Unsupervised Outlier and Anomaly Detection

Gregor Kasieczka
(gregor.kasieczka@uni-hamburg.de)

ErUM Data Community Meeting

1.7.2021

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

PIER
Partnership of
Universität Hamburg and DESY

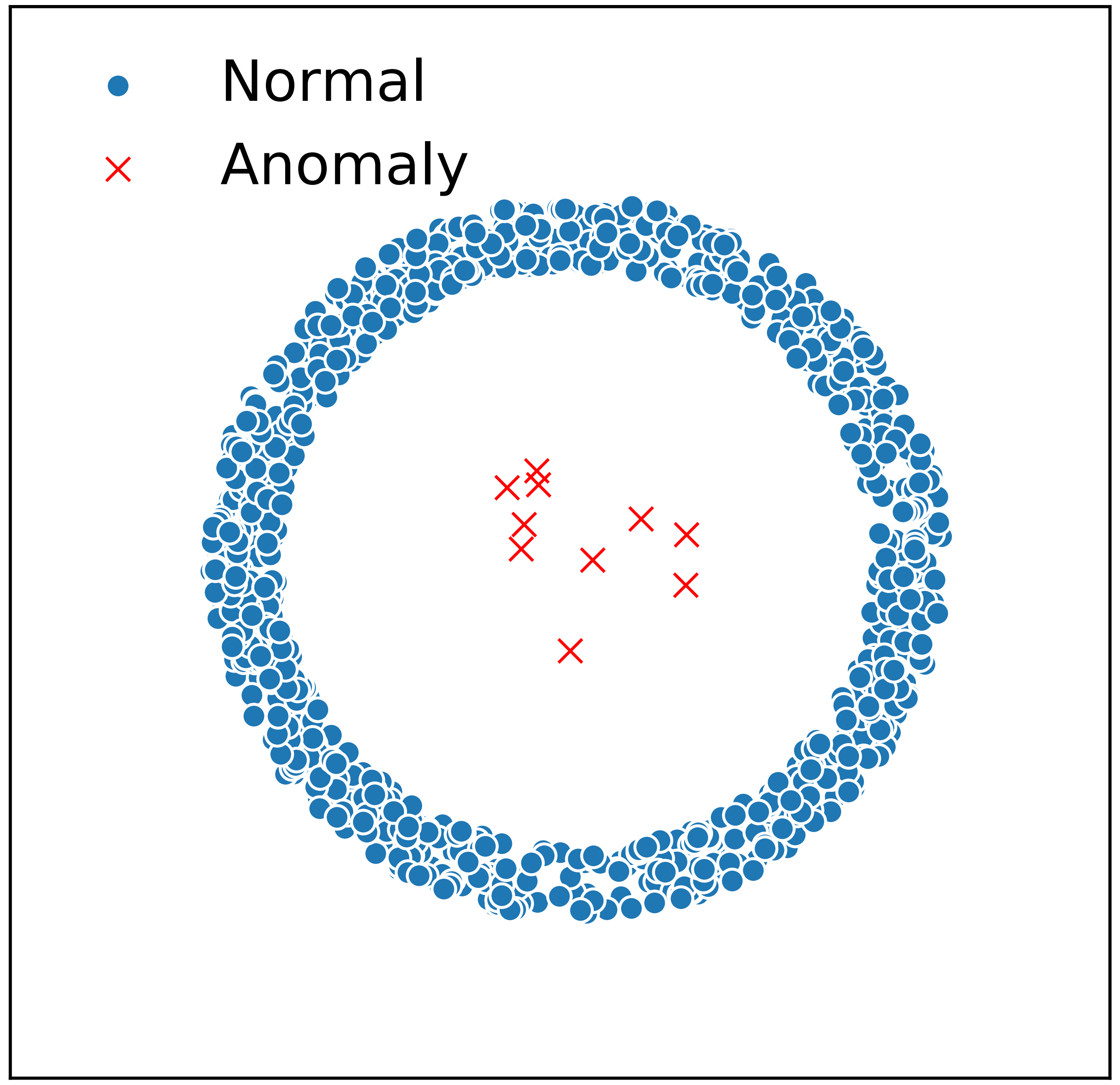
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

Bundesministerium
für Bildung
und Forschung

What is an anomaly?

Point anomaly / Outlier

- Outliers: Datapoints far away from regular distribution
- Examples:
  - Signals that are very different from the background
  - Detector malfunctions



And now?





