iDark
The intelligent dark matter survey

Radboud University

netherlands eScience center

Dark Machines

A research collective of physicists and data scientists to solve dark matter problems using machine learning

LUC HENDRIKS

RADBOUD UNIVERSITY, NIJMEGEN (NL)

WHAT IS DARKMACHINES?

- Research collective of about 200 researchers

WHAT IS DARKMACHINES?

- Research collective of about 200 researchers
- ML & DM experts combining knowledge to solve hard problems
  - Multidisciplinary: eg. ML experts joined from biomedical imaging
  - Challenge based (DM experts deliver data, ML experts deliver solution)
  - Challenges are self-organised by challenge leaders
  - Each challenge produces 1+ papers
  - Anyone can join if interested

WHAT IS DARKMACHINES?

- Research collective of about 200 researchers
- ML & DM experts combining knowledge to solve hard problems
    - Multidisciplinary: eg. ML experts joined from biomedical imaging
    - Challenge based (DM experts deliver data, ML experts deliver solution)
    - Challenges are self-organised by challenge leaders
    - Each challenge produces 1+ papers
    - Anyone can join if interested
    - Yearly workshops (except this year..)

MULTIDISCIPLINARY

- ML methods are “discipline-independent”

MULTIDISCIPLINARY

- ML methods are "discipline-independent"
- When you strip away the (astro)physics, almost any problem becomes a data science problem
  - Interpret satellite data -> computer vision
  - Finding new physics in particle collisions -> anomaly detection
  - Gravitational wave detection -> time-series analysis
  - …
- DarkMachines was founded with this in mind:
  experts in one field can contribute their methods in another

TOPICS

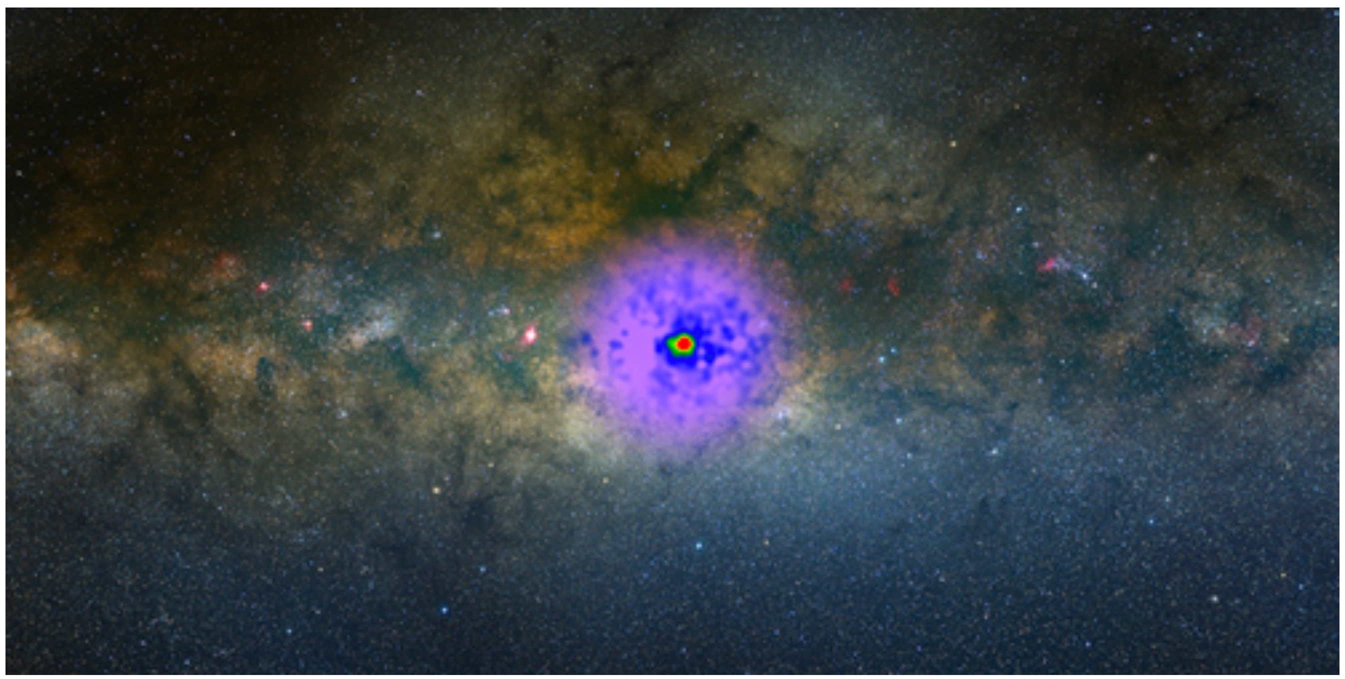
- Exploring high-D parameter spaces
- Unsupervised collider searches
- Generative models as event generators
- Analysis of gamma-ray Galactic Center
- Sampling methods
- Anomaly detection
- VAEs
- Computer vision & Bayesian deep learning

EXPLORING HIGH-D PARAMETER SPACES

- Many problems can be broken down to
  “find optimal set of parameters given some (log-)likelihood”
- For example: which set of parameters in the pMSSM can explain the flux from the Galactic Center excess?

EXPLORING HIGH-D PARAMETER SPACES

- Many problems can be broken down to
  "find optimal set of parameters given some (unknown) (log-)likelihood function"
- For example: which set of parameters in the pMSSM can explain the flux from the Galactic Center excess?



https://arxiv.org/abs/1502.05703

Gamma-ray flux due to DM annihilation

DM candidate

Symbol	Description	number of parameters
$\tan\beta$	the ratio of the vacuum expectation values of the two Higgs doublets	1
M_A	the mass of the pseudoscalar Higgs boson	1
μ	the higgsino mass parameter	1
M_1	the bino mass parameter	1
M_2	the wino mass parameter	1
M_3	the gluino mass parameter	1
$m_{\tilde{q}}, m_{\tilde{u}_R}, m_{\tilde{d}_R}$	the first and second generation squark masses	3
$m_{\tilde{l}}, m_{\tilde{e}_R}$	the first and second generation slepton masses	2
$m_{\tilde{Q}}, m_{\tilde{t}_R}, m_{\tilde{b}_R}$	the third generation squark masses	3
$m_{\tilde{L}}, m_{\tilde{\tau}_R}$	the third generation slepton masses	2
A_t, A_b, A_τ	the third generation trilinear couplings	3

EXPLORING HIGH-D PARAMETER SPACES

- Many problems can be broken down to
  "find optimal set of parameters given some (log-)likelihood"

- For example: which set of parameters in the pMSSM can explain the flux from the Galactic Center excess?

Combination of 19 parameters (=model)

↓

Calculate SUSY model & DM candidate parameters

↓

Check if model is excluded by any SUSY/DM constraints

↓

Calculate gamma-ray flux from DM annihilation

↓

Calculate likelihood of this gamma-ray flux explaining the flux from GCE

<u>Takes about 10 seconds</u>

EXPLORING HIGH-D PARAMETER SPACES

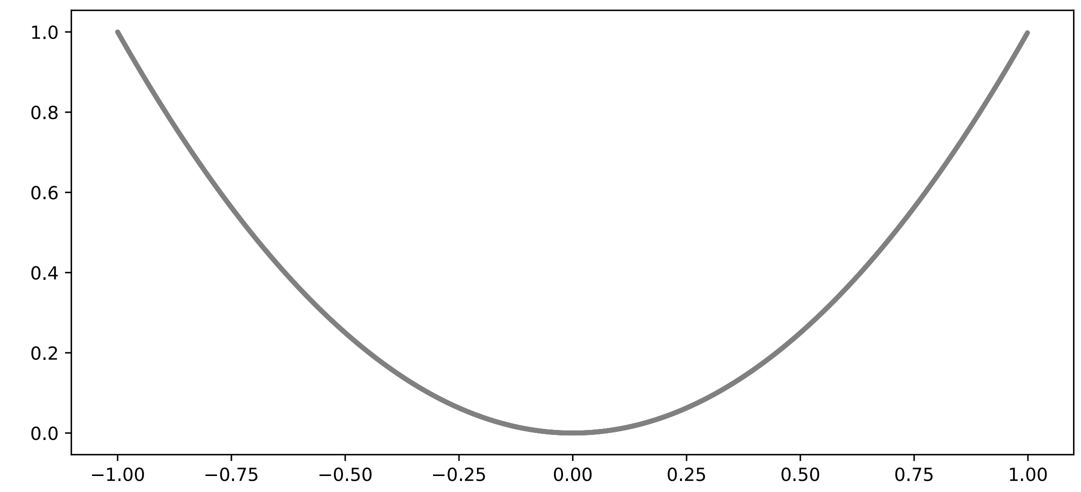
- **Curse of dimensionality**
- Suppose you need 10 different values for every parameter (way too low) to scan the whole parameter space
- Number of combinations is 10^19

EXPLORING HIGH-D PARAMETER SPACES

- **Curse of dimensionality**
- Suppose you need 10 different values for every parameter (way too low) to scan the whole parameter space
- Number of combinations is 10^19
- Total time required to calculate everything is 100x age of the universe!
- Need another way to cleverly scan the parameter space

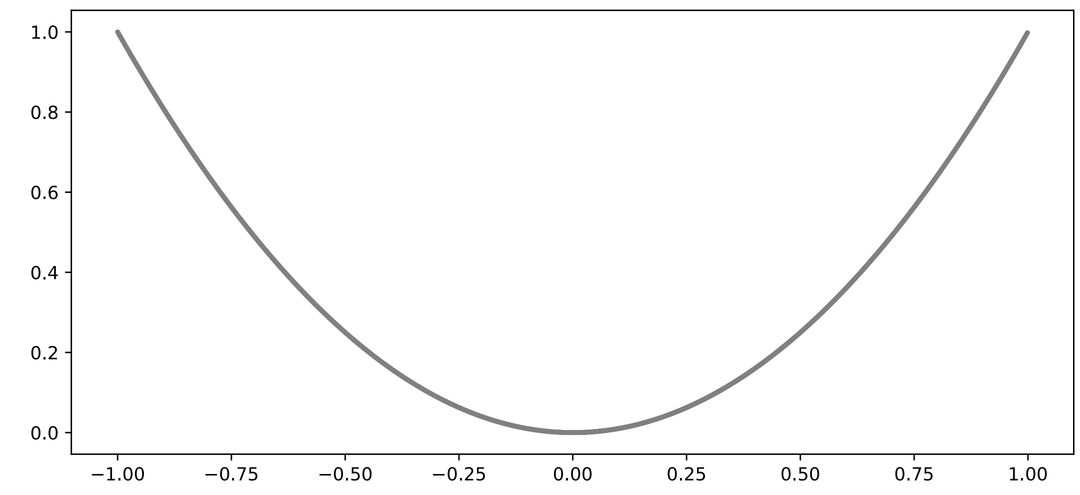
EXPLORING HIGH-D PARAMETER SPACES

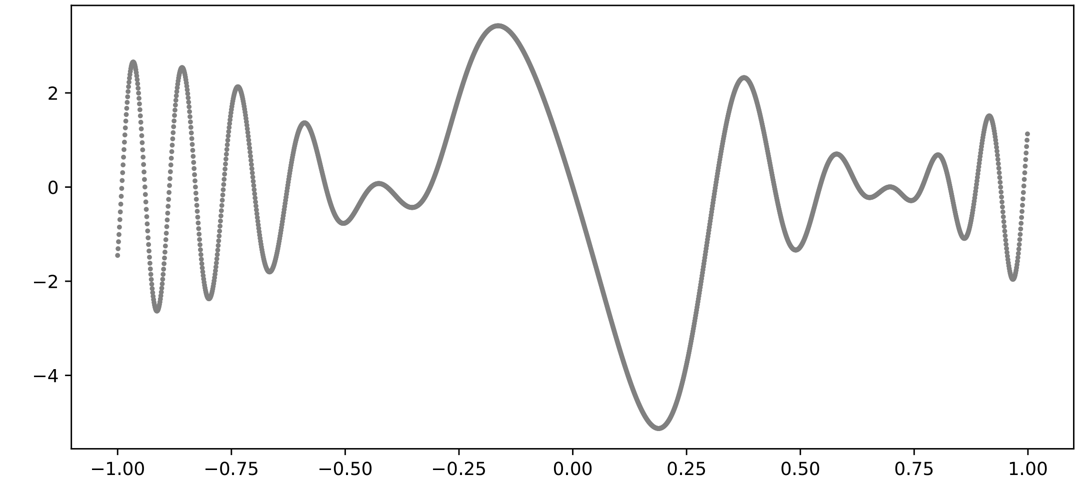
- Visualise: 1D parameter space
- Remember: you don't know the parameter space, you can only sample points (x) and get the likelihood in that point (y)
- Finding the minimum is easy in the top plot (gradient descent from any random starting position)



EXPLORING HIGH-D PARAMETER SPACES

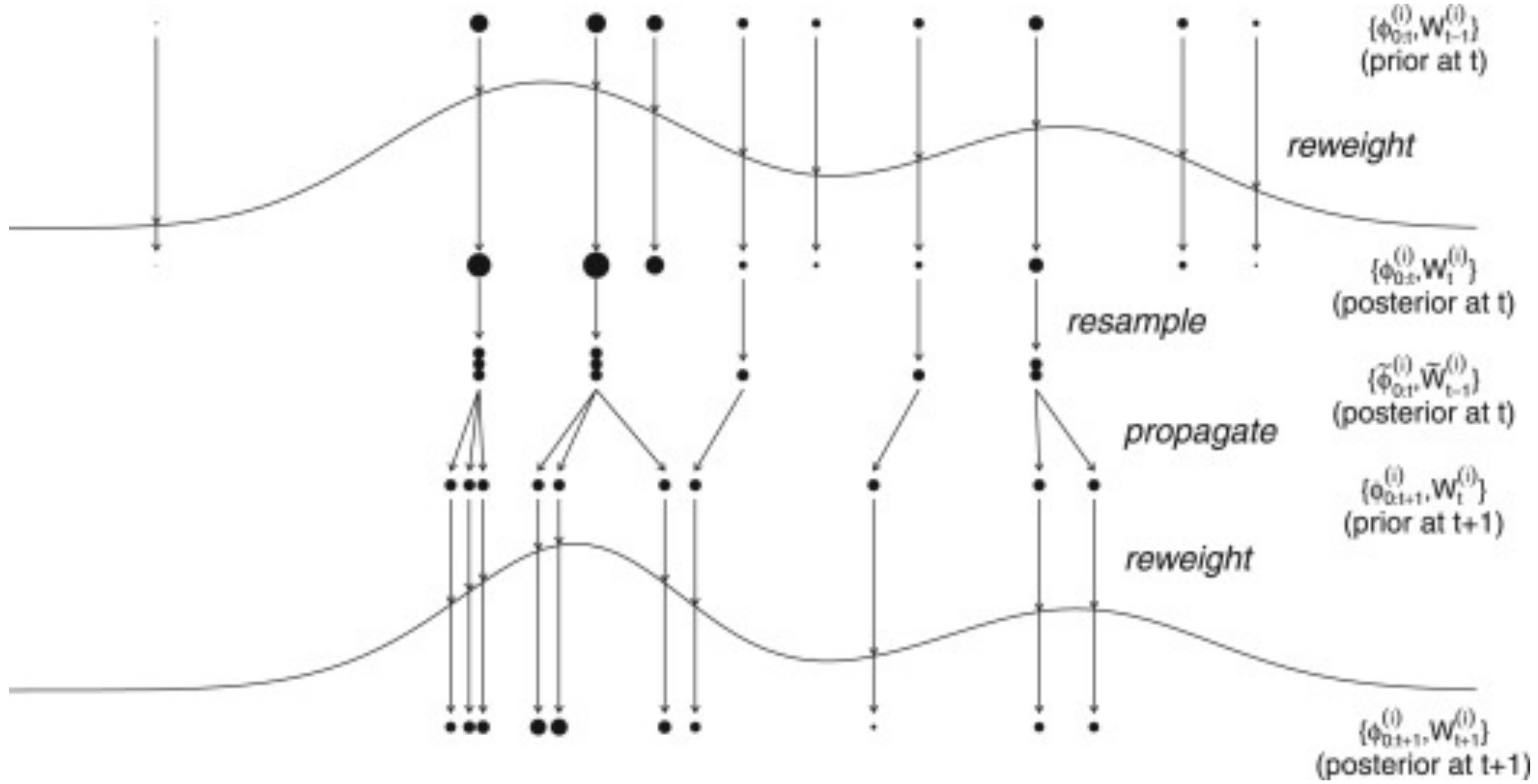
- Visualise: 1D parameter space
- Remember: you don't know the parameter space, you can only sample points (x) and get the likelihood in that point (y)
- Finding the minimum is easy in the top plot (gradient descent from any random starting position)
- Finding the minimum in the second plot is way harder. Depending on the starting position, you end up in different minima





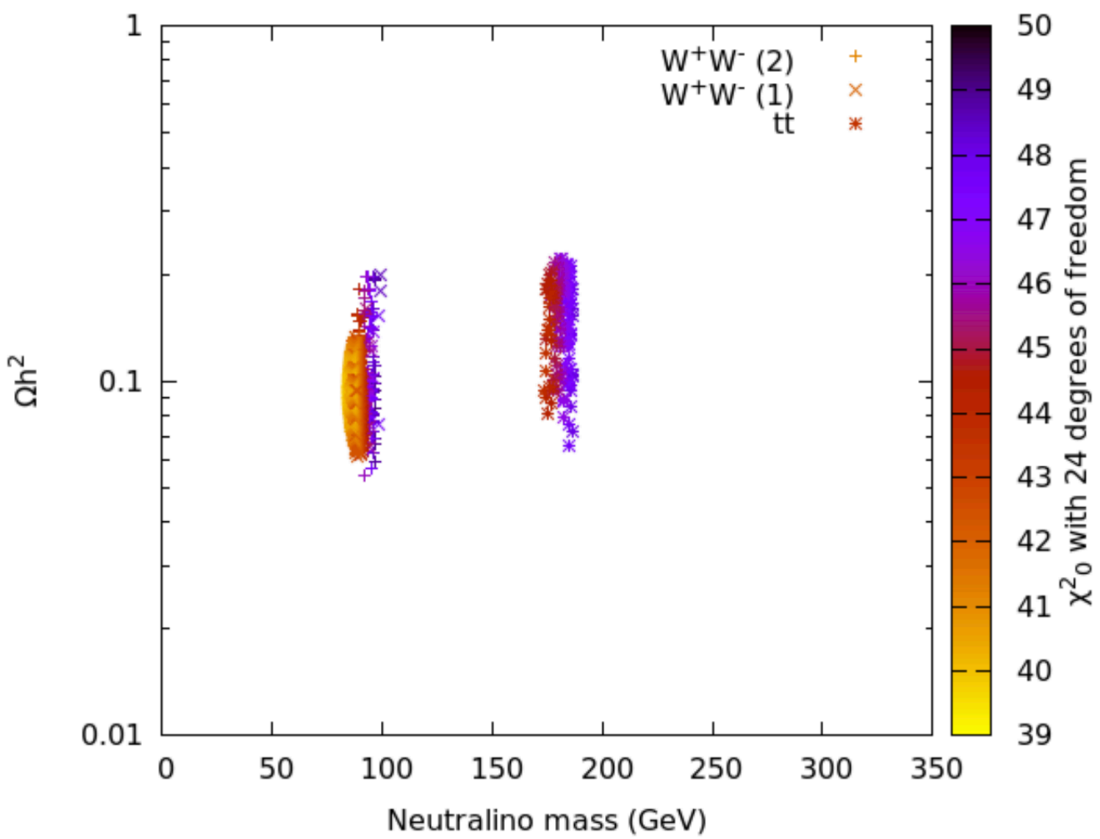
EXPLORING HIGH-D PARAMETER SPACES

- Possible solution: Gaussian particle filtering

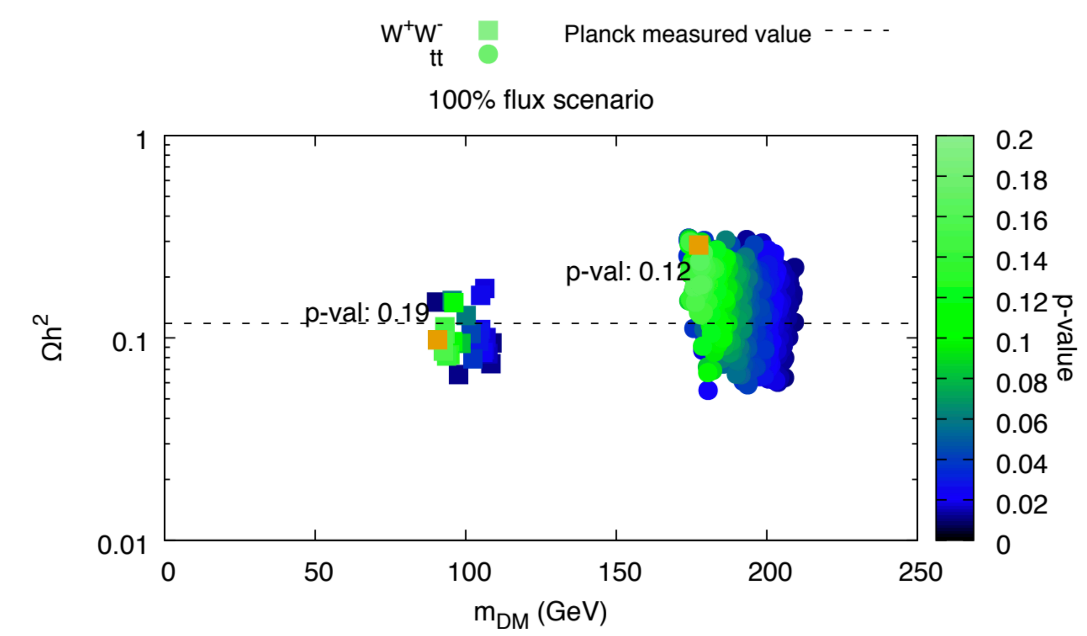


EXPLORING HIGH-D PARAMETER SPACES

- Found region in parameter space of 10^-37 of the total volume
- Not excluded from any experiment, still after 5 years



