Using unsupervised learning in search of new physics

...

Sakarias Garcia de Presno Frette

Outline

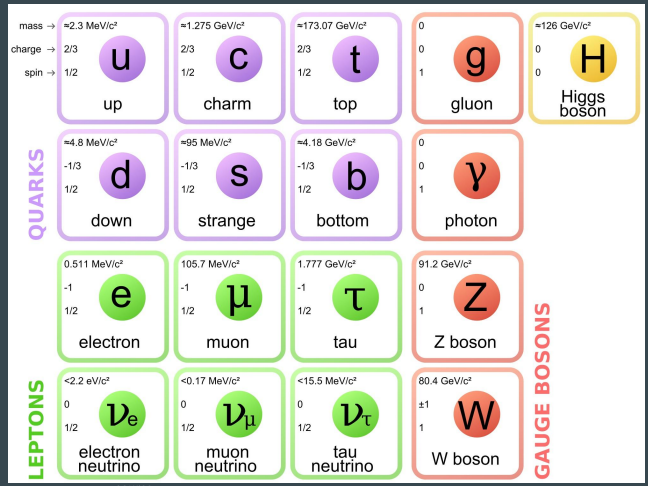
- Motivation for the project
- Machine learning phenomenology
- Implementation
- Results
- Topics for discussion

Using auto encoder for anomaly detection with ATLAS open data

- Attempt to see if the auto encoder can learn, without copying, to reconstruct sm processes and filter out new physics
- Test against new physics models

Motivation for BSM searches

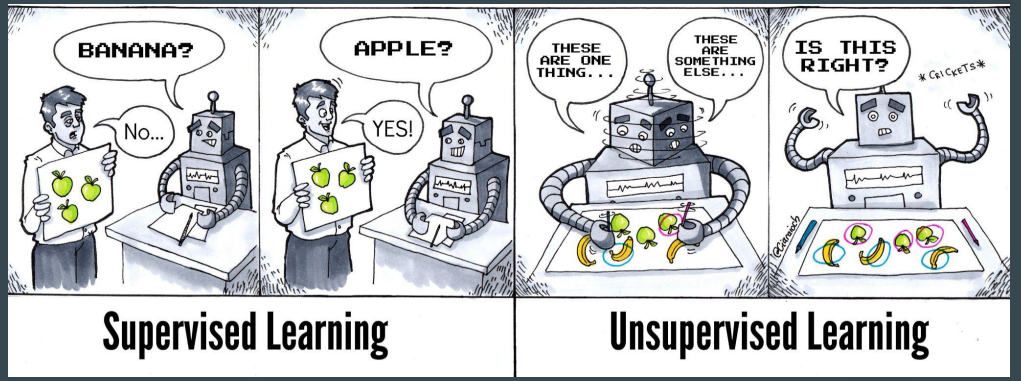
- Success of the standard model
- Why BSM searches
  - No dark matter candidate
  - Gravity



source: https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/

Model independent vs model dependent searches

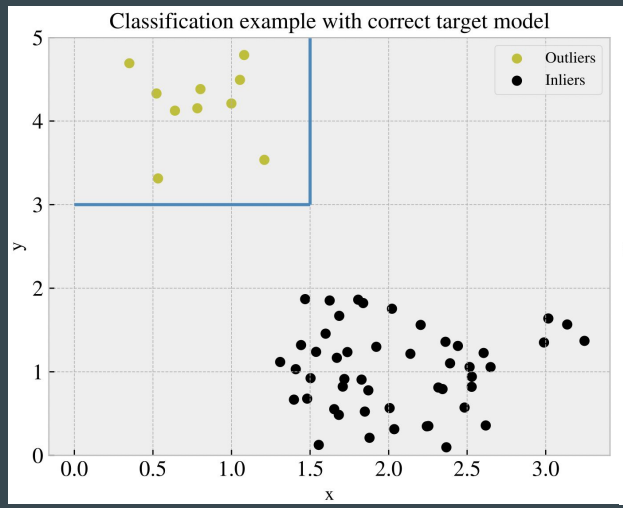
- Bias
- Semi-unsupervised vs unsupervised
- Auto encoder