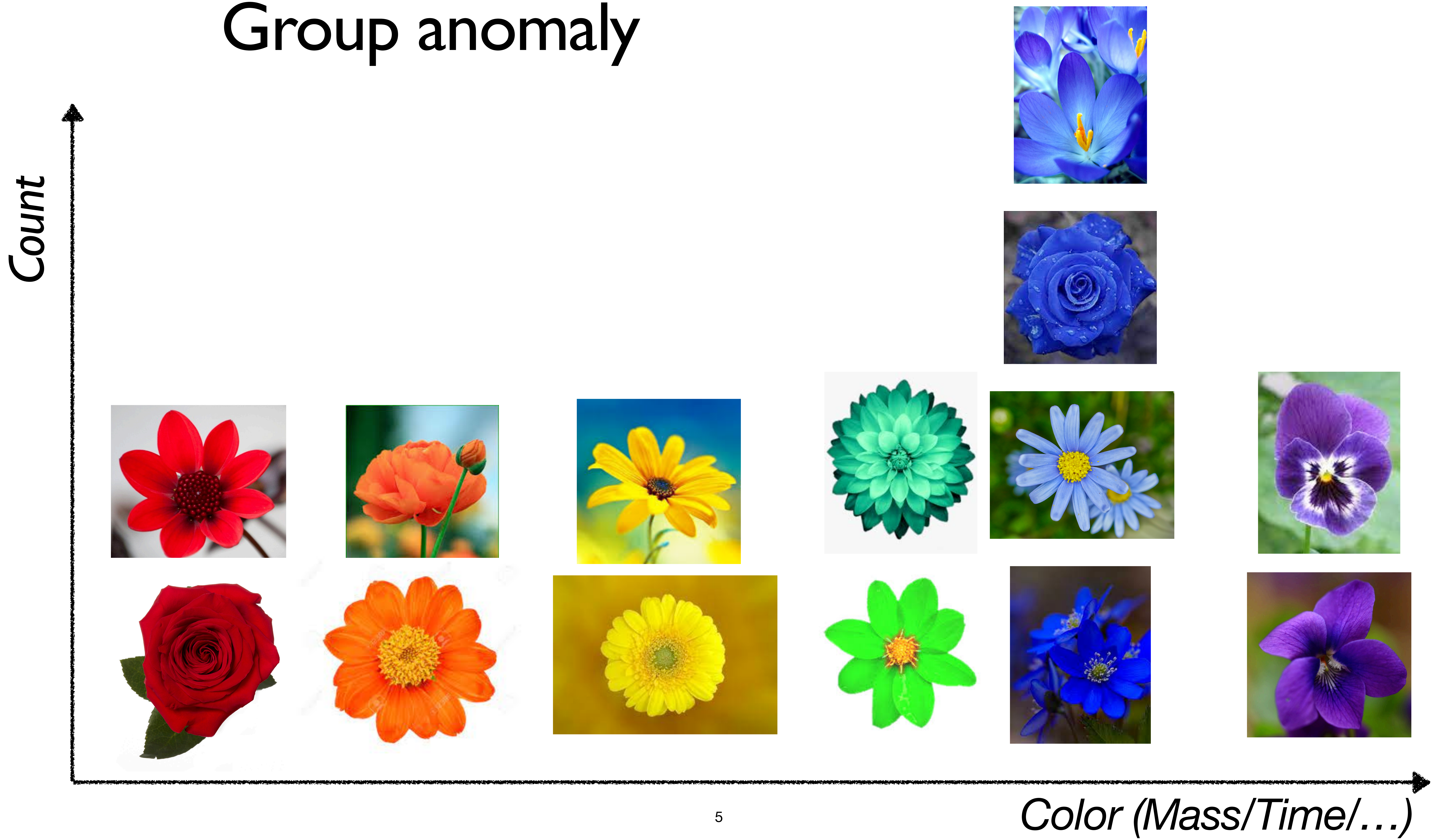






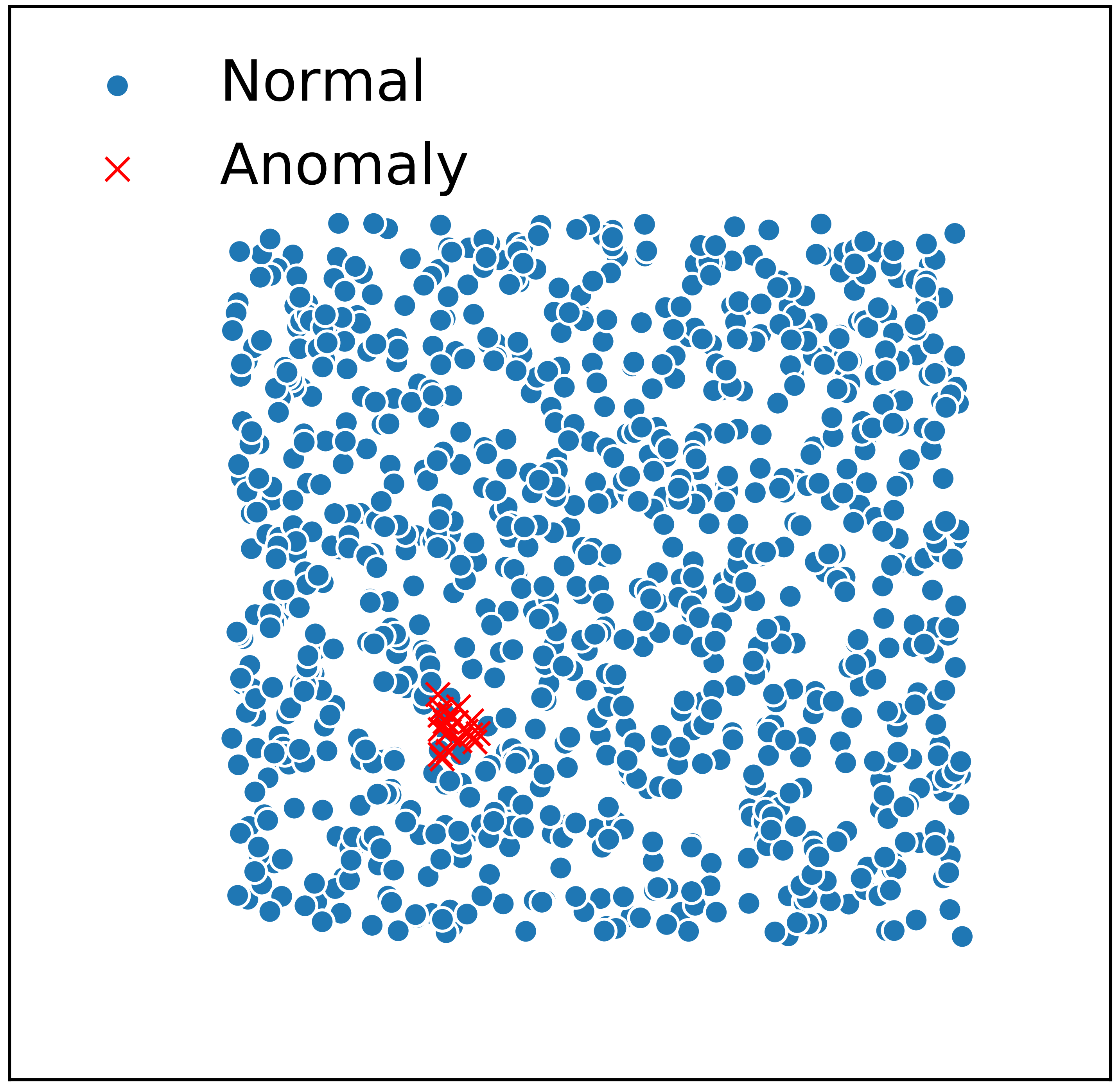


Group anomaly



Group anomaly

- Individual examples not anomalous, but interesting collective behaviour
- Examples:
  - Resonant signals over flat backgrounds (mass)
  - Short-time signals over stationary background (time)



Motivation

- For **discoveries**:
  Gain sensitivity to wide range of potential signals without explicitly specifying

- For **monitoring**:
  Detect malfunctions without known them in advance

- Added bonus:
  Training from data often possible, reduces need for simulation and corresponding uncertainties

- How can we do this?

How to build anomaly score?

- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:

  - $a(x) =$ Supervised

  - $a(x) = 1 / p(x|\text{Background})$

  - $a(x) = p(x|\text{Signal}) / p(x|\text{Background})$

How to build anomaly score?

- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:

    - **a(x) = Supervised**

    - a(x) = 1 / p(x|Background)

    - a(x) = p(x|Signal) / p(x|Background)

Supervised training:

Train a classifier to distinguish a mixture (cocktail) of hypothetical (simulated) signals from (simulated) background

How to build anomaly score?

- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:

  - a(x) = (Semi-) Supervised

  - **a(x) = 1 / p(x|Background)**

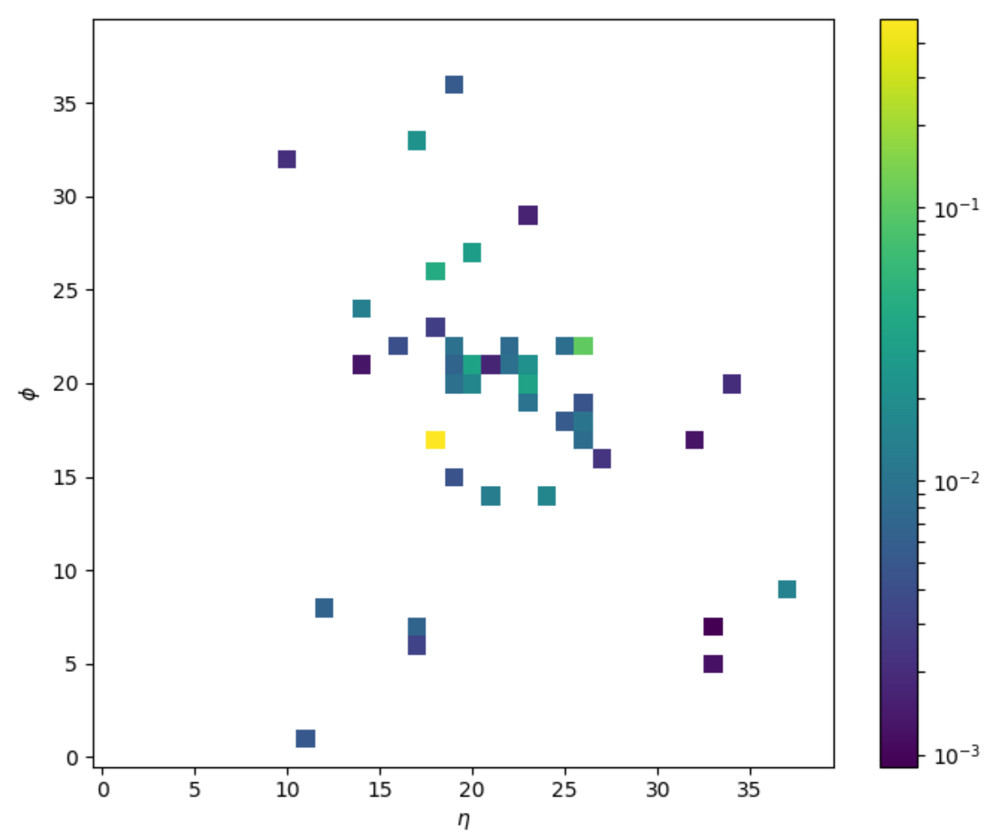
  - a(x) = p(x|Signal) / p(x|Background)

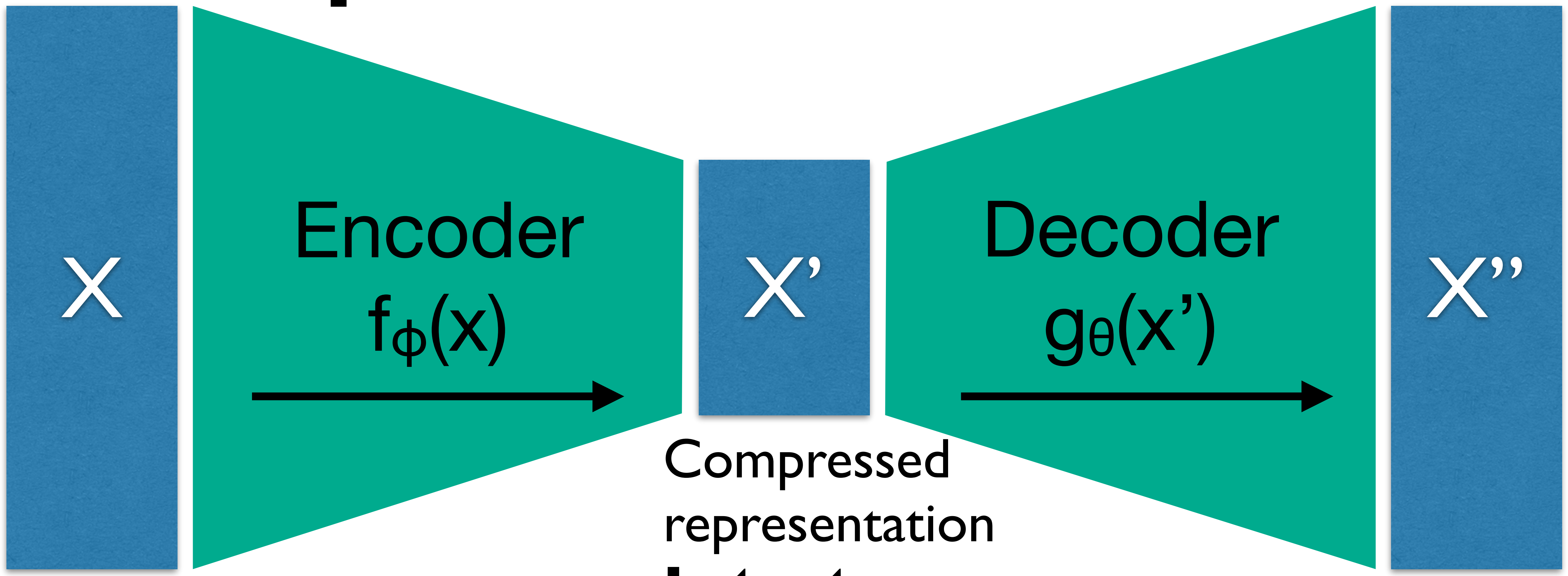
Find regions of low phase-space density for background:

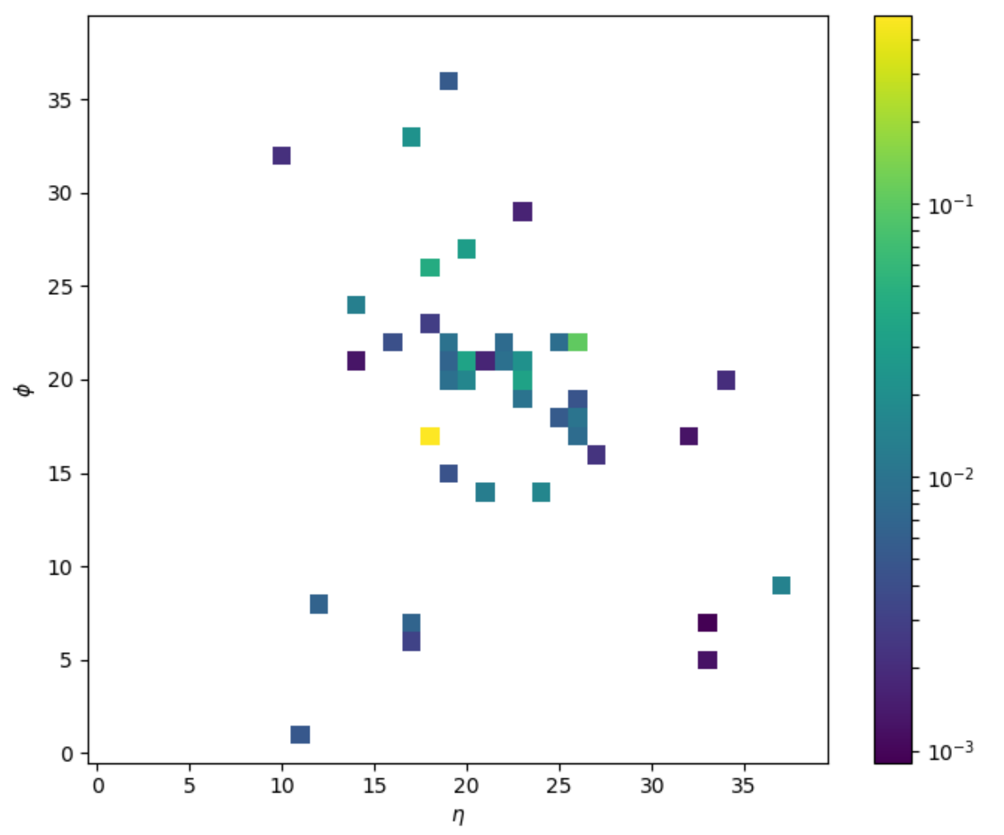
Use simulation to estimate background distribution (e.g. MUSIC, 2010.02984)

Use data to estimate background distribution (e.g. sidebands)

Popular Example: Autoencoder



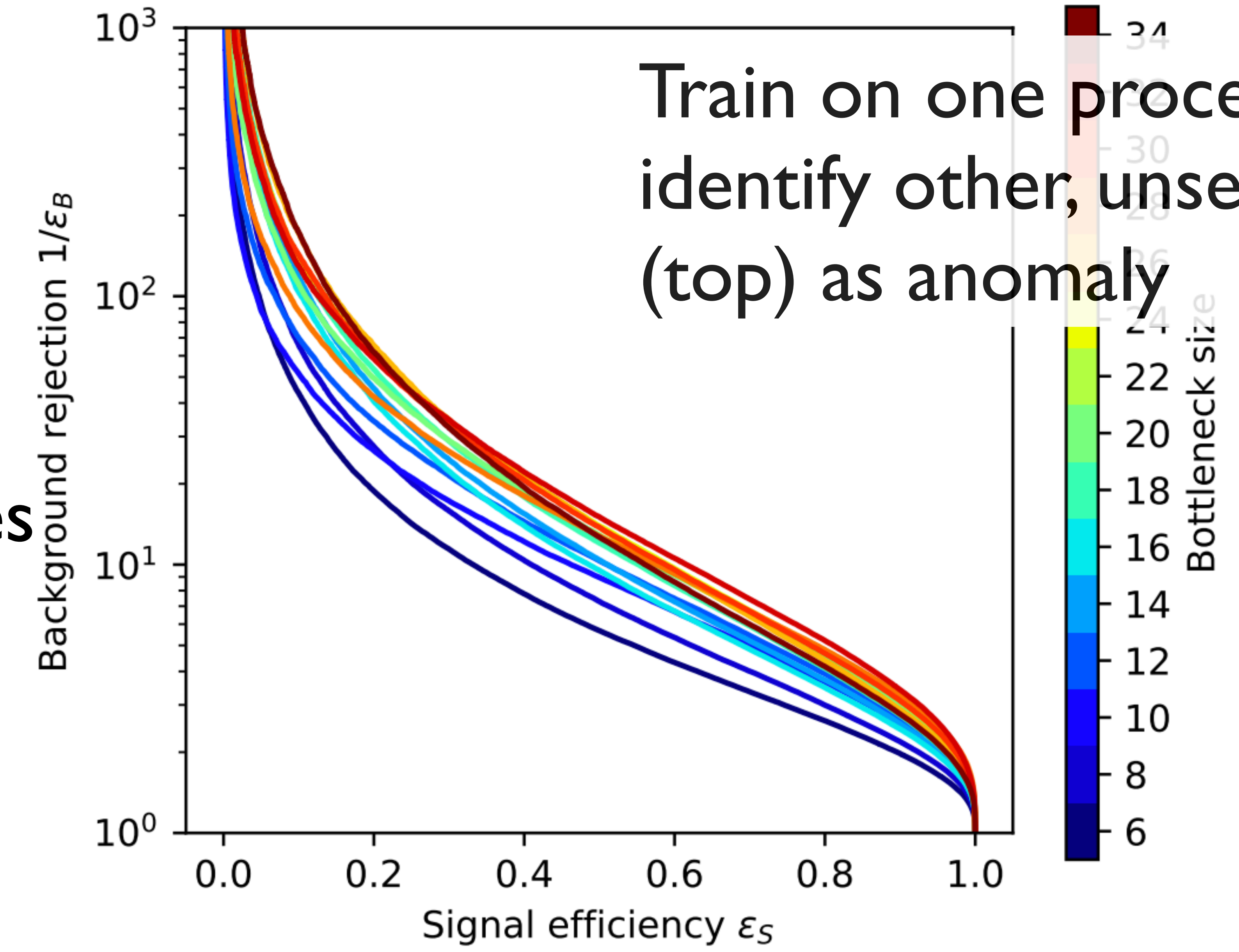




Input data
e.g. images, high level observables, four vectors

Compressed representation
Latent space

Output data
(sane format as input)



Train on one process (QCD), identify other, unseen, process (top) as anomaly

- Core idea:
  - Train lossy compression algorithm on anomaly-free data
  - Apply to data containing potential anomalies
- Expect quality to decrease for atypical examples:
  anomaly score

Heimel, GK, Plehn, Thompson, QCD or What?, 1808.08979 (also Shih et al 1808.08992); See Krämer et al (2104.09051) and Plehn et al (2104.08291) for recent studies of bias; Additional potential difficulty if data space has a non-trivial global topology (2102.08380)

