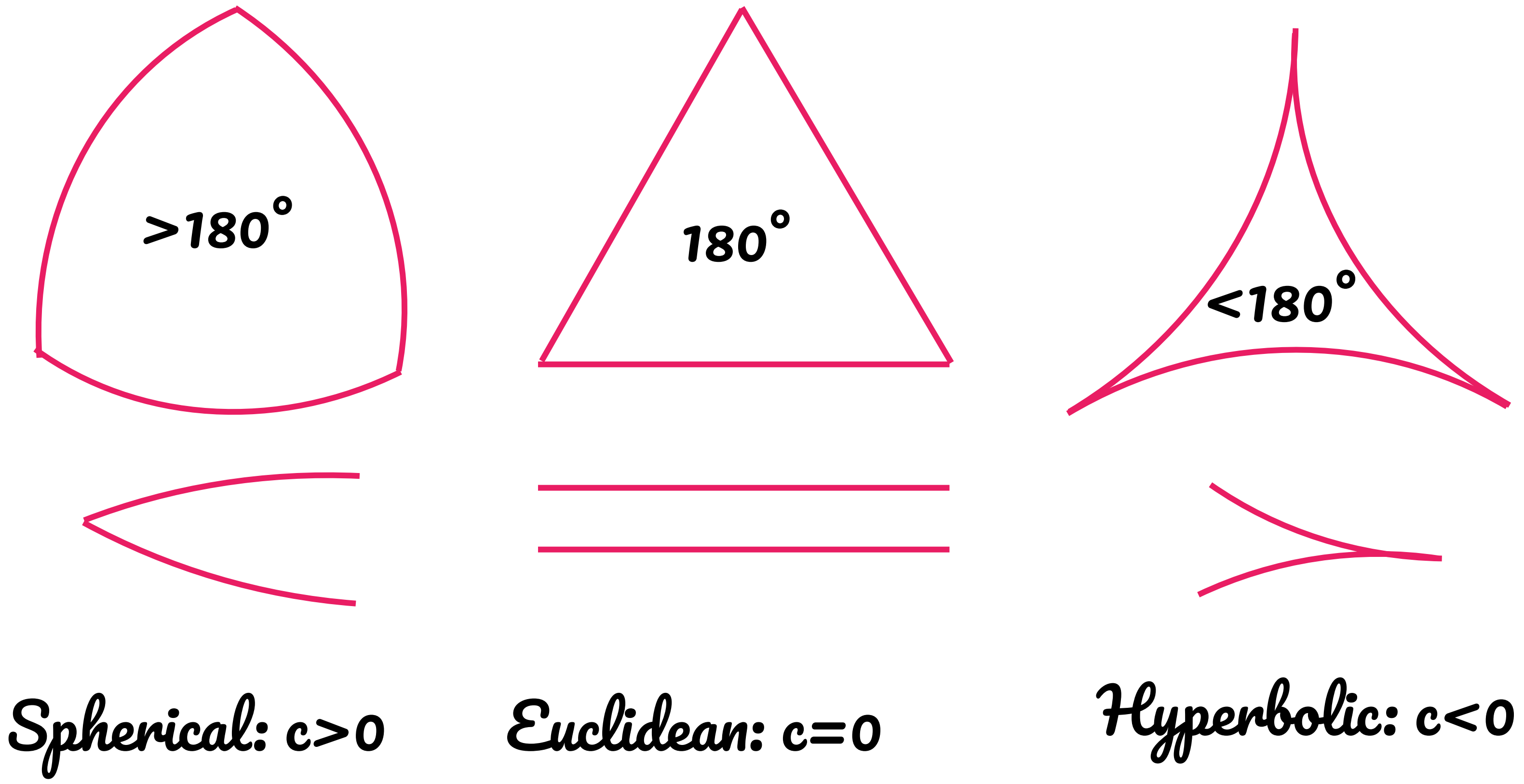
Poincaré Variational Auto-Encoders

E. Mathieu, C. Le Lan, C. Maddison, R. Tomioka, Y.W Teh

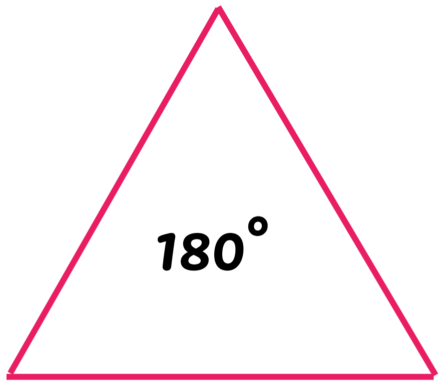
UNIVERSITY OF OXFORD

Hyperbolic Geometry

(Gaussian) curvature and triangles

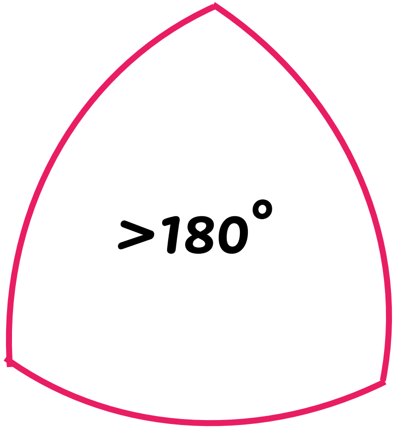