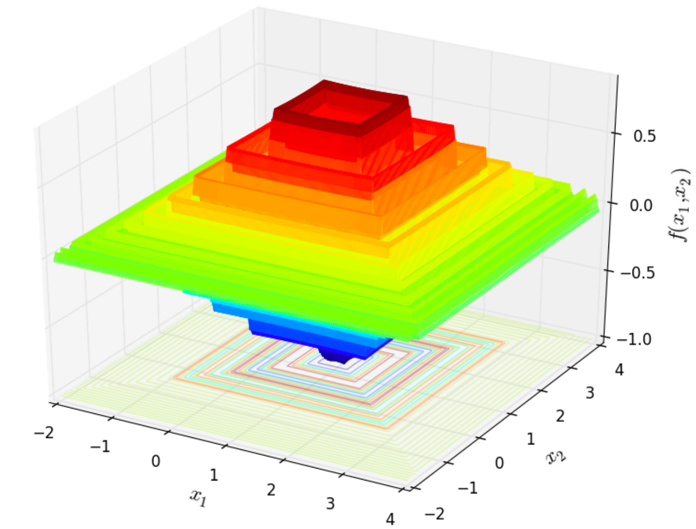
https://arxiv.org/abs/1502.05703

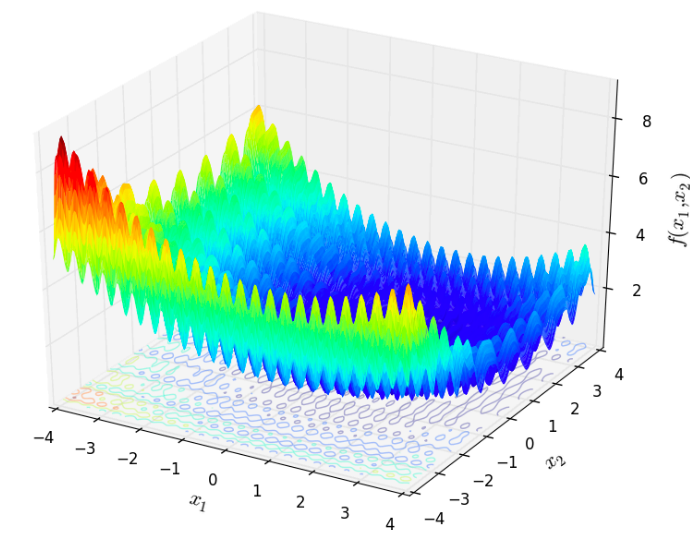


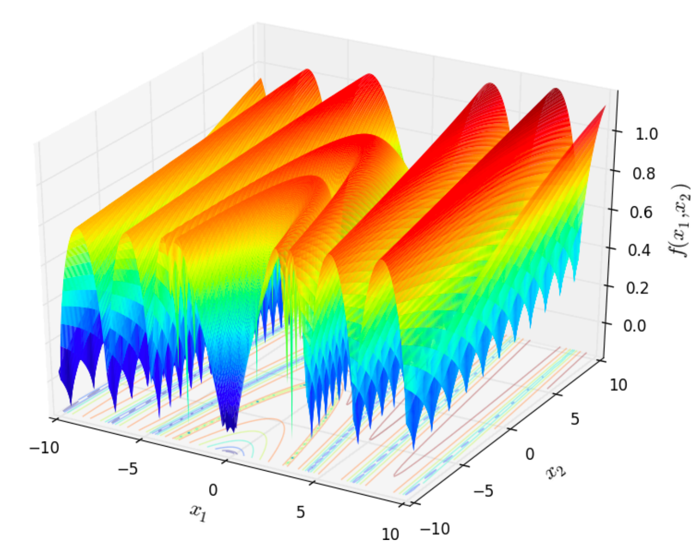
https://arxiv.org/abs/1709.10429

EXPLORING HIGH-D PARAMETER SPACES

- Gaussian particle filter is just one method to scan parameter spaces
  Super parallelizable but does not use gradient information
  - Genetic algorithms
  - Nested sampling (multinest)
  - Tunneling methods
- When does which method work best?
- Challenge: hidden test function like on the right
  - Try different methods and see which one performs
  - Results will be published soon







http://infinity77.net/global_optimization

UNSUPERVISED COLLIDER SEARCHES

- Second topic: find new physics in particle colliders

UNSUPERVISED COLLIDER SEARCHES

- Second topic: find new physics in particle colliders
- Mainly about BSM theories and the LHC
- Typical: theory finds a particle candidate, then experiment tries to find or exclude it

UNSUPERVISED COLLIDER SEARCHES

- Second topic: find new physics in particle colliders
- Mainly about BSM theories and the LHC
- Typical: theory finds a particle candidate, then experiment tries to find or exclude it
- Curse of dimensionality!
  You cannot exclude the whole pMSSM ever, and that's just one theory…

UNSUPERVISED COLLIDER SEARCHES

- Second topic: find new physics in particle colliders
- Mainly about BSM theories and the LHC
- Typical: theory finds a particle candidate, then experiment tries to find or exclude it
- Curse of dimensionality!
  You cannot exclude the whole pMSSM ever, and that's just one theory…
- Alternative:
  - The experiment records data
  - Compare with expectation from only SM hypothesis
  - If rejected -> look at the events that reject that hypothesis and try to explain

UNSUPERVISED COLLIDER SEARCHES

- Second topic: find new physics in particle colliders
- Mainly about BSM theories and the LHC
- Typical: theory finds a particle candidate, then experiment tries to find or exclude it
- Curse of dimensionality!
  You cannot exclude the whole pMSSM ever, and that's just one theory…
- Alternative:
  - The experiment records data
  - Compare with expectation from only SM hypothesis
  - If rejected -> look at the events that reject that hypothesis and try to explain
  - (=unsupervised search of new physics)

UNSUPERVISED COLLIDER SEARCHES

- Typical setup of the experiment:
  - Compare experiment data (real data) to expected data from only SM (simulated data)
  - Real data contains SM plus possible, but unknown, signal
  - Two datasets:
    - SM only (from simulation)
    - SM + possible signal (from real data)

UNSUPERVISED COLLIDER SEARCHES

- For evaluating performance, simulate also signals and pretend you don't know. Gives two datasets:
  - Train on SM only simulated data
  - Test on SM+signal simulated data

UNSUPERVISED COLLIDER SEARCHES

- For evaluating performance, simulate also signals and pretend you don't know. Gives two datasets:
  - Train on SM only simulated data
  - Test on SM+signal simulated data
- Counting experiment:
  - From SM only hypothesis you expect λ events
  - You measure k events

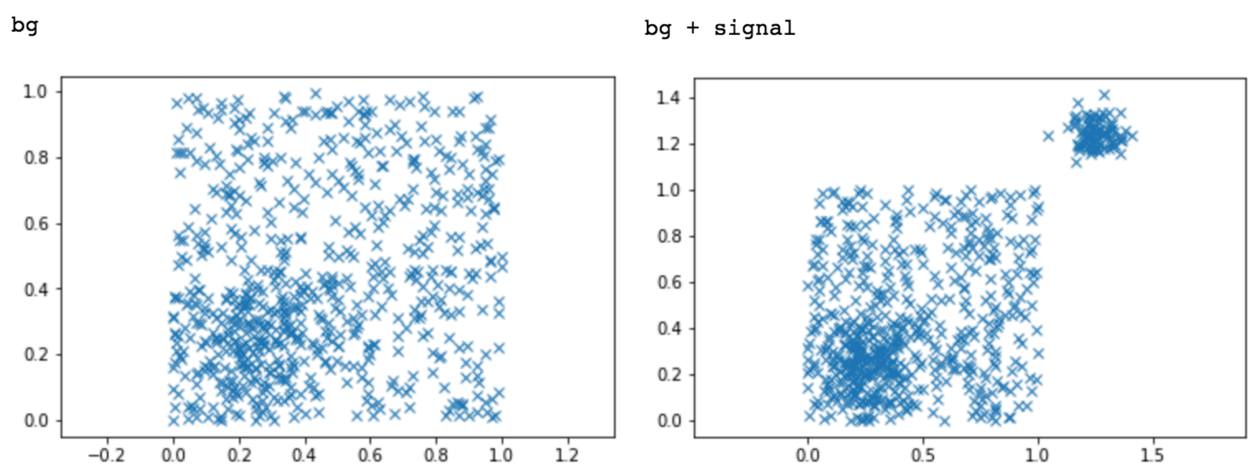
$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

  - (actual statistics a bit more complex, but this is the general idea)
- Filter the data such that you "cut" out the SM so only signal is left, using only SM as your training data

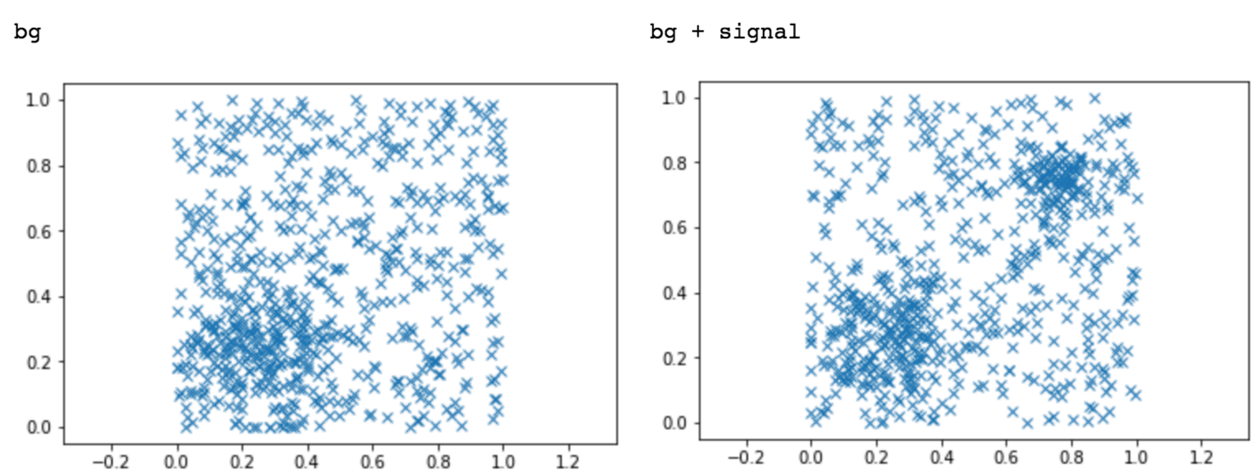
UNSUPERVISED COLLIDER SEARCHES

Two types of signals:

Outlier detection



Density estimation

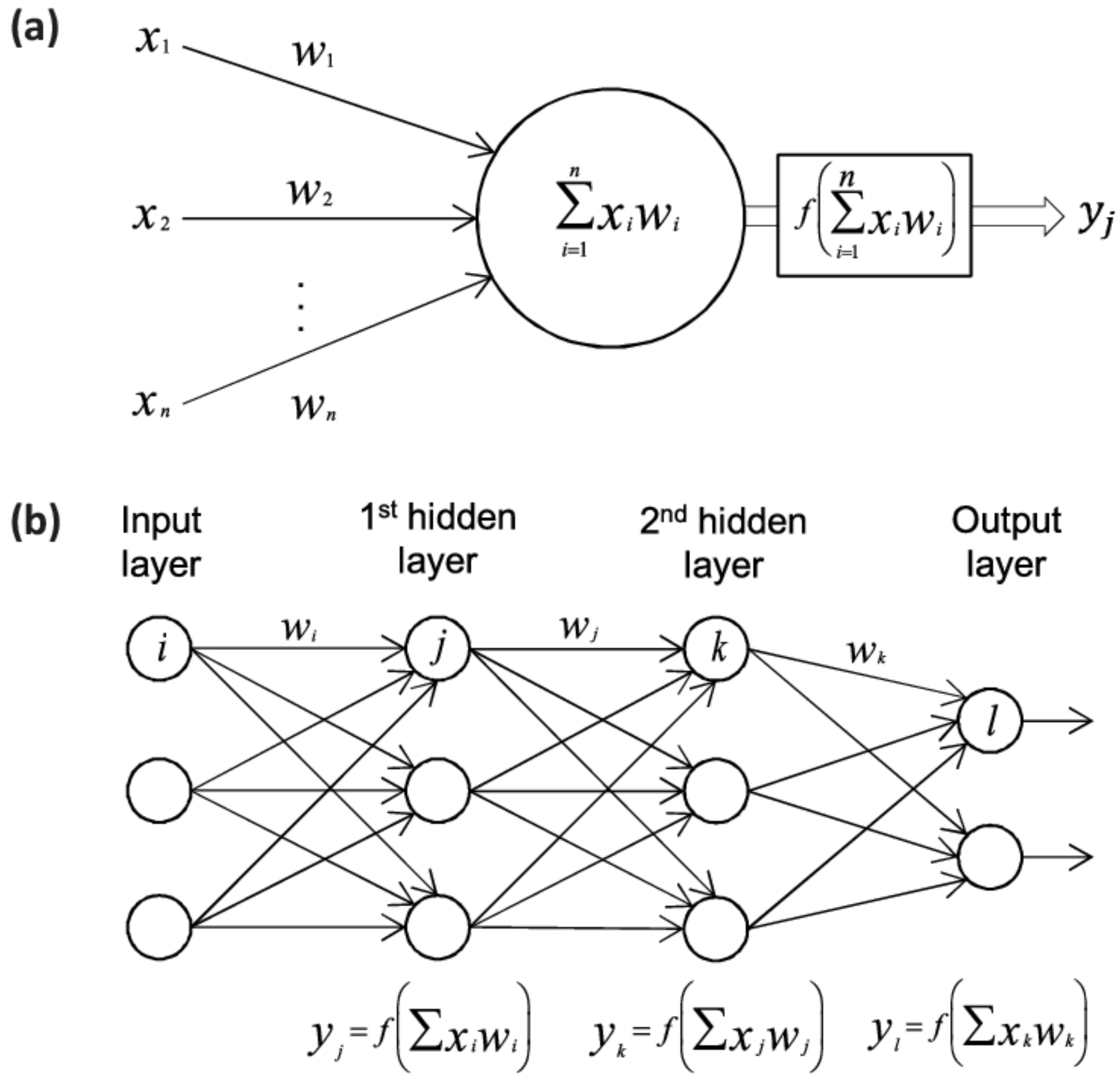


UNSUPERVISED COLLIDER SEARCHES

- Focus on outlier detection
- Huge field with many methods, focus on autoencoders

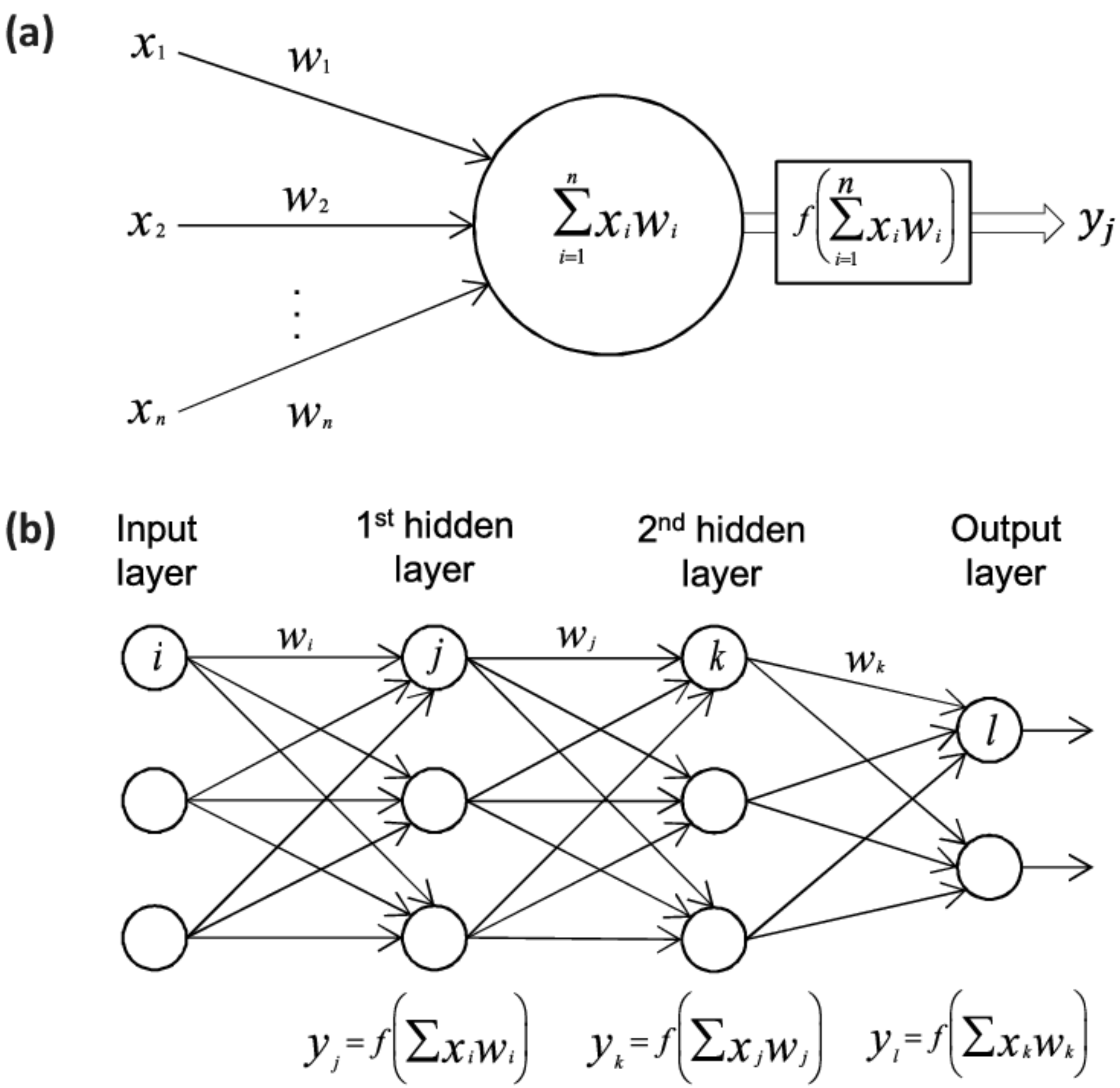
UNSUPERVISED COLLIDER SEARCHES

- Focus on outlier detection
- Huge field with many methods, focus on autoencoders
- One slide neural network introduction:



UNSUPERVISED COLLIDER SEARCHES

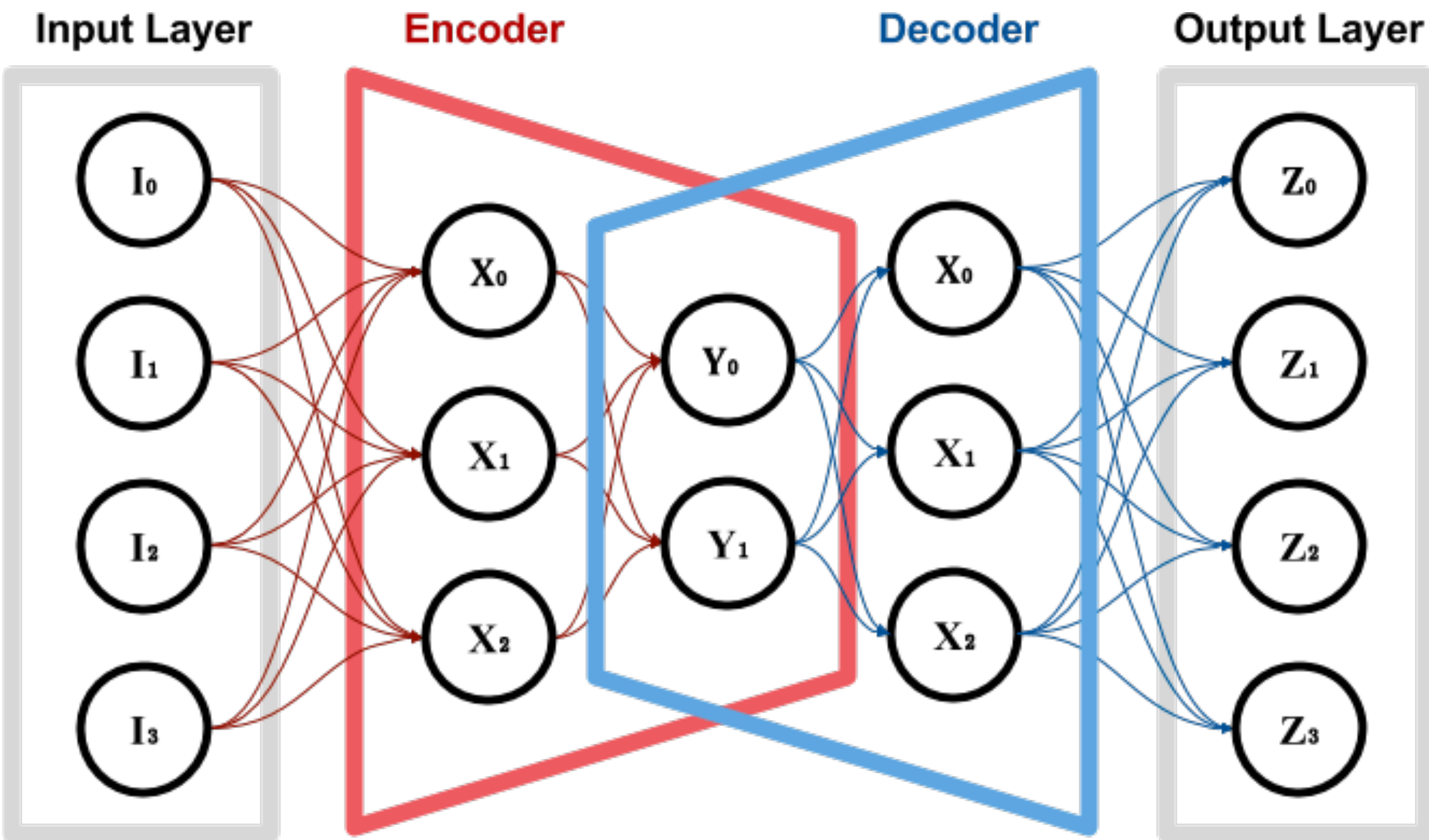
- Focus on outlier detection
- Huge field with many methods, focus on autoencoders
- One slide neural network introduction:



- (training a neural network is a high-D parameter optimisation problem too!)