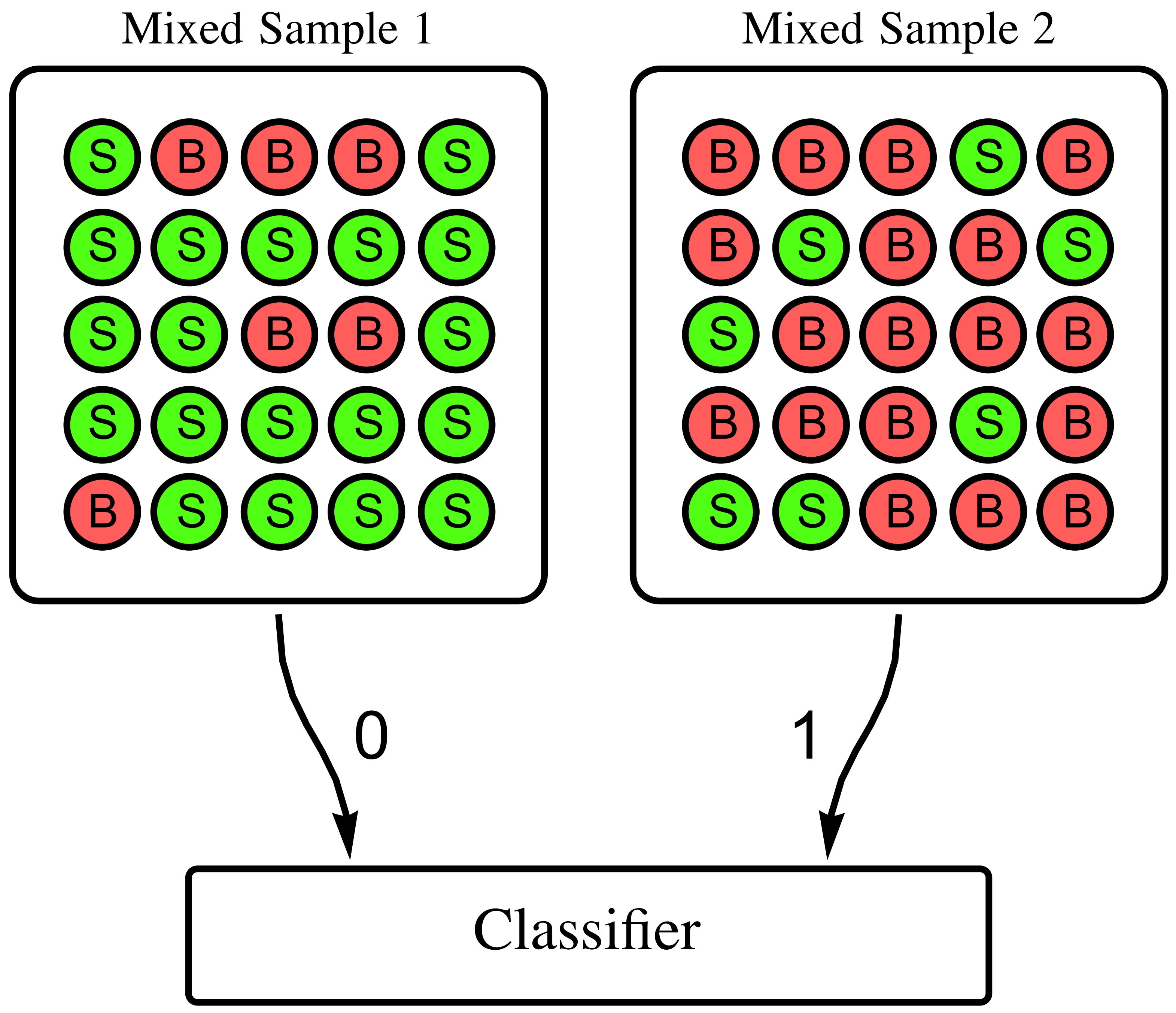
How to build anomaly score?

- Anomaly score **a** should be high for anomalous (signal-like)
  and low for background-like events

- Some options:
  - a(x) = (Semi-) Supervised
  - a(x) = 1 / p(x|Background)
  - **a(x) = p(x|Signal) / p(x|Background)**

Construct a likelihood-ratio:

Mixed sample training



$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1-f_1)\, p_B}{f_2\, p_S + (1-f_2)\, p_B} = \frac{f_1\, L_{S/B} + (1-f_1)}{f_2\, L_{S/B} + (1-f_2)}$$

Metodiev, Nachman, Thaler, *Classification without labels: Learning from mixed samples in high energy physics, 1708.02949*
Collins, Howe, Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning, 1805.02664*

How to build anomaly score?

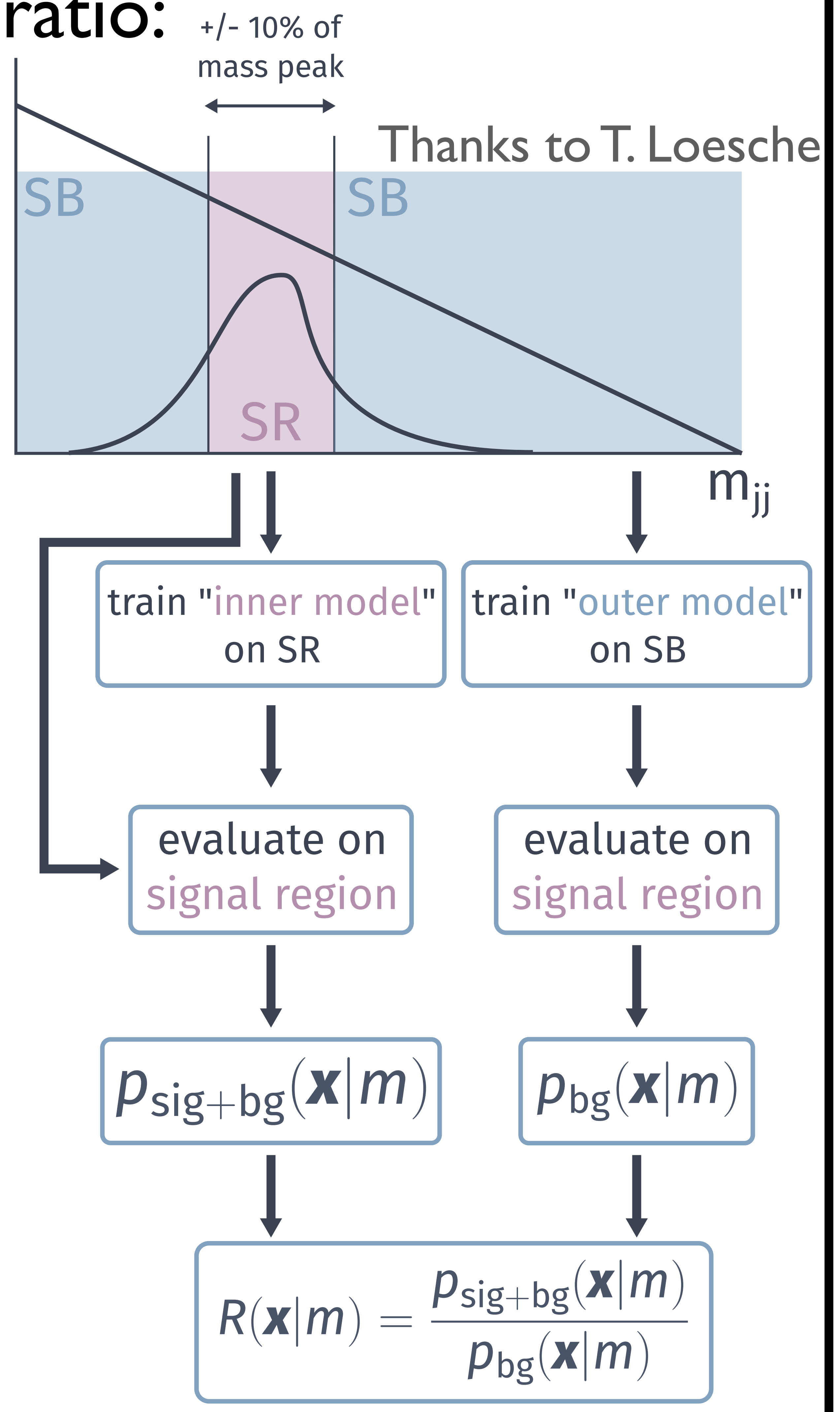
- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:
  - a(x) = (Semi-) Supervised
  - a(x) = 1 / p(x|Background)
  - **a(x) = p(x|Signal) / p(x|Background)**

Construct a likelihood-ratio:

Density estimation using flow-based models

Compare learned density of data in signal region and sidebands



Nachman, Shih, Anomaly Detection with Density Estimation 2001.04990

How to build anomaly score?