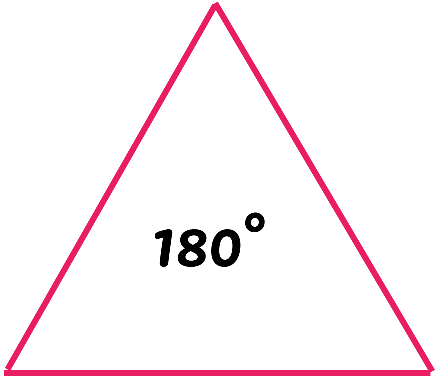
(Gaussian) curvature and triangles



Euclidean: c=0

(Gaussian) curvature and triangles

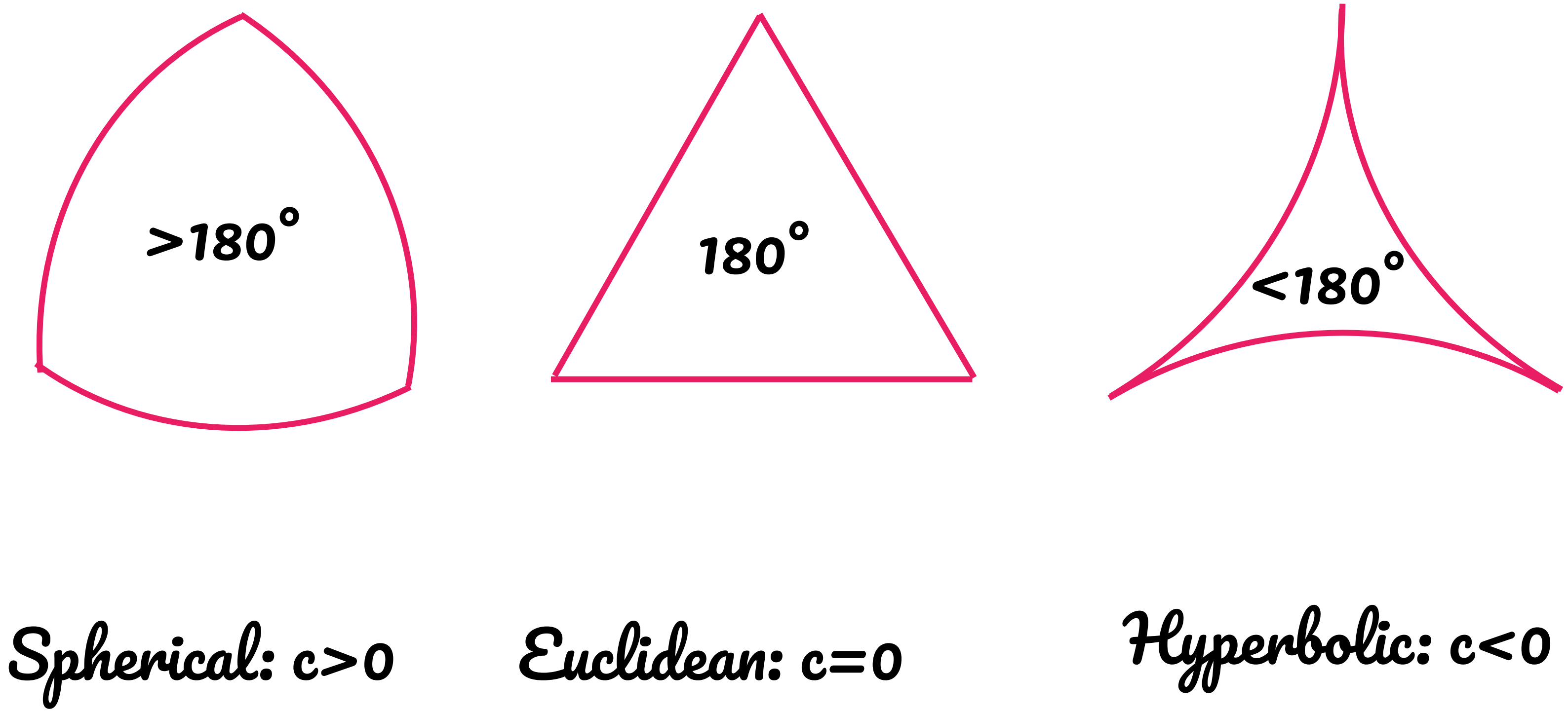




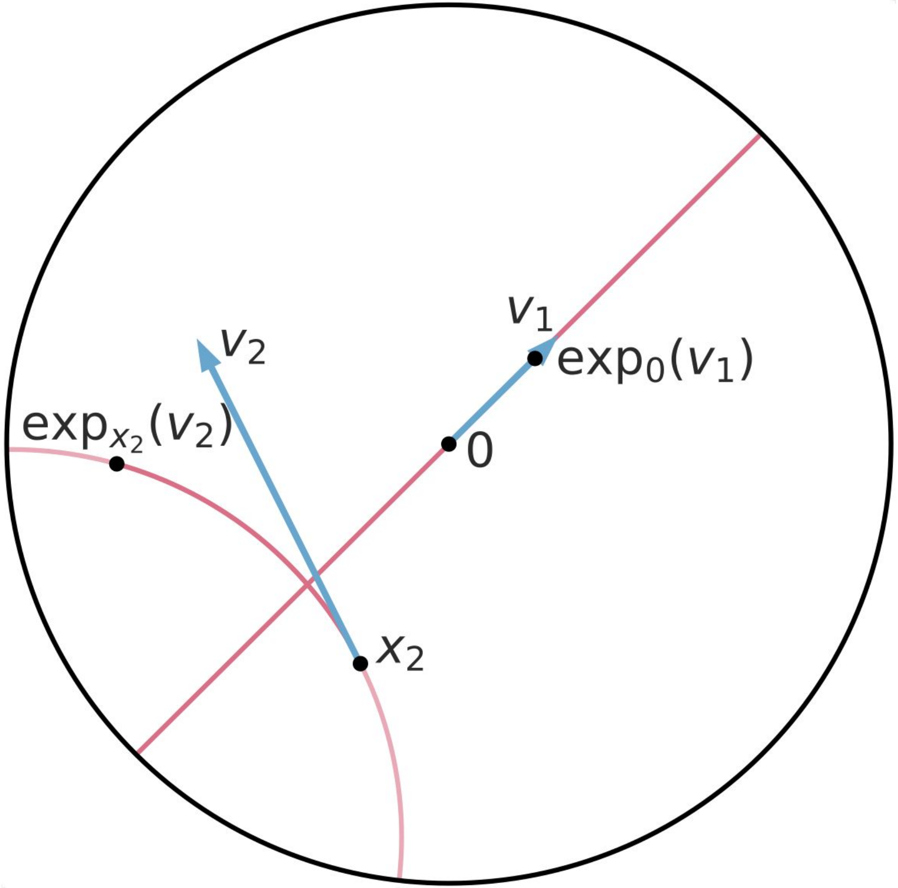
Spherical: c>0

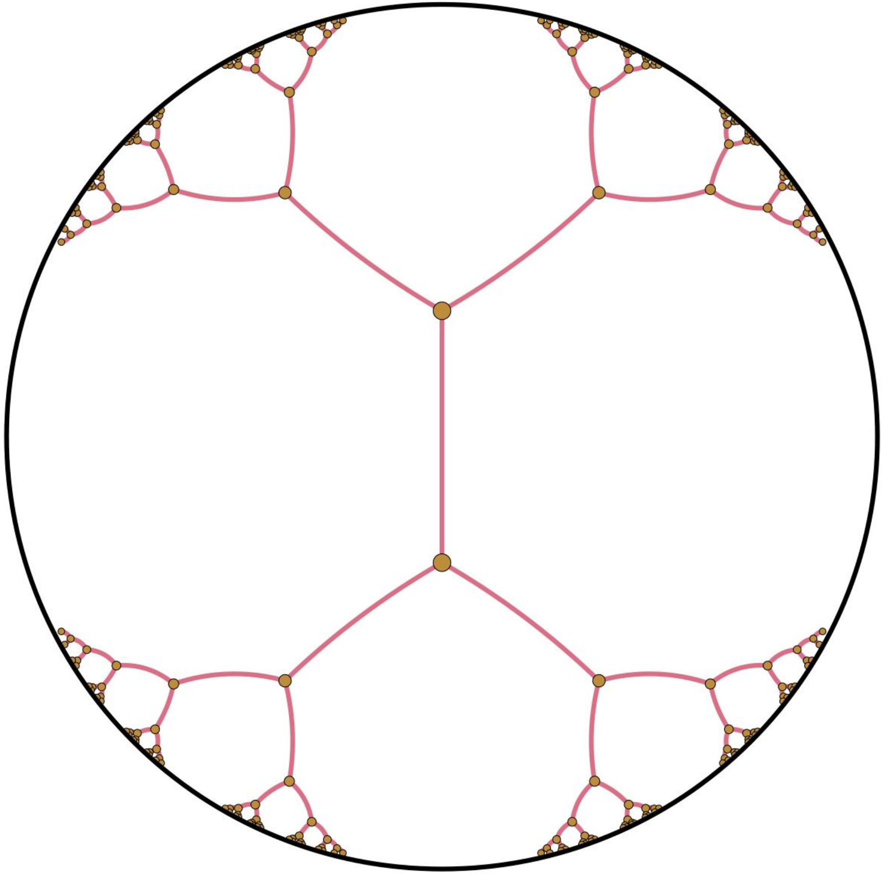
Euclidean: c=0

(Gaussian) curvature and triangles