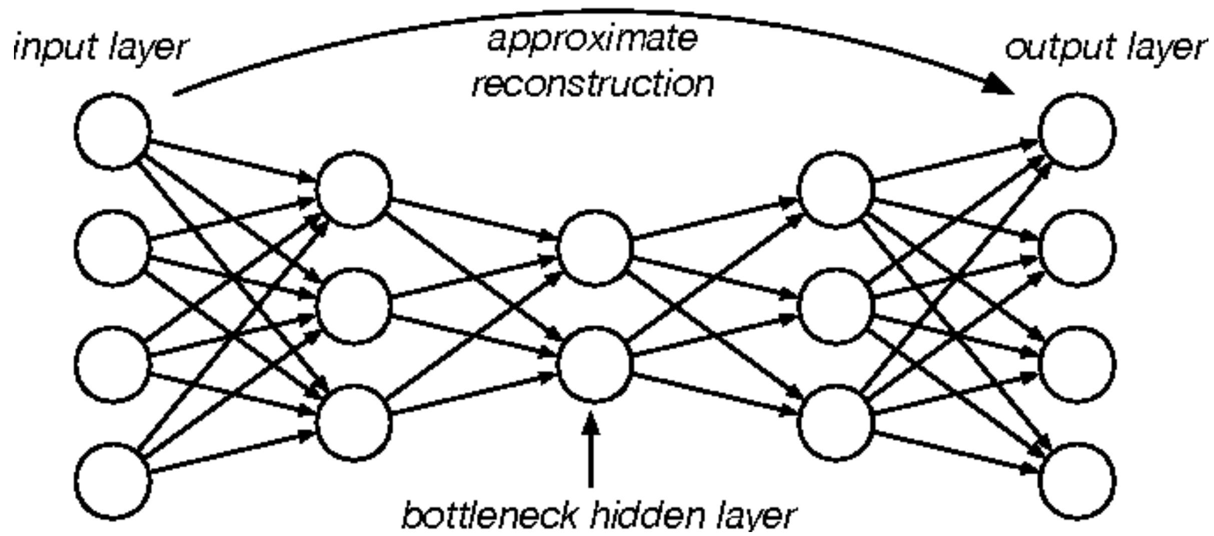
UNSUPERVISED COLLIDER SEARCHES

- Autoencoder

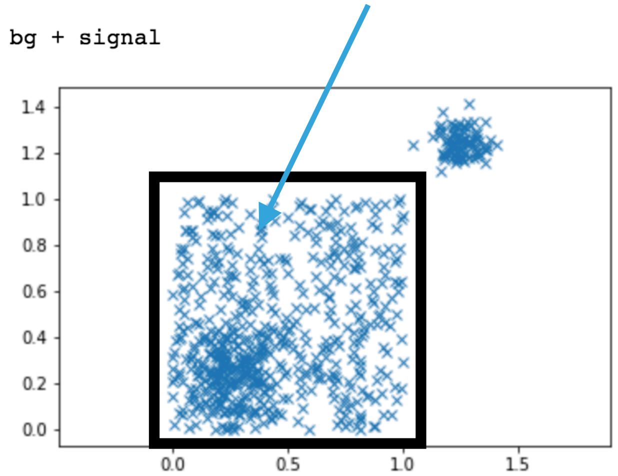


Anomaly score = normalised reconstruction loss (eg MSE)

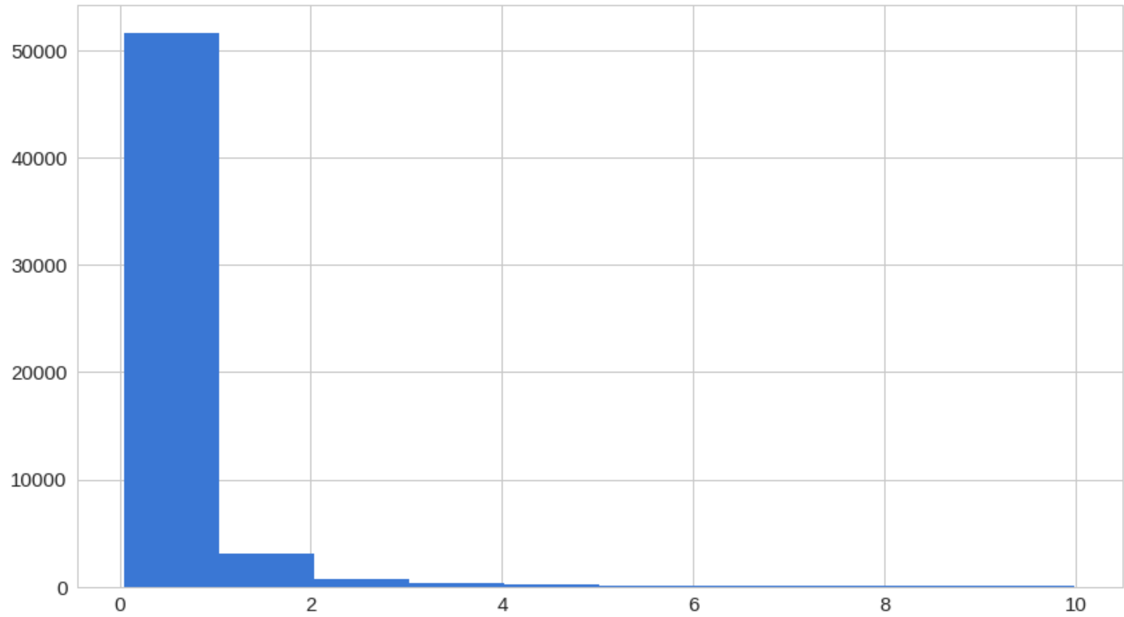
- Example: credit card fraud detection with autoencoder



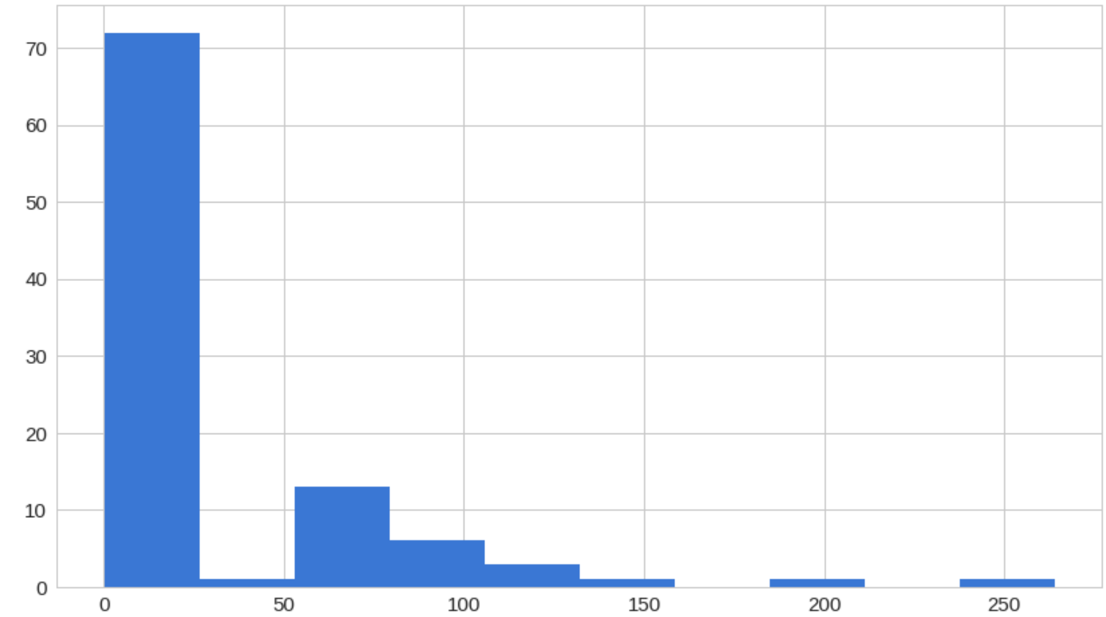
Can only reconstruct well inside the box



No fraud



Fraud

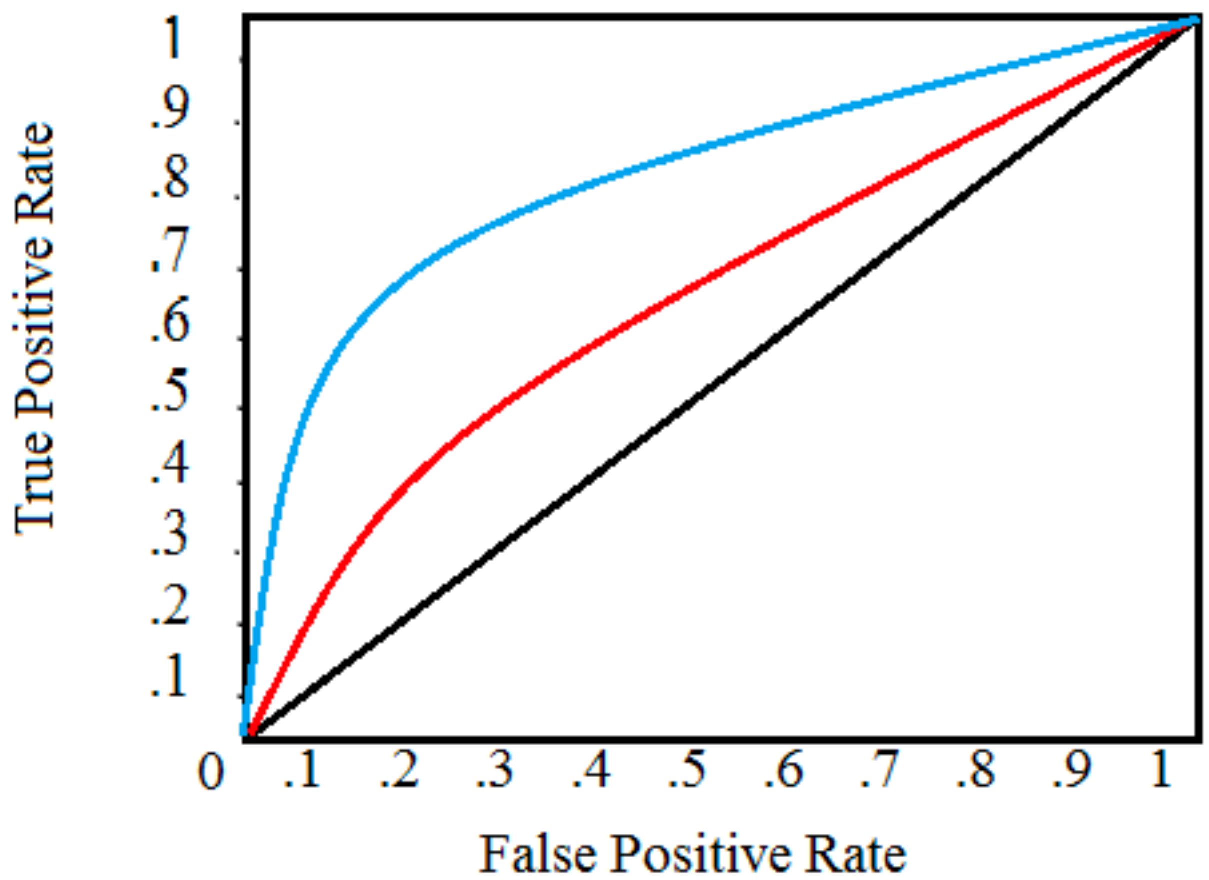


https://medium.com/@curiousily/credit-card-fraud-detection-using-autoencoders-in-keras-tensorflow-for-hackers-part-vii-20e0c85301bd

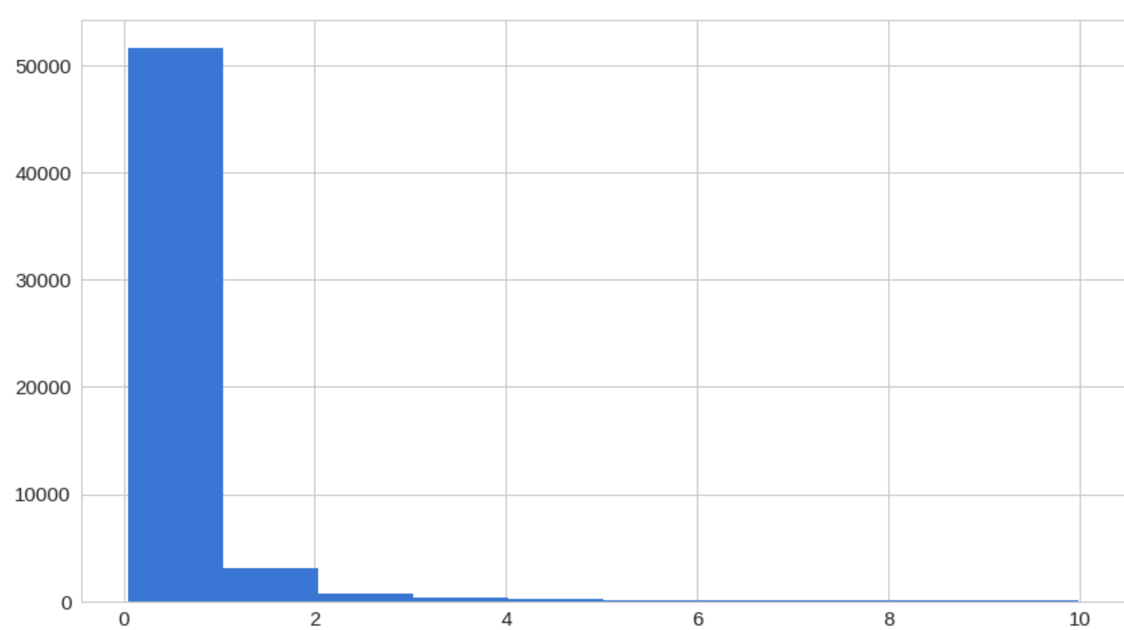
UNSUPERVISED COLLIDER SEARCHES

- Dataset: www.phenomldata.org
- Contains >30GB of simulated LHC events
- Separated in background and various signals

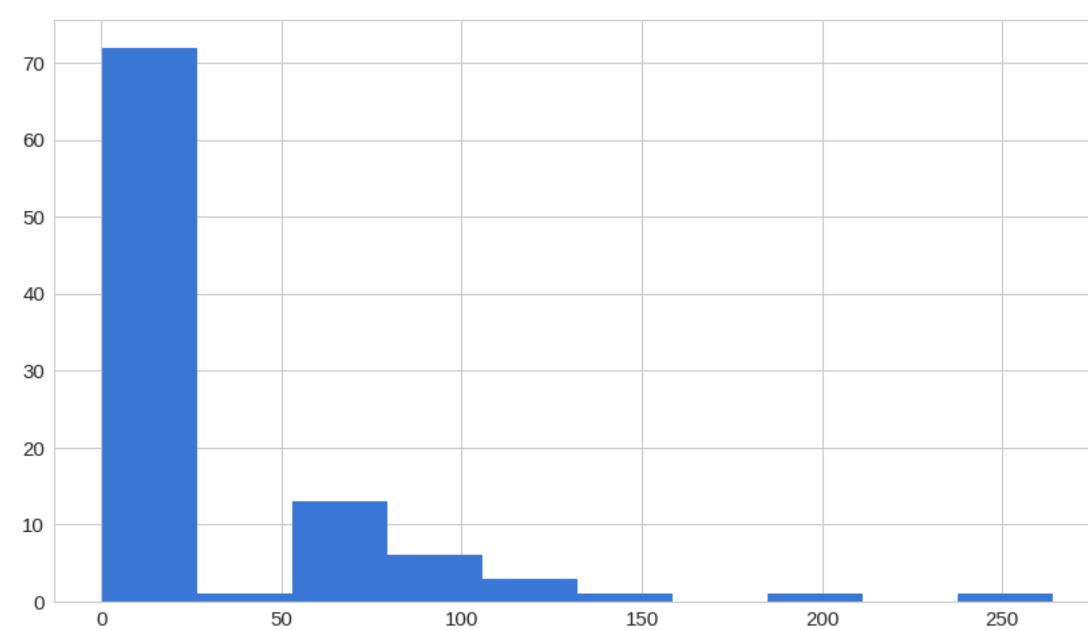
- You can use a ROC curve and the AUC to determine how well an algorithm does