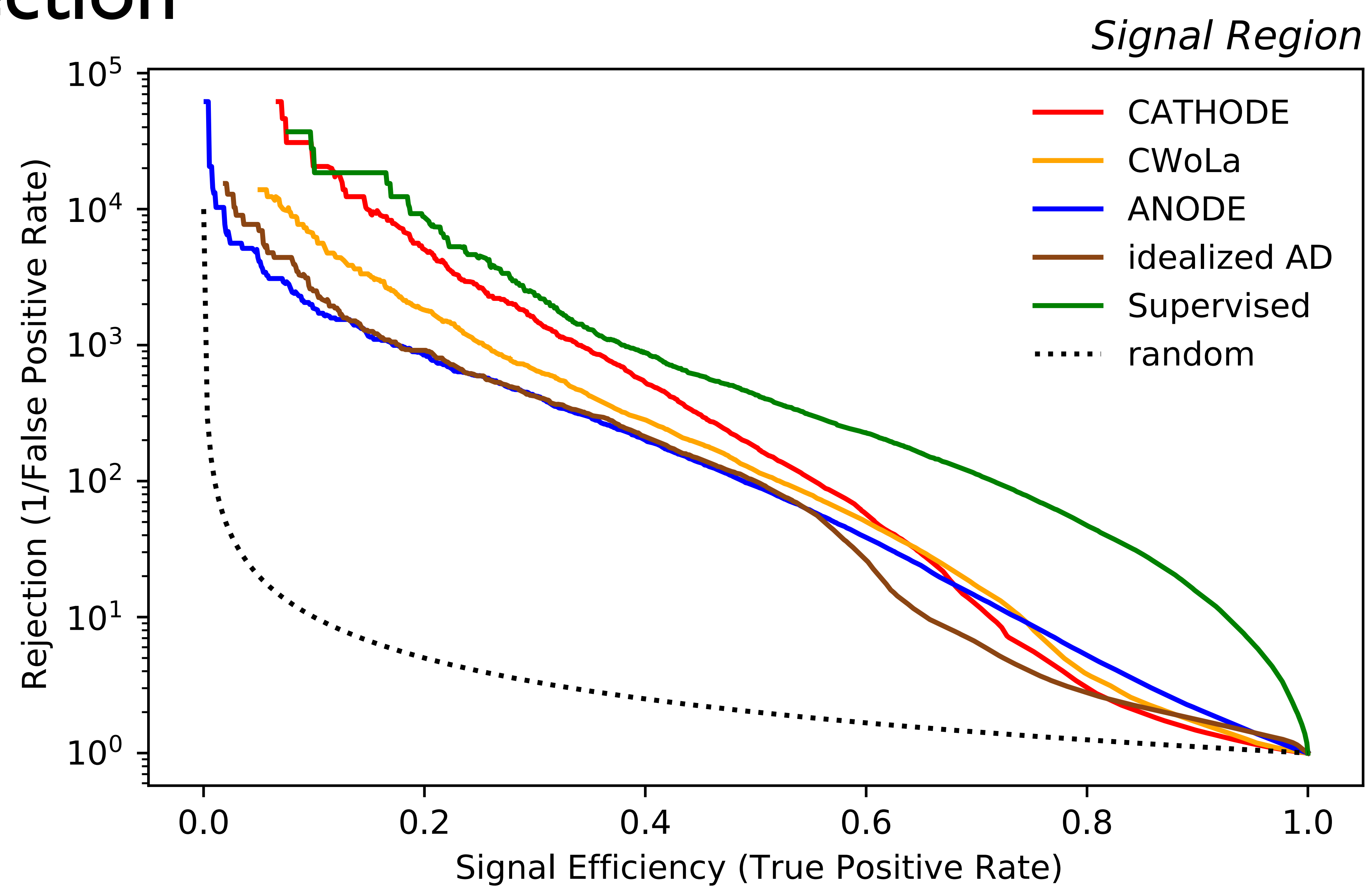
- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:
    - a(x) = (Semi-) Supervised
    - a(x) = 1 / p(x|Background)
    - **a(x) = p(x|Signal) / p(x|Background)**

Construct a likelihood-ratio: **Both**

Classifier-based **A**nomaly detection **TH**rough **O**uter **D**ensity **E**stimation (CATHODE)

Use flow to transport sideband into signal region, mixed sample training for anomaly detection



*Will show results first at **ML4Jets in Heidelberg next week**, paper coming soon*
https://indico.cern.ch/event/980214/

Closing

Just scratching the surface. For much more see https://lhco2020.github.io/homepage/ and our community paper (2101.08320):

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

Gregor Kasieczka (ed),[1] Benjamin Nachman (ed),[2,3] David Shih (ed),[4] Oz Amram,[5] Anders Andreassen,[6] Kees Benkendorfer,[2,7] Blaz Bortolato,[8] Gustaaf Brooijmans,[9] Florencia Canelli,[10] Jack H. Collins,[11] Biwei Dai,[12] Felipe F. De Freitas,[13] Barry M. Dillon,[8,14] Ioan-Mihail Dinu,[5] Zhongtian Dong,[15] Julien Donini,[16] Javier Duarte,[17] D. A. Faroughy[10] Julia Gonski,[9] Philip Harris,[18] Alan Kahn,[9] Jernej F. Kamenik,[8,19] Charanjit K. Khosa,[20,30] Patrick Komiske,[21] Luc Le Pottier,[2,22] Pablo Martín-Ramiro,[2,23] Andrej Matevc,[8,19] Eric Metodiev,[21] Vinicius Mikuni,[10] Inês Ochoa,[24] Sang Eon Park,[18] Maurizio Pierini,[25] Dylan Rankin,[18] Veronica Sanz,[20,26] Nilai Sarda,[27] Uroš Seljak,[2,3,12] Aleks Smolkovic,[8] George Stein,[2,12] Cristina Mantilla Suarez,[5] Manuel Szewc,[28] Jesse Thaler,[21] Steven Tsan,[17] Silviu-Marian Udrescu,[18] Louis Vaslin,[16] Jean-Roch Vlimant,[29] Daniel Williams,[9] Mikaeel Yunus[18]

- Unsupervised anomaly and outlier detection
  - Find new signals **without specifying** what to look for
  - Trade-off between specific sensitivity and coverage
- Also interesting for data quality monitoring, detector control, computing monitoring, online event selection, stellar structures (2104.12789), …
- Vibrant field, active development of new ideas and first applications in data

Thank you!