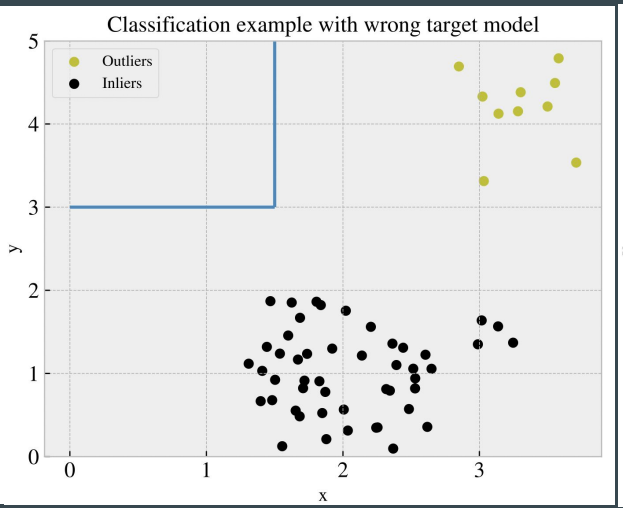


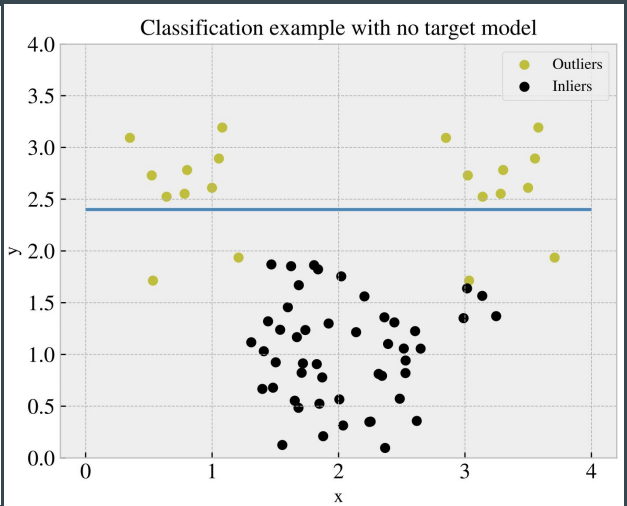
Source: https://twitter.com/athena_schools/status/1063013435779223553