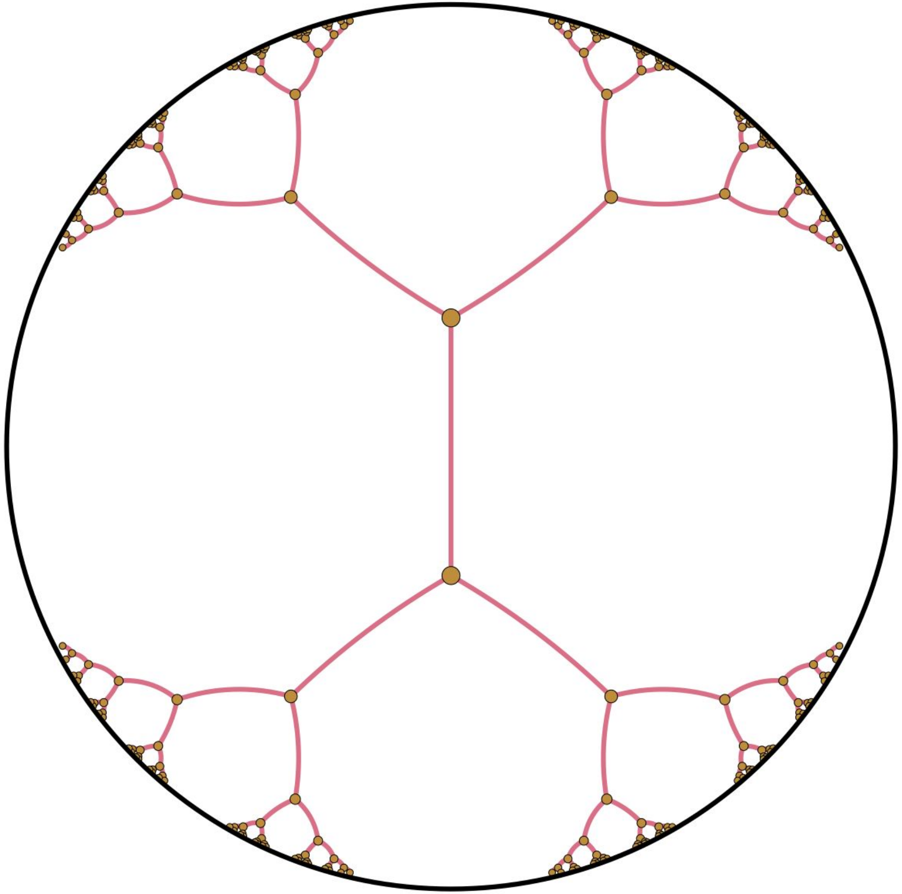


Poincaré Ball: A model of Hyperbolic geometry



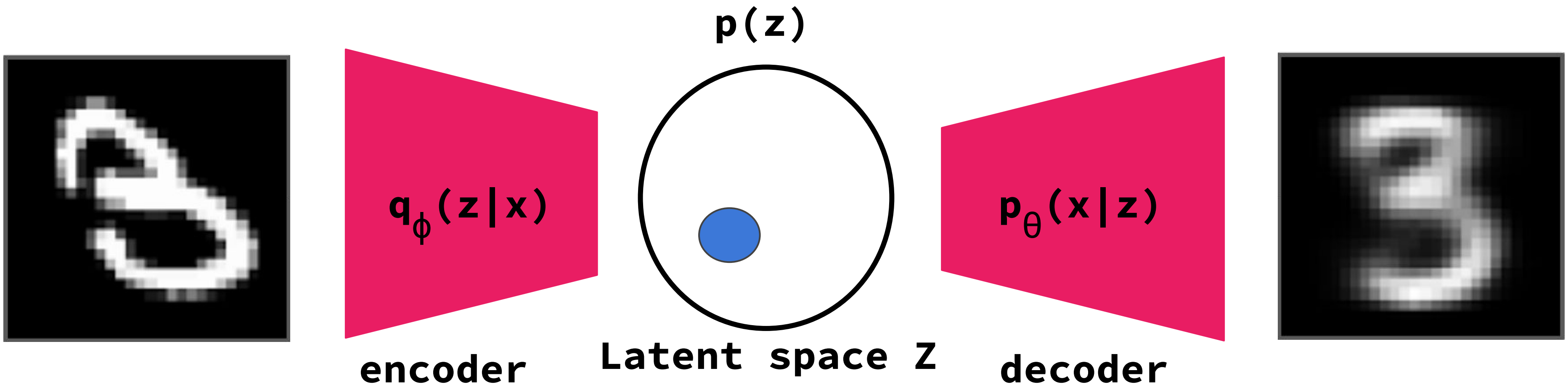


Poincaré Ball: A model of Hyperbolic geometry

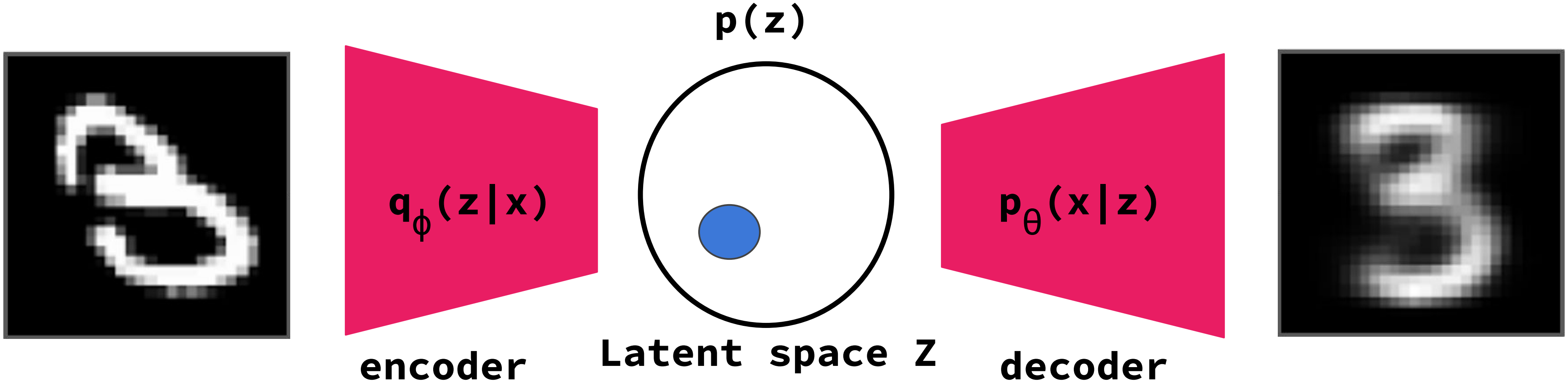


Model

Poincaré Variational Auto-Encoder

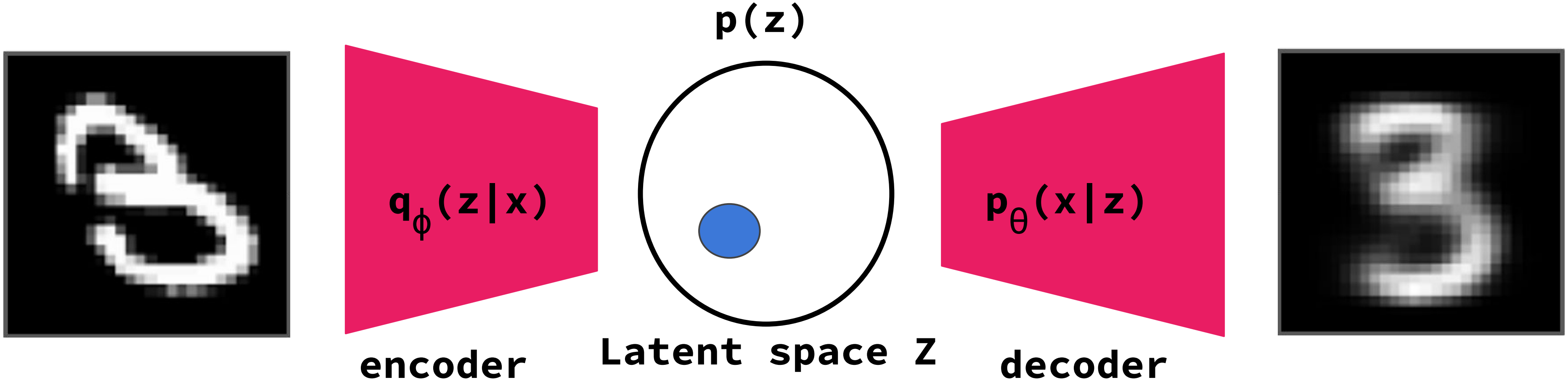


