Representation Learning of Collider Events

Jack Collins

SLAC NATIONAL ACCELERATOR LABORATORY

IAS Program on High Energy Physics 2021

How Much Information is in a Jet?

Kaustuv Datta and Andrew Larkoski

Physics Department, Reed College, Portland, OR 97202, USA

How Much Information is in a Jet?

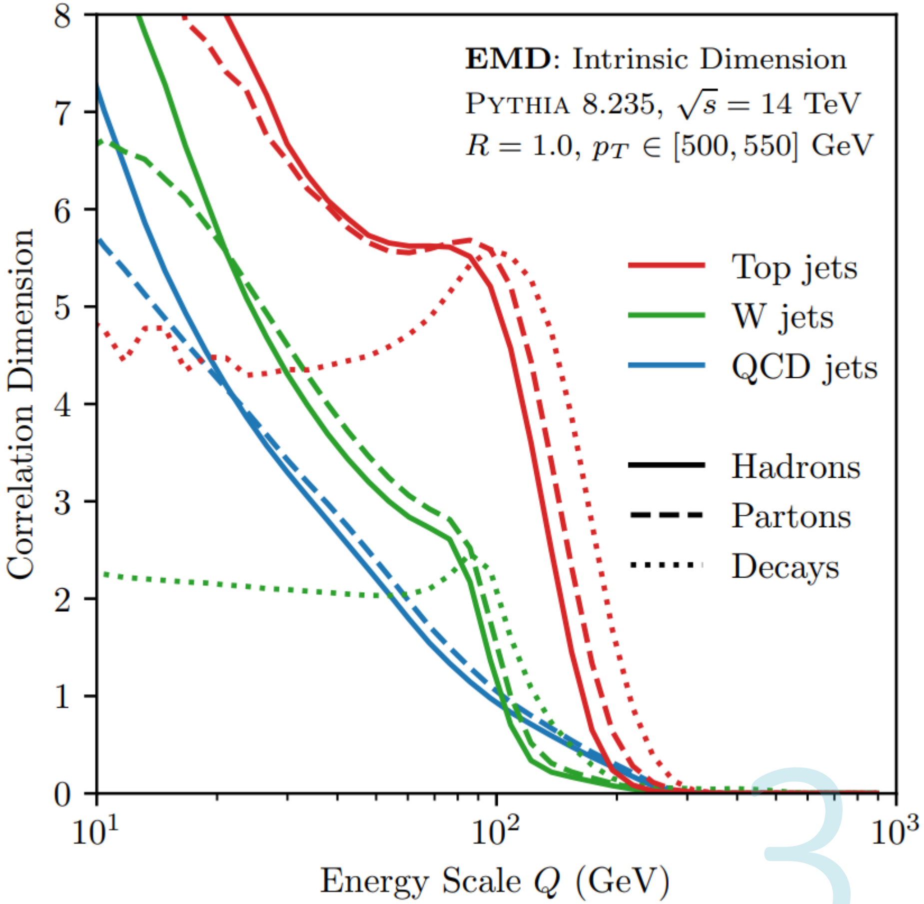
Kaustuv Datta and Andrew Larkoski

Physics Department, Reed College, Portland, OR 97202, USA

The Metric Space of Collider Events

Patrick T. Komiske,* Eric M. Metodiev,† and Jesse Thaler‡

Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and Department of Physics, Harvard University, Cambridge, MA 02138, USA



Conclusions

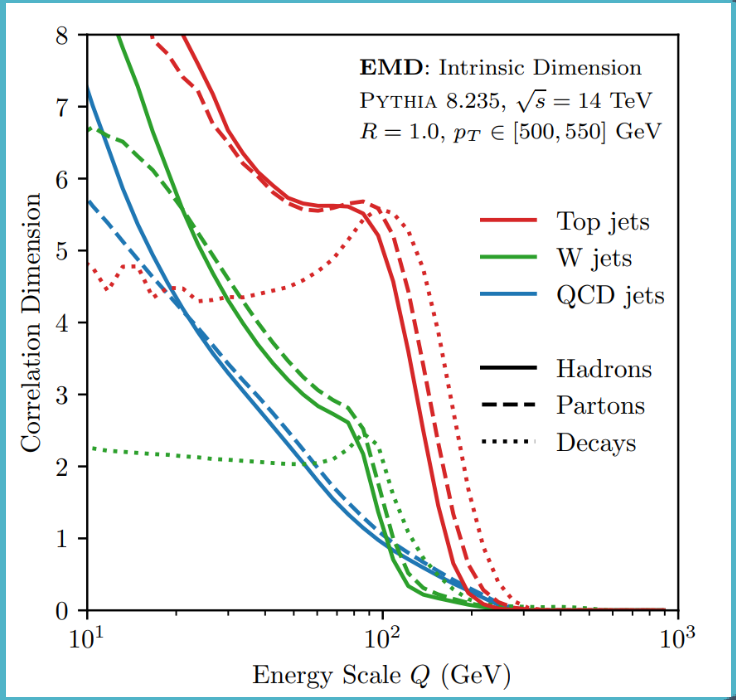
I have been training Variational Autoencoders to reconstruct jets or collider events using Earth Movers Distance as the reconstruction metric.

The learnt representation:

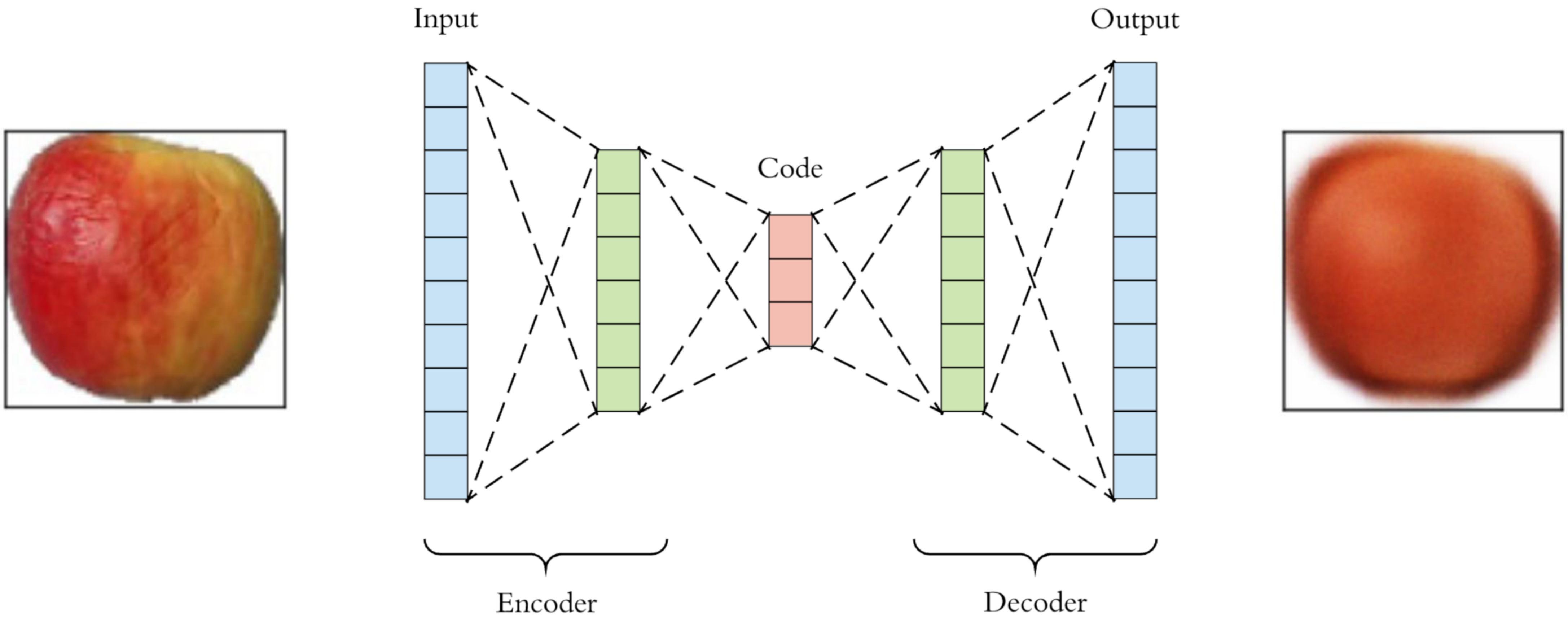
- Is scale dependent
- Is orthogonalized
- Is hierarchically organized by scale
- Has fractal dimension which relates to that of the data manifold

This is because:

- The VAE is trained to be parsimonious with information
- The metric space is physically meaningful and structured

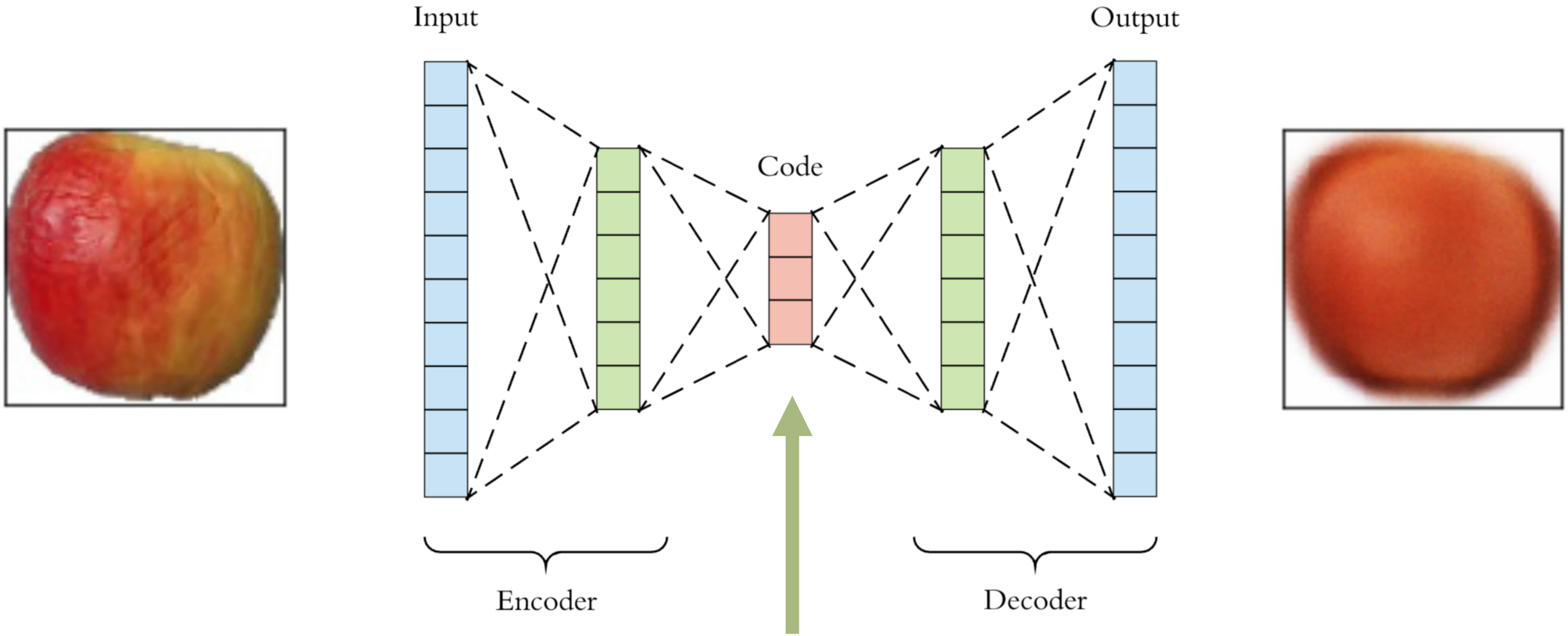


The Plain Autoencoder



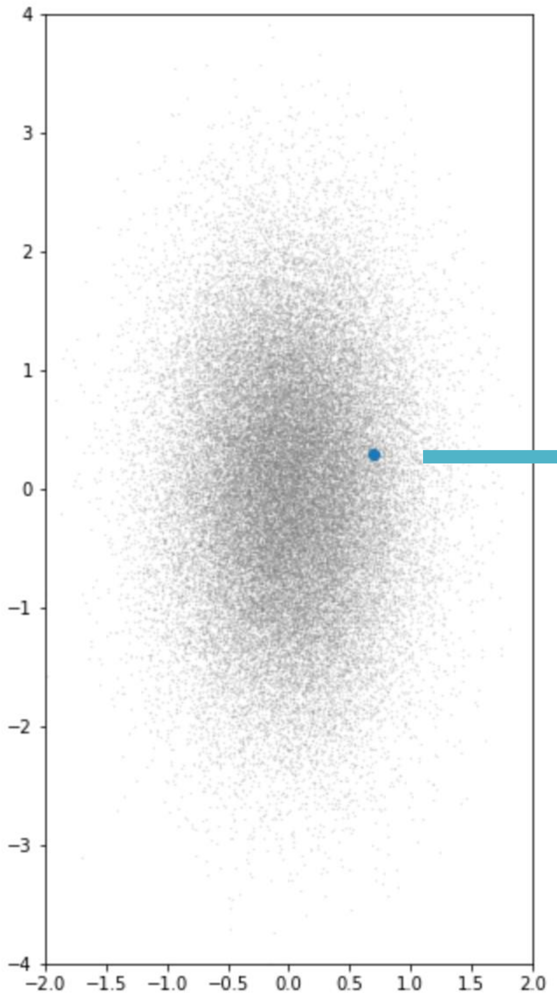
Loss = |Output – Input| *(what is this for jets?)*

