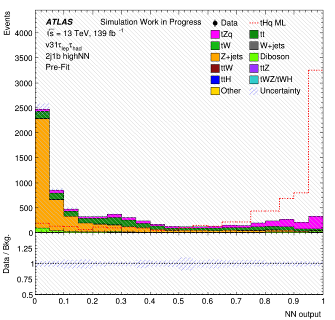
tHq
A binary classifier

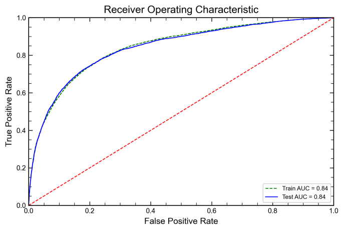
Hyperparameters

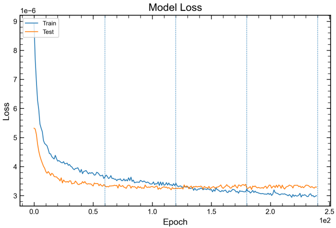
- Optimised by small grid search

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



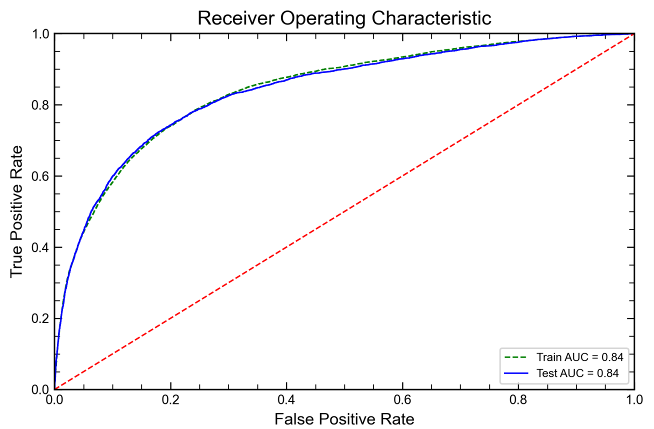




- Decent separation
- Stable agreement

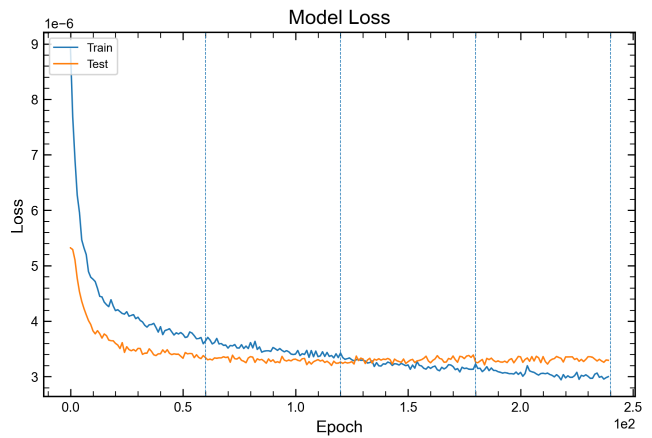
tHq
A closer look

The ROC curve



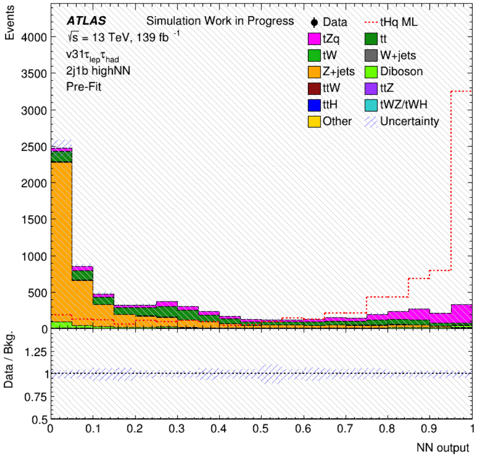
- AUC = Are Under the Curve
- Good estimator for model quality