Poincaré Variational Auto-Encoder



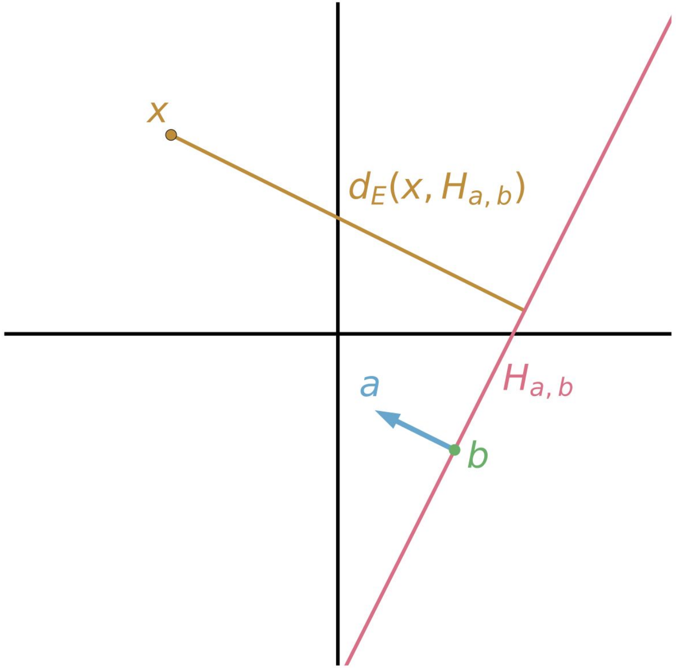
$$\log p(\boldsymbol{x}) \geq \int_{\mathcal{M}} \ln\left(\frac{p_\theta(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right) q_\phi(\boldsymbol{z}|\boldsymbol{x})\, d\mathcal{M}(\boldsymbol{z}).$$