The Plain Autoencoder



Latent space =?= Learnt representation

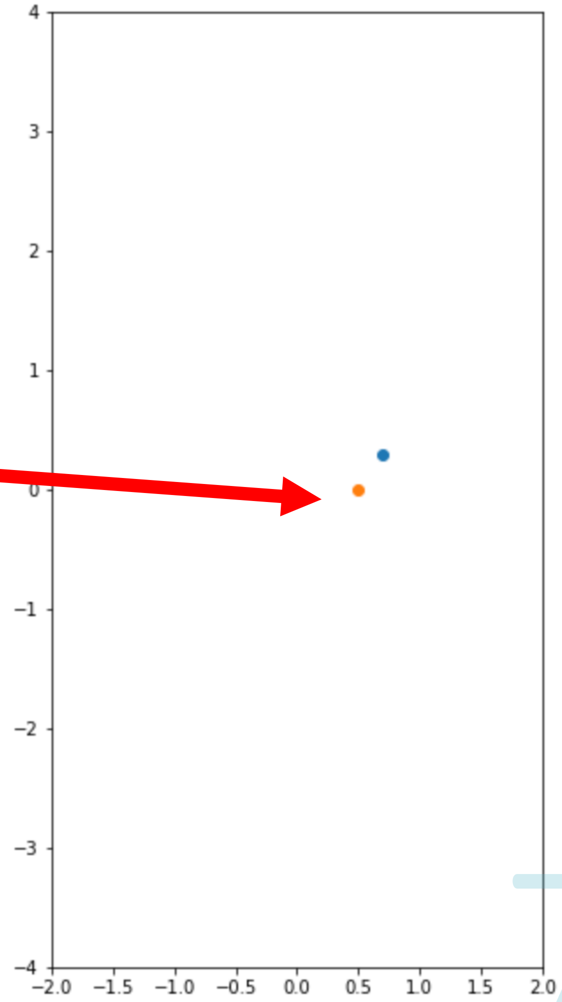
The Plain Autoencoder: *a toy example*



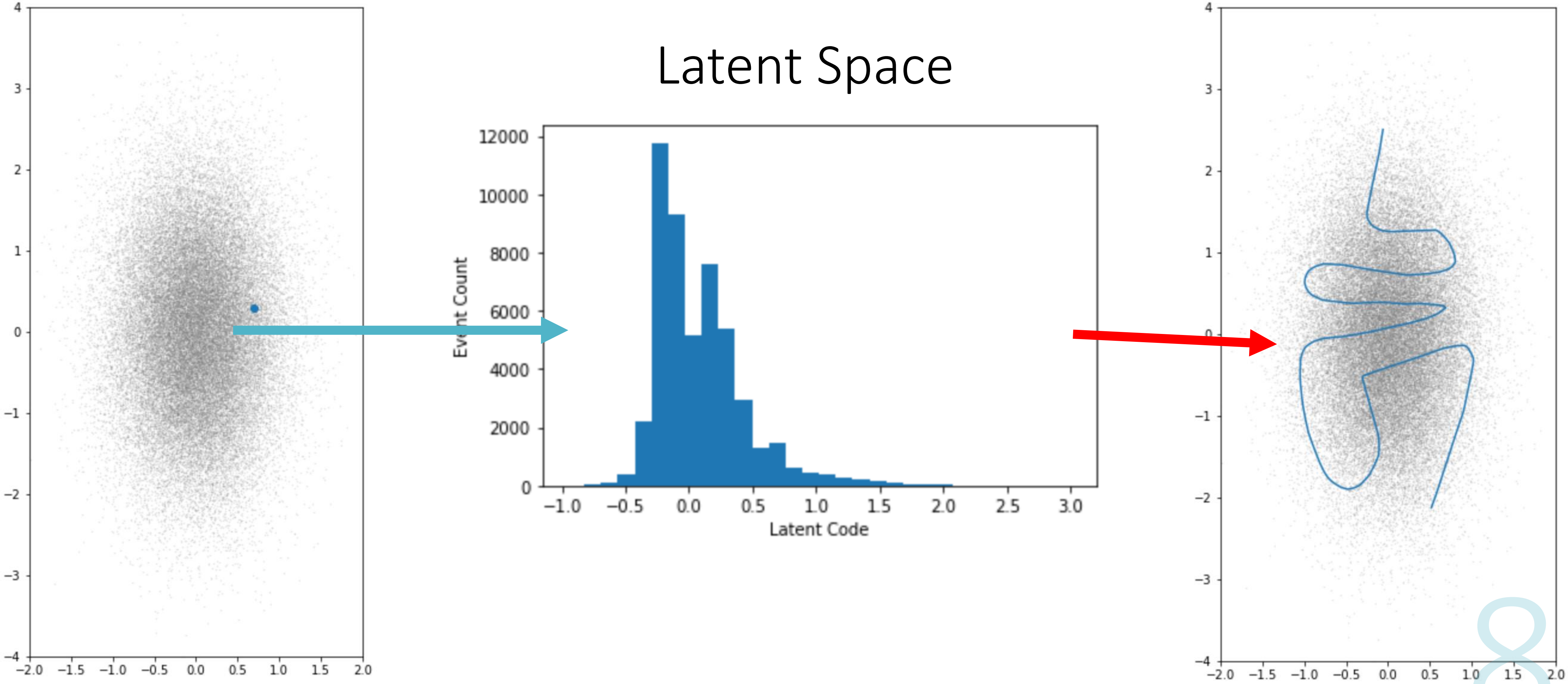
AE

(1D latent space)

Rec. Loss = $|\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^{(2)}$



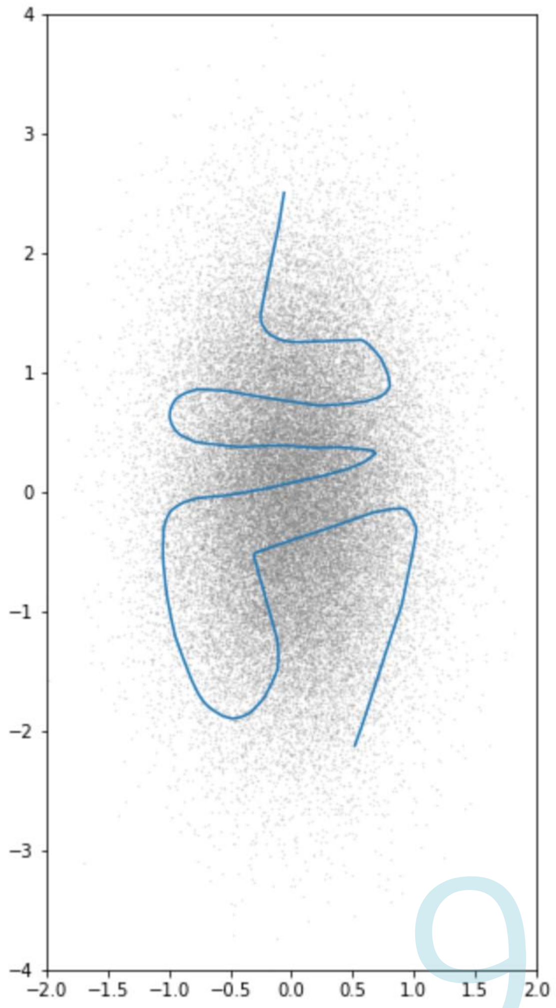
The Plain Autoencoder: *a toy example*



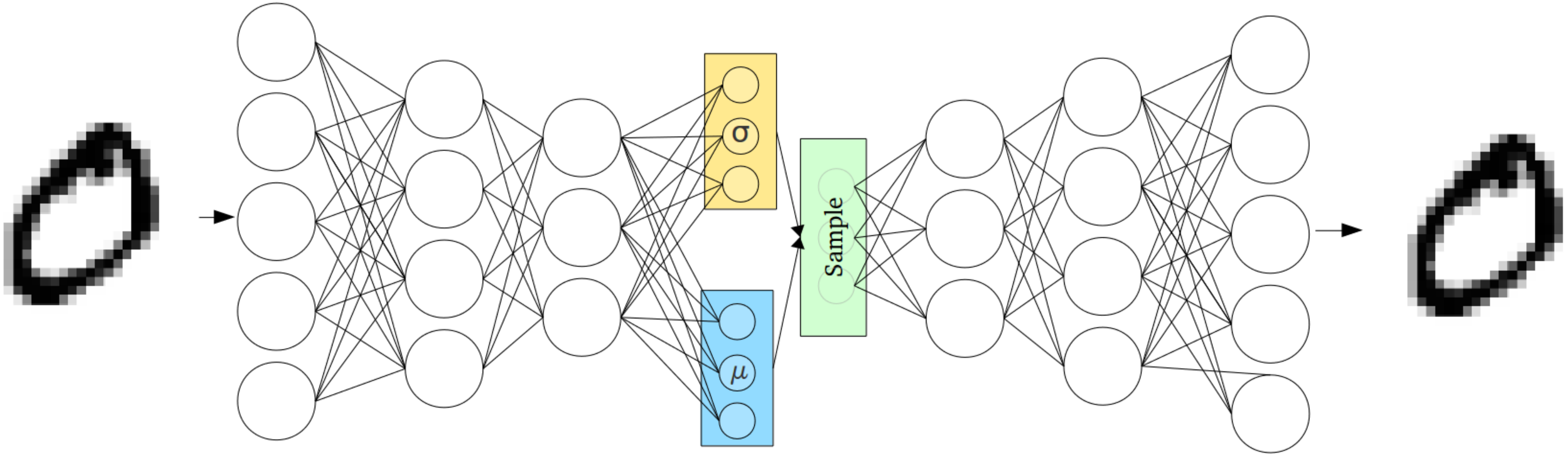
The Plain Autoencoder: *a toy example*

1. The AE learns some dense packing of the data space

2. The latent representation is highly coupled with the expressiveness of the network architecture of the encoder and decoder



The Variational Autoencoder



$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2/2\beta^2 - \sum_i \frac{1}{2}\left(1 + \log\sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

Reconstruction error

KL(q(z|x)||p(z)) ~ "Information cost"

The Variational Autoencoder:

Information and the loss function

$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2 / 2\beta^2 - \sum_i \frac{1}{2}\left(1 + \log \sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2 - 2\beta^2 \sum_i \frac{1}{2}\left(1 + \log \sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

The Variational Autoencoder:

Information and the loss function

$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2/2\beta^2 - \sum_i \frac{1}{2}\left(1 + \log\sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

1) $\boldsymbol{\beta}$ is dimensionful!

The same dimension as the distance metric,
e.g. GeV.

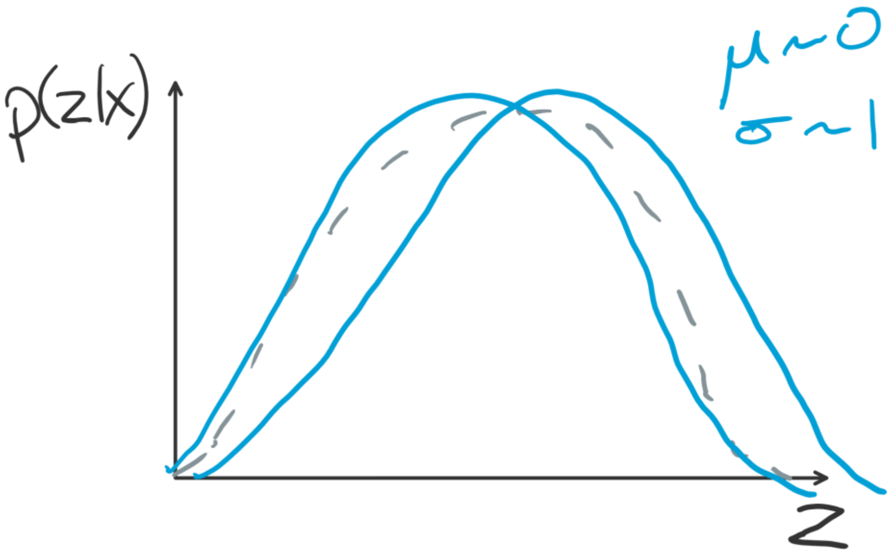
$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2 - 2\beta^2 \sum_i \frac{1}{2}\left(1 + \log\sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

The Variational Autoencoder:

Information and the loss function

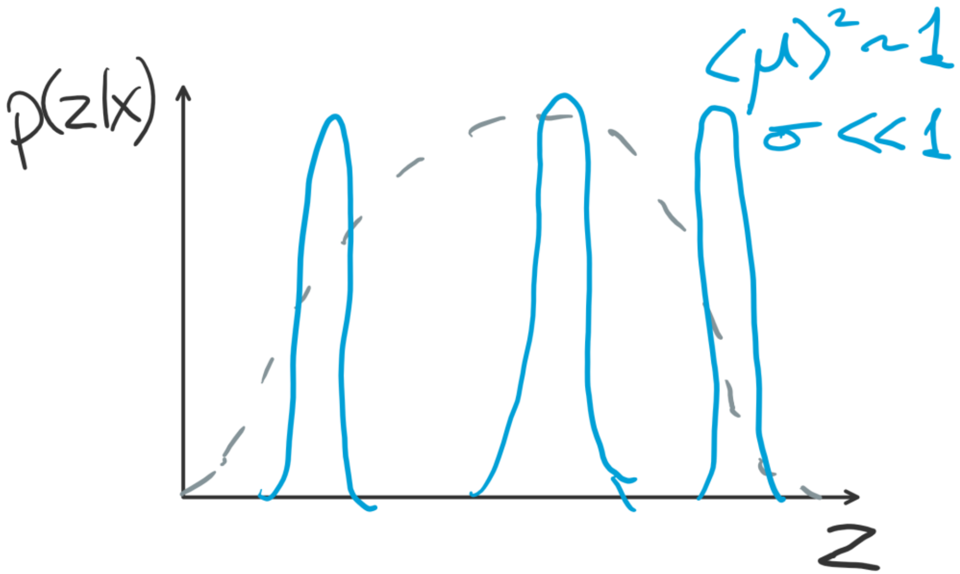
$\beta \rightarrow \infty$

No info encoded in latent space



$\beta \ll$ *Lengthscale*

Info encoded in latent space



$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2 - 2\beta^2 \sum_i \frac{1}{2}\left(1 + \log \sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

The Variational Autoencoder:

Information and the loss function

$\boldsymbol{\beta} \rightarrow \infty$

No info encoded in latent space

$\boldsymbol{\beta} \ll$ ***Lengthscale***

Info encoded in latent space

2) $\boldsymbol{\beta}$ is the cost for encoding information

The encoder will only encode information about the input to the extent that its usefulness for reconstruction is sufficient to justify the cost.