The loss curve



- Should decrease
- Should not diverge

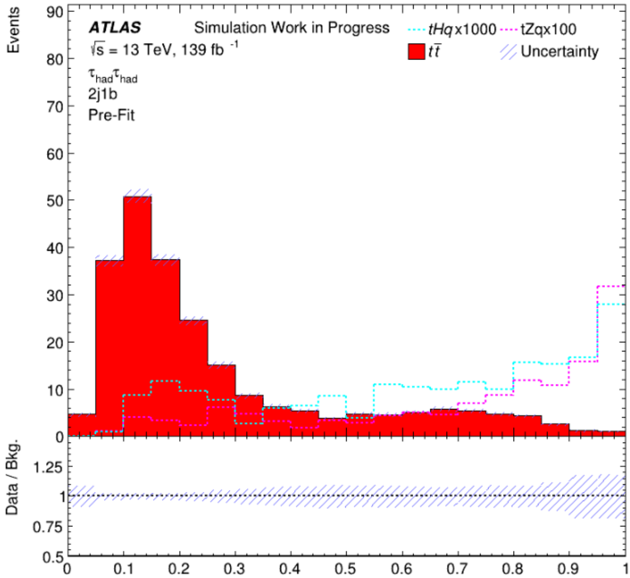
The response



- Seeing a separation is good
- Signal should be right

tZq
Weakness in a binary separation

Response


ATLAS Simulation Work in Progress
√s = 13 TeV, 139 fb -1
τhadτhad
2j1b
Pre-Fit
tHq x1000
tZqx100
tt̄
Uncertainty
Events
Data / Bkg.

Categorical Neural Networks

The idea of categorical neural networks

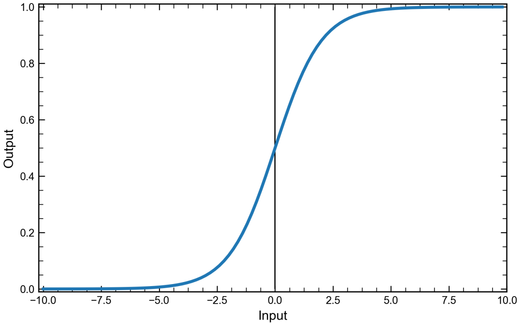
- Difficult backgrounds $\rightarrow$ additional targets

Target handling

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

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

Final node activation



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

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

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

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

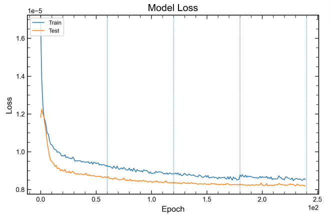
- Measures model quality for multiple classes

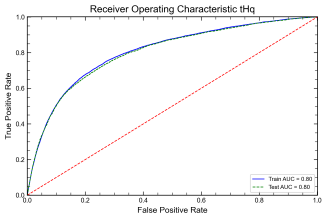
Hyperparameters

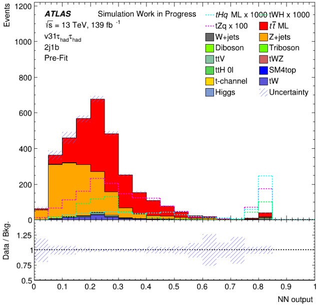
- Optimised by small grid search

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


Model Loss
1e−5
Train
Test
Loss
Epoch
1e2

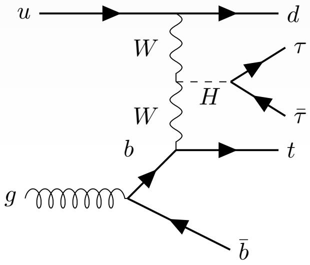

Receiver Operating Characteristic tHq
True Positive Rate
False Positive Rate
Train AUC = 0.80
Test AUC = 0.80


ATLAS Simulation Work in Progress
√s = 13 TeV, 139 fb⁻¹
v31τhadτhad
2j1b
Pre-Fit
tHq ML x 1000
tWH x 1000
tZq x 100
tt̄ ML
W+jets
Z+jets
Diboson
Triboson
ttV
tWZ
ttH 0l
SM4top
t-channel
tW
Higgs
Uncertainty
Events
Data / Bkg.
NN output

Conclusion

- No excuses for ignoring other MVA talks
- Categorical neural networks are a promising approach
- More testing and better plots needed
- Testing for more backgrounds easily possible

General ML selection



- n-jets: 2 (b-jets: **1**)
- b-jet WP: 70 DL1r
- nLeptons & nTaus: $\mathbf{2e}/\mu\ \mathbf{1}\tau_{\text{had}}$
- $E_{\text{T,miss}}$: 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 and weight setup

- Absolute weights for training because →best and most stable results
- Preliminary selection of variables

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
HvisMass	mass of LorentzV sum of hadronic taus
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

deltaRTau	Delta R of the hadronic taus
deltaPhiTau	Delta phi of the 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