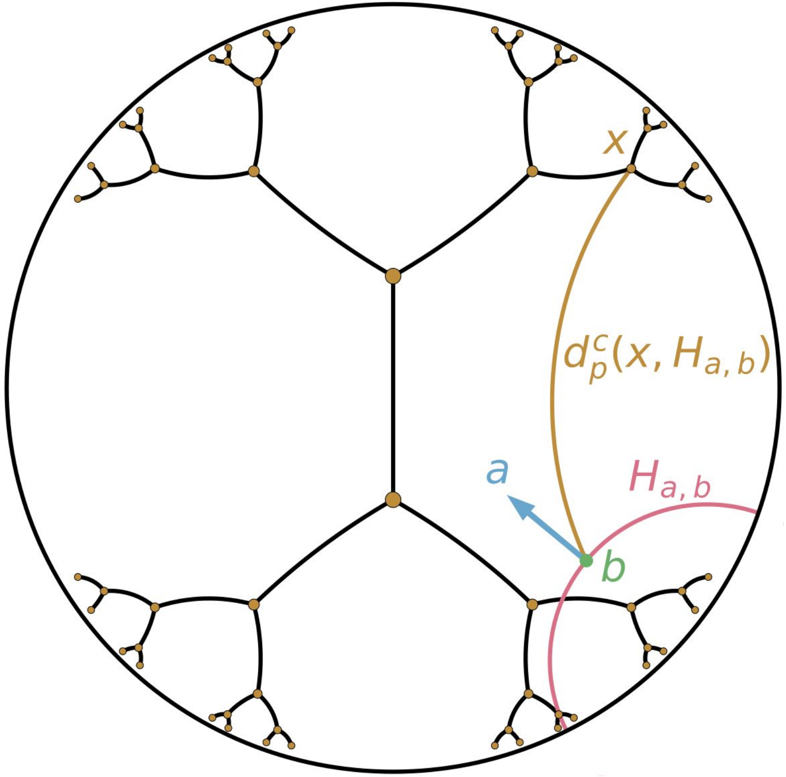
Poincaré Variational Auto-Encoder



$$\log p(\boldsymbol{x}) = \ln \int_{\mathcal{M}} p_\theta(\boldsymbol{x}|\boldsymbol{z}) p(\boldsymbol{z}) d\mathcal{M}(\boldsymbol{z})$$
$$\geq \mathcal{L}_{\mathcal{M}}(\boldsymbol{x}; \theta, \phi) \triangleq \int_{\mathcal{M}} \ln \left(\frac{p_\theta(\boldsymbol{x}|\boldsymbol{z}) p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})} \right) q_\phi(\boldsymbol{z}|\boldsymbol{x}) \, d\mathcal{M}(\boldsymbol{z}).$$
$$\approx \sum_k \ln p_\theta(\boldsymbol{x}|\boldsymbol{z}^k) + \ln p(\boldsymbol{z}^k) - \ln q_\phi(\boldsymbol{z}^k|\boldsymbol{x})$$

Decoder: From latents to observations



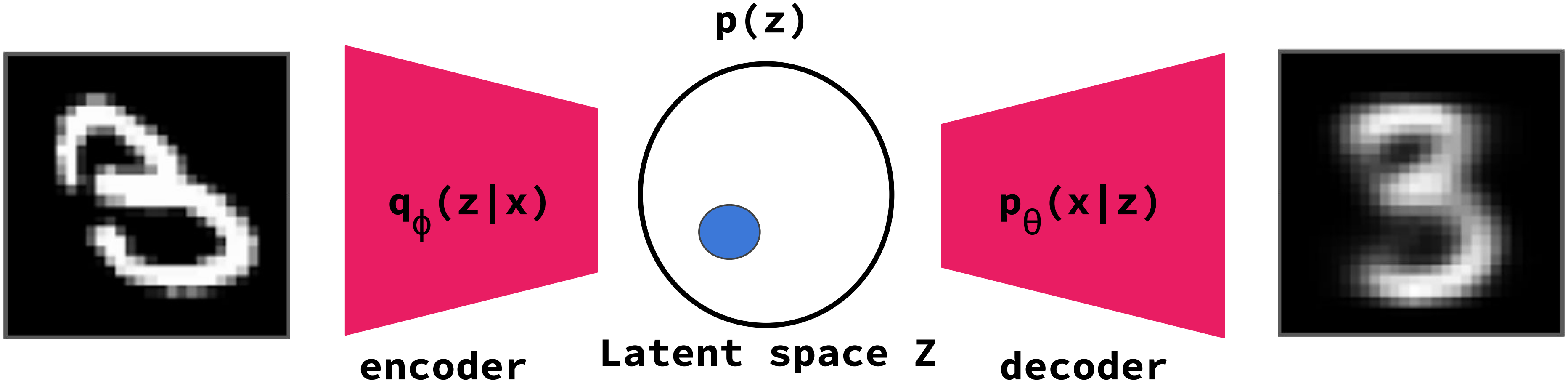


$$f_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x} - \boldsymbol{b} \rangle$$

$$f_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{a}, \boldsymbol{x} - \boldsymbol{b} \rangle) \, \|\boldsymbol{a}\| \, d_E(\boldsymbol{x}, H^c_{\boldsymbol{a},\boldsymbol{b}})$$

$$f^c_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{a}, \log^c_{\boldsymbol{x}}(\boldsymbol{b}) \rangle_{\boldsymbol{b}}) \, \|\boldsymbol{a}\|_{\boldsymbol{b}} \, d^c_p(\boldsymbol{x}, H^c_{\boldsymbol{a},\boldsymbol{b}})$$

Poincaré Variational Auto-Encoder




```
V, sigma = enc(x)
mu = exp0(v)
```

Generalizing the Normal distribution

Riemannian Normal

$$\frac{d\nu^{\mathrm{R}}(\boldsymbol{x}|\boldsymbol{\mu},\sigma^2)}{d\mathcal{M}(\boldsymbol{x})} \propto \exp\left(-\frac{d_p^c(\boldsymbol{\mu},\boldsymbol{x})^2}{2\sigma^2}\right)$$