$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2 - 2\beta^2 \sum_i \frac{1}{2}\left(1 + \log\sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

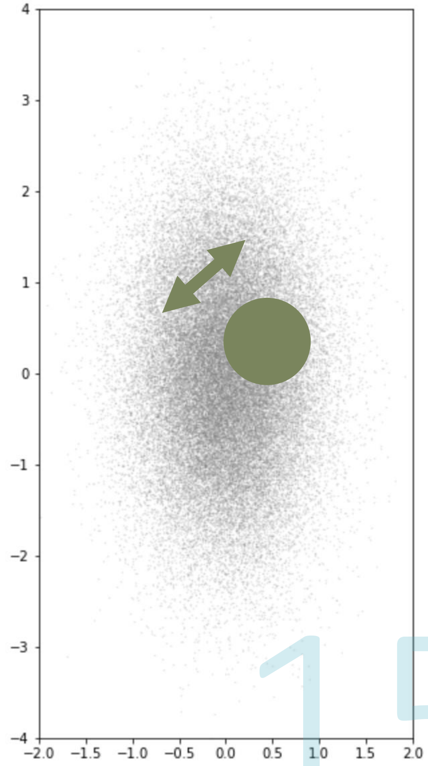
The Variational Autoencoder:

Information and the loss function

$$\text{Loss} = |\boldsymbol{x}_{out} - \boldsymbol{x}_{in}|^2/2\beta^2 - \sum_i \frac{1}{2}\left(1 + \log \sigma_i^2 - \mu_i^2 - \sigma_i^2\right)$$

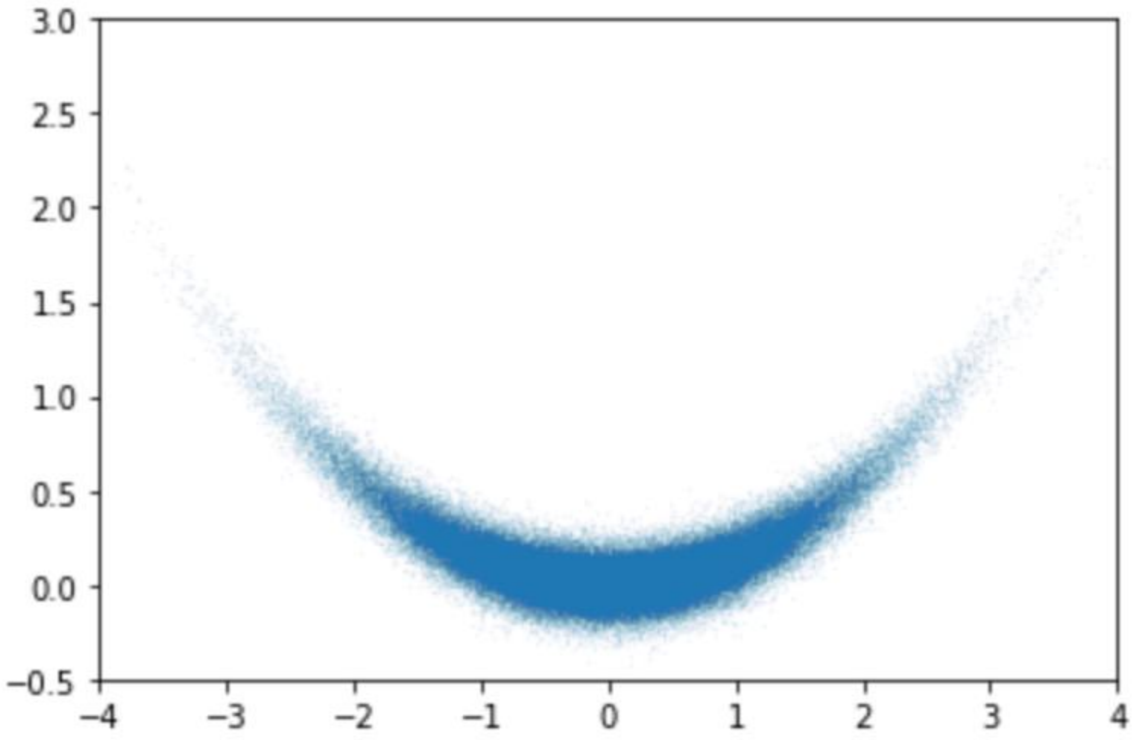
3) $\boldsymbol{\beta}$ is the distance resolution in reconstruction space

The stochasticity of the latent sampling will smear the reconstruction at scale $\sim \beta$



The Variational Autoencoder:

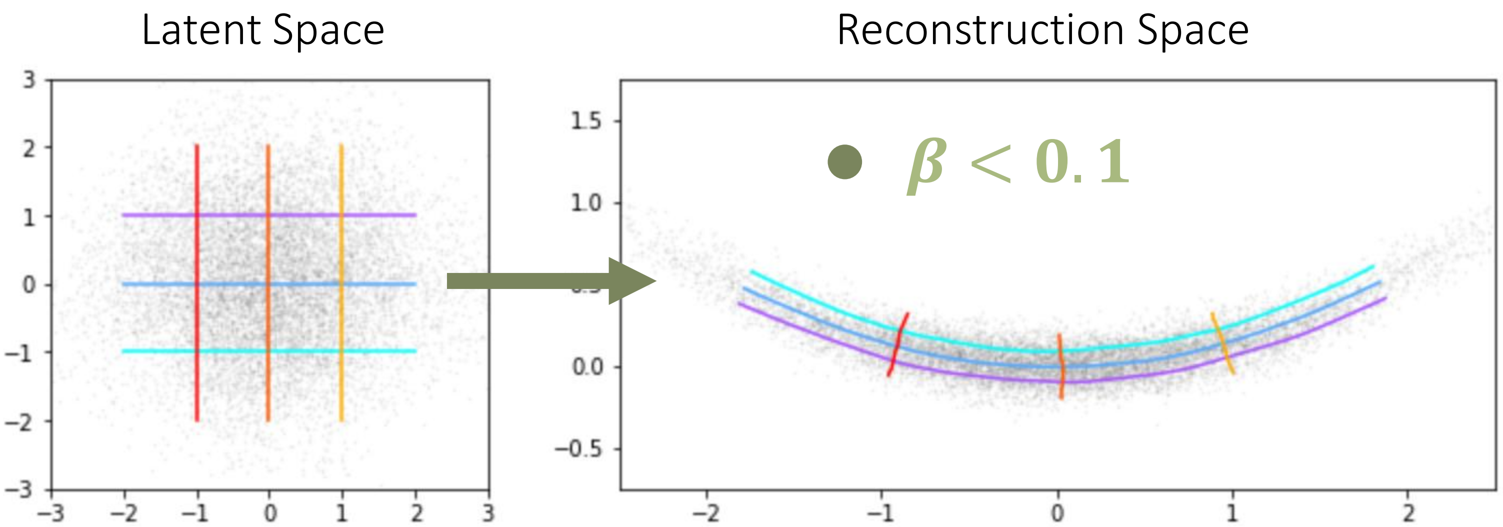
Bananas

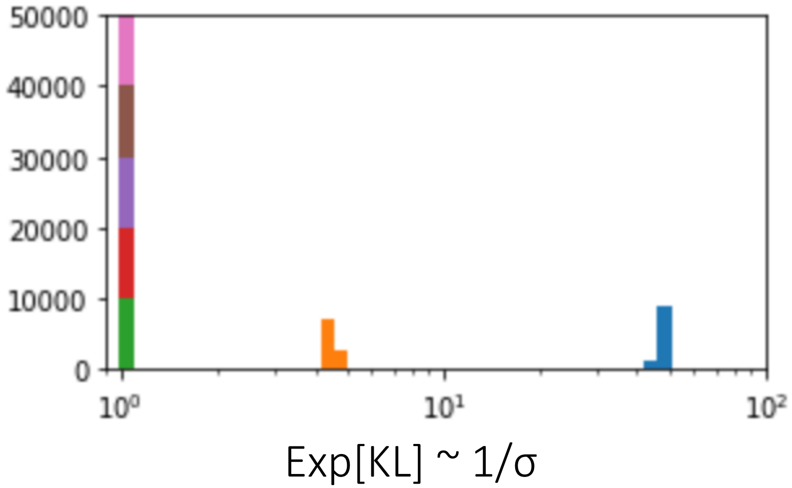


Train VAE with a 10 dim latent space on this 2D dataset

The Variational Autoencoder:

Bananas

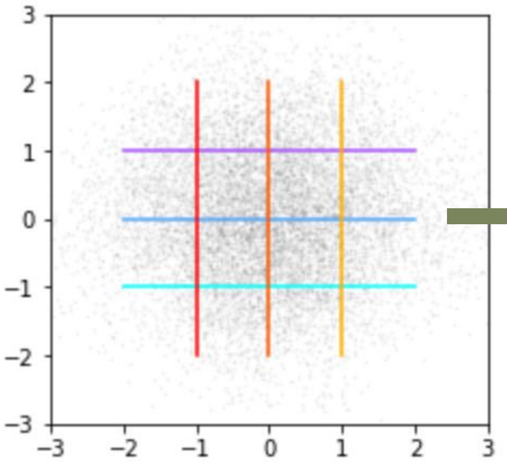




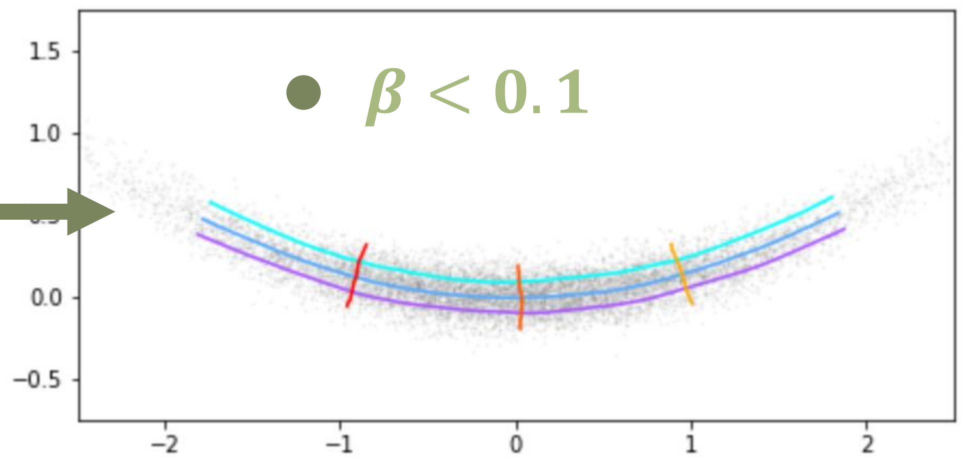
The Variational Autoencoder:

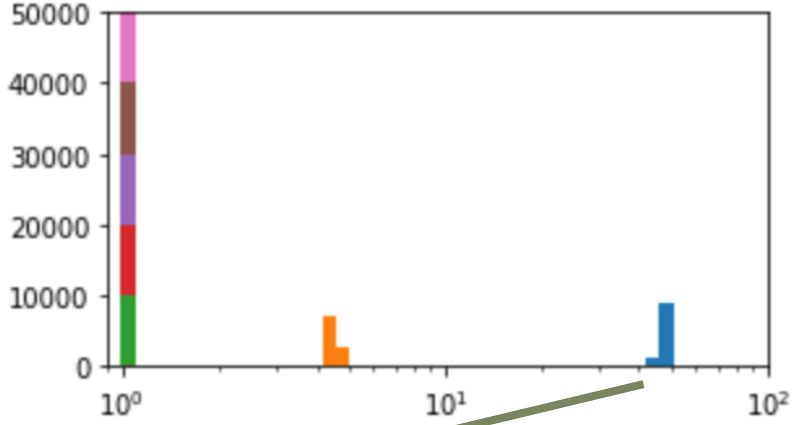
Bananas

Latent Space



Reconstruction Space



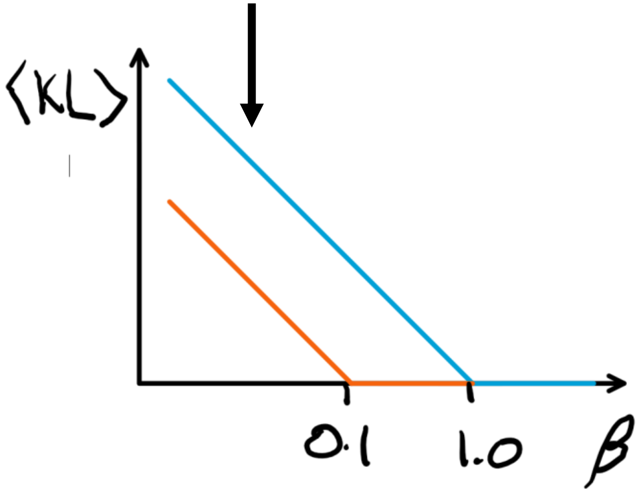


The VAE is doing non-linear PCA

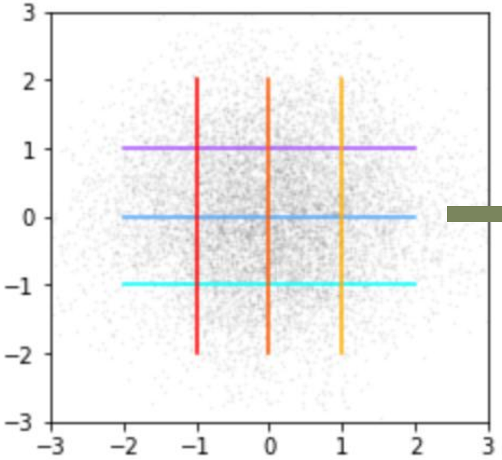
Size = $^{\beta}/_{\sigma}$

The Variational Autoencoder:

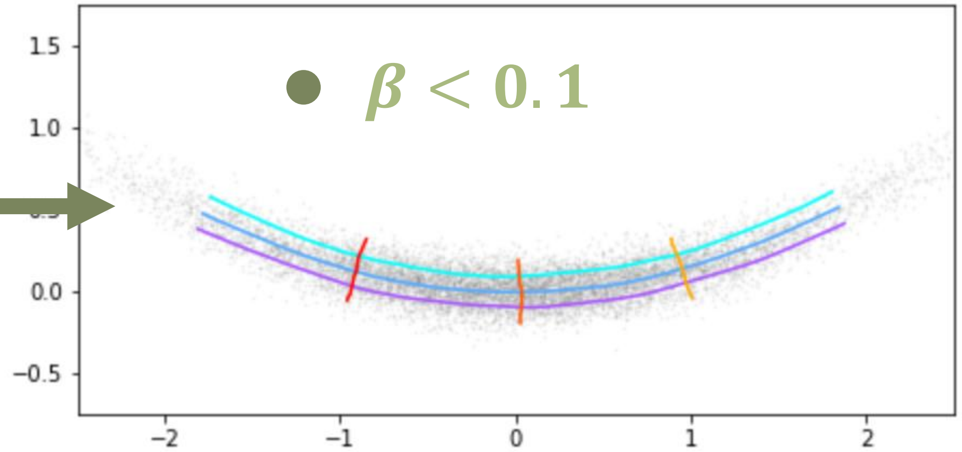
Bananas

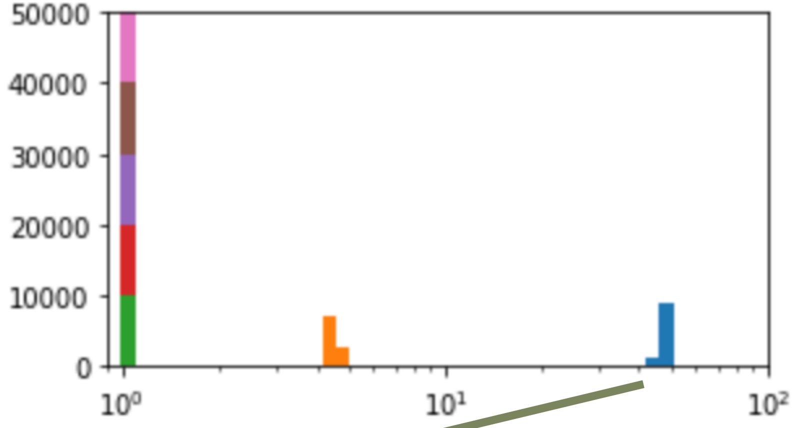


Latent Space



Reconstruction Space





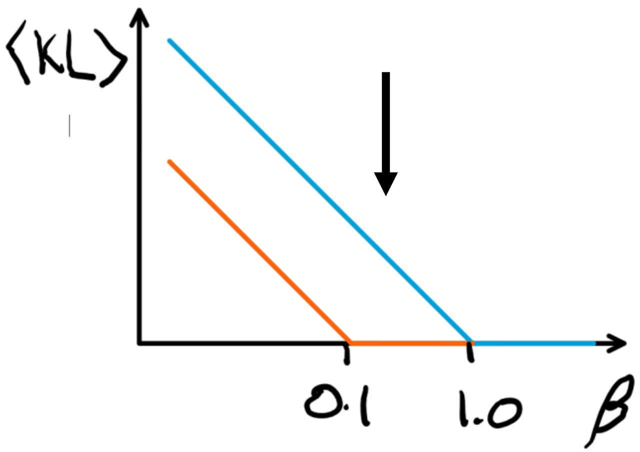
Exp[KL] ~ 1/σ

The VAE is doing non-linear PCA

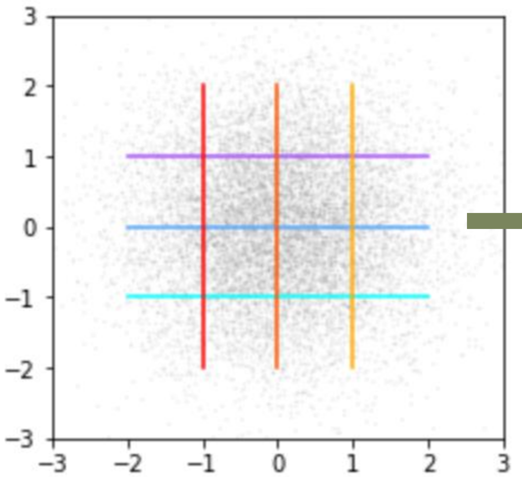
Size = $^{\beta}/_{\sigma}$

The Variational Autoencoder:

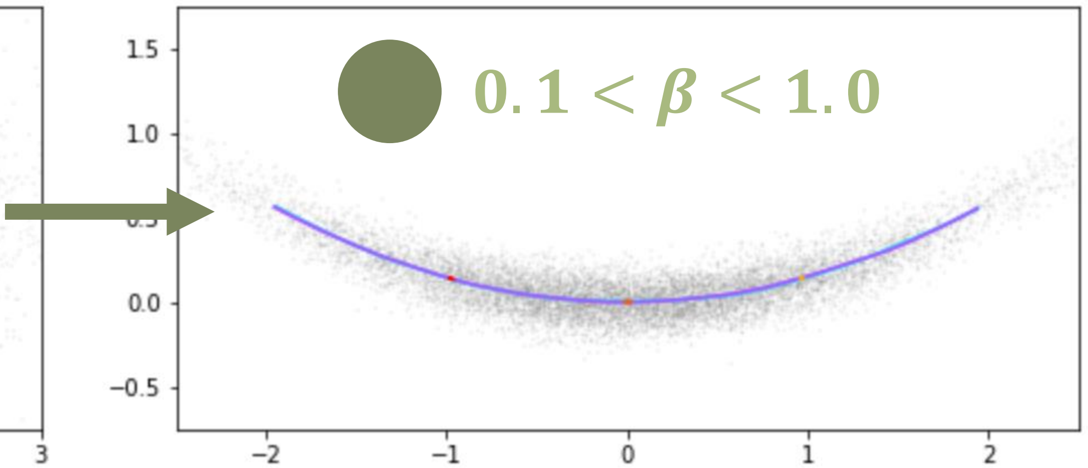
Bananas

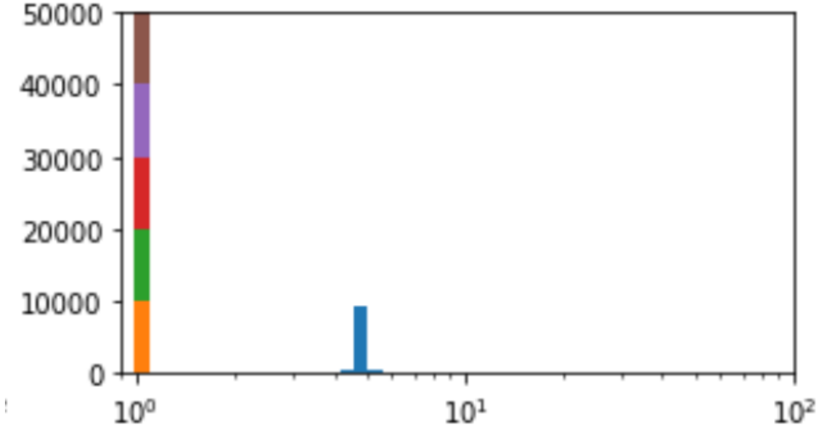


Latent Space



Reconstruction Space

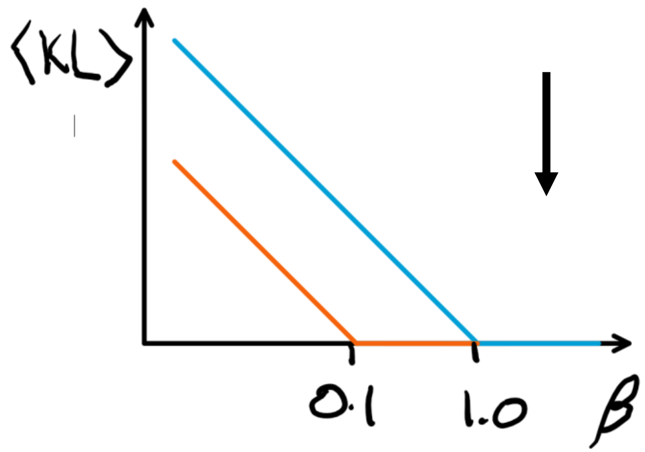




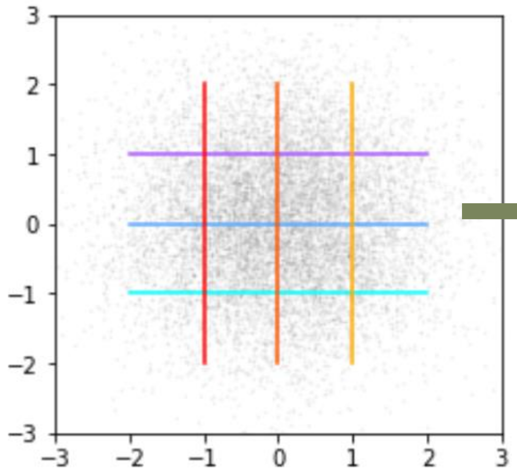
Exp[KL] ~ 1/σ

The Variational Autoencoder:

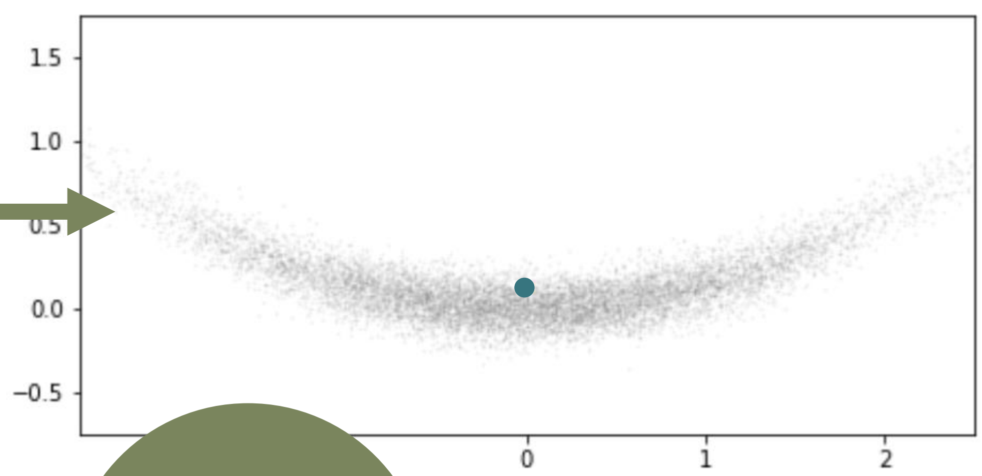
Bananas

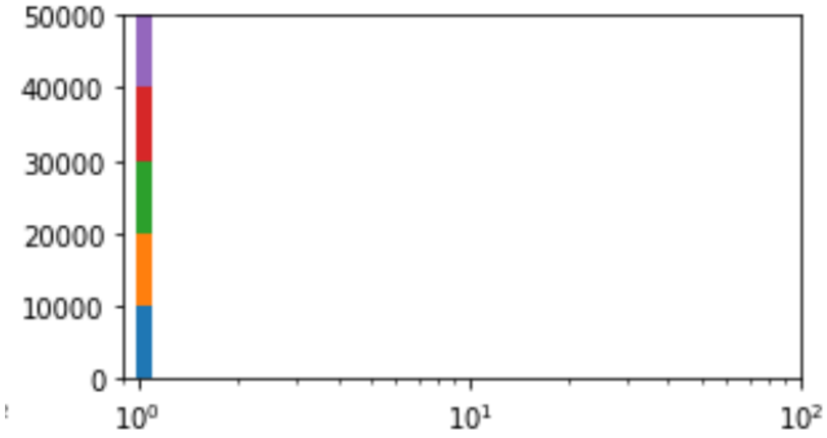


Latent Space



Reconstruction Space



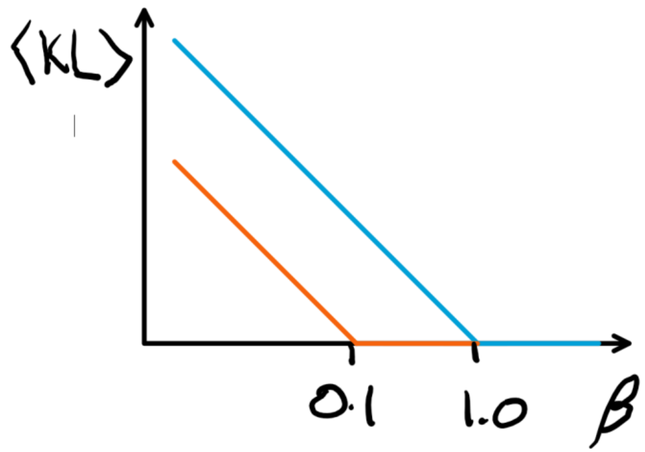


Exp[KL] ~ 1/σ

$\beta \gg 1$

The Variational Autoencoder:

Dimensionality

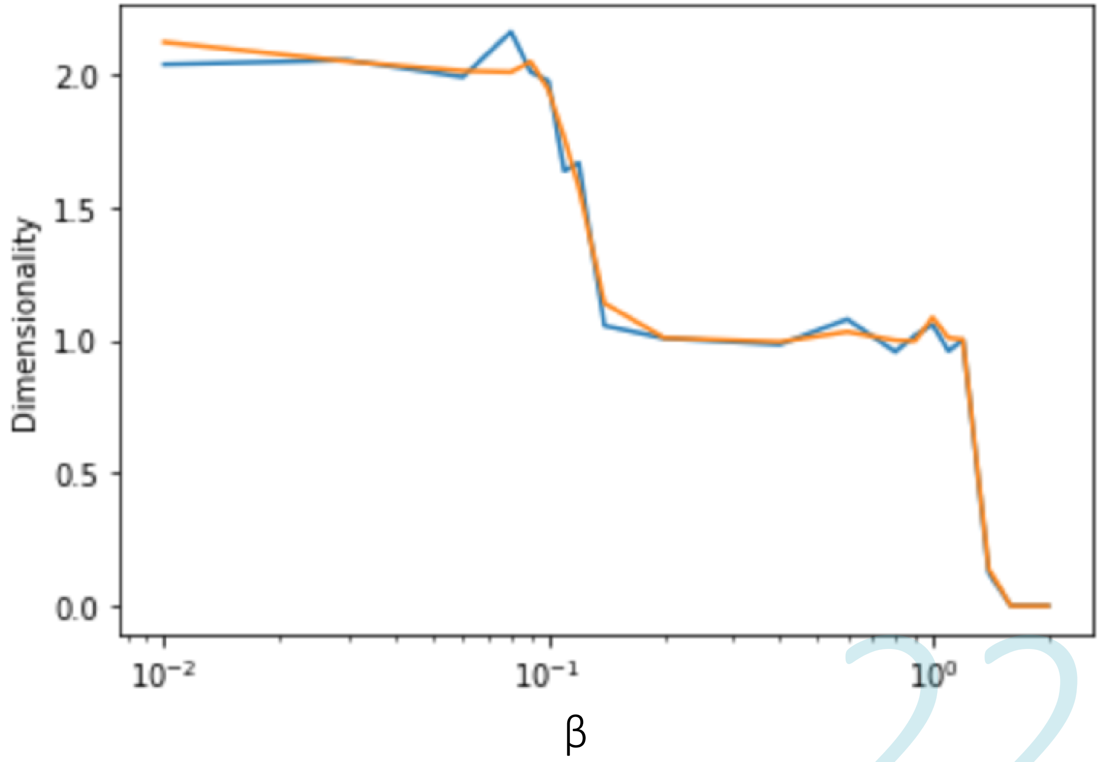


The scaling of KL with beta suggests a notion of dimensionality, that relates to how tightly small gaussians are being packed into the latent space.