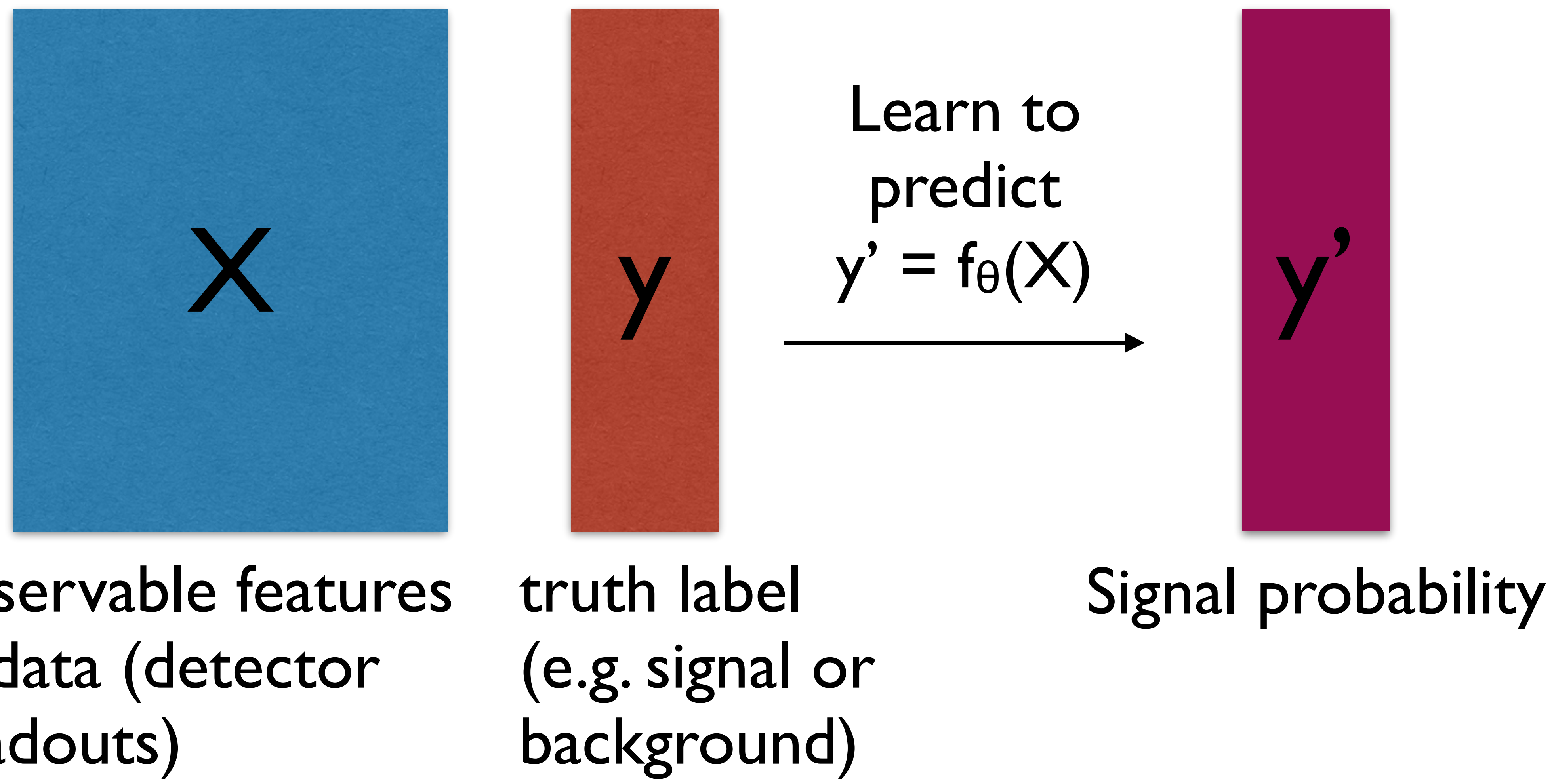
Backup

Two* types of problem:

Supervised Learning

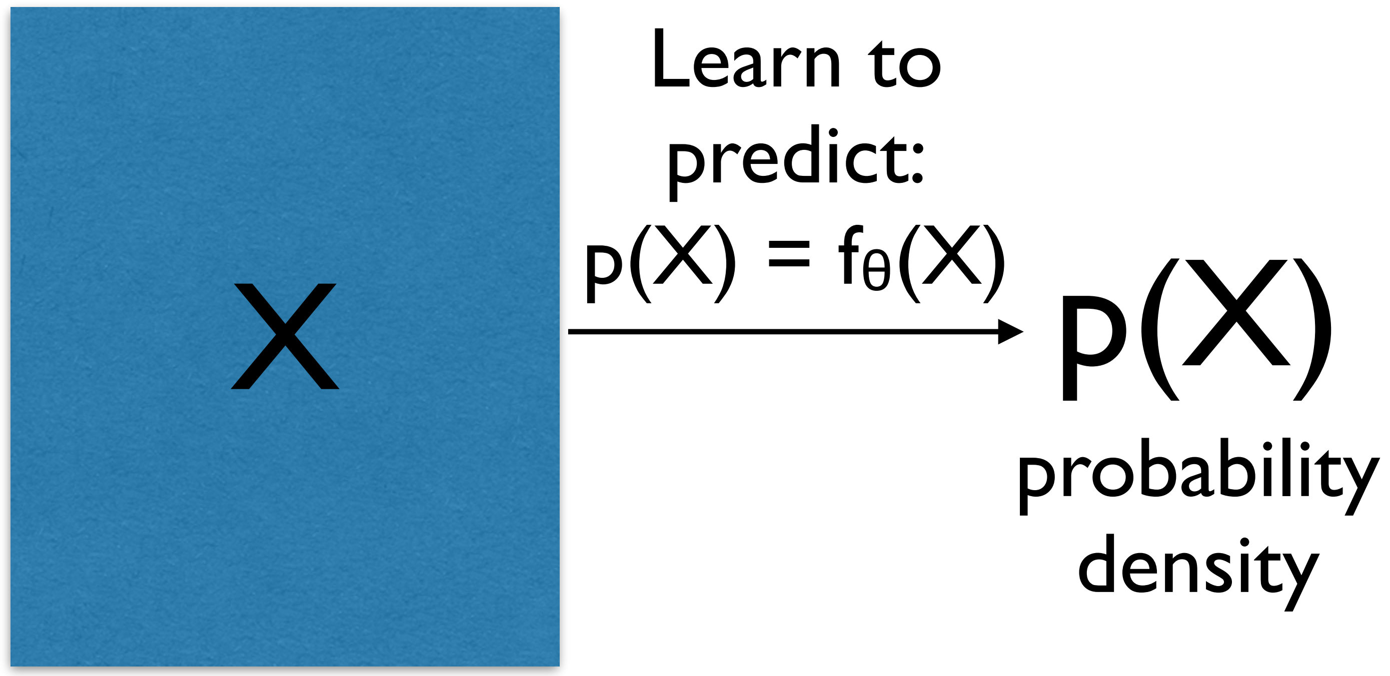
Attempt to infer some target *(truth label)*: classification or regression

Use training data with known labels (often from simulation)



Unsupervised

No target, learn the probability distribution (directly from data)



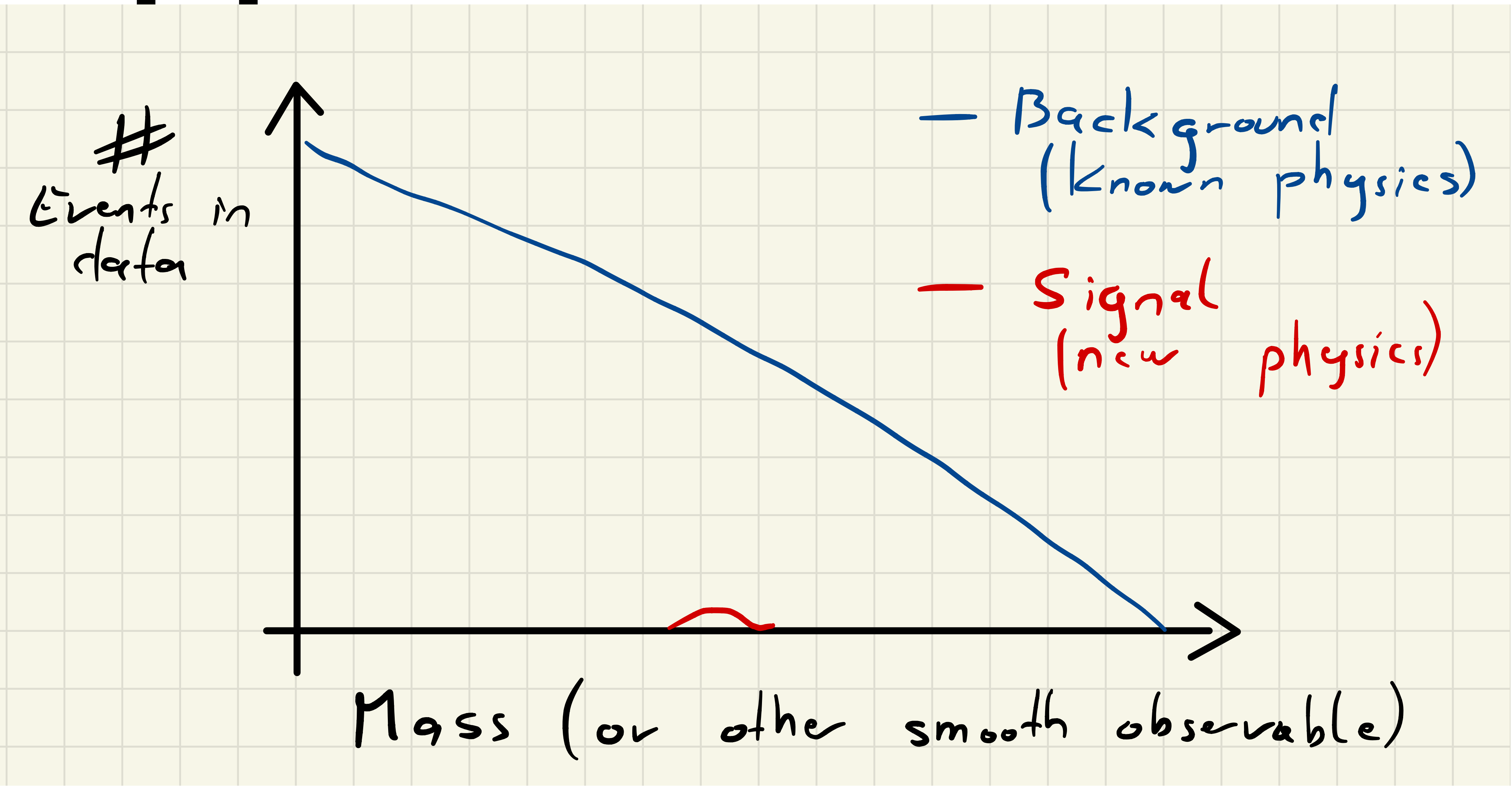
Maximize likelihood p(X)
(minimize -log p(x))