No fraud

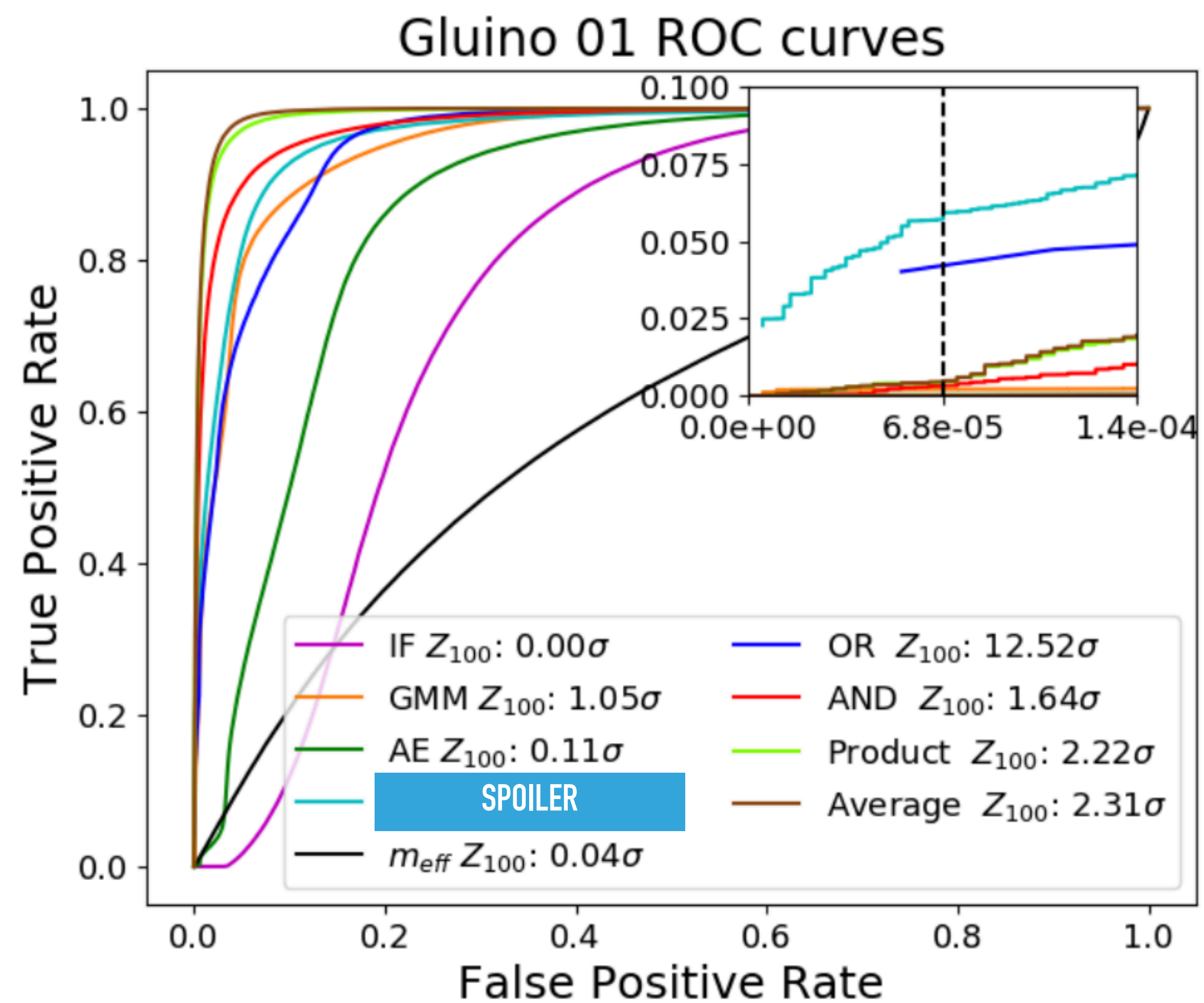


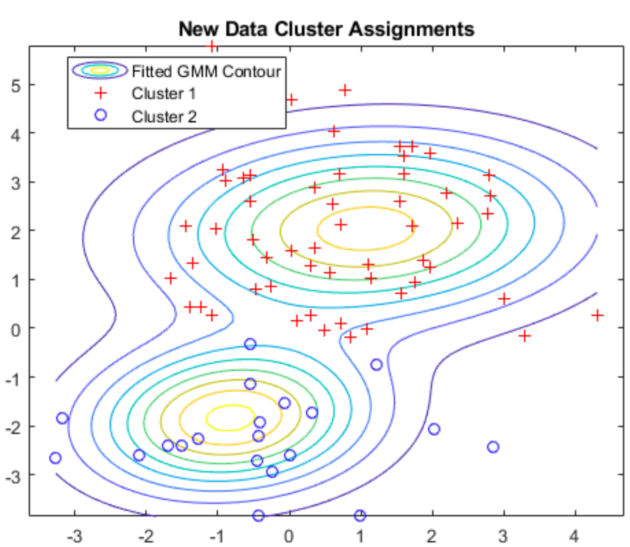
Fraud



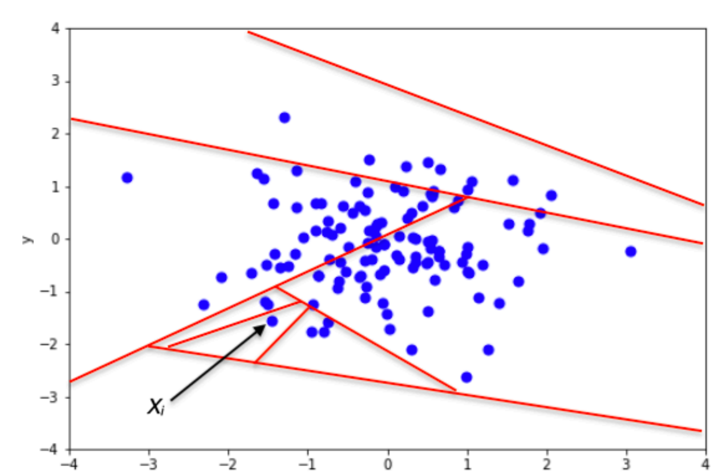
UNSUPERVISED COLLIDER SEARCHES

- You can use a ROC curve and the AUC to determine how well an algorithm does
- Additionally, determine signal efficiency at a predetermined background efficiency





GMM

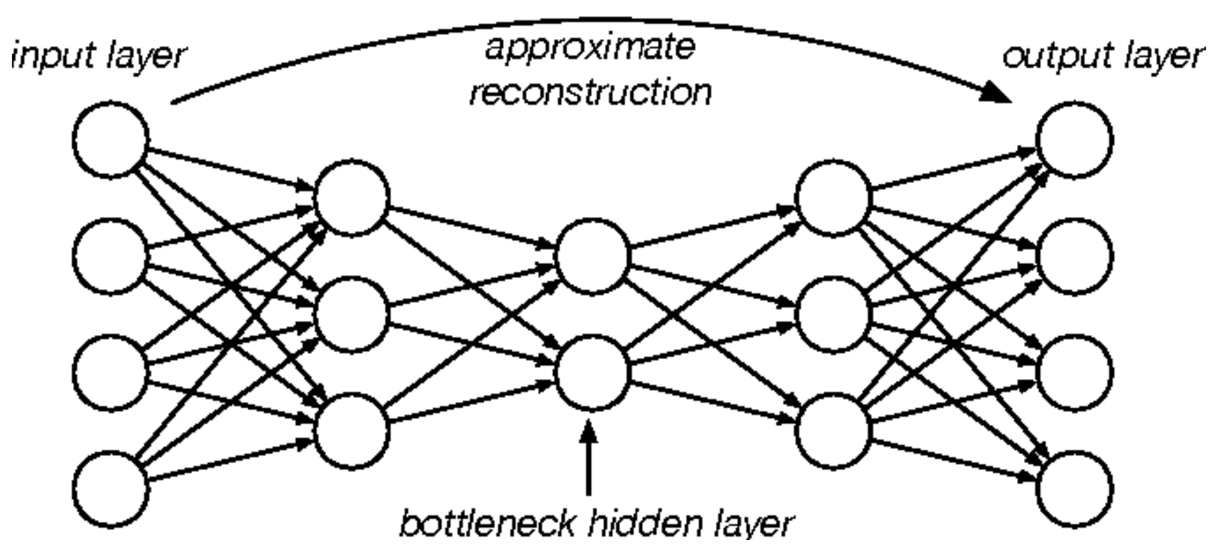


IF

https://arxiv.org/abs/2010.07940

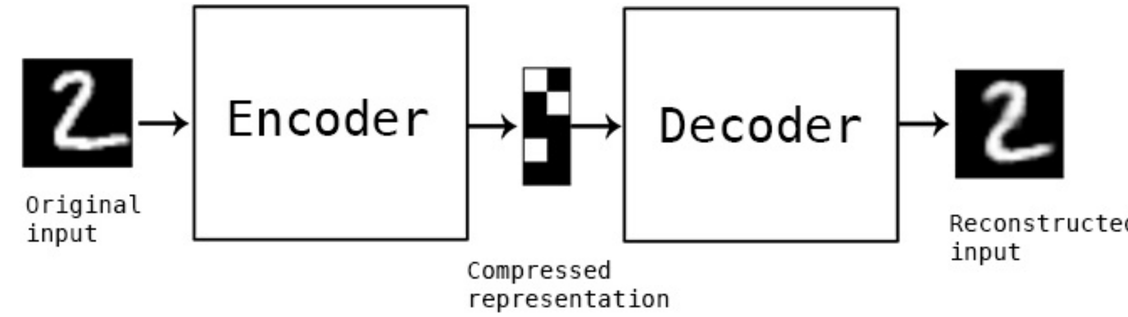
AUTOENCODERS

- If autoencoders are bad, why are they so popular?
- The bottleneck layer interesting
- Transforms 4D to 2D
- Latent space



AUTOENCODERS

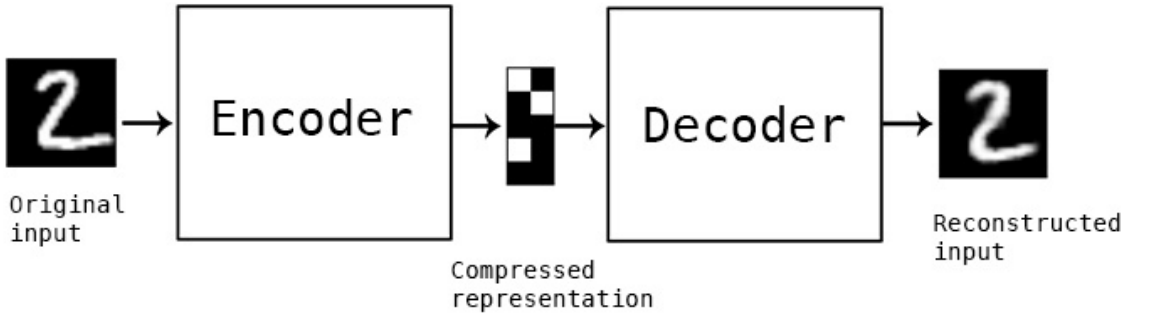
- Autoencoders have no ordering in latent space



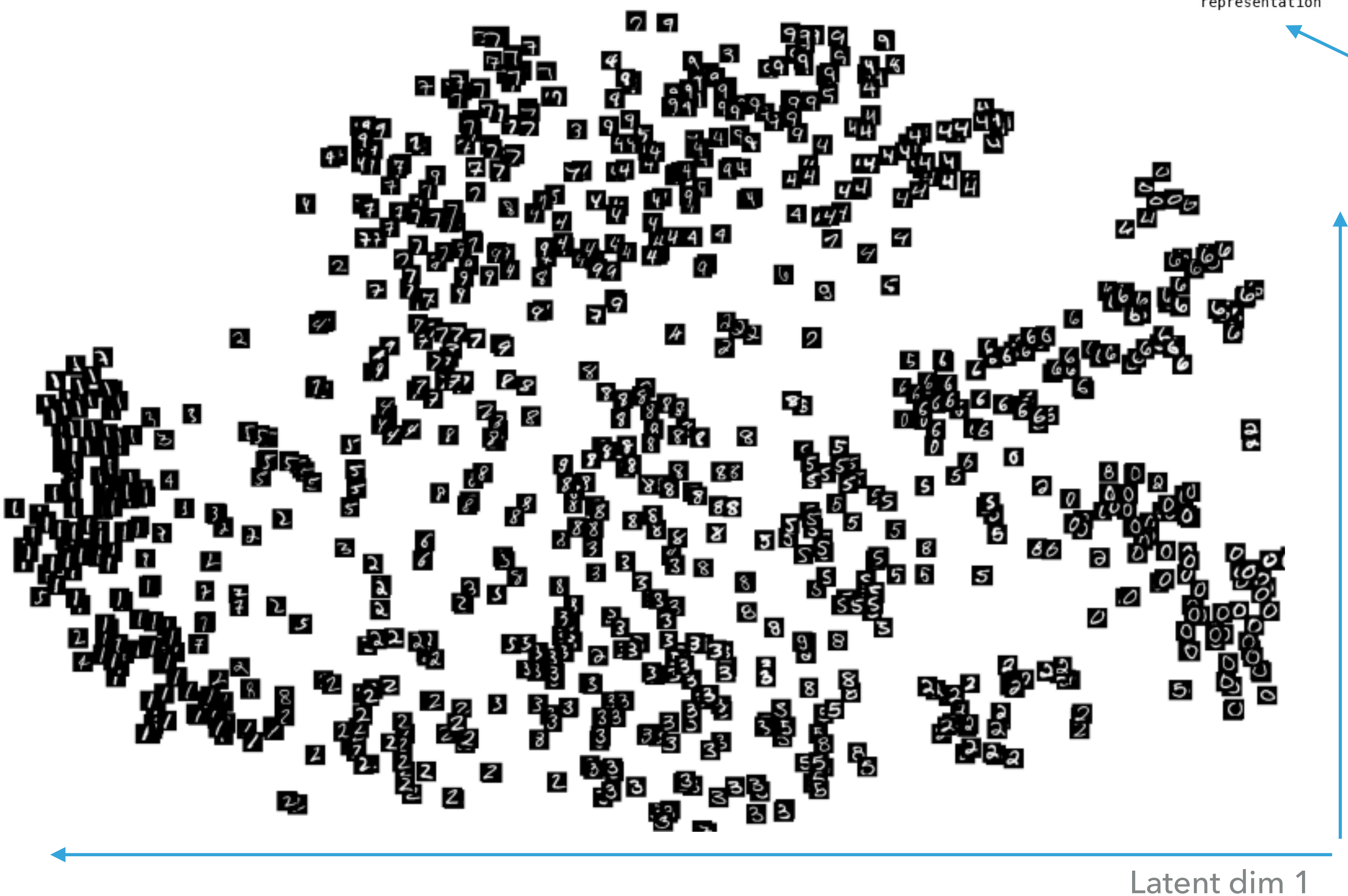
Assume 2D easy viz.

AUTOENCODERS

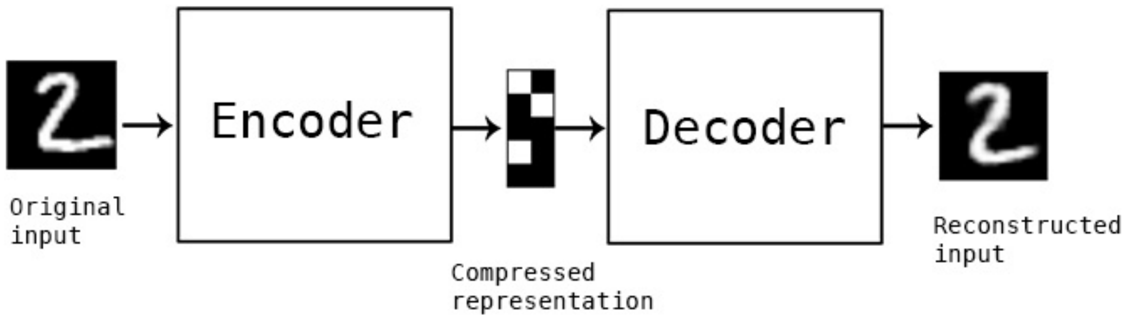
- Autoencoders have no ordering in latent space



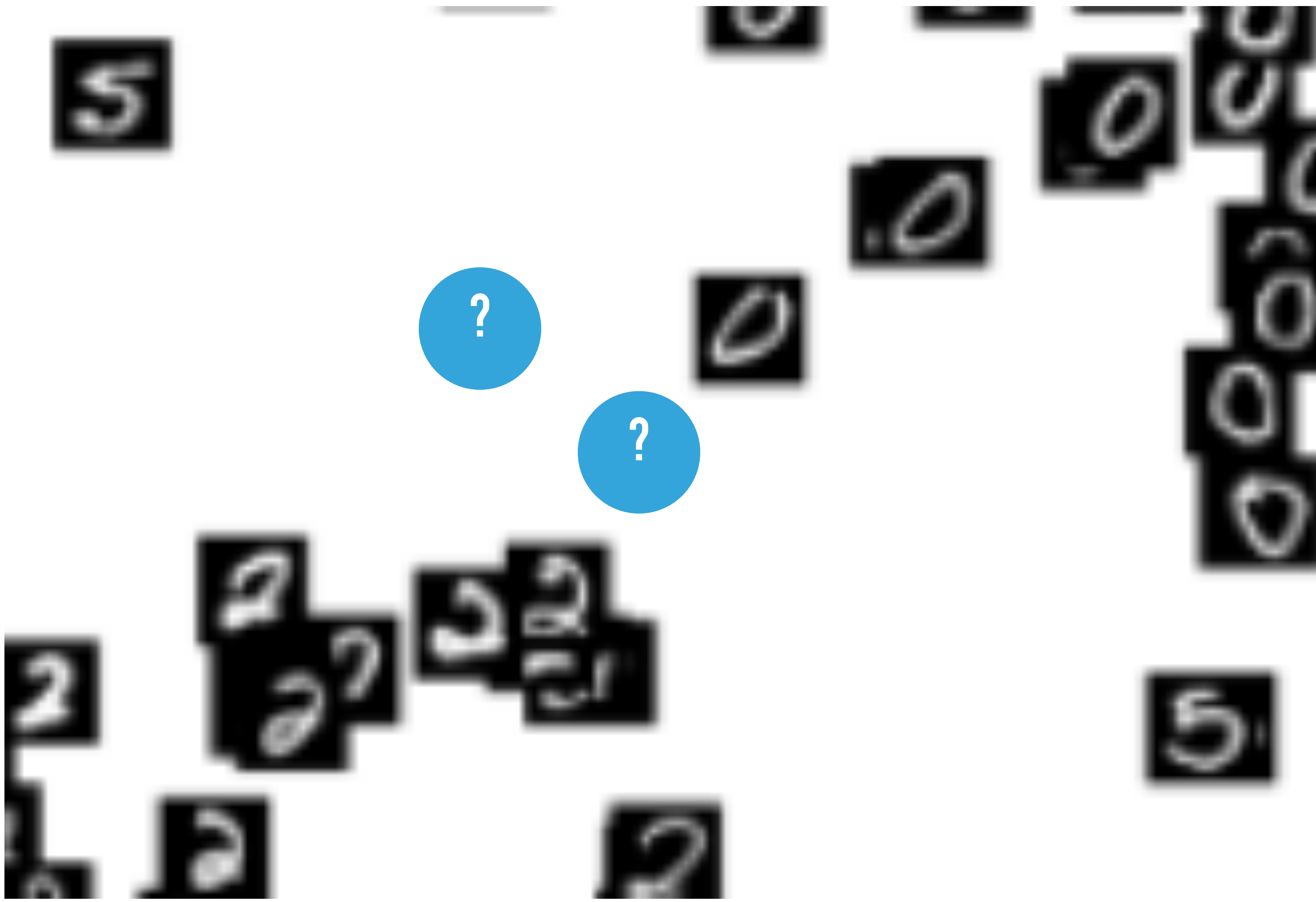
Assume 2D easy viz.



AUTOENCODERS

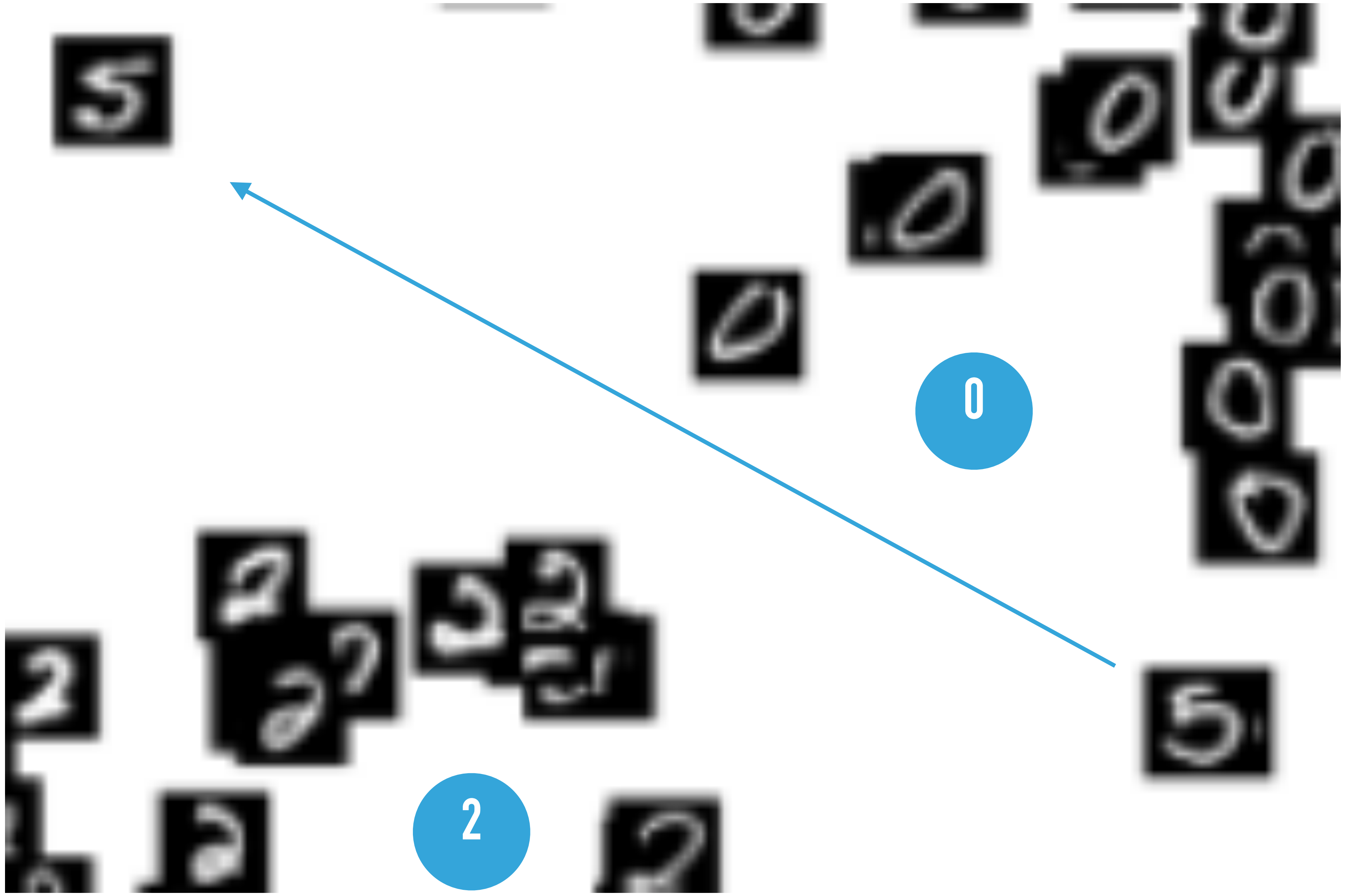


- Input slightly different than training set –> reconstruction loss high, because latent space is ill-defined there
- Not robust
- What is between the data points?