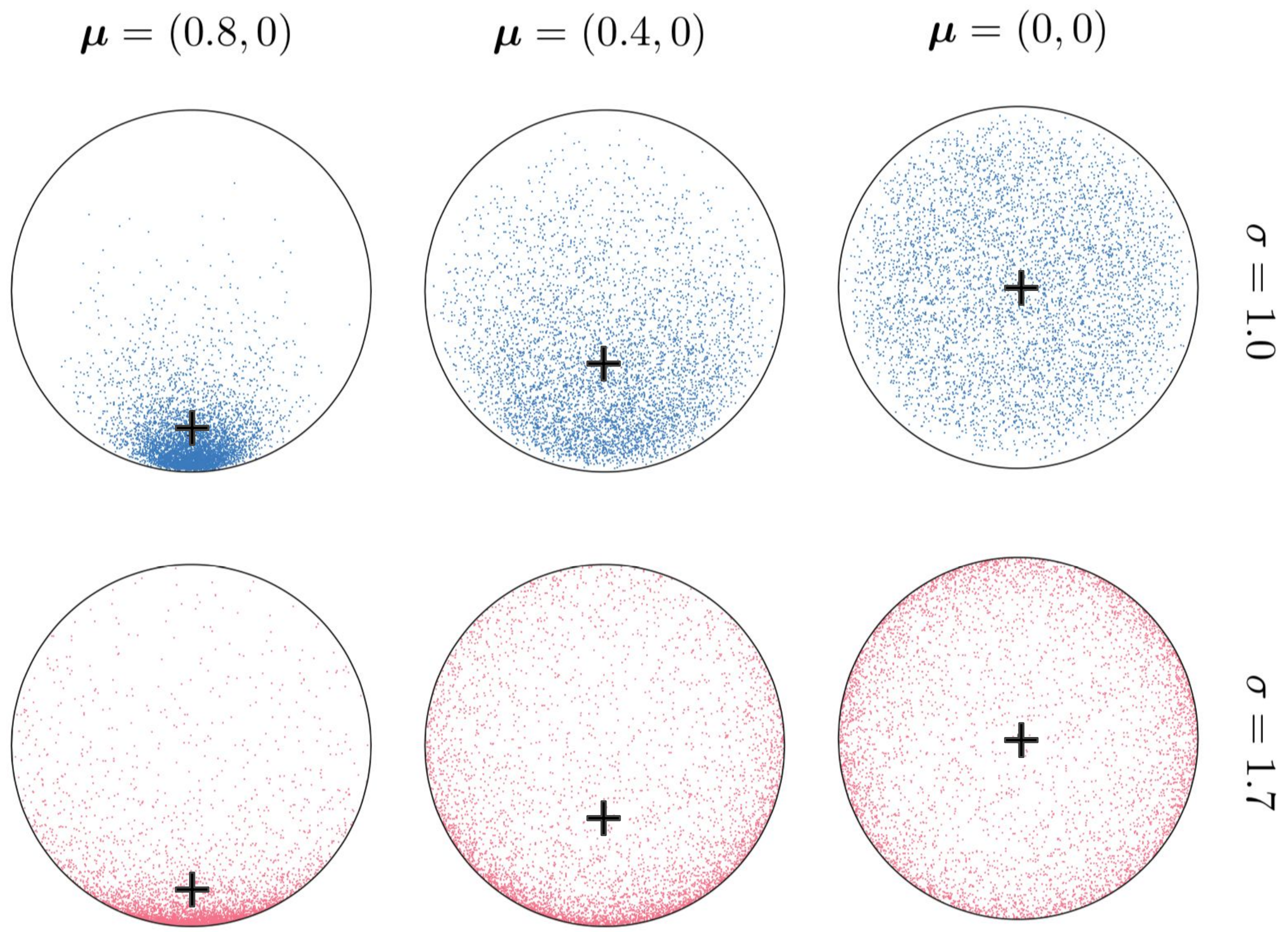
Generalizing the Normal distribution

Riemannian Normal

$$\frac{d\nu^{\mathrm{R}}(\boldsymbol{x}|\boldsymbol{\mu},\sigma^2)}{d\mathcal{M}(\boldsymbol{x})} \propto \exp\left(-\frac{d_p^c(\boldsymbol{\mu},\boldsymbol{x})^2}{2\sigma^2}\right)$$

Wrapped Normal

$$\boldsymbol{x} = \exp_{\boldsymbol{\mu}}^c\left(\frac{\boldsymbol{v}}{\lambda_{\boldsymbol{\mu}}^c}\right) \quad \boldsymbol{v} \sim \mathcal{N}(\cdot|\mathbf{0},\sigma^2)$$


$\boldsymbol{\mu} = (0.8, 0)$
$\boldsymbol{\mu} = (0.4, 0)$
$\boldsymbol{\mu} = (0, 0)$
$\sigma = 1.0$
$\sigma = 1.7$

Reparametrization "trick"

Riemannian Normal

$$\frac{d\nu^{\mathrm{R}}(\boldsymbol{x}|\boldsymbol{\mu}, \sigma^2)}{d\mathcal{M}(\boldsymbol{x})} \propto \exp\left(-\frac{d_p^c(\boldsymbol{\mu}, \boldsymbol{x})^2}{2\sigma^2}\right)$$

Wrapped Normal

$$\boldsymbol{x} = \exp_{\boldsymbol{\mu}}^c\left(\frac{\boldsymbol{v}}{\lambda_{\boldsymbol{\mu}}^c}\right) \quad \boldsymbol{v} \sim \mathcal{N}(\cdot|\mathbf{0}, \sigma^2)$$

Reparametrization

$$\boldsymbol{x} = \exp_{\boldsymbol{\mu}}^c\left(\frac{r}{\lambda_{\boldsymbol{\mu}}^c}\boldsymbol{\alpha}\right) \quad \texttt{with} \quad r = d_p^c(\boldsymbol{\mu}, \boldsymbol{x})$$

Reparametrization "trick"

Riemannian Normal

Wrapped Normal

$$\frac{d\nu^{\mathrm{R}}(\boldsymbol{x}|\boldsymbol{\mu},\sigma^2)}{d\mathcal{M}(\boldsymbol{x})} \propto \exp\left(-\frac{d_p^c(\boldsymbol{\mu},\boldsymbol{x})^2}{2\sigma^2}\right)$$

$$\boldsymbol{x} = \exp_{\boldsymbol{\mu}}^c\left(\frac{\boldsymbol{v}}{\lambda_{\boldsymbol{\mu}}^c}\right) \quad \boldsymbol{v} \sim \mathcal{N}(\cdot|\mathbf{0},\sigma^2)$$

Reparametrization

$$\boldsymbol{x} = \exp_{\boldsymbol{\mu}}^c\left(\frac{r}{\lambda_{\boldsymbol{\mu}}^c}\boldsymbol{\alpha}\right) \quad \texttt{with} \quad r = d_p^c(\boldsymbol{\mu},\boldsymbol{x})$$