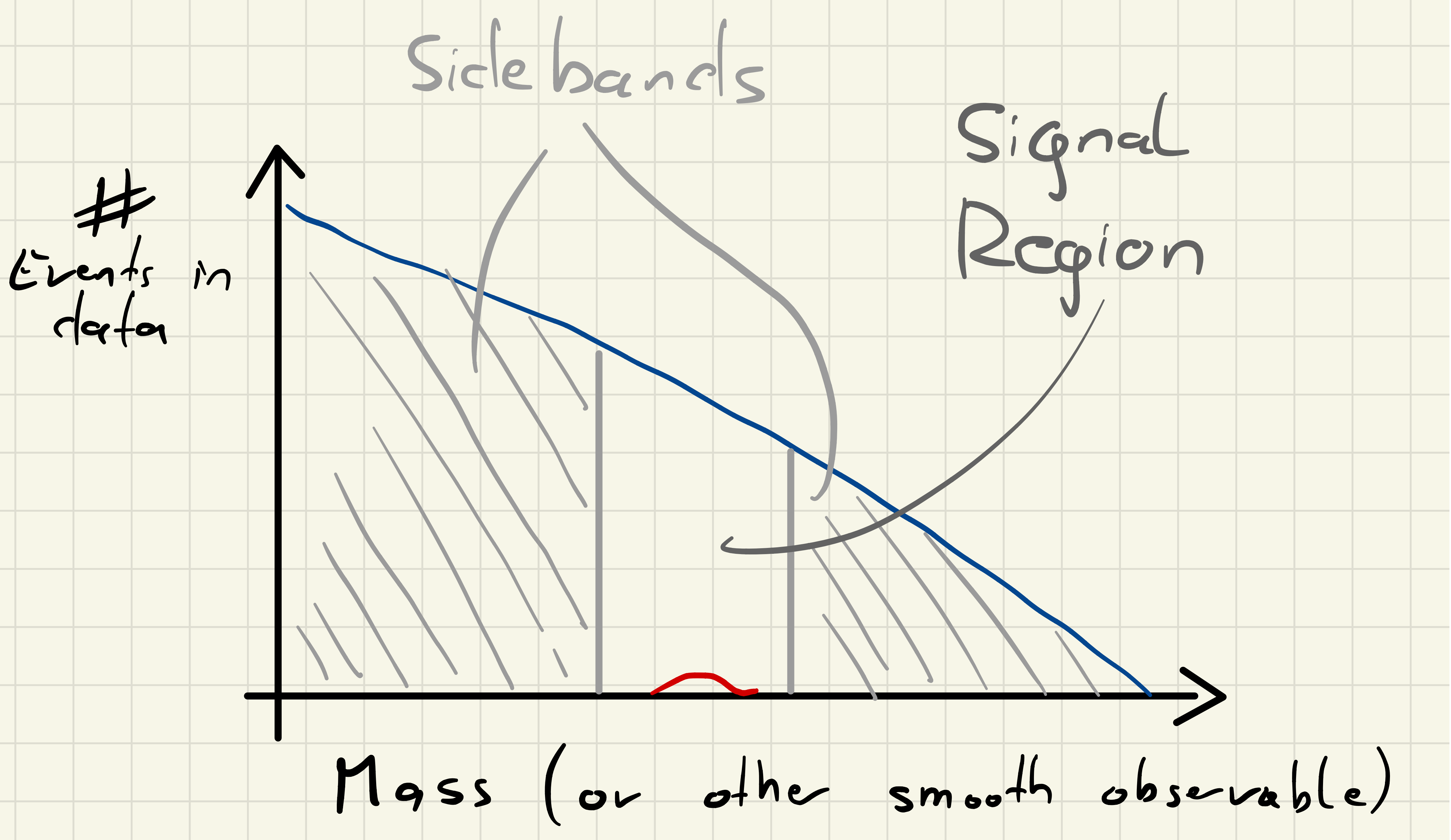
*and many more in between

Sideband approach

Key assumptions:

- There exists one feature (e.g. mass) so that:
  - Background distribution is smooth
  - Signal distribution is localised (and very small wrt/ background)





- Use sidebands to train anomaly score.
- Test signal region for new physics.
- Scan over different signal regions (trial factor)
- (Other ways to define anomaly-free regions in data possible as well. Not thoroughly explored yet)

Thanks to T. Loesche

Normalising Flows

- **Goal:** assign probability density to each datapoint
- Learn bijective transformation between data and a latent space with tractable probability
- Build from simple invertible transformations, tractable Jacobian

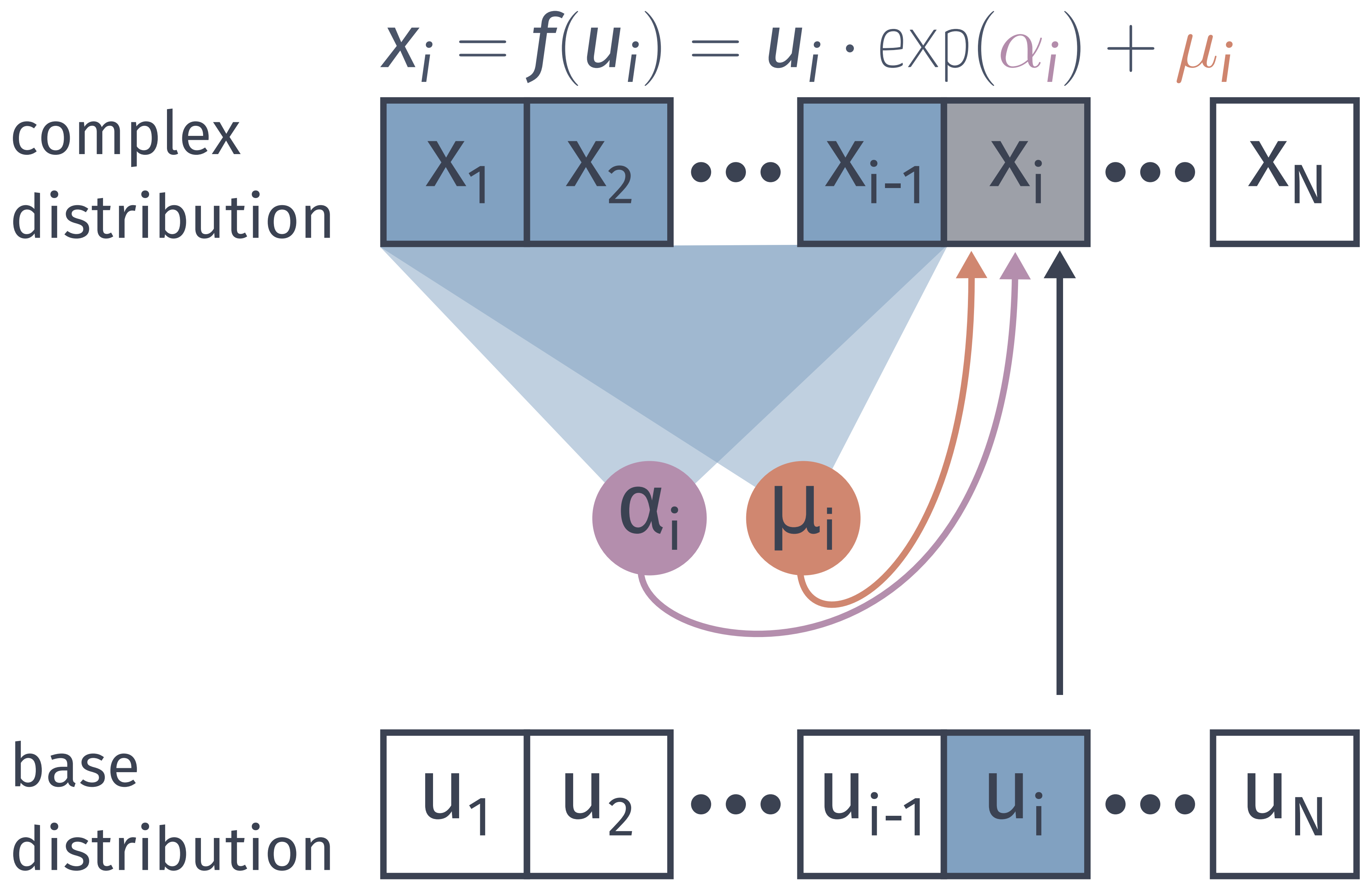
$$p(\boldsymbol{x}) = p(\boldsymbol{f}^{-1}(\boldsymbol{x})) \prod_i \left| \det \left(\frac{\partial \boldsymbol{f}_i^{-1}}{\partial \boldsymbol{x}} \right) \right| = p(\boldsymbol{u}) \prod_i \left| \det \left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{u}} \right) \right|^{-1}$$