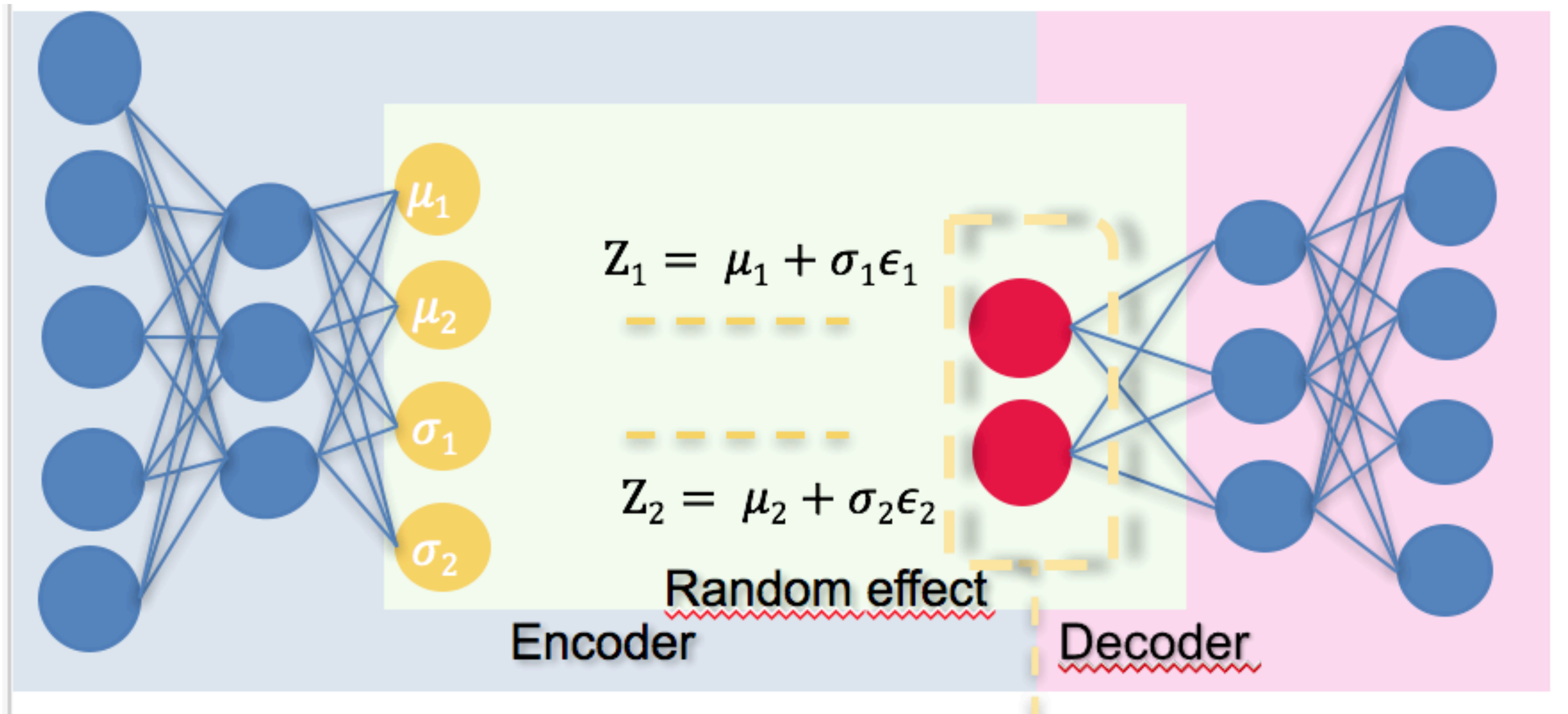


AUTOENCODERS

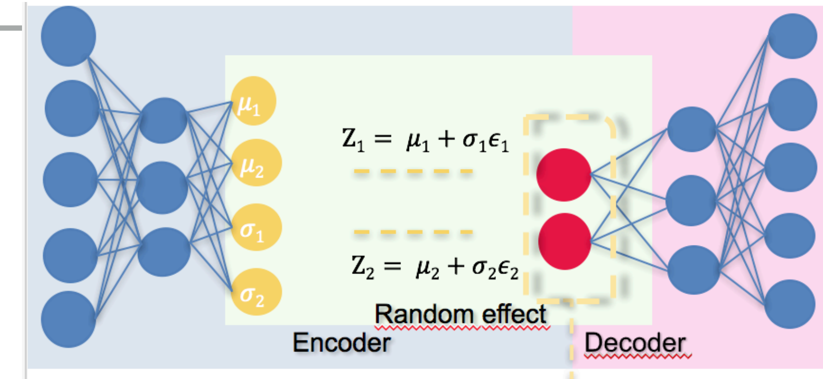
- If only the points could be grouped together…
- Unsupervised clustering, interpolation between data points …



VARIATIONAL AUTOENCODERS

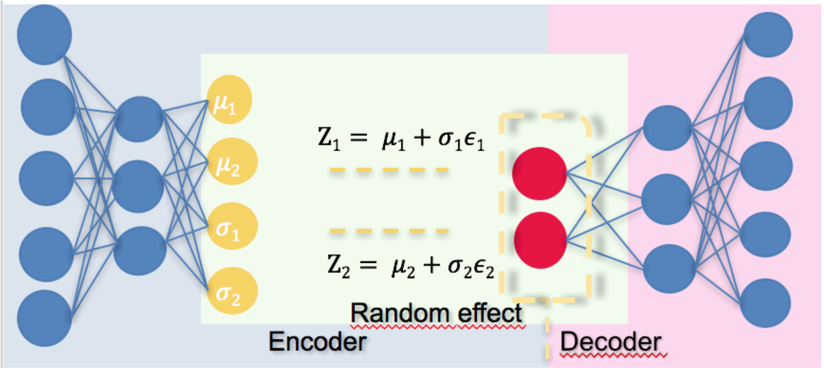


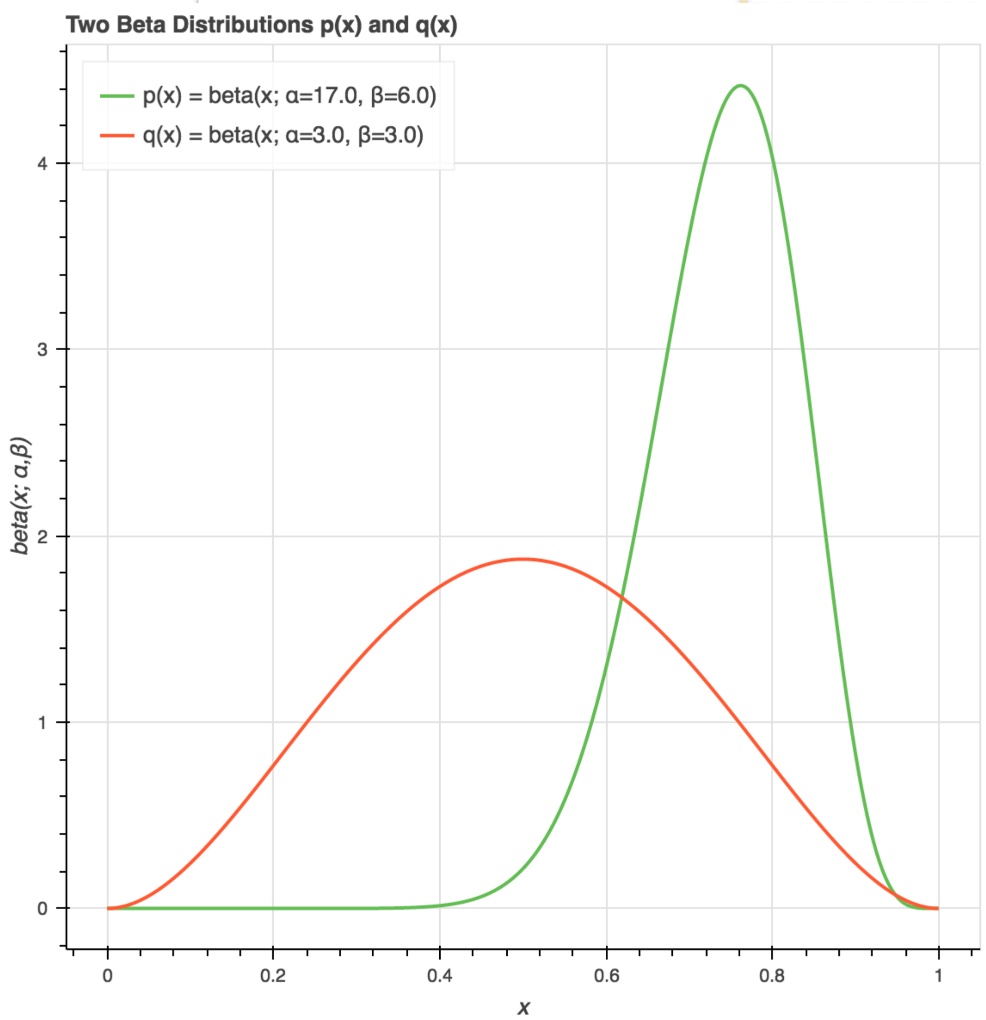
VAE



- Force ordering in latent space
- During training, you are minimising some loss function
- For regression (normal AE): MSE(*output - input*)

VAE



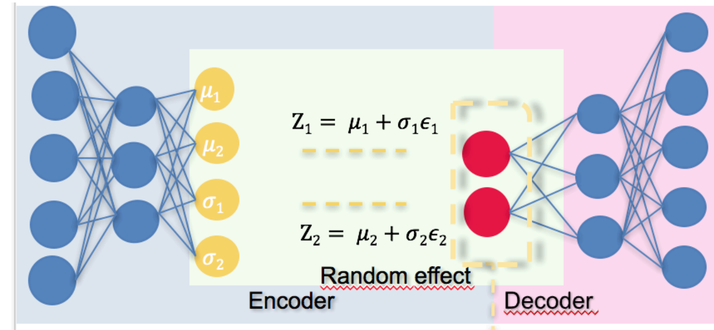


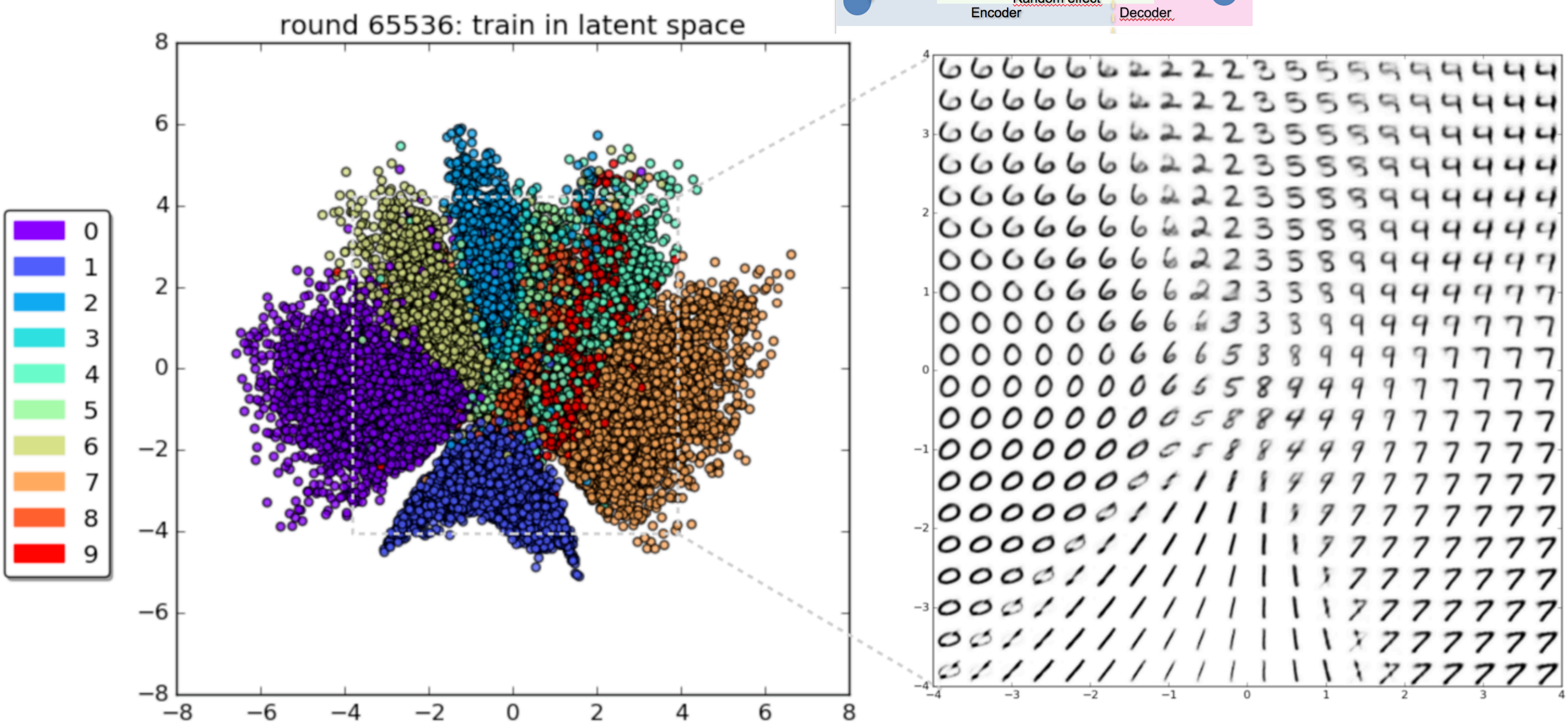
- Force ordering in latent space
- During training, you are minimising some loss function
- For regression (normal AE): MSE(*output - input*)
- Add KL-divergence term: $\Sigma_i \mathrm{KL}(\mathcal{N}(\mu_i, \sigma_i), \mathcal{N}(0,1)) := \mathrm{KL}(\mu,\sigma)$
- So $\mathscr{L} = \mathrm{MSE}(\mathit{output} - \mathit{input}) + \mathrm{KL}(\mu,\sigma)$

- The KL divergence punishes latent space values far away from the center
- Also, every point has a variance that is pushed to 1
- Balance MSE and KL –> group similar structures around the center while keeping RL in check

LATENT SPACE

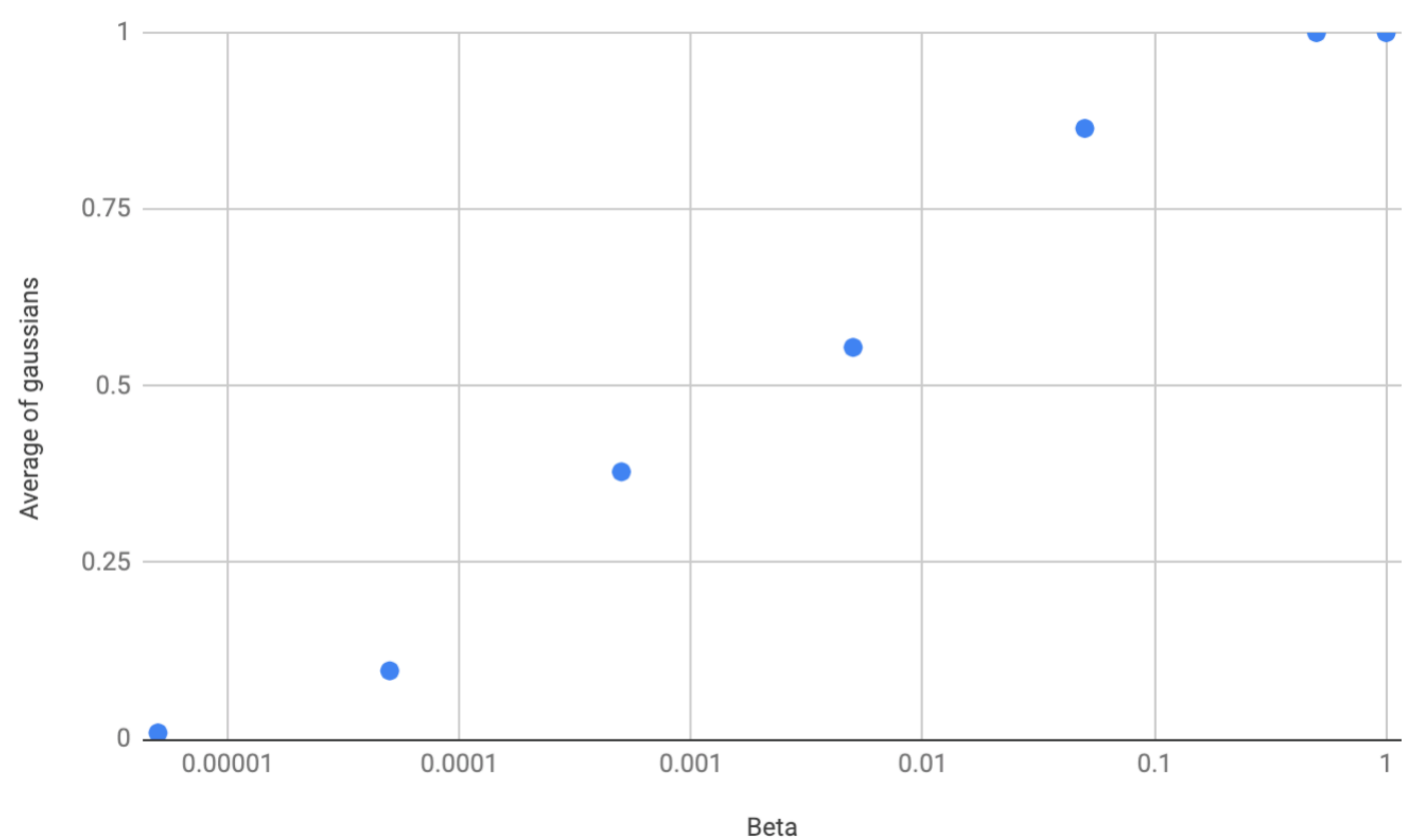
- Same example, but now a VAE





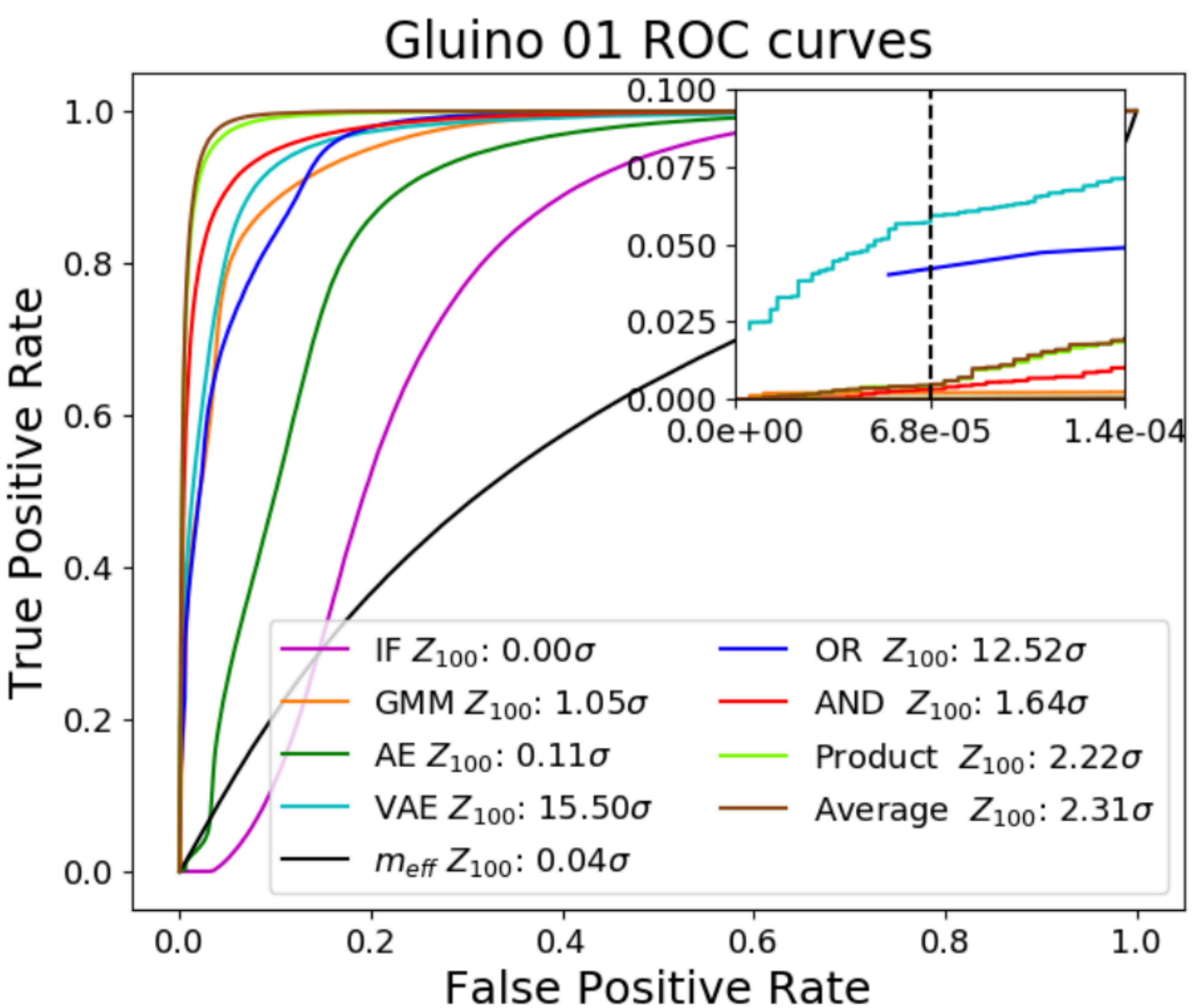
- Balancing MSE and KL is tricky
- Balance using another hyperparameter **β**
  - $\mathscr{L}$ = (1-β) * MSE(*output* - *input*) + β * KL(μ, σ)
- β-VAE