Anomaly detection

- Point anomalies
- Contextual anomalies
- Collective anomalies

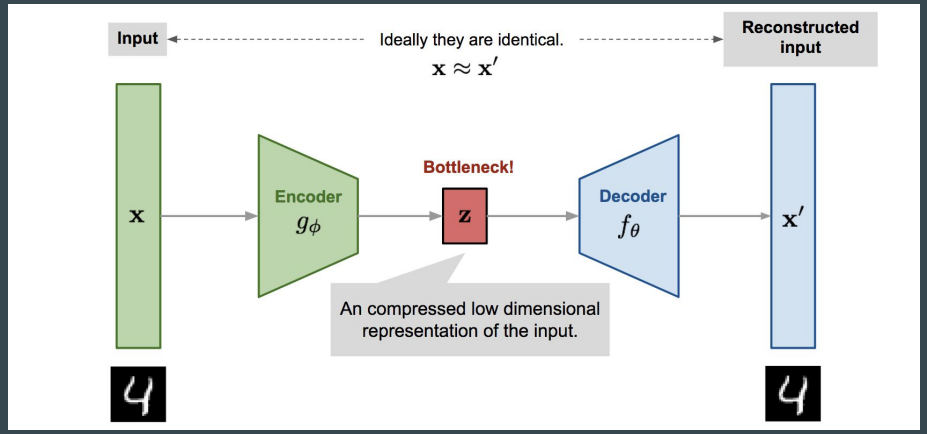




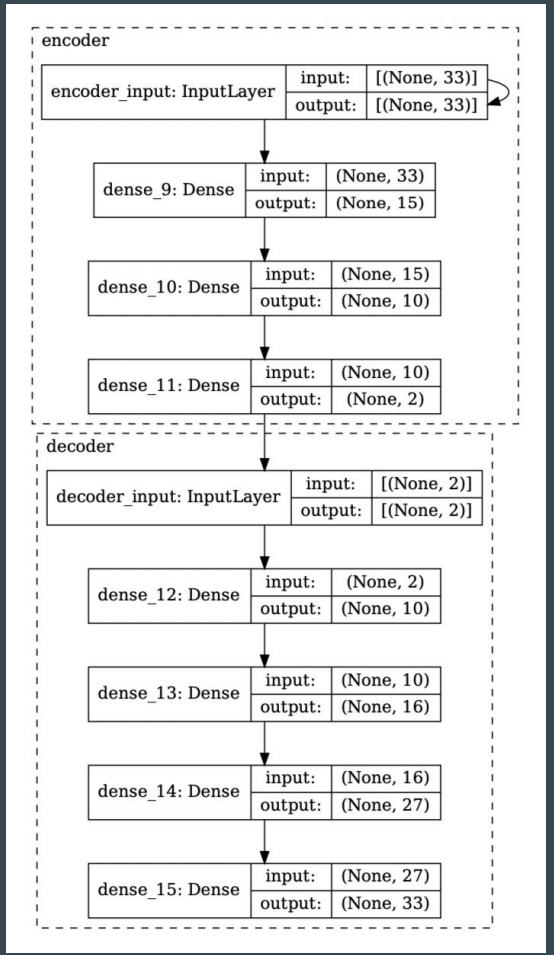


Auto Encoders

- Reconstruction, not copying
- Hyper parameters



Source: https://lilianweng.github.io/posts/2018-08-12-vae/

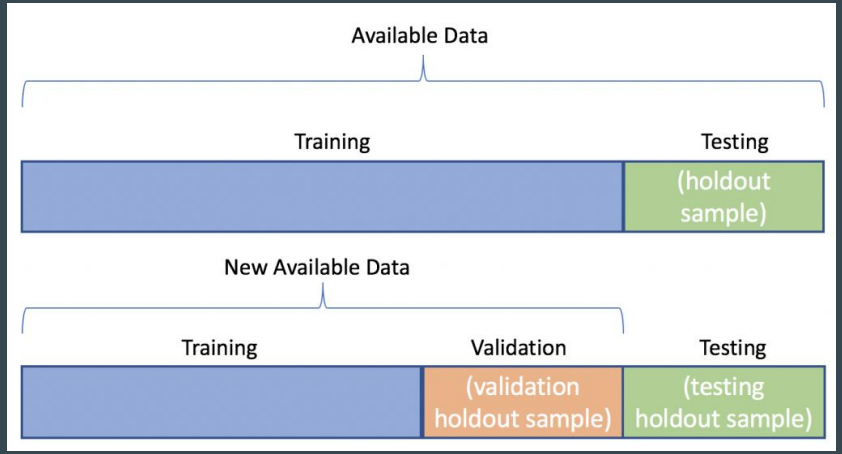


The data sets

- Background samples
- Signals
  - Supersymmetric models
  - Vector boson candidate
  - Randall-Sundrum Graviton
  - Dark matter candidate
- Training, validation and test data sets