$$D_1 = \sum_i \frac{d\langle KL_i \rangle}{d \log \beta}$$

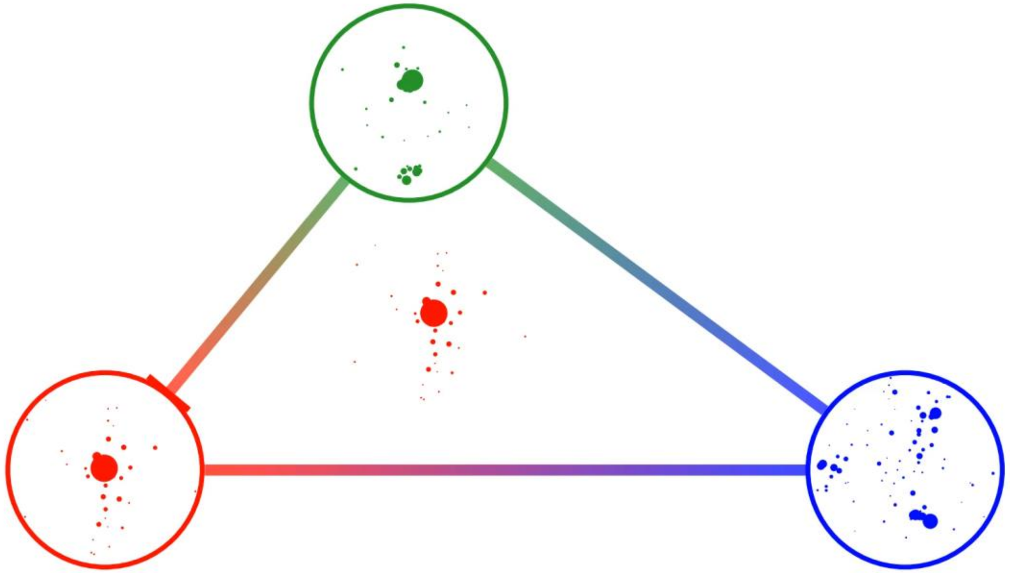
A similar notion of dimensionality can be derived from the packing of gaussians into the data-space.

$$D_2 = \frac{d\langle |\Delta \boldsymbol{x}|^2 \rangle}{d\beta^2}$$



Distance between Jets:

*EMD: Cost to transform one jet into another = Energy * distance*



$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \min_{\{f_{ij}\}} \sum_{ij} f_{ij}\frac{\theta_{ij}}{R} + \left|\sum_i E_i - \sum_j E'_j\right|,$$

$$f_{ij} \geq 0, \quad \sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = E_{\min},$$

Defines a metric space in which jets or collider events form a geometric manifold.

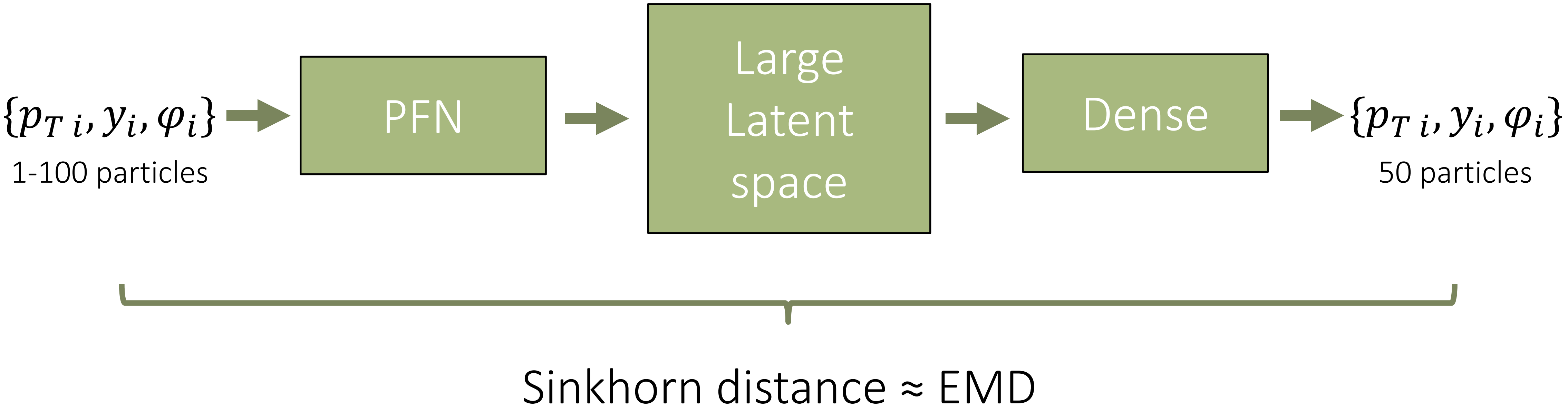
arXiv:1902.02346

Video taken from https://energyflow.network/docs/emd/, Eric Metediov, Patrick Komiske III, Jesse Thaler

In practice I use a tractable approximation to EMD called Sinkhorn distance

arXiv:1306.0895 [stat.ML] M. Cuturi

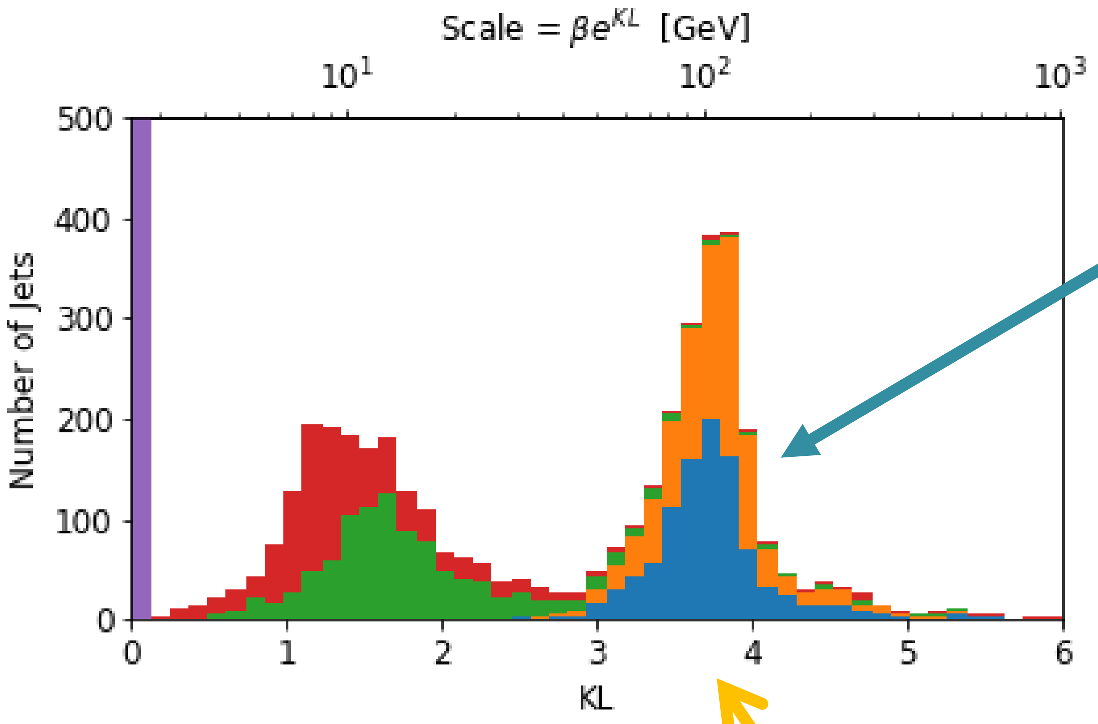
Jet VAE

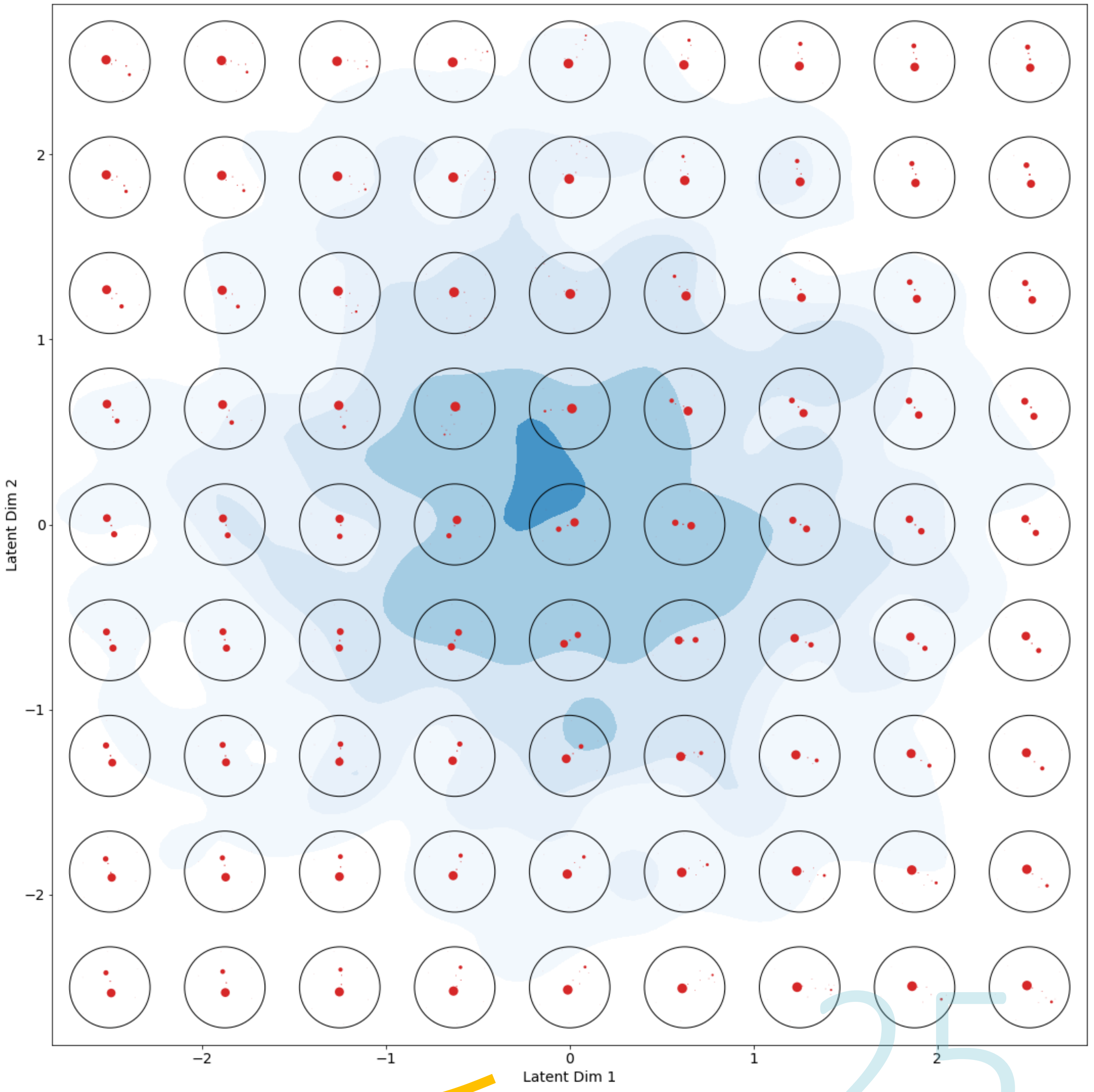


$\{p_{T\,i}, y_i, \varphi_i\}$
1-100 particles
PFN
Large Latent space
Dense
$\{p_{T\,i}, y_i, \varphi_i\}$
50 particles
Sinkhorn distance ≈ EMD