β	Avg var	Avg mean
1	1	1.89E-09
5E-01	0.99999905	2.35E-07
5E-02	0.86448085	...
5E-03	0.554529	
5E-04	0.3784553	
5E-05	0.09676677	
5E-06	0.008932933	
0	0.0000442	

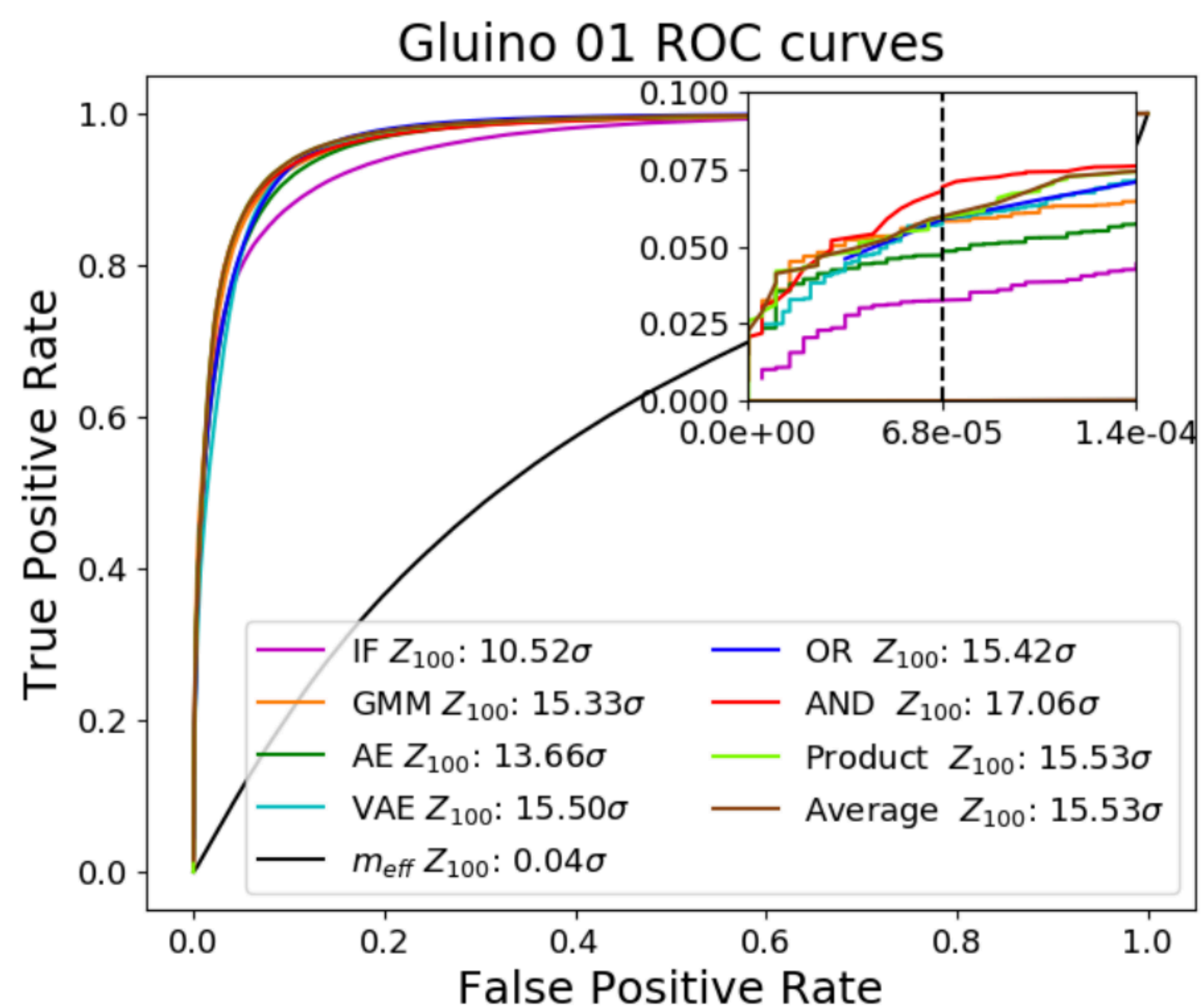


UNSUPERVISED COLLIDER SEARCHES

- Now apply the AE/IF/GMM from before on the latent space of a VAE trained on the background events



4-vector space



Latent space of VAE

https://arxiv.org/abs/2010.07940

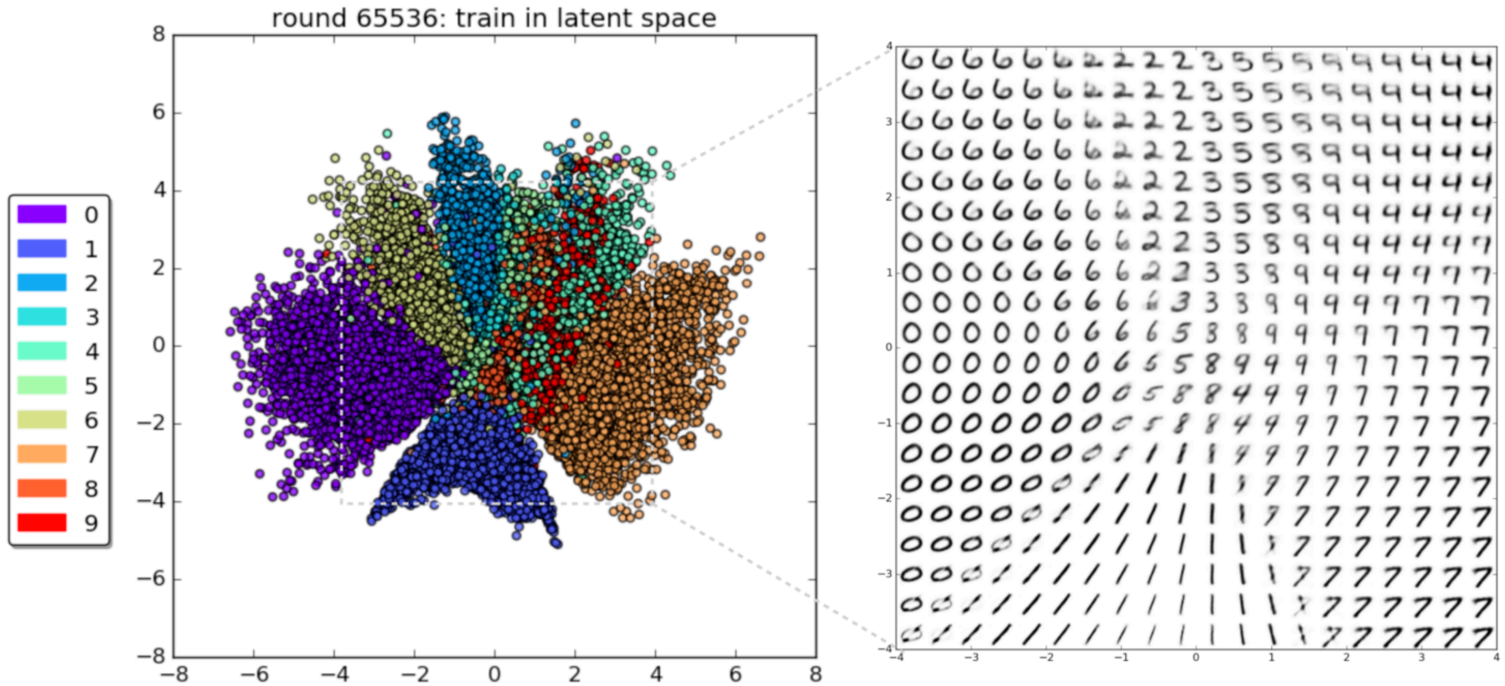
GENERATIVE MODELS AS EVENT GENERATORS

- Topic: generative models as event generators

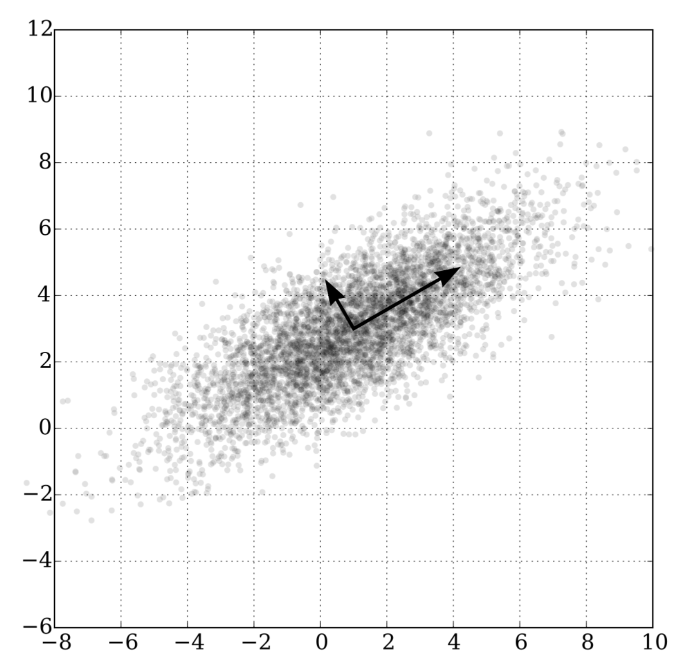
GENERATIVE MODELS AS EVENT GENERATORS

- Topic: generative models as event generators
- To be able to do the previous, need lots of events
- Event generation is slow, especially if you need billions of events and need to run the whole LHC simulation pipeline

- Use the latent space and decoder as generative model
- Explore the latent space

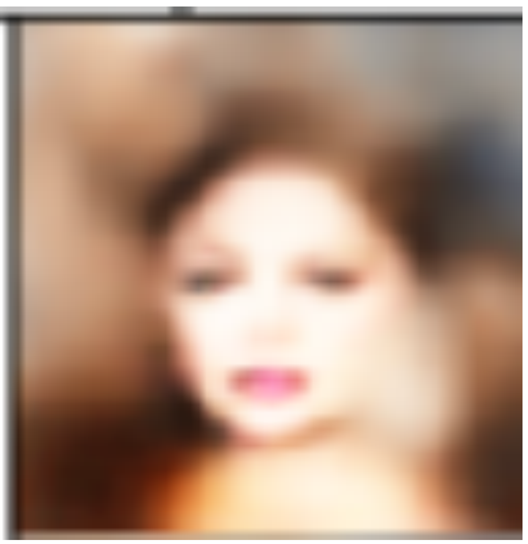


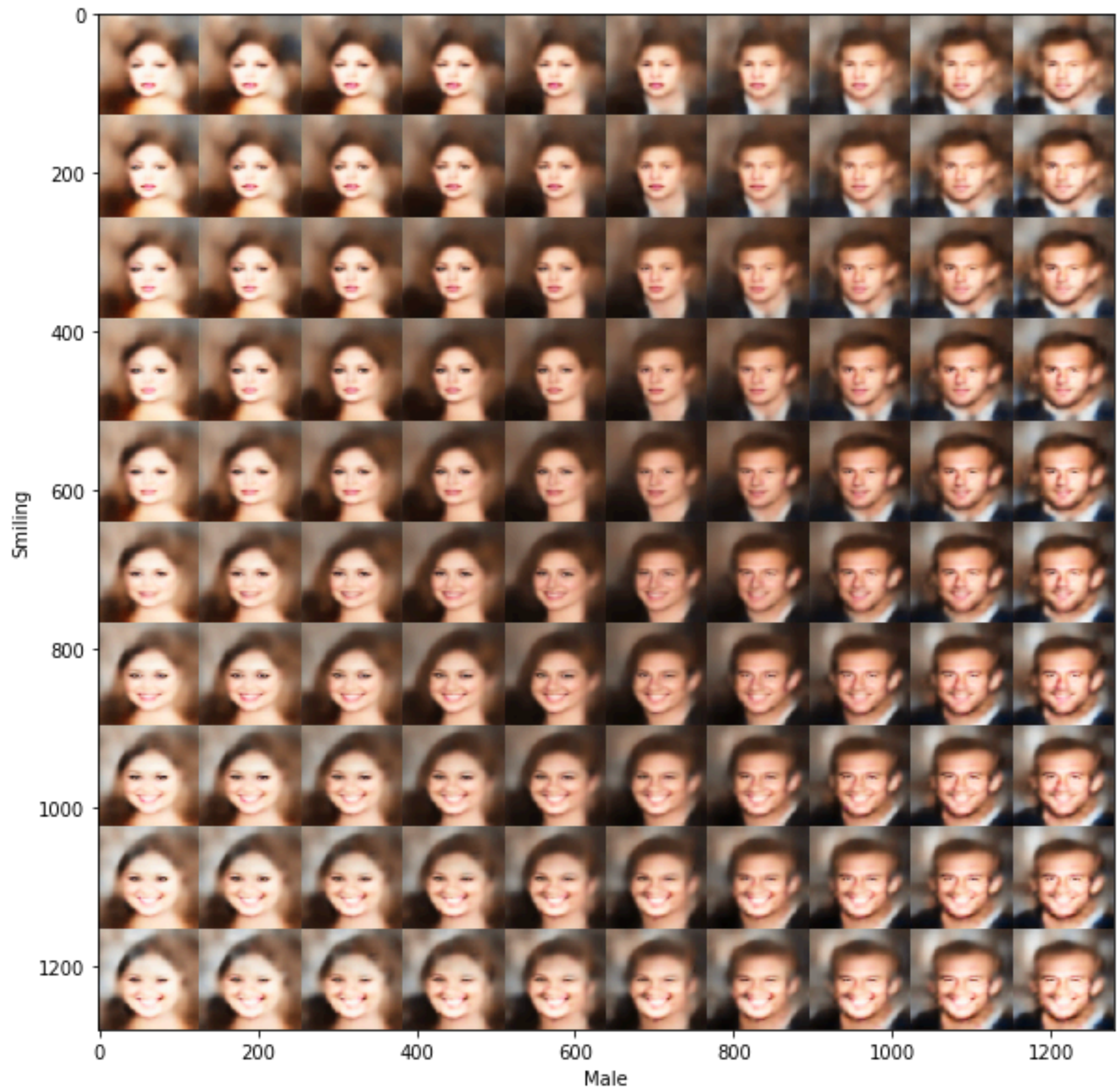
PCA on the latent variables

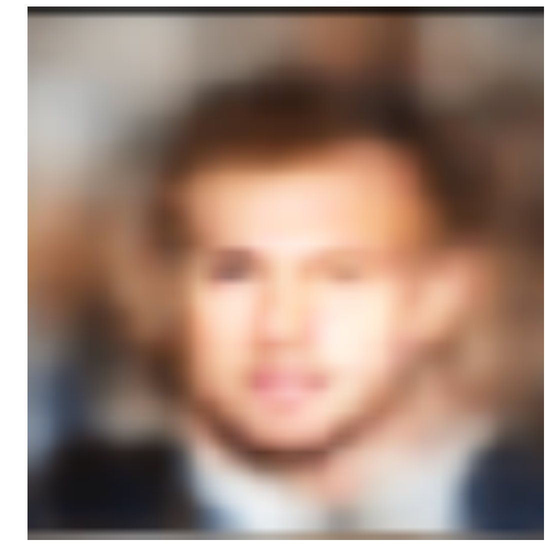


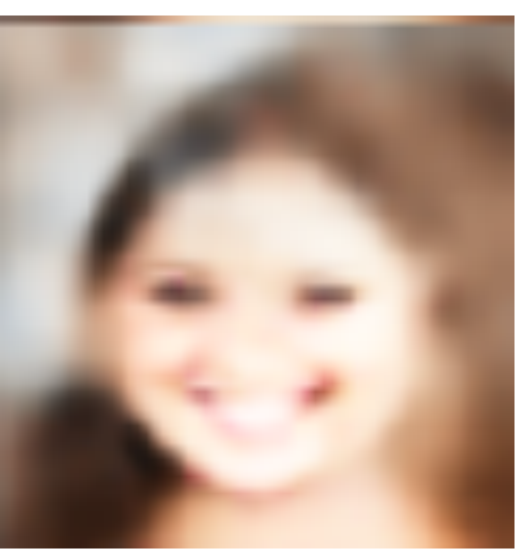
PLAYING WITH LATENT SPACES

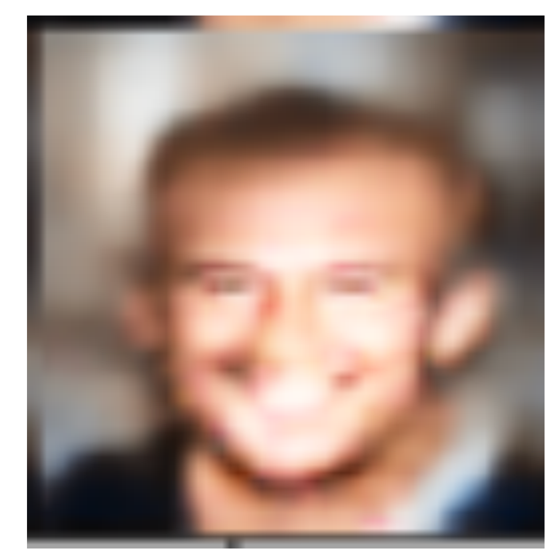
- Train VAE on face images
- Change the latent space variables