Process	Unique "channelNumber"	Generator, hadronisation	Additional information
Top-quark production			
$t\bar{t}$+jets	410000	PowHeg-Box v2 [68] + Pythia 8 [69]	only 1ℓ and 2ℓ decays of $t\bar{t}$-system
single (anti)top t-channel	(410012) 410011	PowHeg-Box v1 + Pythia 6 [70]	
single (anti)top Wt-channel	(410014) 410013	PowHeg-Box v2 + Pythia 6	
single (anti)top s-channel	(410026) 410025	PowHeg-Box v2 + Pythia 6	
W/Z (+ jets) production			
$Z \to ee$, $\mu\mu$, $\tau\tau$	361106 – 361108	PowHeg-Box v2 + Pythia 8	LO accuracy up to $N_{\text{jets}} = 1$
$W \to e\nu$, $\mu\nu$, $\tau\nu$	361100 – 361105	PowHeg-Box v2 + Pythia 8	LO accuracy up to $N_{\text{jets}} = 1$
$W \to e\nu$, $\mu\nu$, $\tau\nu$ + jets	364156 – 364197	Sherpa 2.2 [71]	LO accuracy up to 3-jets final states
$Z \to ee$, $\mu\mu$, $\tau\tau$ + jets	364100 – 364141	Sherpa 2.2	LO accuracy up to 3-jets final states
Diboson production			
WW	363359, 363360	Sherpa 2.2	$qq'\ell\nu$ final states
WW	363492	Sherpa 2.2	$\ell\nu\ell'\nu'$ final states
ZZ	363356	Sherpa 2.2	$qq'\ell^+\ell^-$ final states
ZZ	363490	Sherpa 2.2	$\ell^+\ell^-\ell'^+\ell'^-$ final states
WZ	363358	Sherpa 2.2	$qq'\ell^+\ell^-$ final states
WZ	363489	Sherpa 2.2	$\ell\nu qq'$ final states
WZ	363491	Sherpa 2.2	$\ell\nu\ell^+\ell^-$ final states
WZ	363493	Sherpa 2.2	$\ell\nu\nu\nu'$ final states
SM Higgs production ($m_{\text{H}} = 125$ GeV)			
ggF, $H \to WW$	345324	PowHeg-Box v2 + Pythia 8	$\ell\nu\ell'\nu'$ final states
VBF, $H \to WW$	345323	PowHeg-Box v2 + Pythia 8	$\ell\nu\ell'\nu'$ final states
ggF, $H \to ZZ$	345060	PowHeg-Box v2 + Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
VBF, $H \to ZZ$	344235	PowHeg-Box v2 + Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
ZH, $H \to ZZ$	341947	Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
WH, $H \to ZZ$	341964	Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
ggF, $H \to \gamma\gamma$	343981	PowHeg-Box v2 + Pythia 8	
VBF, $H \to \gamma\gamma$	345041	PowHeg-Box v2 + Pythia 8	
$WH(ZH)$, $H \to \gamma\gamma$	345318, 345319	PowHeg-Box v2 + Pythia 8	
$t\bar{t}H$, $H \to \gamma\gamma$	341081	aMC@NLO [72] + Pythia 8	