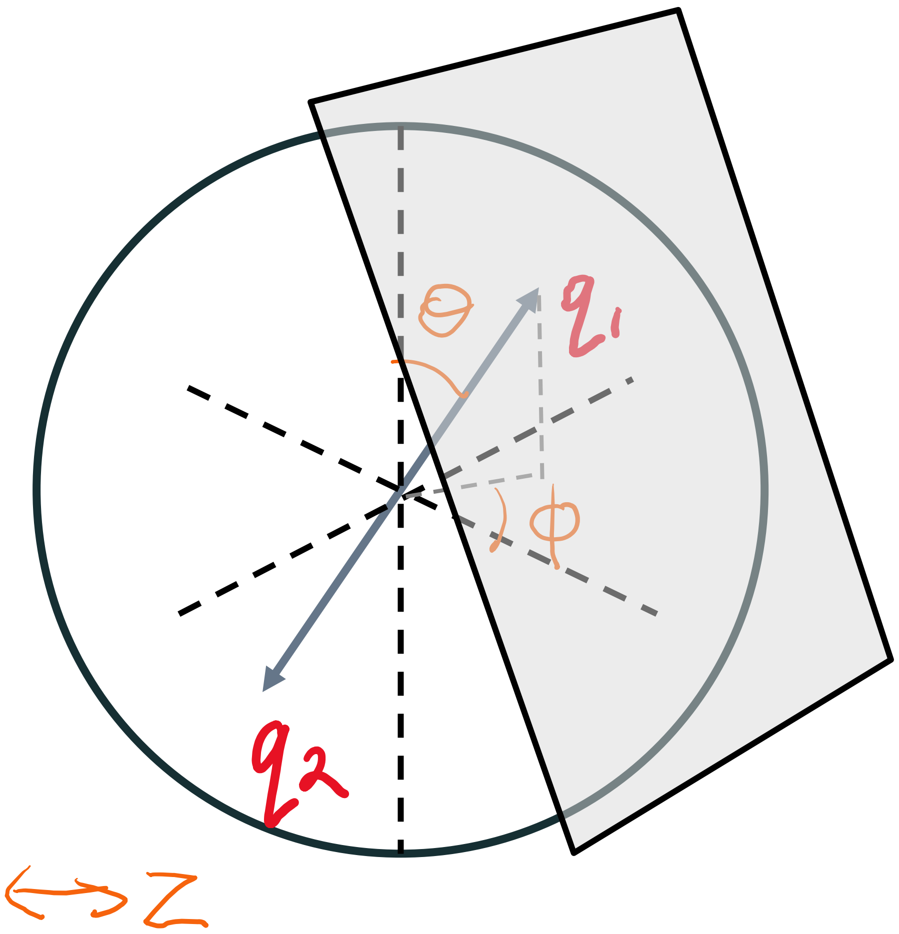
W Jets

$\beta = 2.5\ GeV$

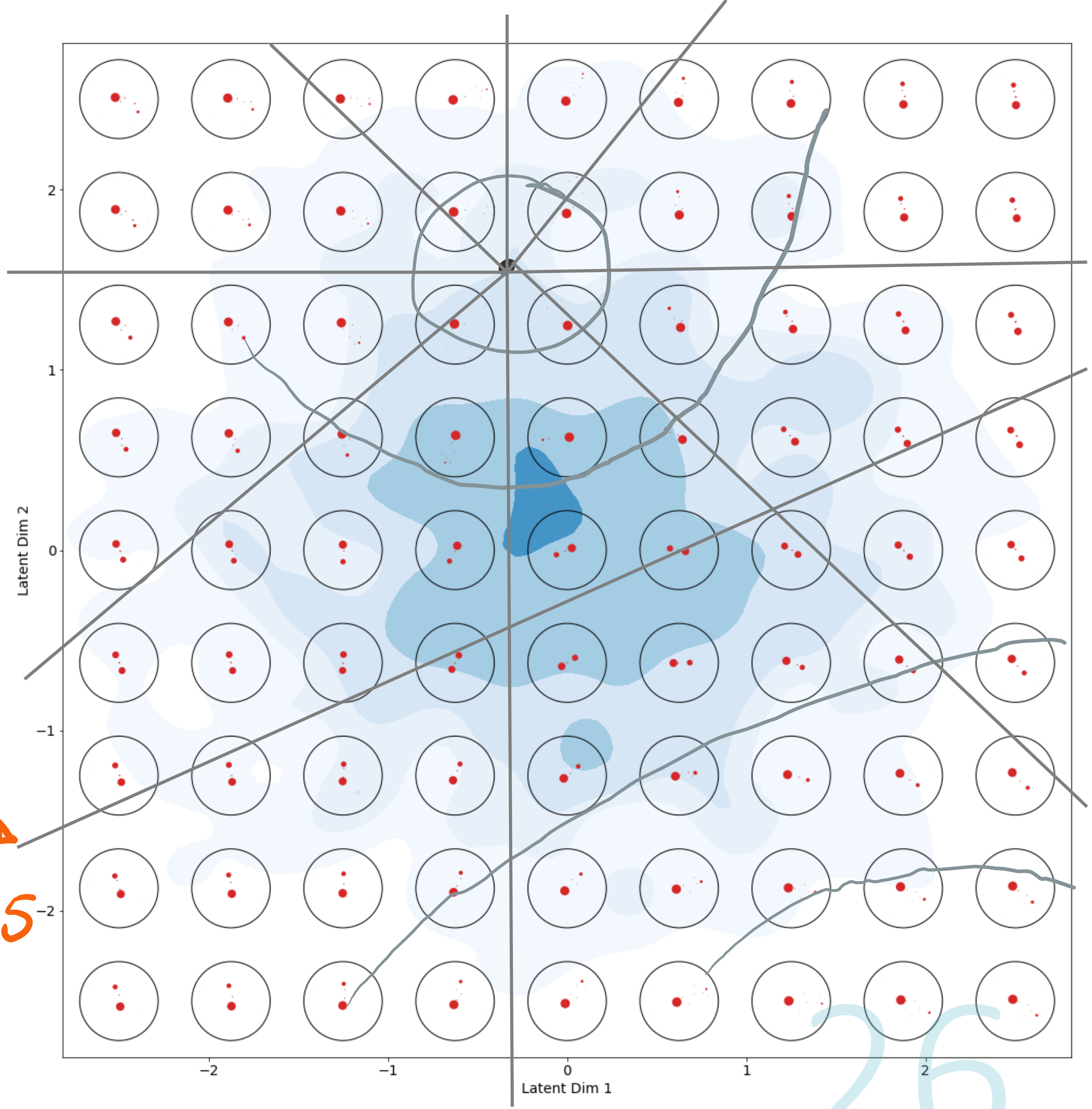




W Jets

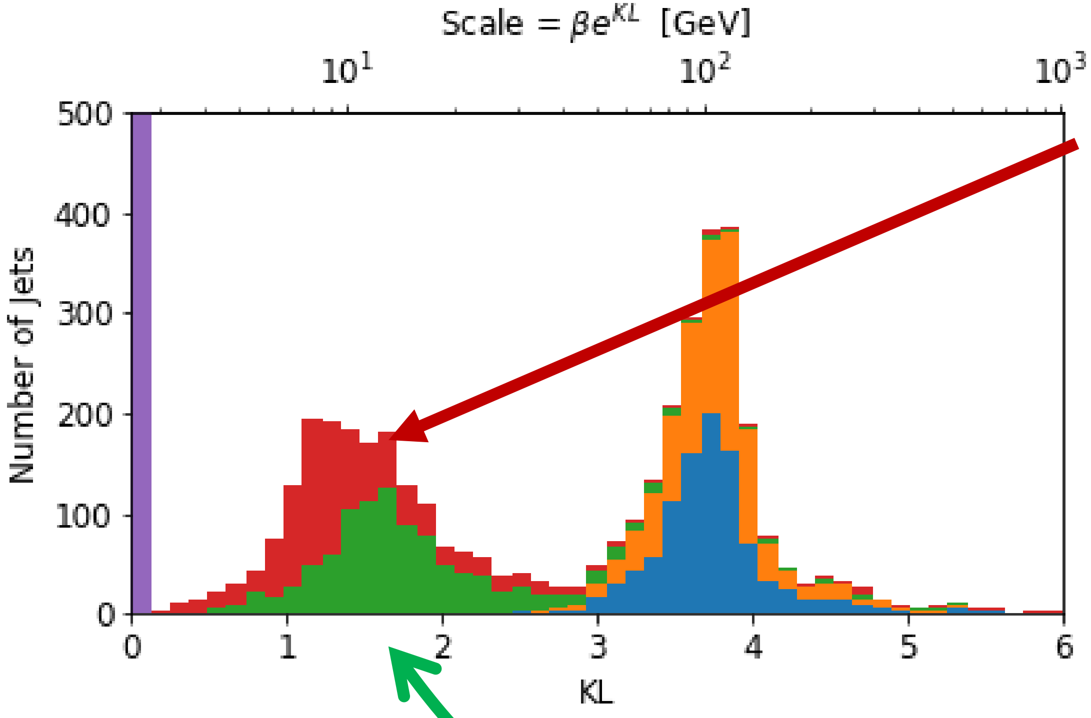


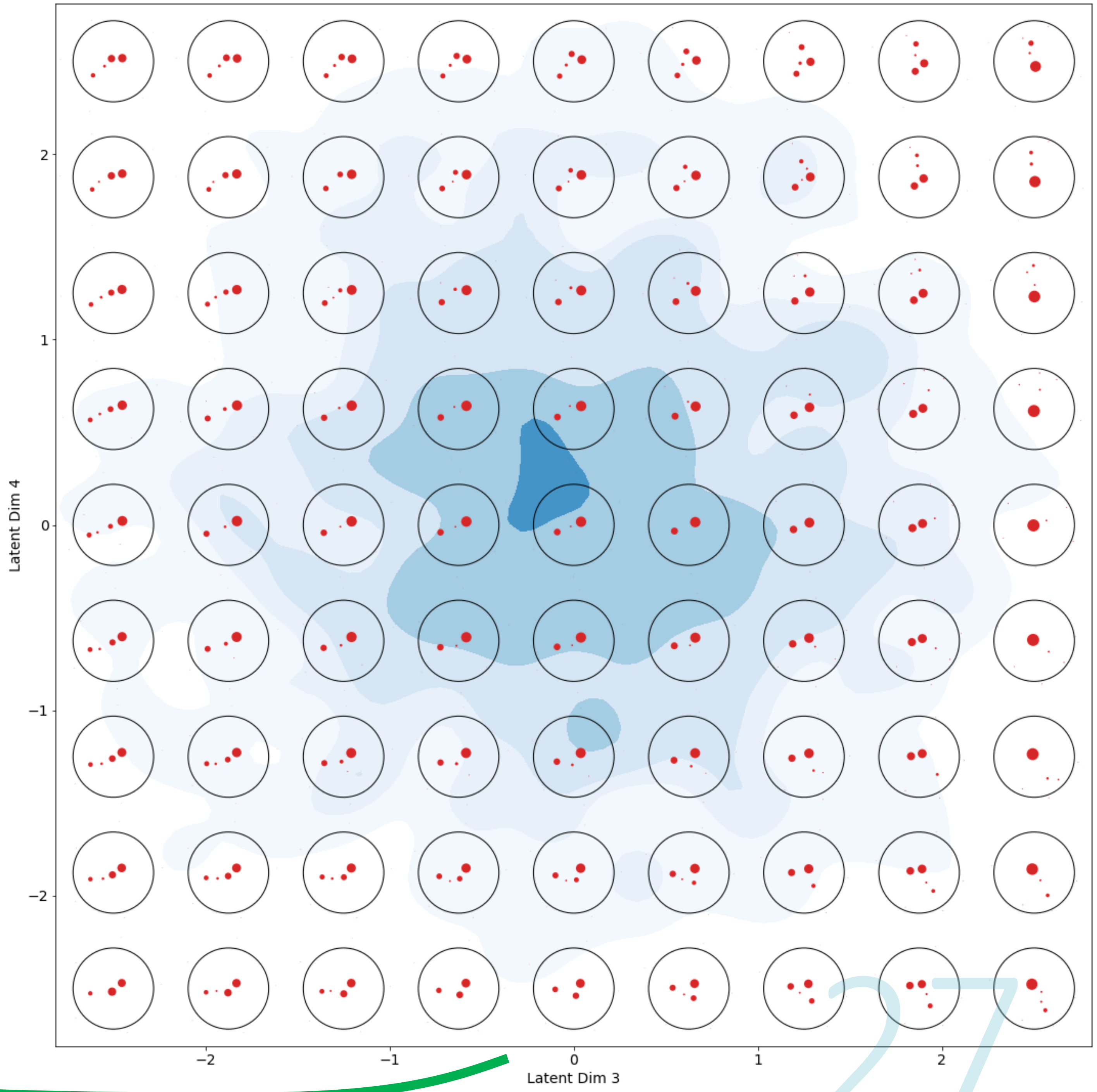
Boosted Frame:



W Jets

$\beta = 2.5\ GeV$

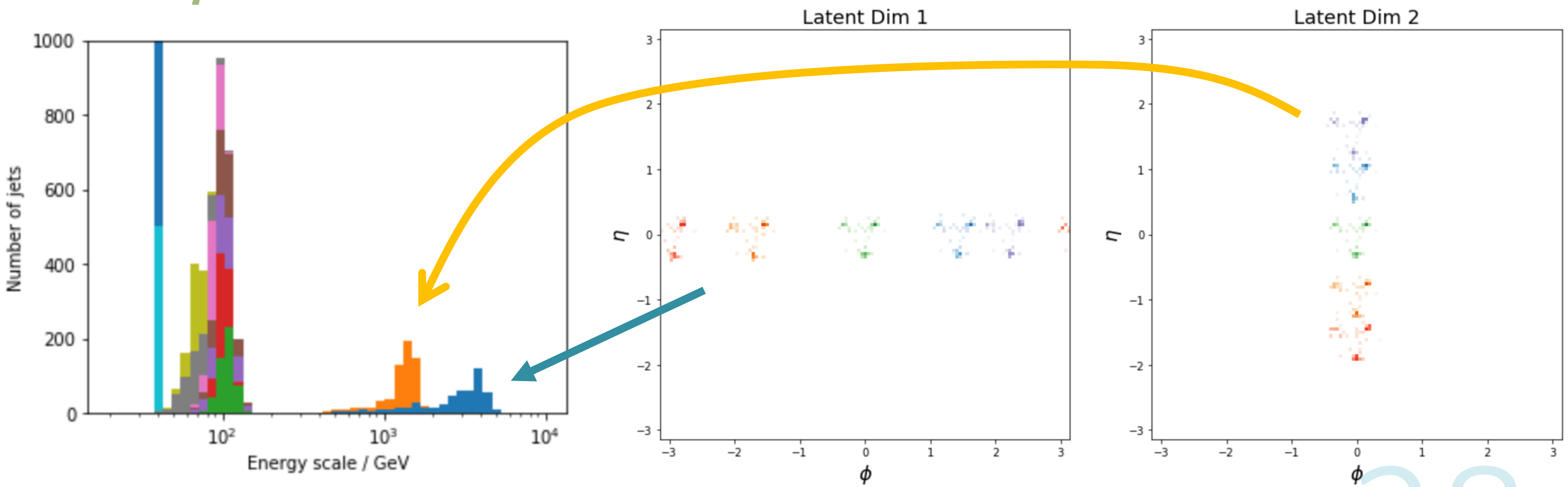




Exploring the Learnt Representation:

Top Jets

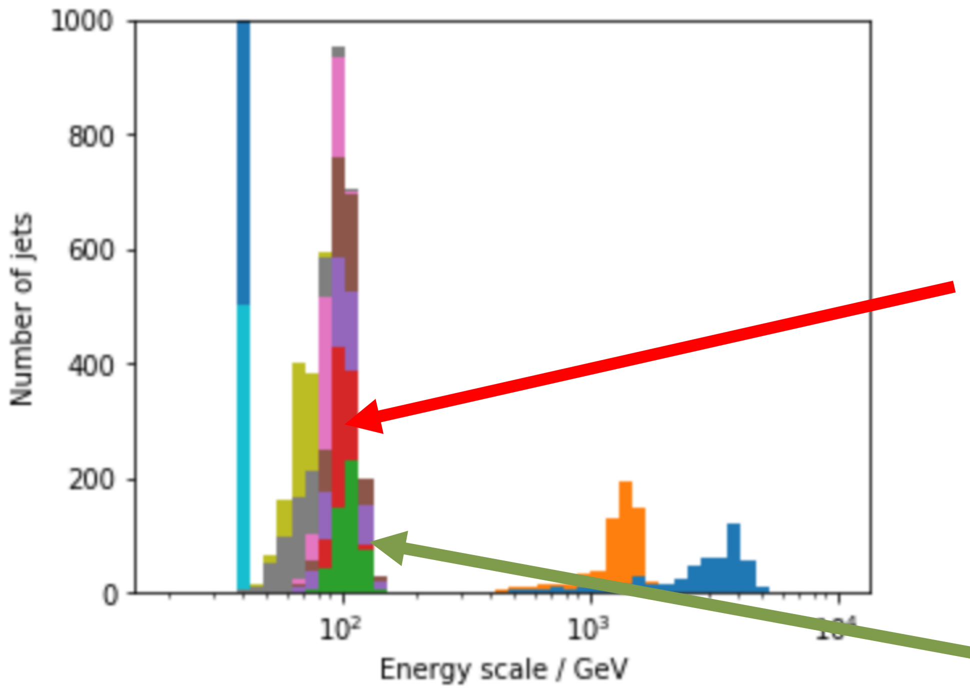
$\beta = 40\ GeV$

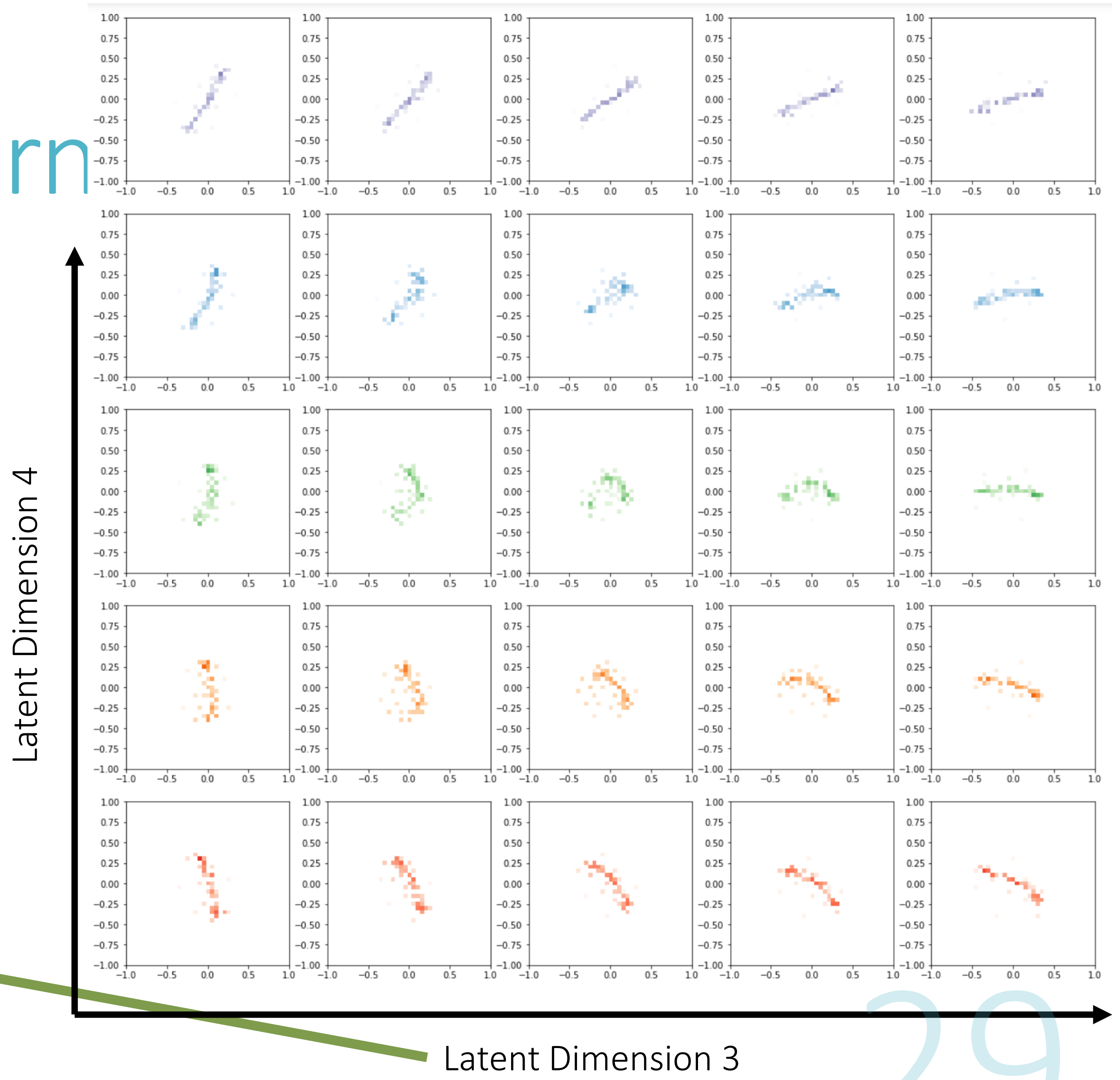


Exploring the Learn

Top Jets

$\beta = 40\ GeV$

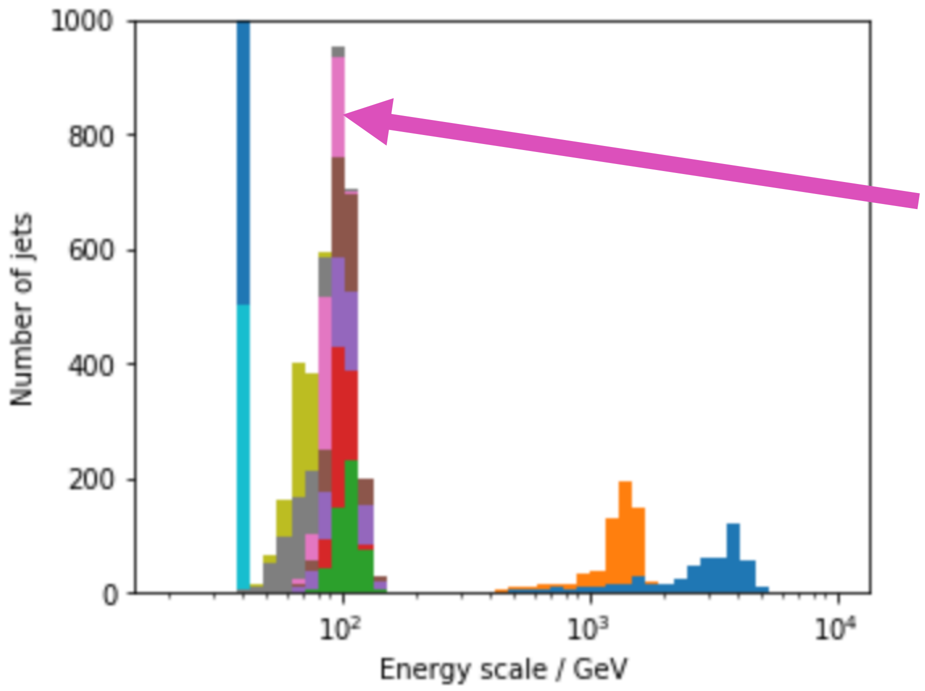


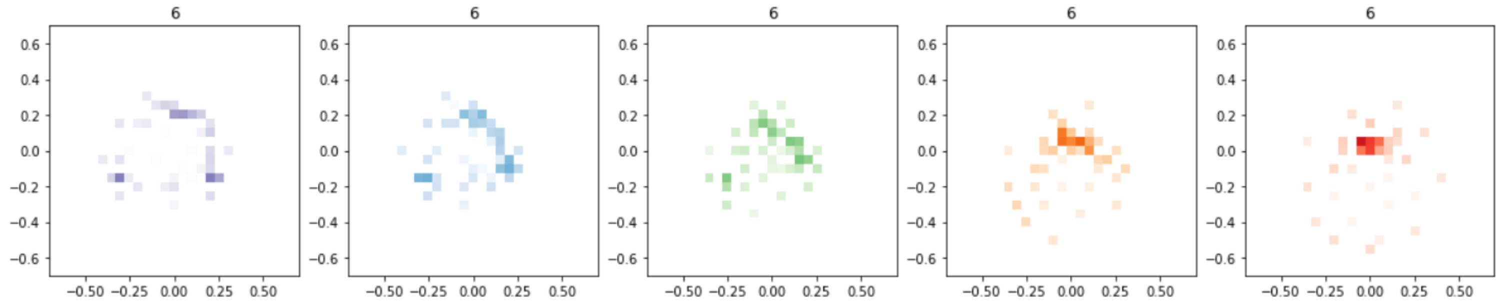


Exploring the Learnt Representation:

Top Jets

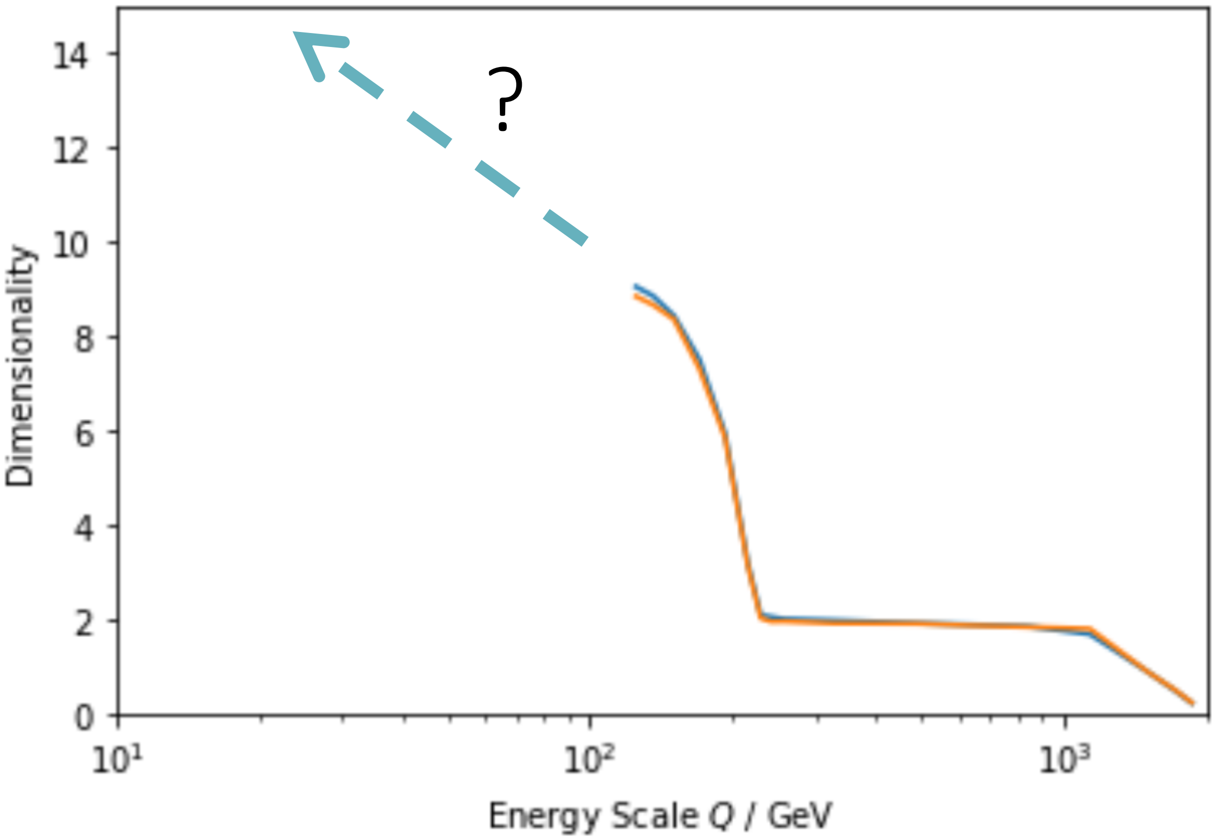
$\beta = 40\ GeV$

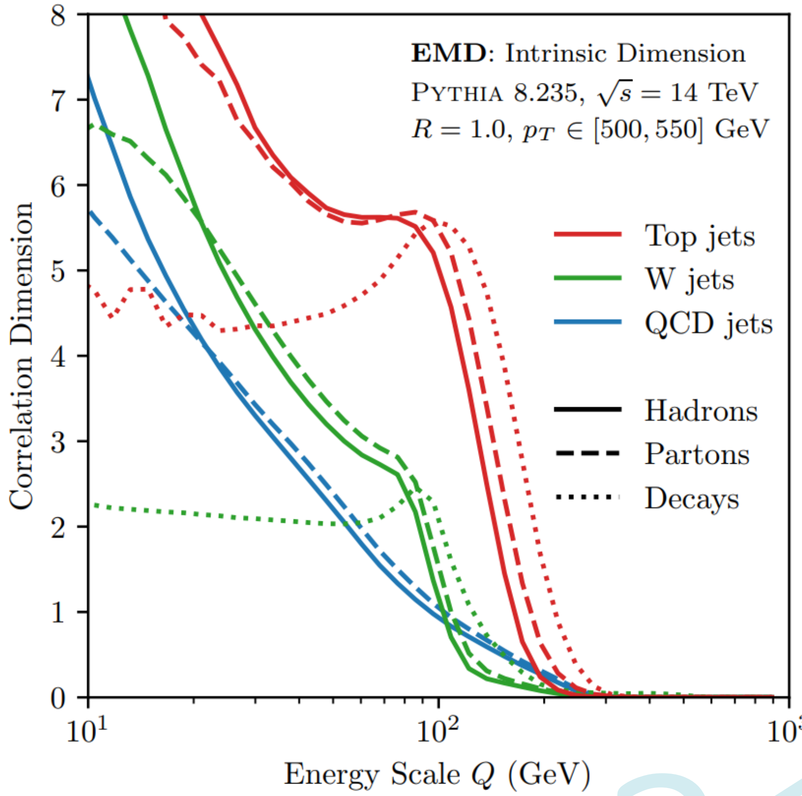




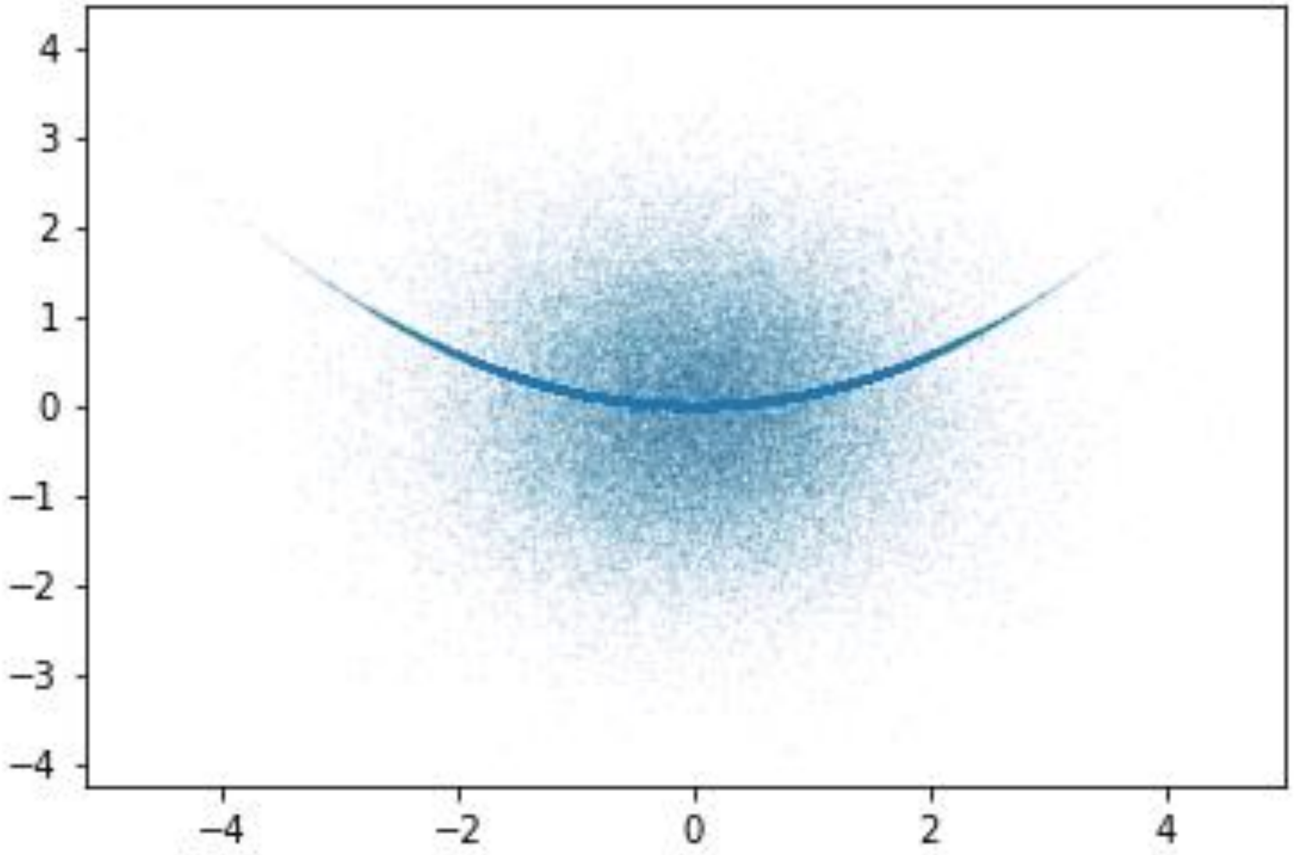
Exploring the Learnt Representation:

Dimensionality



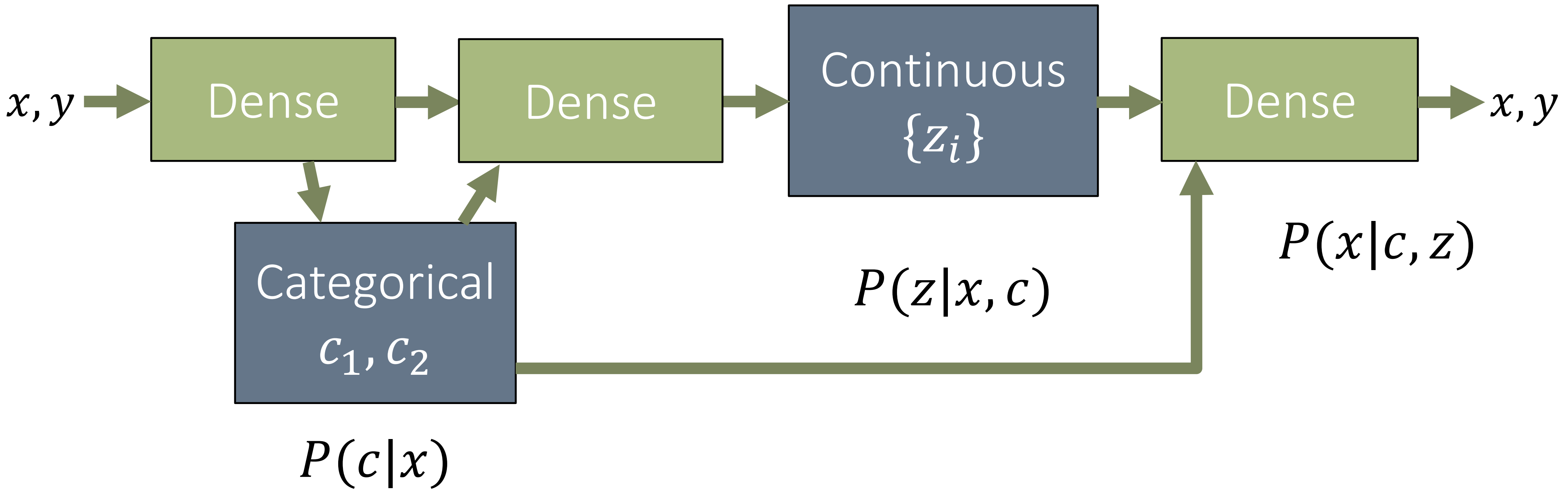


A Mixed Sample



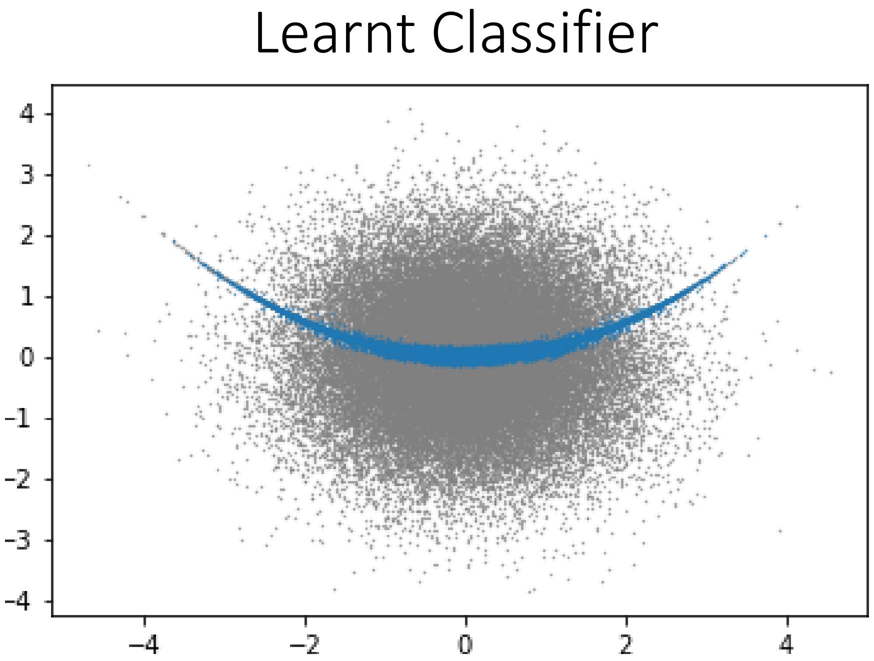
A Mixed Sample

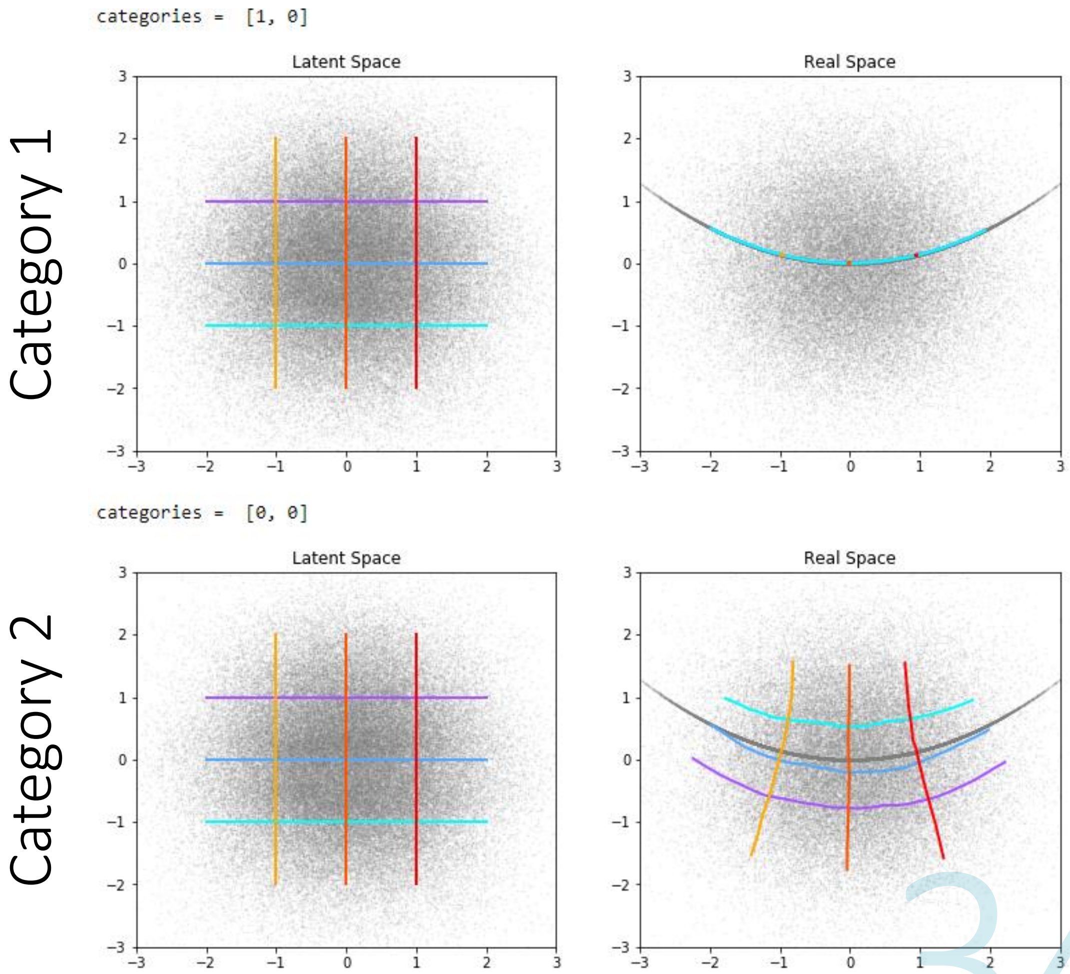
VAE structure



A Mixed Sample

VAE structure





Conclusions

VAE latent spaces learn concrete representations of the manifolds on which they are trained.

A meaningful distance metric which encodes interesting physics at different scales leads to a meaningful learnt representation which encodes interesting physics at different scales.

For a sufficiently simple manifold, the VAE learnt representation is:

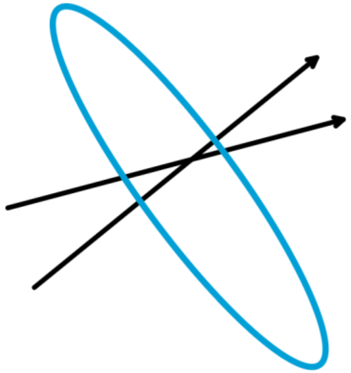
- *Orthogonalized*
- *Hierarchically organized*
- *Has a scale-dependent fractal dimension which directly relates to that of the true data manifold*

These properties are due to the demand to be ***parsimonious*** with information.

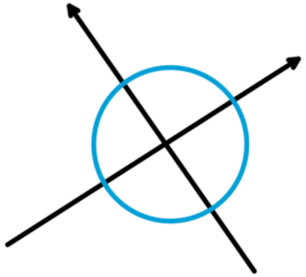
The Variational Autoencoder:

Orthogonalization and Organization is Information-Efficient

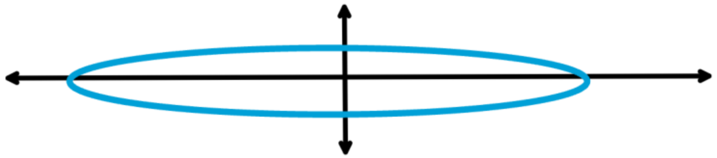
Orthogonalization:



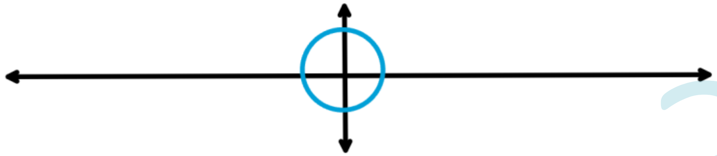
vs



Organization:



vs



Exploring the Learnt Representation:

Top Jets

$\beta = 400\ GeV$