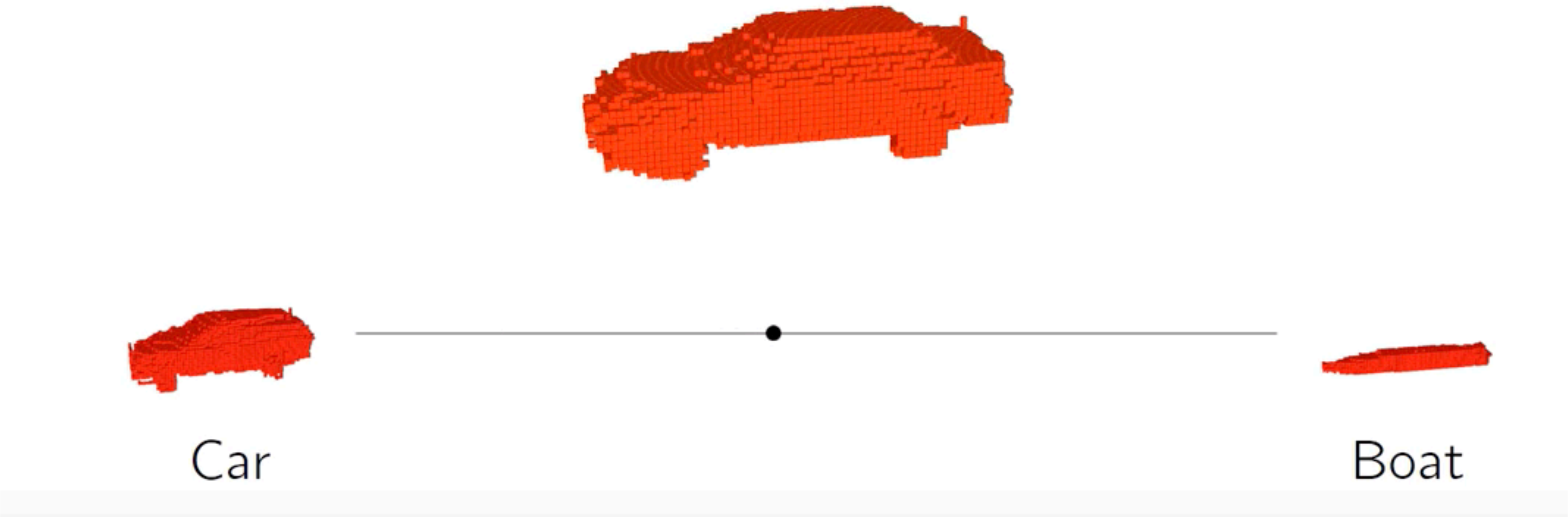




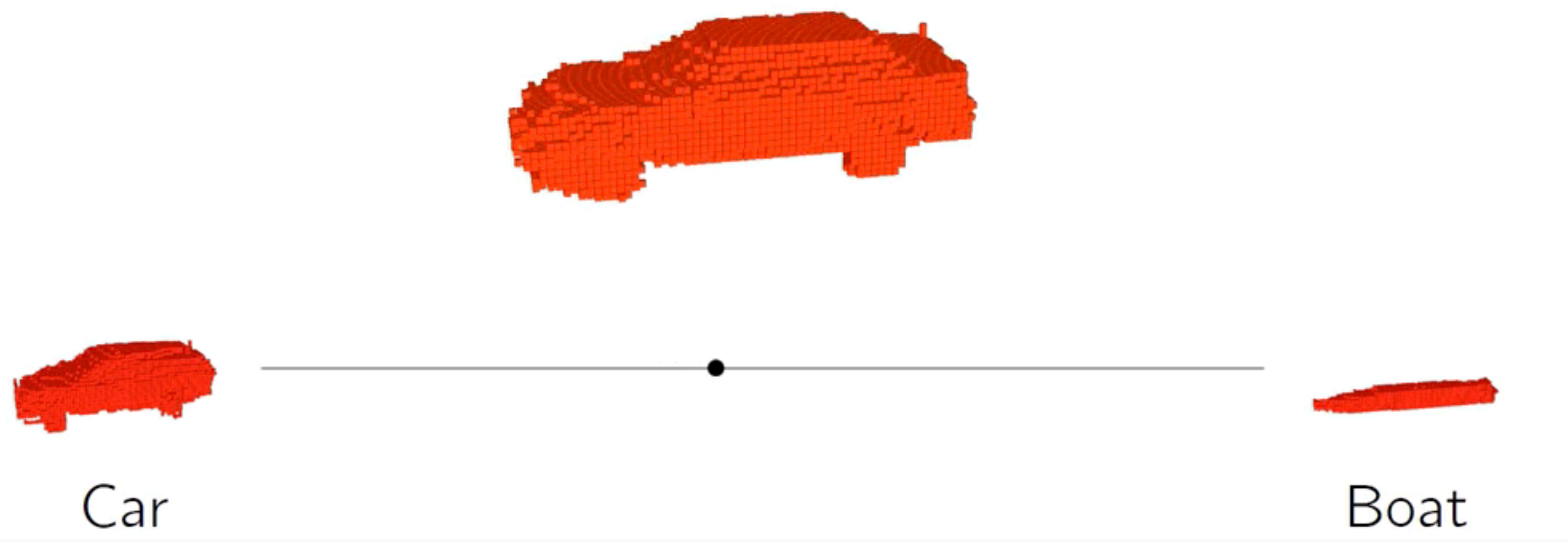


PLAYING WITH LATENT SPACES

- Or 3D objects



- Or 3D objects



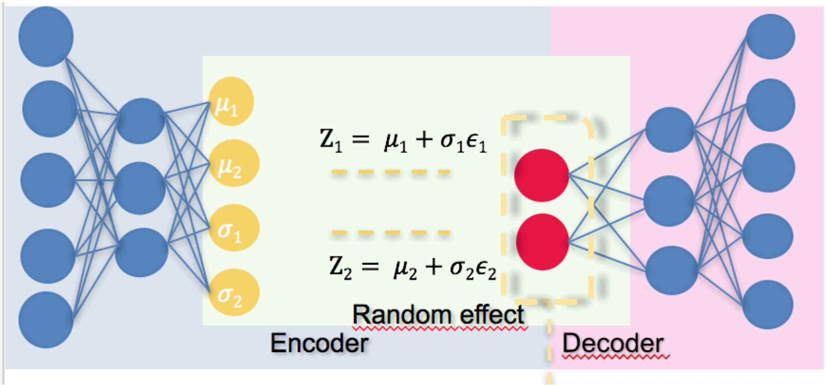
- Latent space = abstract representation of your data
- Encoder maps input to gaussians in latent space
  = Gaussian mixture –> you can do lots of stuff

EVENT GENERATORS USING VAE

- Set up a VAE, train on the events you want to generate

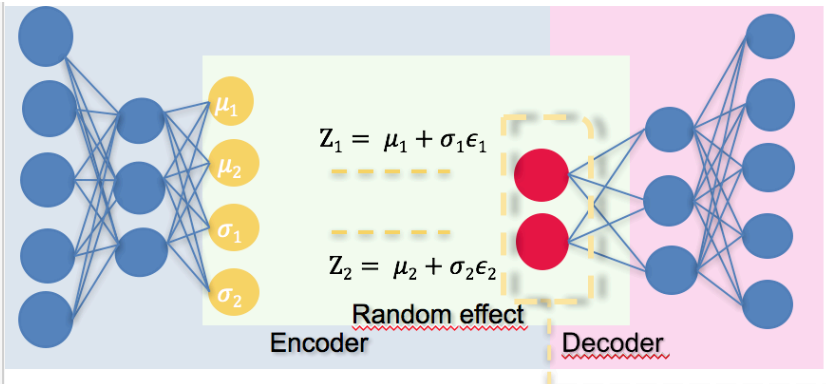
EVENT GENERATORS USING VAE

- Set up a VAE, train on the events you want to generate
- Run representative set through trained encoder to get PDF of the dataset in latent space
  - (=sum of gaussians)



EVENT GENERATORS USING VAE

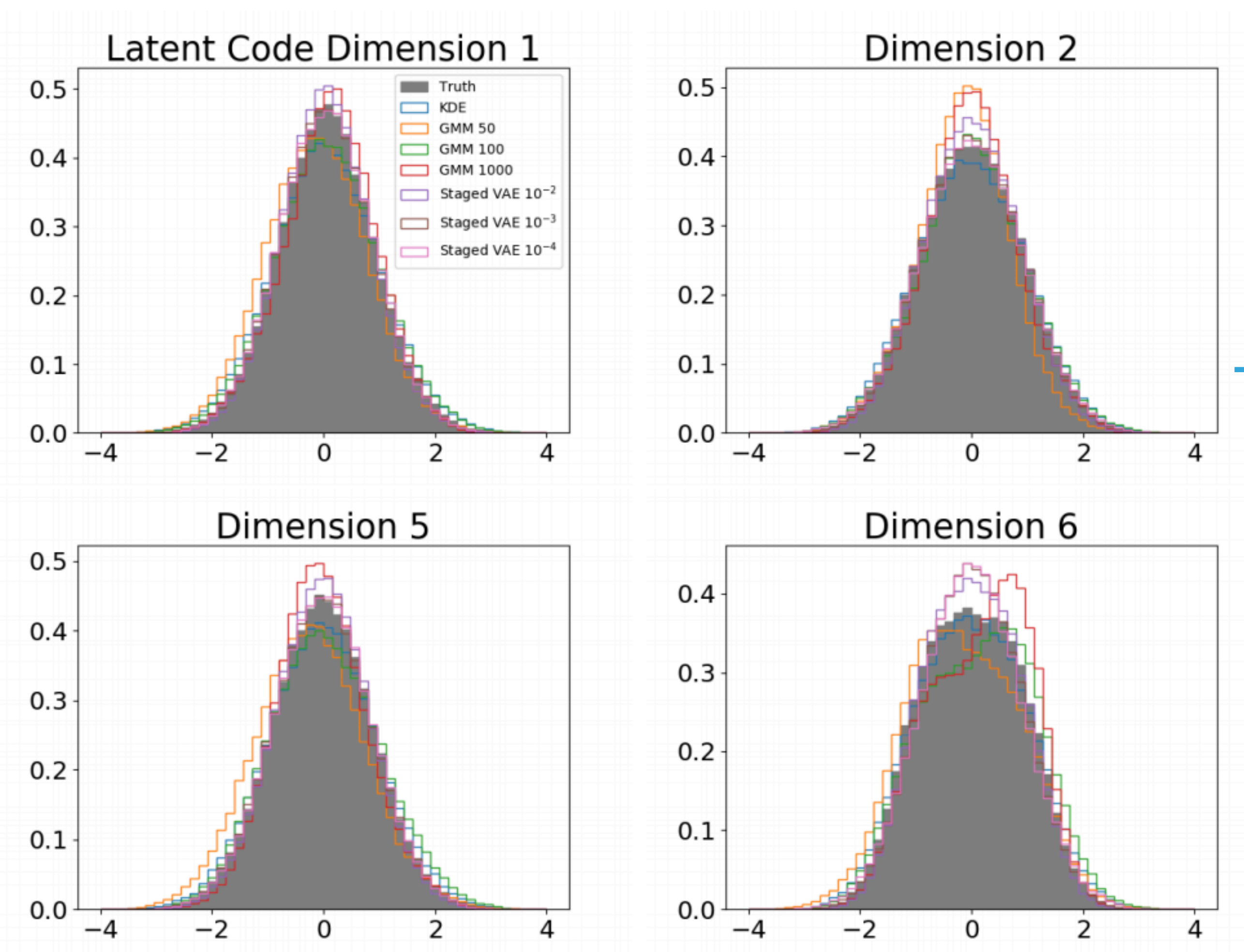
- Set up a VAE, train on the events you want to generate
- Run representative set through trained encoder to get PDF of the dataset in latent space
  - (=sum of gaussians)

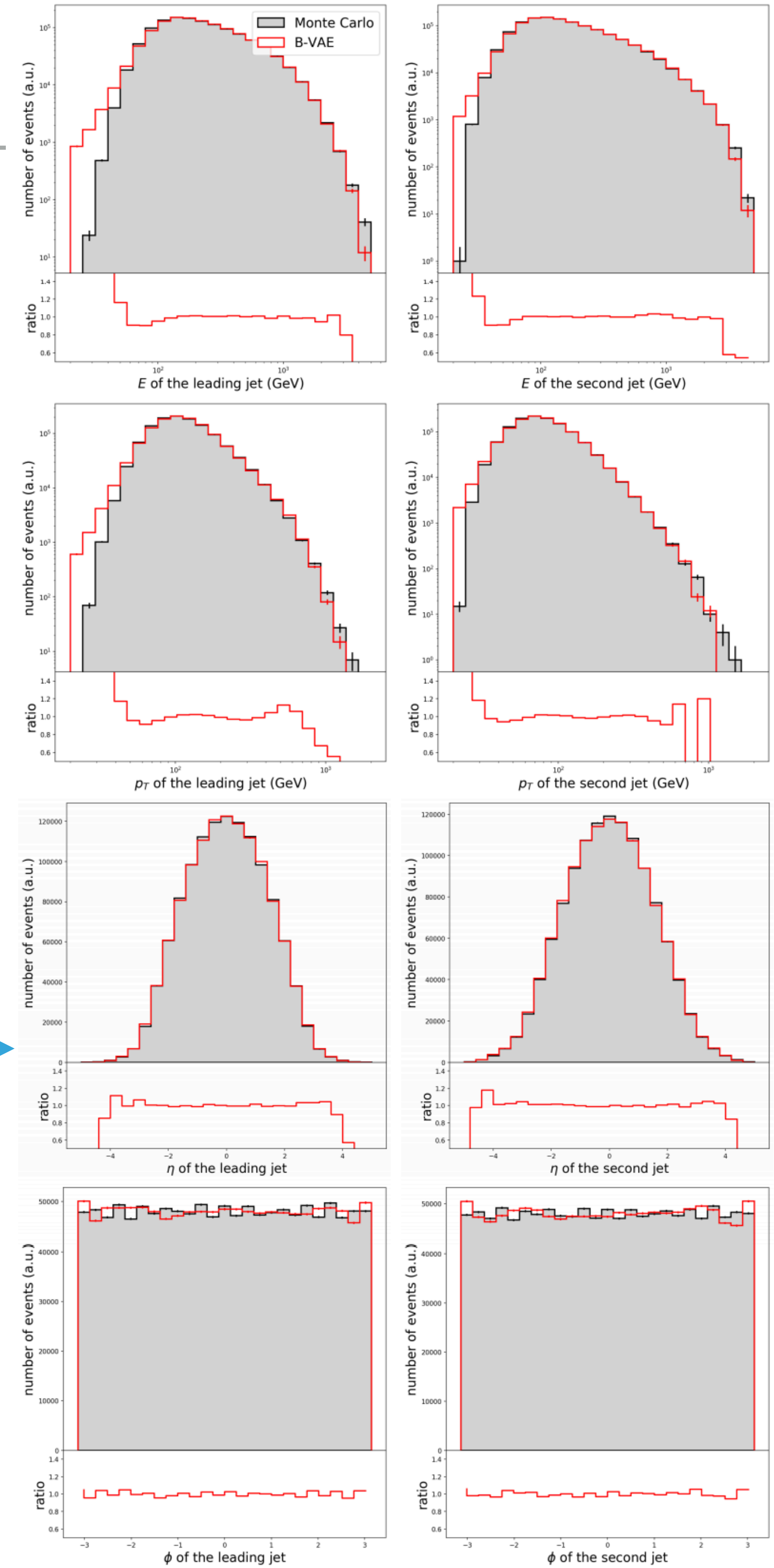


- Sample from the PDF, run through decoder

EVENT GENERATORS USING VAE

- It generates events in 28D, show 8
- Z=20, show 4
- Using B-VAE is orders of magnitude faster (10 million events in 3 minutes)



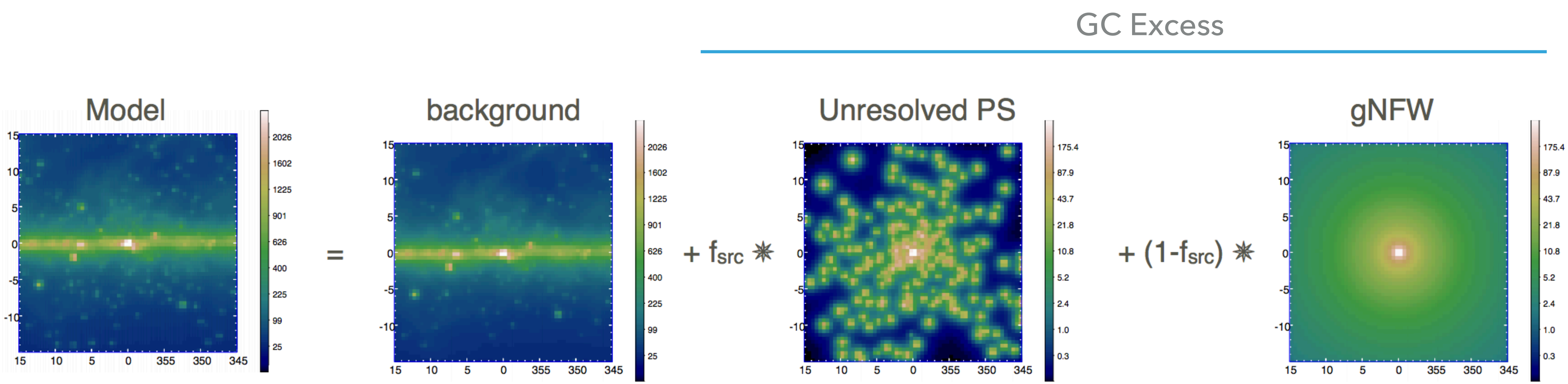


ANALYSE THE GALACTIC CENTER

- Topic: analyse the GC and the possible DM nature of the GCE

ANALYSE THE GALACTIC CENTER

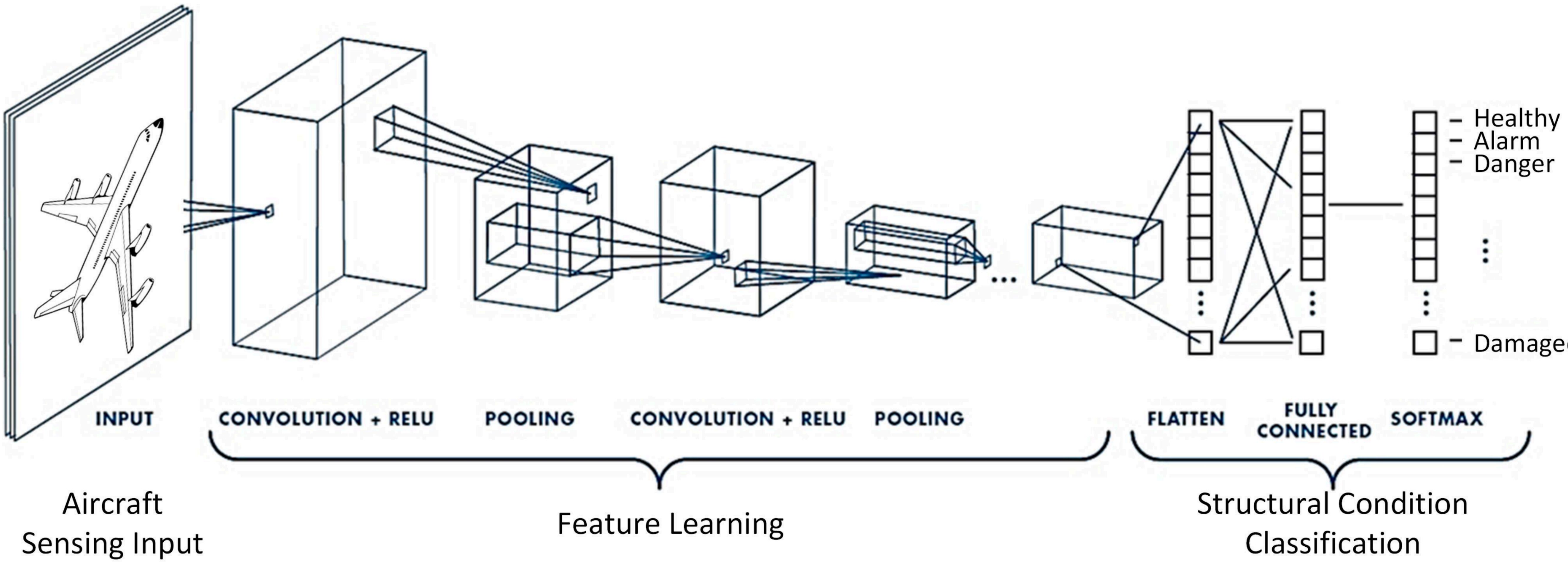
- Topic: analyse the GC and the possible DM nature of the GCE
- V1: what is the fraction of diffuse (dm) and point source (msp) in the GC excess



https://arxiv.org/abs/1708.06706

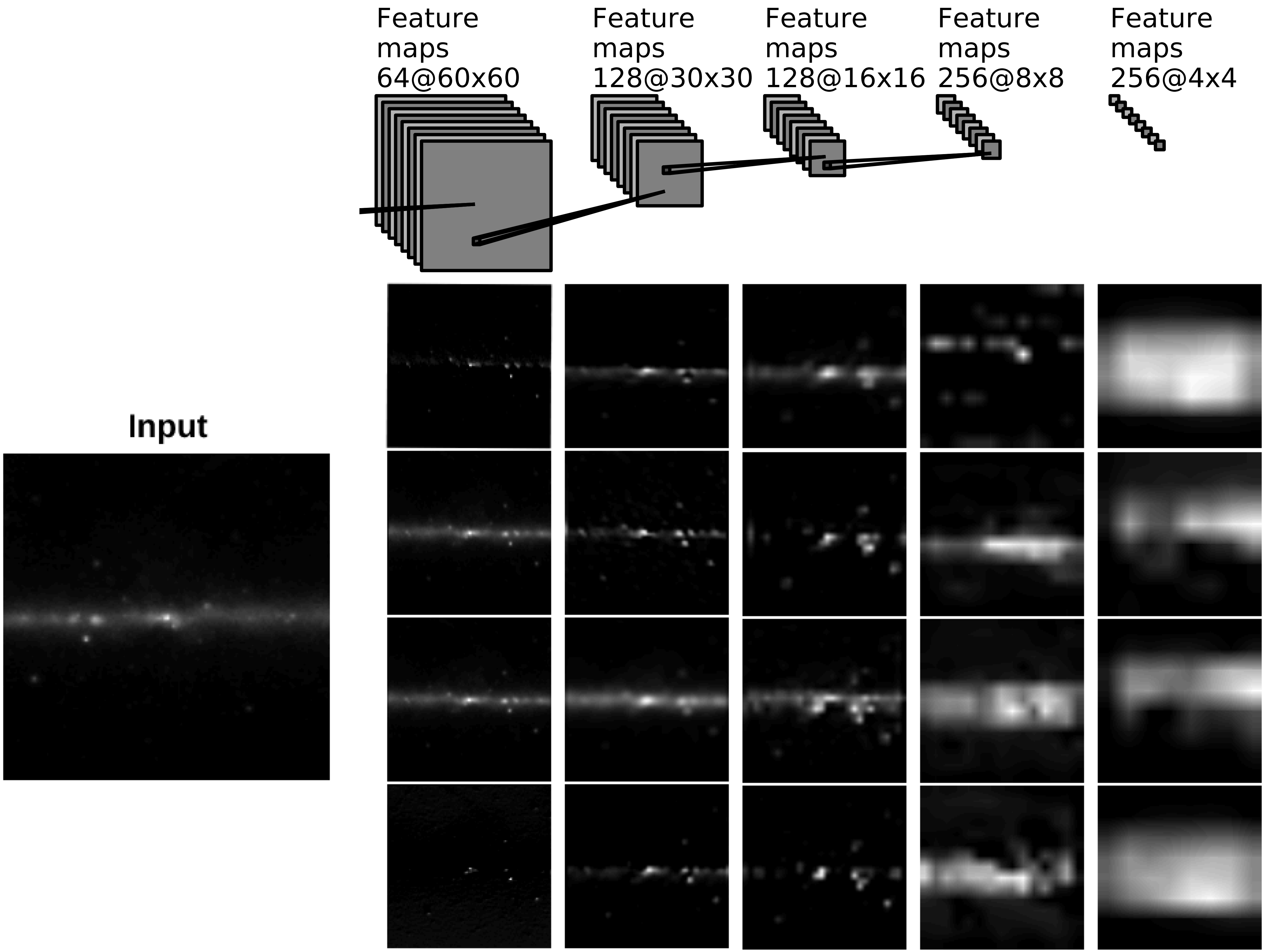
ANALYSE THE GALACTIC CENTER

- Use convolutional neural networks



- Utilise spatial short-range correlations to lower number of trainable weights
- Translation independent