source: http://opendata.atlas.cern/release/2020/documentation/datasets/mc.html

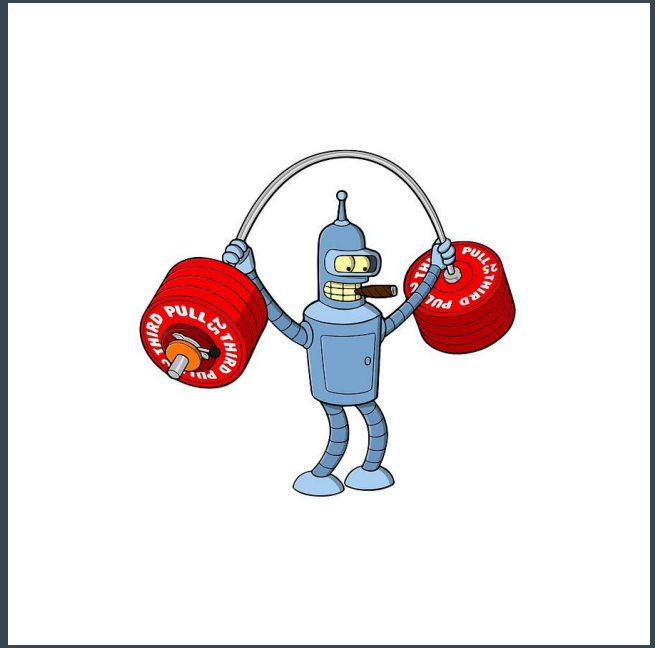


source: https://algotrading101.com/learn/train-test-split/

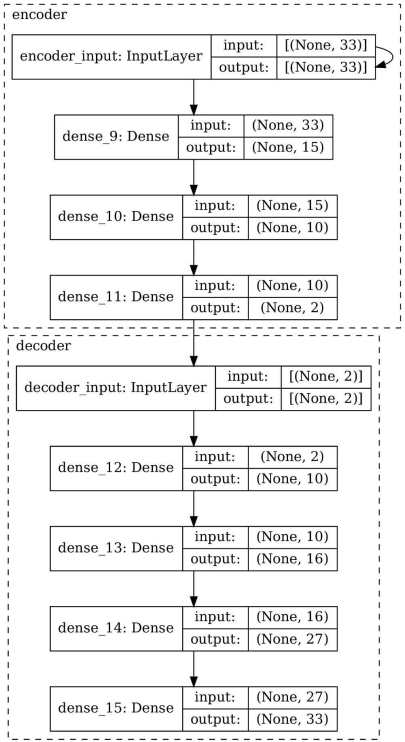
Training of the model

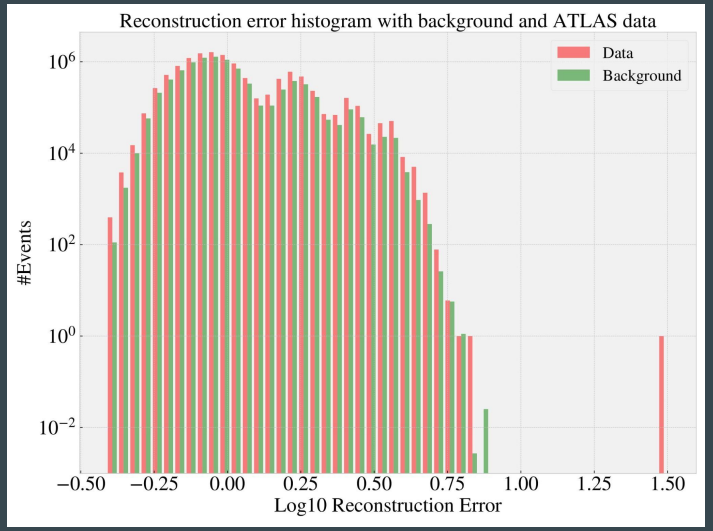
- Scaling
- Padding of features
- Tuning and training (Keras-Tuner, hyperband)