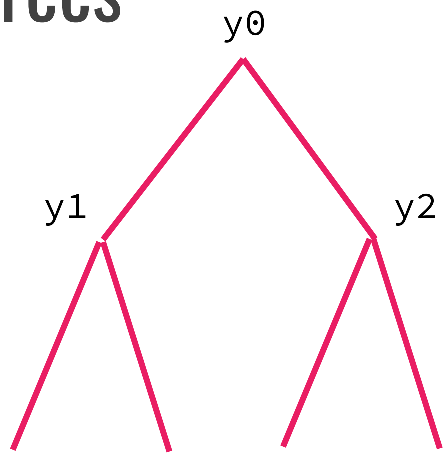
$$\rho^{\mathrm{R}}(r) \propto \mathbb{1}_{\mathbb{R}_+}(r)e^{-\frac{r^2}{2\sigma^2}}\left(\frac{\sinh(\sqrt{c}r)}{\sqrt{c}}\right)^{d-1}$$

$$\rho^{\mathrm{W}}(r) \propto \mathbb{1}_{\mathbb{R}_+}(r)\, e^{-\frac{r^2}{2\sigma^2}}\, r^{d-1}$$

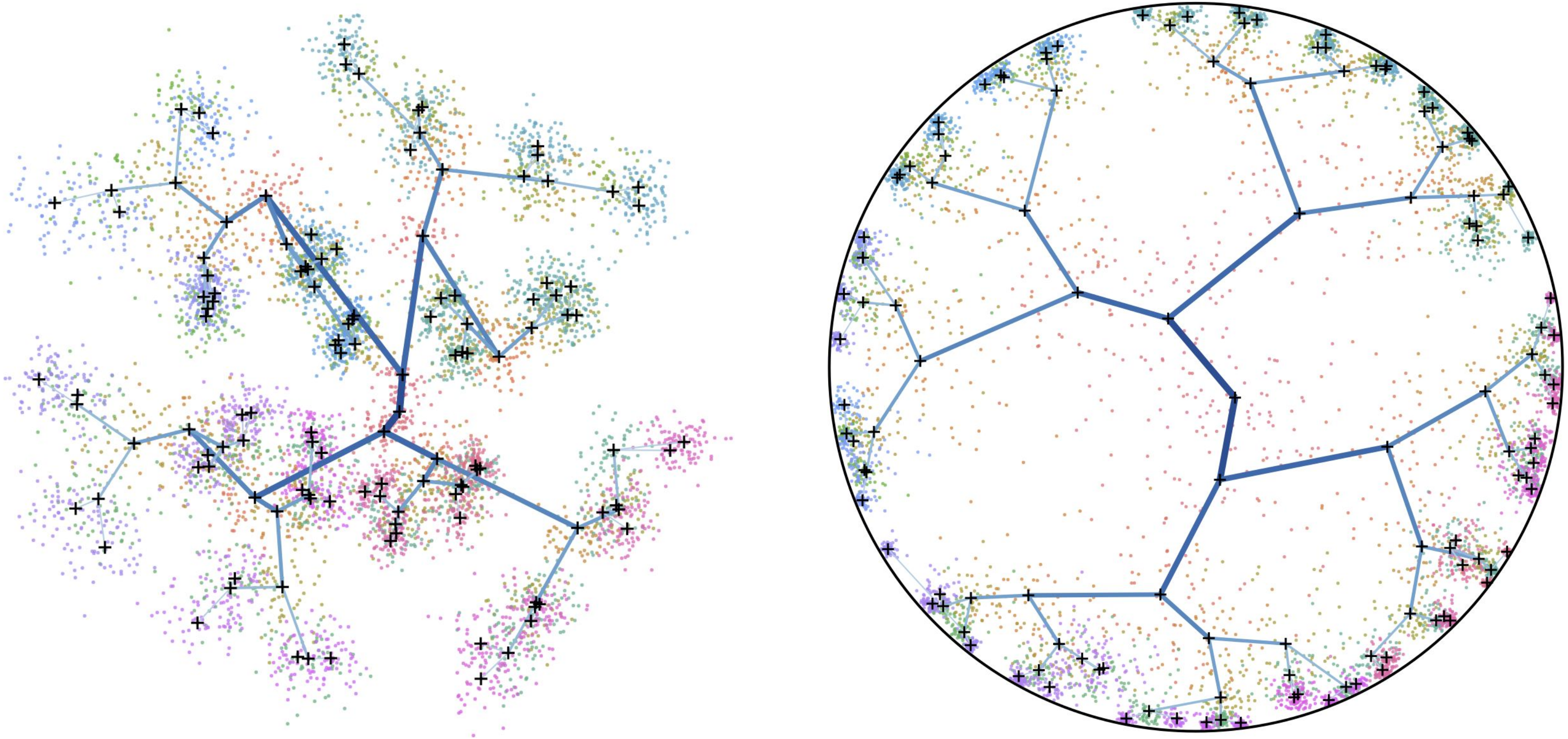
Experiments

Synthetic dataset: Branching diffusion trees

$$\boldsymbol{y}_i \sim \mathcal{N}\left(\cdot \,|\boldsymbol{y}_{\pi(i)}, \sigma_0^2\right) \quad \forall i \in 1, \ldots, N$$



		Models					
	σ	$\mathcal{N}$-VAE	$\mathcal{P}^{0.1}$-VAE	$\mathcal{P}^{0.3}$-VAE	$\mathcal{P}^{0.8}$-VAE	$\mathcal{P}^{1.0}$-VAE	$\mathcal{P}^{1.2}$-VAE
$\mathcal{L}_{\text{IWAE}}$	1	$57.14 \pm .20$	$57.10 \pm .18$	$57.16 \pm .18$	$56.88 \pm .20$	$56.71 \pm .19$	$56.58 \pm .22$
$\mathcal{L}_{\text{IWAE}}$	1.3	$57.03 \pm .20$	$56.91 \pm .18$	$56.91 \pm .18$	$56.39 \pm .19$	$56.21 \pm .20$	$56.07 \pm .22$
$\mathcal{L}_{\text{IWAE}}$	1.7	$57.00 \pm .18$	$56.77 \pm .18$	$56.62 \pm .16$	$55.90 \pm .22$	$55.70 \pm .19$	$\mathbf{55.60 \pm .18}$