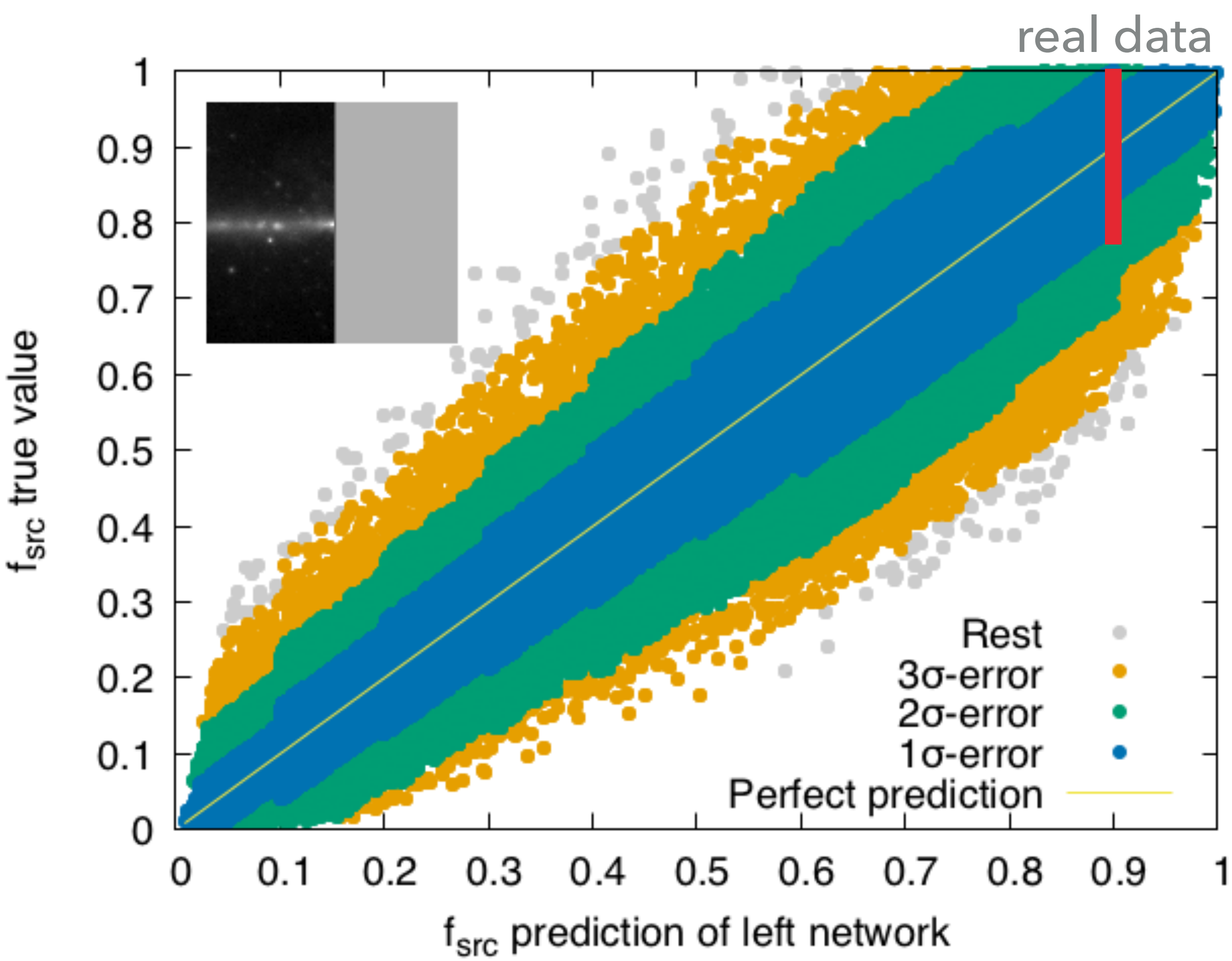
A LOOK INSIDE THE NETWORK

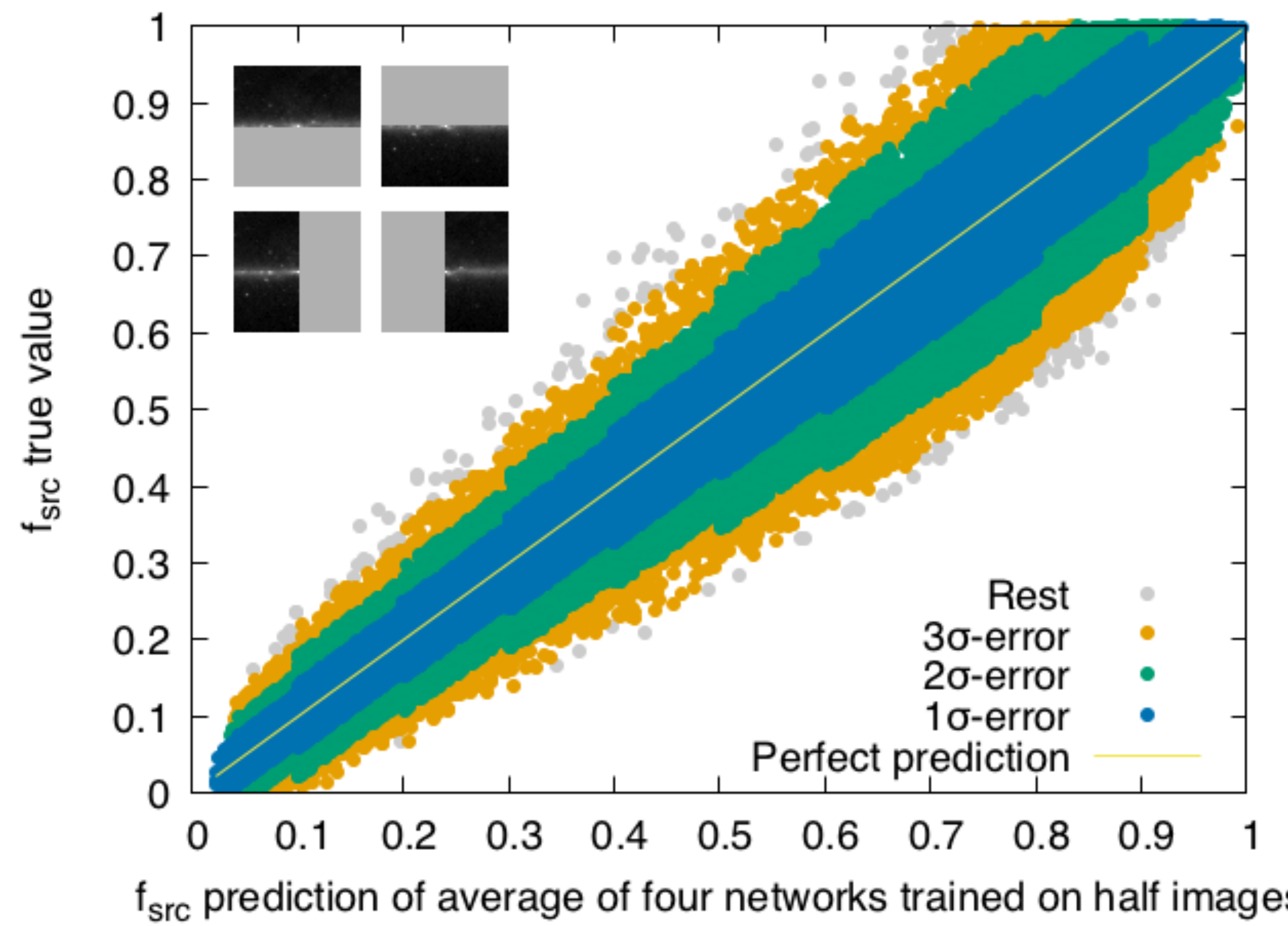


Prediction: 0.86
Truth: 0.82

RESULTS

- Train using 3 background models, test on 2 others
- Test data: 2x30000 test points



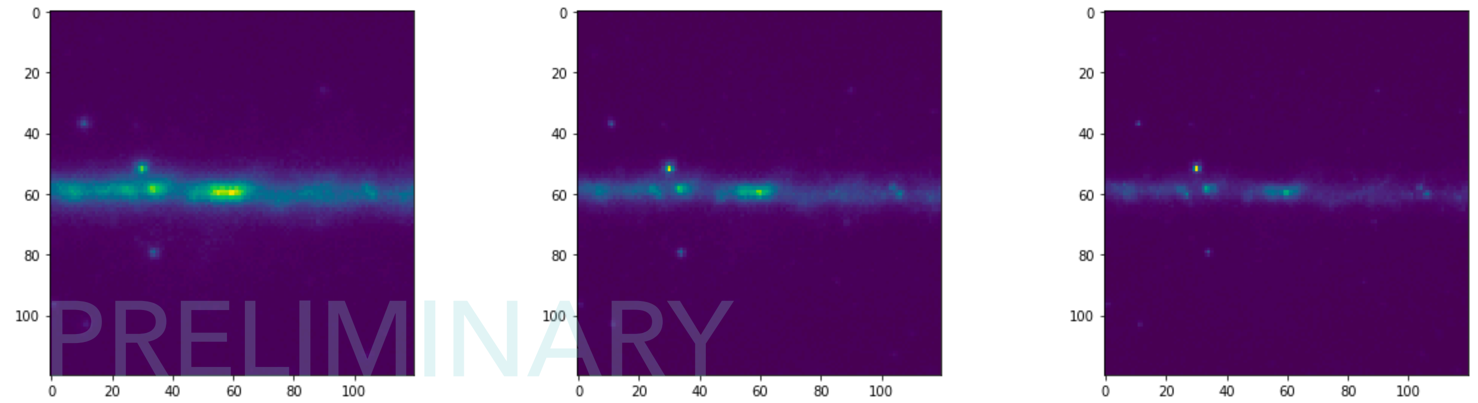


CAVEATS

- Used 5 different background models, but were very similar
- Made some assumption on parameters that are not so sure (eg steepness parameter gamma=0.8 in gFNW profile while real value [0.5,1.2])
- Variance on the test set is nog a good uncertainty measure

NEW APPROACH

- More realistic modelling
    - Use 17-25 parameters that together make up the background and the GC excess, instead of fixed background models
    - Use a range of uncertain parameters (eg gamma)
    - Use 5 energy bins instead of 1
    - Use network that can quantify uncertainty



BAYESIAN DEEP LEARNING

- Use Bayesian neural networks to quantify uncertainties

BAYESIAN DEEP LEARNING

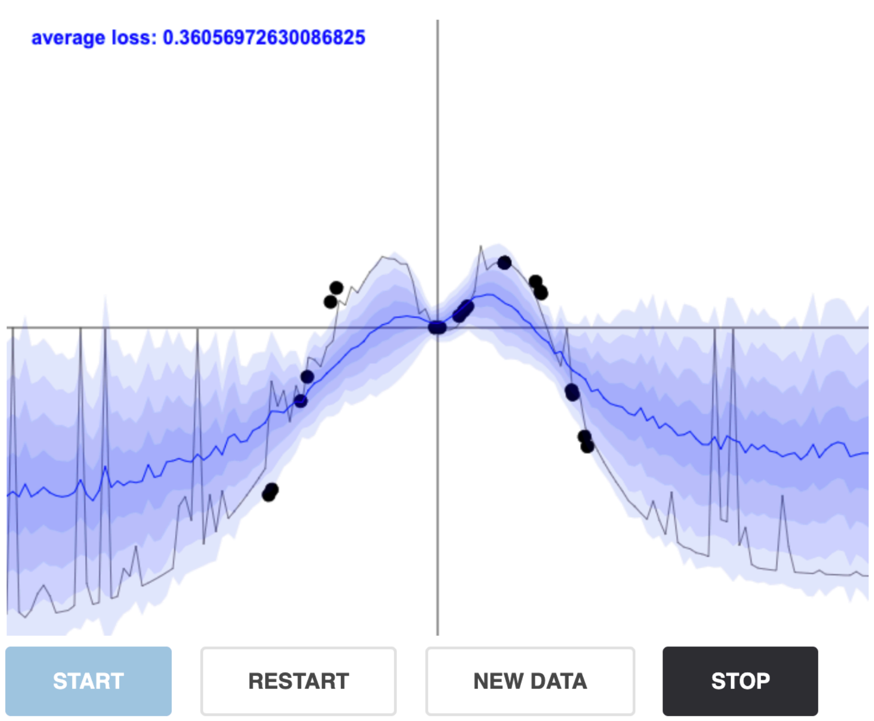
- Use Bayesian neural networks to quantify uncertainties
  - Aleatoric uncertainty: "noise in the data"

$$\mathcal{L}_{\text{NN}}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2\sigma(\mathbf{x}_i)^2}||\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)||^2 + \frac{1}{2}\log\sigma(\mathbf{x}_i)^2$$

- Use Bayesian neural networks to quantify uncertainties
  - Aleatoric uncertainty: “noise in the data”

$$\mathcal{L}_{\text{NN}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2\sigma(\mathbf{x}_i)^2}||\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)||^2 + \frac{1}{2}\log\sigma(\mathbf{x}_i)^2$$

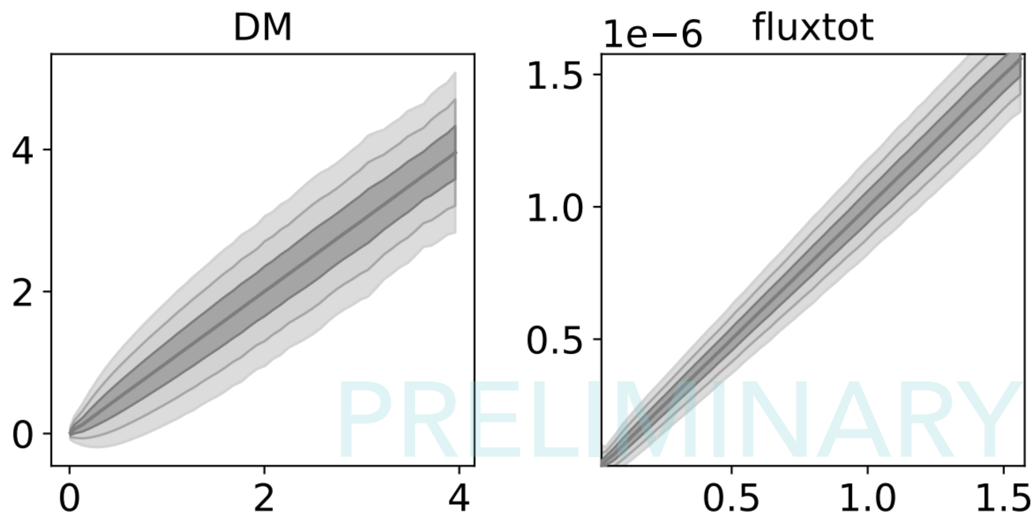
  - Epistemic uncertainty: “NN uncertainty – imperfect training”
    - Monte Carlo dropout



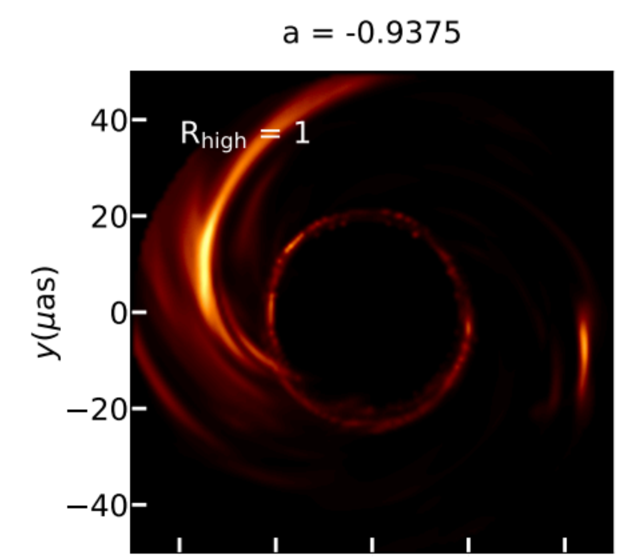
http://mlg.eng.cam.ac.uk/yarin/blog_3d801aa532c1ce.html

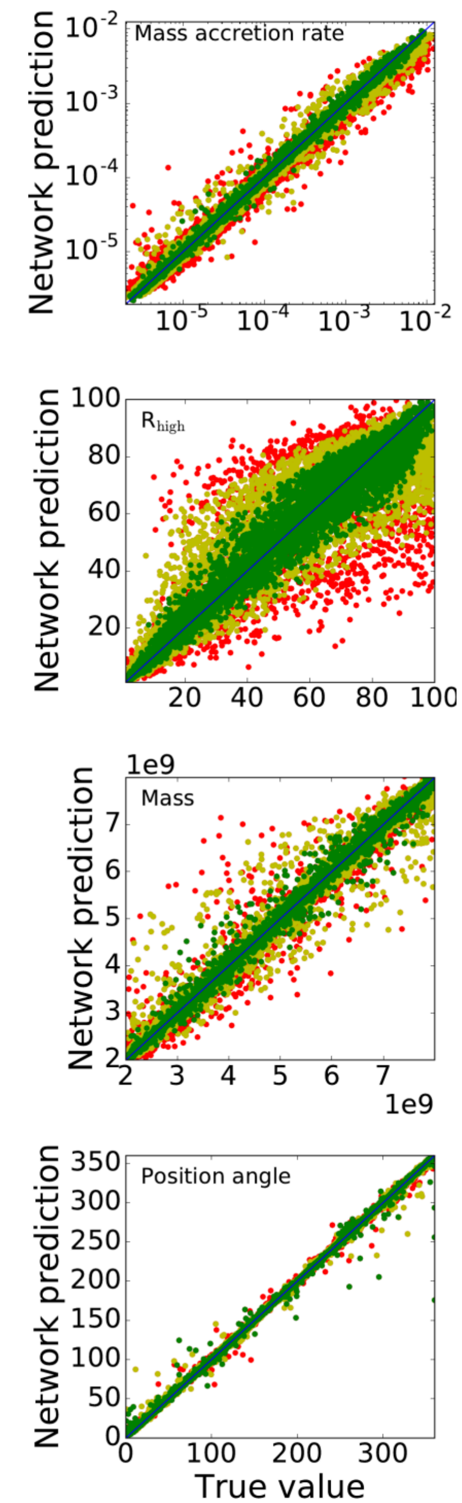
BAYESIAN DEEP LEARNING

Teaser of the result:



We applied the same method on predicting parameters from BH images from EHT simulations





https://arxiv.org/abs/1910.13236

SUMMARY

- DarkMachines is a research collective to tackle important and interesting problems in DM using ML
  - Brings together experts from both fields
- Explained topics
  - High-D parameter optimisation
  - Anomaly detection using AE/VAE
  - Event generation using B-VAE
  - Parameter inference using Bayesian CNNs
- There are more active challenges:
  - Gravitational lensing, gamma-ray point source detection, …

JOIN?

- Many different DM applications (HEP, astro, detectors, theoretical)
- Many different ML approaches (regression, classification, generative modelling, outlier detection, …)
- For ML everything is just data – learn from each other and from other fields!
- **Interested? Join: darkmachines.org**
  Challenges can be joined via CERN mailing lists
  or contacting the challenge coordinators