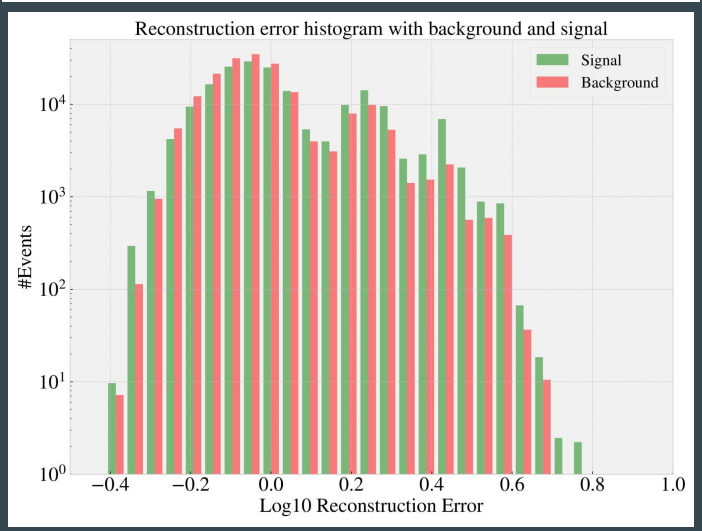
```
Output exceeds the size limit. Open the full output data in a text editor
Trial 24 Complete [00h 00m 46s]
val_mse: 0.015464218333363533

Best val_mse So Far: 0.015463903546333313
Total elapsed time: 00h 15m 25s

Search: Running Trial #25

Value             |Best Value So Far |Hyperparameter
0.1               |0.5               |Kernel_reg
0.5               |0.05              |Atc_reg
0.5               |0.05              |alpha
11                |12                |num_of_neurons1
linear            |linear            |1_act
8                 |9                 |num_of_neurons2
relu              |relu              |2_act
4                 |6                 |num_of_neurons3
leakyrelu         |leakyrelu         |3_act
2                 |2                 |lat_num
linear            |relu              |4_act
5                 |4                 |num_of_neurons5
relu              |linear            |5_act
7                 |8                 |num_of_neurons6
leakyrelu         |leakyrelu         |6_act
11                |11                |num_of_neurons7
...
1221/1221 [==============================] - 9s 7ms/step - loss: 0.0202 - mse: 0.0155 - val_loss: 0.0202 - val_mse: 0.0155
Epoch 9/15
1221/1221 [==============================] - 9s 7ms/step - loss: 0.0202 - mse: 0.0155 - val_loss: 0.0198 - val_mse: 0.0155
Epoch 10/15
1048/1221 [========================>.....] - ETA: 1s - loss: 0.0202 - mse: 0.0155
```