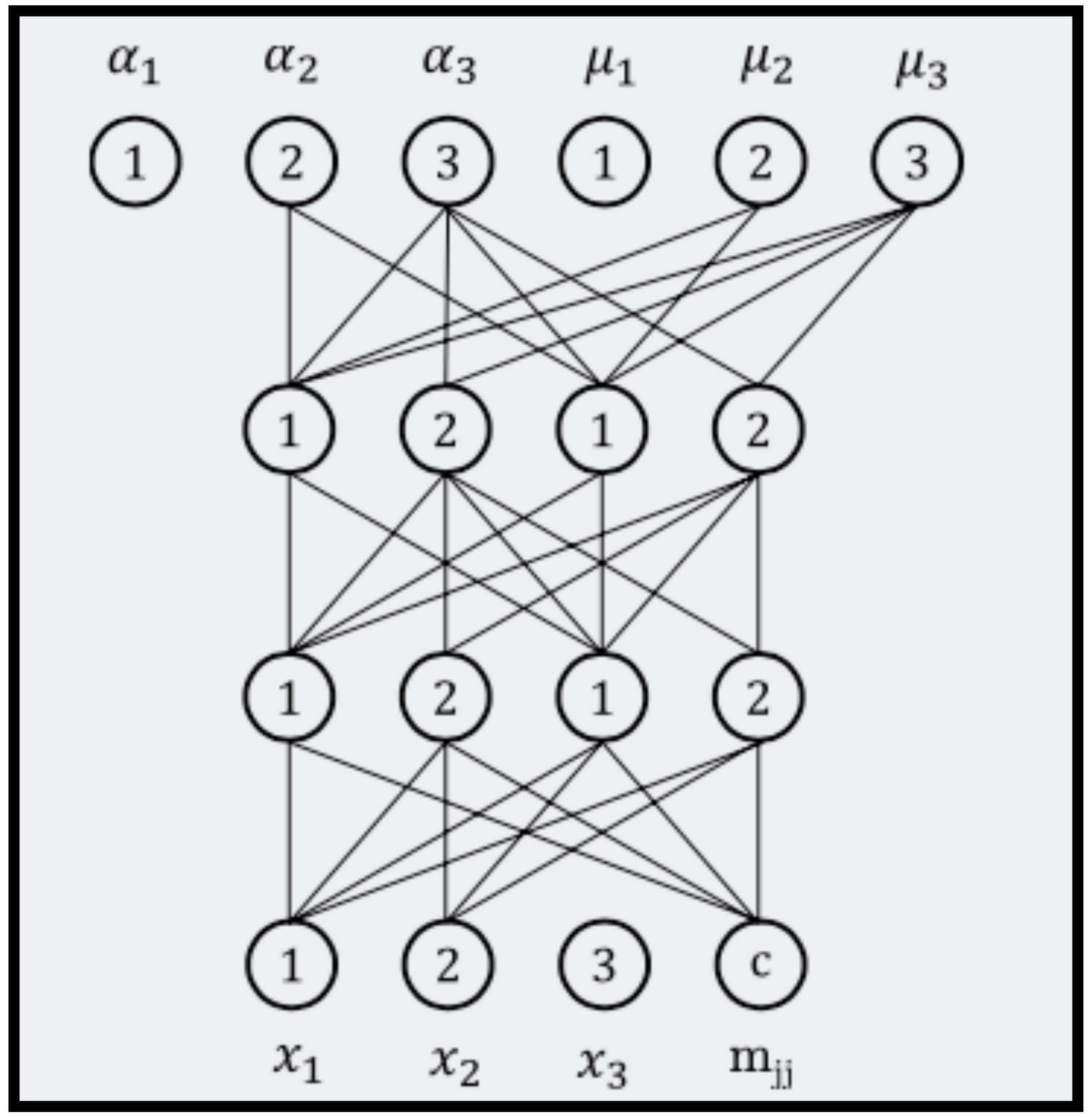
MADE/MAF

- Masked Autoregressive flow (1502.03509/1705.07057)
- Start with fully connected network, but drop connections so output a_j / mu_j are only connected to input x_1,..,x_j-1
- Autoregressive: no dependence of early features on late features
  - -> Jacobian is upper triangular matrix and easily invertible
- Combine multiple such blocks

$$p(x) = \prod_i p(x_i|x_{1:i-1})$$

$$p(x_i|x_{1:i-1}) = \mathcal{N}(x_i|\mu_i, (\exp \alpha_i)^2)$$
$$\mu_i = f_{\mu_i}(x_{1:i-1})$$
$$\alpha_i = f_{\alpha_i}(x_{1:i-1})$$



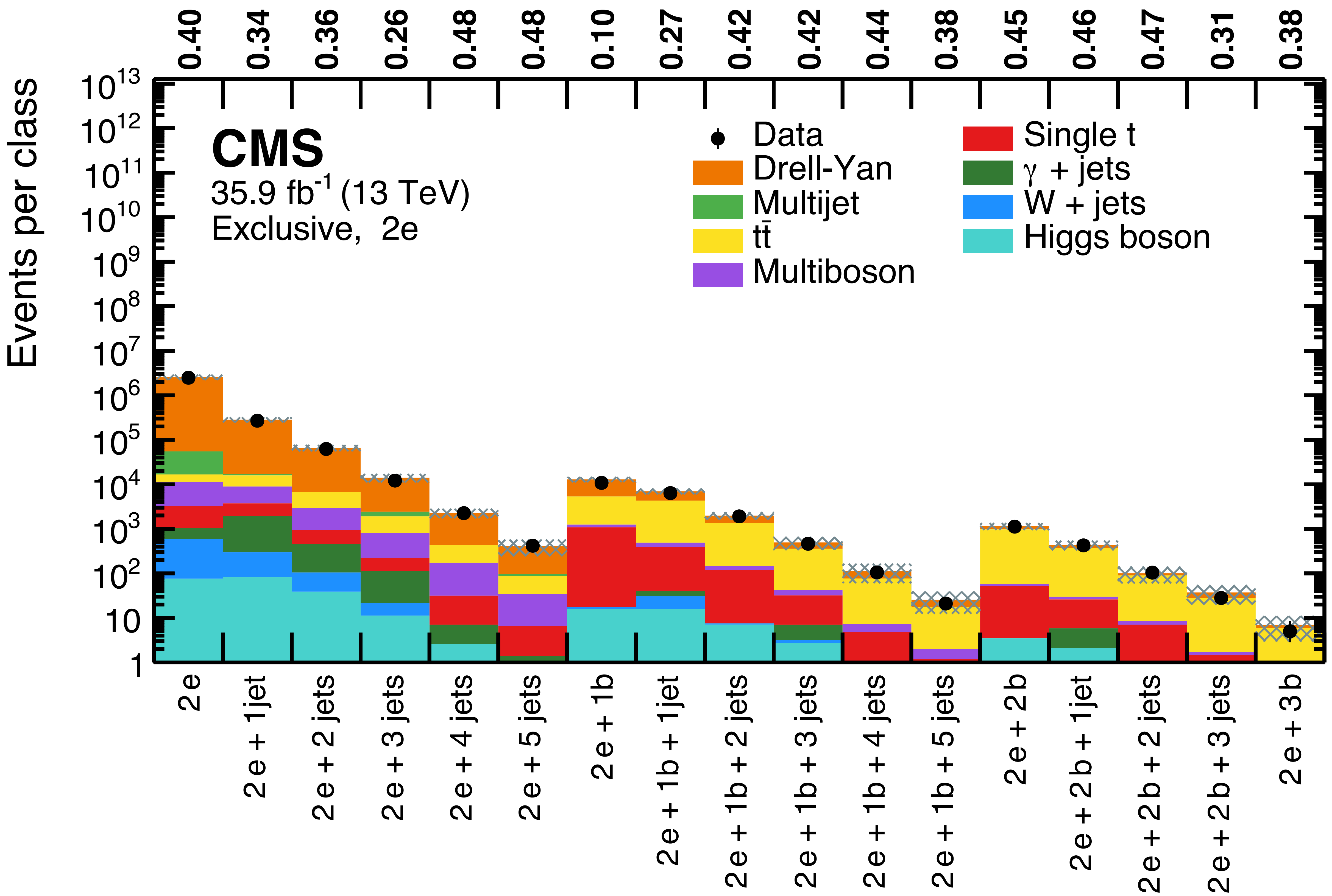


How to build anomaly score?

- Anomaly score **a** should be high for anomalous (signal-like) and low for background-like events

- Some options:
  - a(x) = (Semi-) Supervised
  - **a(x) = 1 / p(x|Background) *(from simulation)***

- Systematically look for differences between background simulation and data
- MUSIC / General search



CMS Collaboration, *MUSiC: a model unspecific search for new physics in proton-proton collisions at sqrt(s)=13 TeV*, 2010.02984