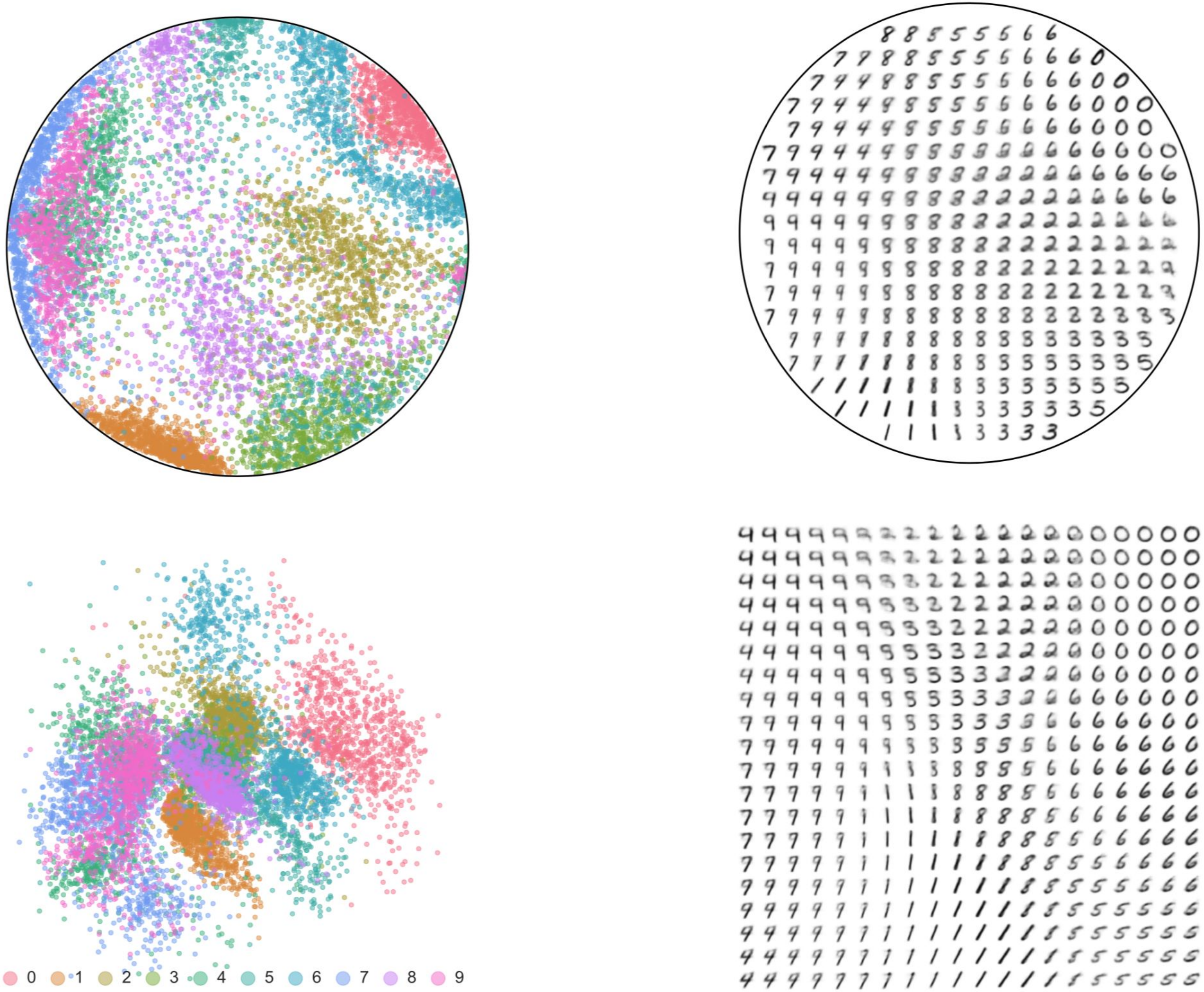
MNIST: Handwritten digits

	c	Dimensionality 2	5	10	20
$\mathcal{N}$-VAE	(0)	$144.51 \pm .37$	$114.66 \pm .08$	$100.16 \pm .08$	$97.64 \pm .04$
$\mathcal{P}$-VAE (Wrapped)	0.1	$143.87 \pm .47$	$115.49 \pm .27$	$100.22 \pm .06$	$97.22 \pm .05$
	0.2	$144.22 \pm .54$	$115.29 \pm .28$	$100.06 \pm .13$	$97.15 \pm .04$
	0.7	$143.82 \pm .56$	$115.08 \pm .31$	$100.19 \pm .11$	$97.49 \pm .04$
	1.4	$143.97 \pm .61$	$114.73 \pm .16$	$100.69 \pm .13$	$97.99 \pm .06$
$\mathcal{P}$-VAE (Riemannian)	0.1	$143.67 \pm .59$	$115.18 \pm .20$	$99.90 \pm .14$	$\mathbf{96.97 \pm .05}$
	0.2	$143.77 \pm .42$	$114.69 \pm .25$	$\mathbf{99.68 \pm .15}$	$97.43 \pm .12$
	0.7	$143.05 \pm .41$	$\mathbf{114.13 \pm .21}$	$101.16 \pm .18$	*
	1.4	$\mathbf{142.46 \pm .40}$	$115.48 \pm .32$	*	*

MNIST: Handwritten digits

	c	Dimensionality 2	5	10	20
$\mathcal{N}$-VAE	(0)	$144.51 \pm .37$	$114.66 \pm .08$	$100.16 \pm .08$	$97.64 \pm .04$
$\mathcal{P}$-VAE (Wrapped)	0.1	$143.87 \pm .47$	$115.49 \pm .27$	$100.22 \pm .06$	$97.22 \pm .05$
	0.2	$144.22 \pm .54$	$115.29 \pm .28$	$100.06 \pm .13$	$97.15 \pm .04$
	0.7	$143.82 \pm .56$	$115.08 \pm .31$	$100.19 \pm .11$	$97.49 \pm .04$
	1.4	$143.97 \pm .61$	$114.73 \pm .16$	$100.69 \pm .13$	$97.99 \pm .06$
$\mathcal{P}$-VAE (Riemannian)	0.1	$143.67 \pm .59$	$115.18 \pm .20$	$99.90 \pm .14$	$\mathbf{96.97 \pm .05}$
	0.2	$143.77 \pm .42$	$114.69 \pm .25$	$\mathbf{99.68 \pm .15}$	$97.43 \pm .12$
	0.7	$143.05 \pm .41$	$\mathbf{114.13 \pm .21}$	$101.16 \pm .18$	*
	1.4	$\mathbf{142.46 \pm .40}$	$115.48 \pm .32$	*	*

Digits	0	1	2	3	4	5	6	7	8	9	Avg
$\mathcal{N}$-VAE	89	97	81	75	59	43	89	**78**	68	**57**	73.6
$\mathcal{P}^{1.4}$-VAE	**94**	97	**82**	**79**	**69**	**47**	**90**	77	68	53	**75.6**


0
1
2
3
4
5
6
7
8
9

Conclusion

- Deep generative model for hierarchical data
- Make use of Hyperbolic geometry
- Currently experimenting with biological data

Thank you! Questions?