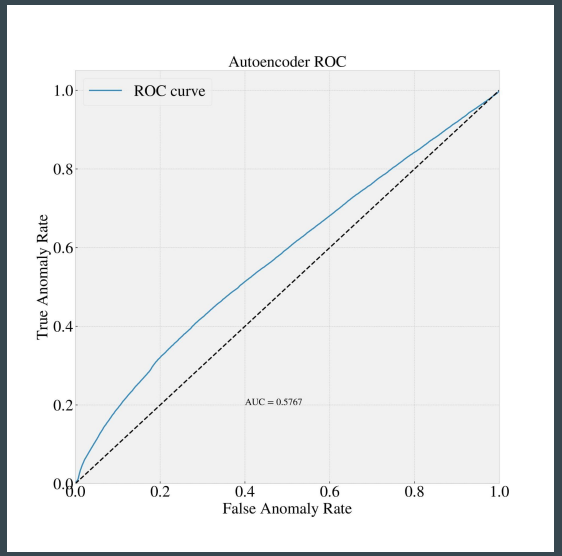


Results

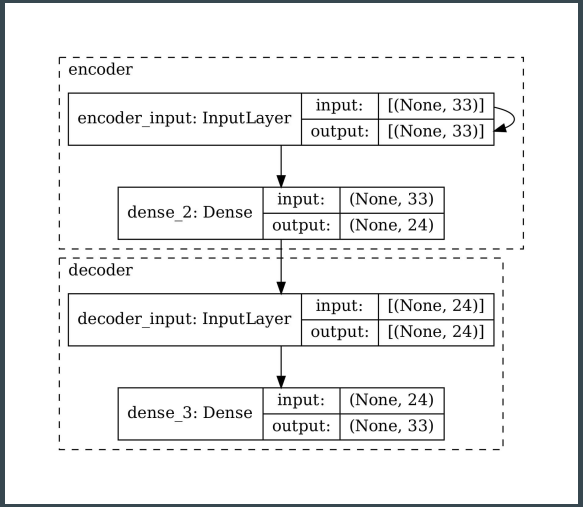

encoder
encoder_input: InputLayer
input: [(None, 33)]
output: [(None, 33)]
dense_9: Dense
input: (None, 33)
output: (None, 15)
dense_10: Dense
input: (None, 15)
output: (None, 10)
dense_11: Dense
input: (None, 10)
output: (None, 2)
decoder
decoder_input: InputLayer
input: [(None, 2)]
output: [(None, 2)]
dense_12: Dense
input: (None, 2)
output: (None, 10)
dense_13: Dense
input: (None, 10)
output: (None, 16)
dense_14: Dense
input: (None, 16)
output: (None, 27)
dense_15: Dense
input: (None, 27)
output: (None, 33)

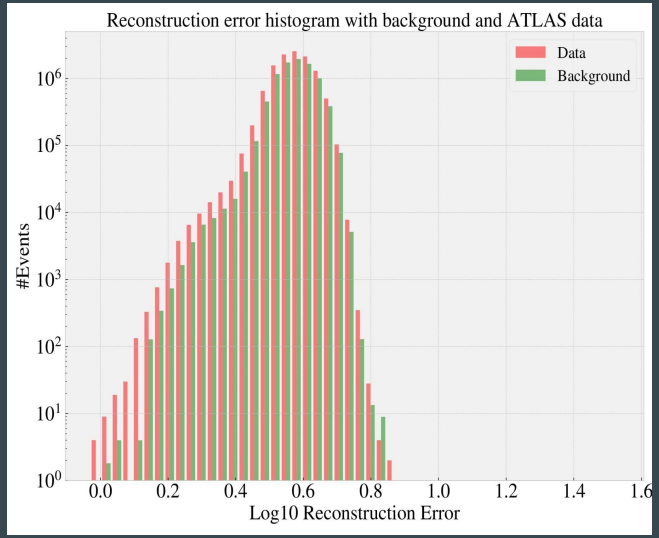

Reconstruction error histogram with background and ATLAS data
Data
Background
#Events
Log10 Reconstruction Error

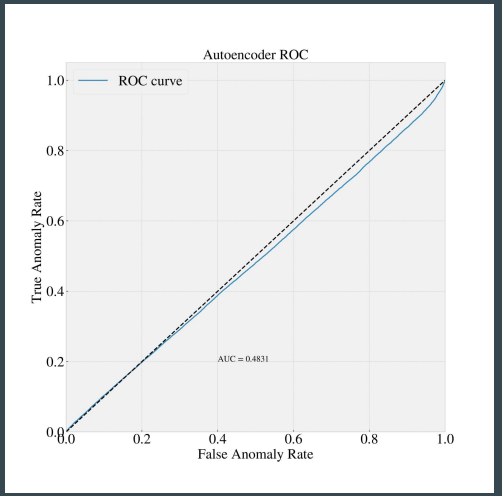

Reconstruction error histogram with background and signal
Signal
Background
#Events
Log10 Reconstruction Error


Autoencoder ROC
ROC curve
AUC = 0.5767
True Anomaly Rate
False Anomaly Rate

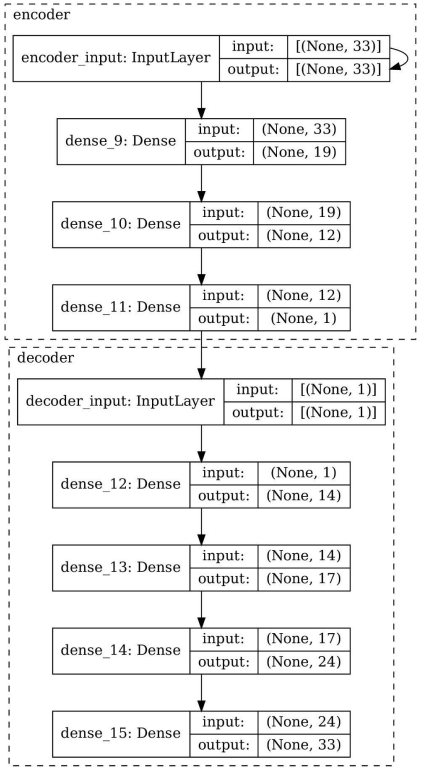
Small autoencoder

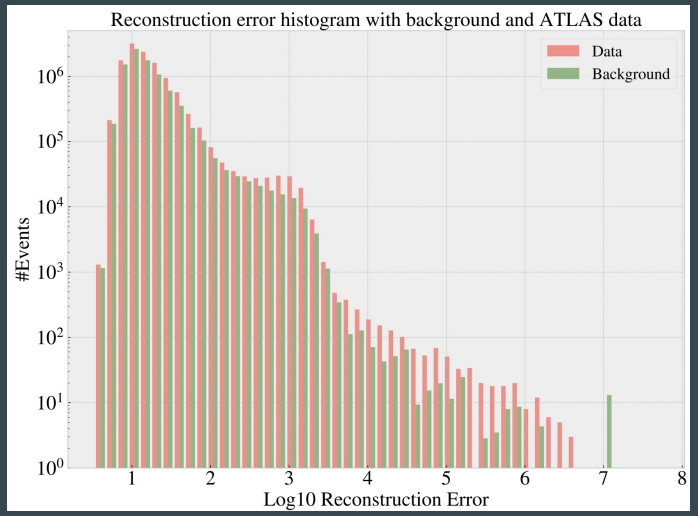


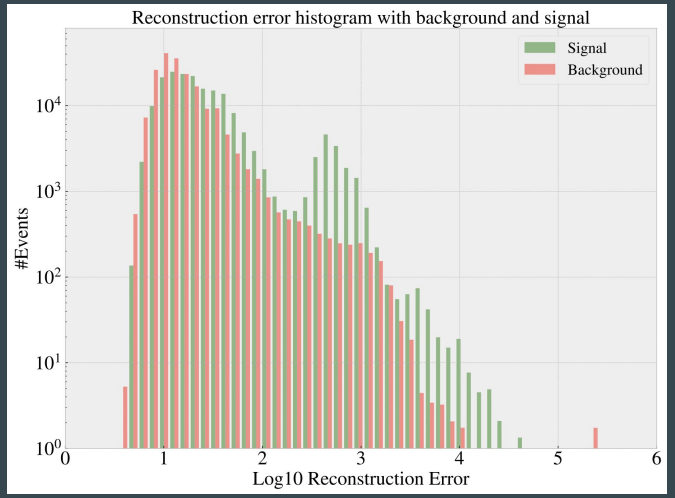


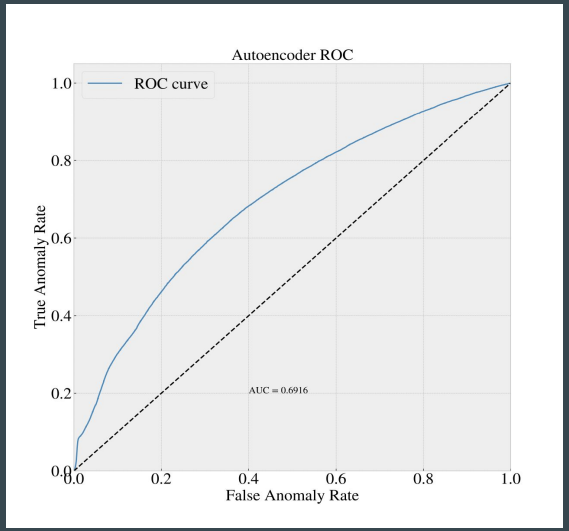


Standard scaling









Topics for discussion

- Extended gridsearch
  - More hyper parameters
  - More computation time required
  - Better sampling for more accurate representation of sm processes
- Reevaluating the metric for learning (weighted reconstruction error per feature)
- Rethinking construction of higher level features

Conclusion