Is Simple Better?: Revisiting Simple Generative Models for Unsupervised Clustering

Jhosimar Arias Figueroa   Adín Ramírez Rivera

jhosimar.figueroa@students.ic.unicamp.br, adin@ic.unicamp.br

Institute of Computing, University of Campinas, Brazil

Outline

Problem

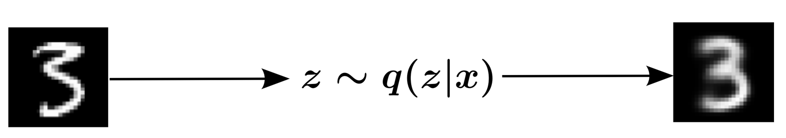
- Hierarchical stochastic variables are difficult to train.
- Auxiliary clustering algorithms are commonly required.
- Complex priors like mixture of Gaussians.

Approach

- Avoid the problem of hierarchical stochastic variables by using deterministic ones.
- Use of Gumbel-softmax to model our clusters.
- Simple priors like normal gaussian.

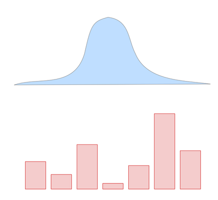
Background

Variational Autoencoder

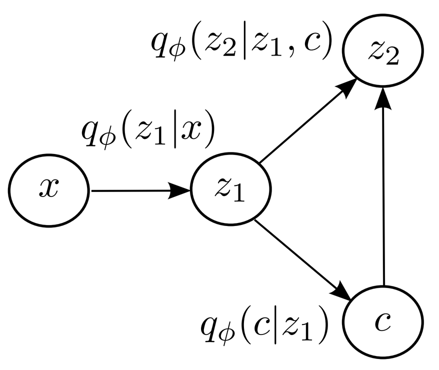


$z \sim \mathcal{N}(0,1)$

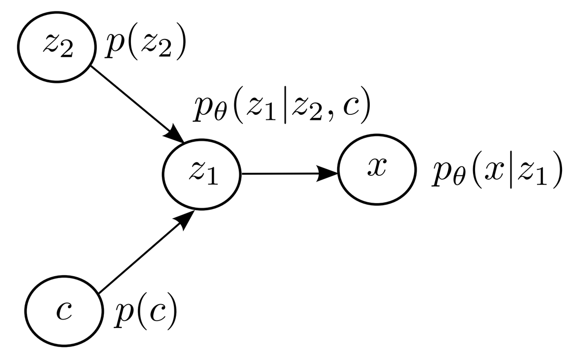
$z \sim Gumbel(0,1)$



Stacked Generative Semi-Supervised Model (M1+M2)



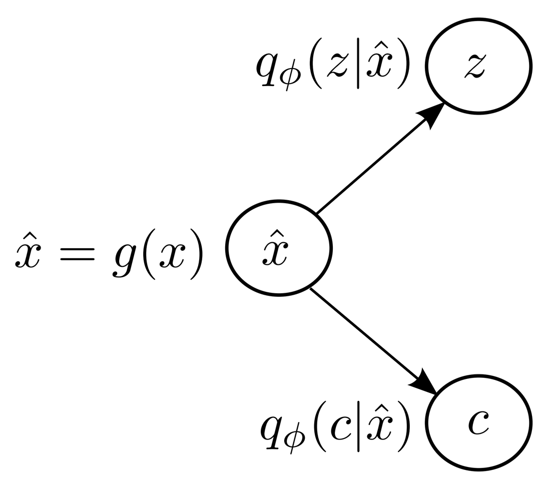
Inference Model



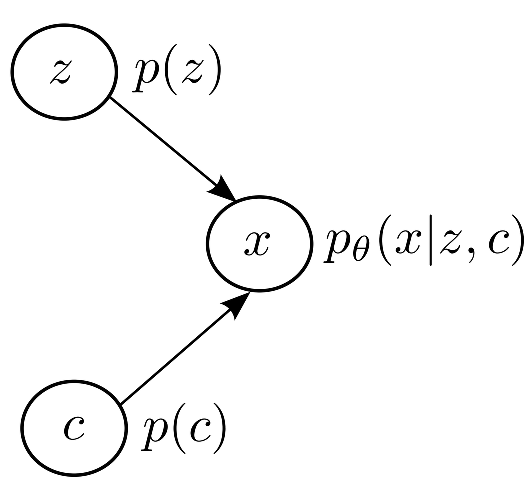
Generative Model

Deep Generative Models for Clustering

Probabilistic Graphical Model



Inference Model



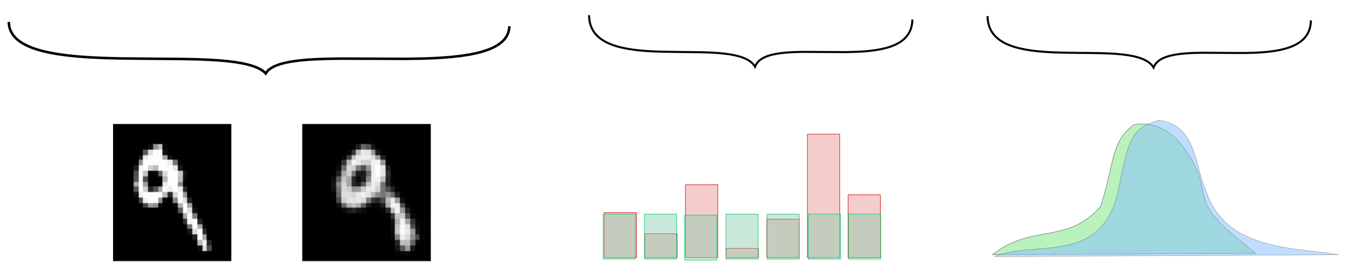
Generative Model

$$q_\phi(c|\hat{x}) = Cat(c|\pi_\phi(\hat{x})),$$
$$q_\phi(z|\hat{x}) = \mathcal{N}\left(z|\mu_\phi(\hat{x}), \text{diag}\left(\sigma^2_\phi(\hat{x})\right)\right)$$

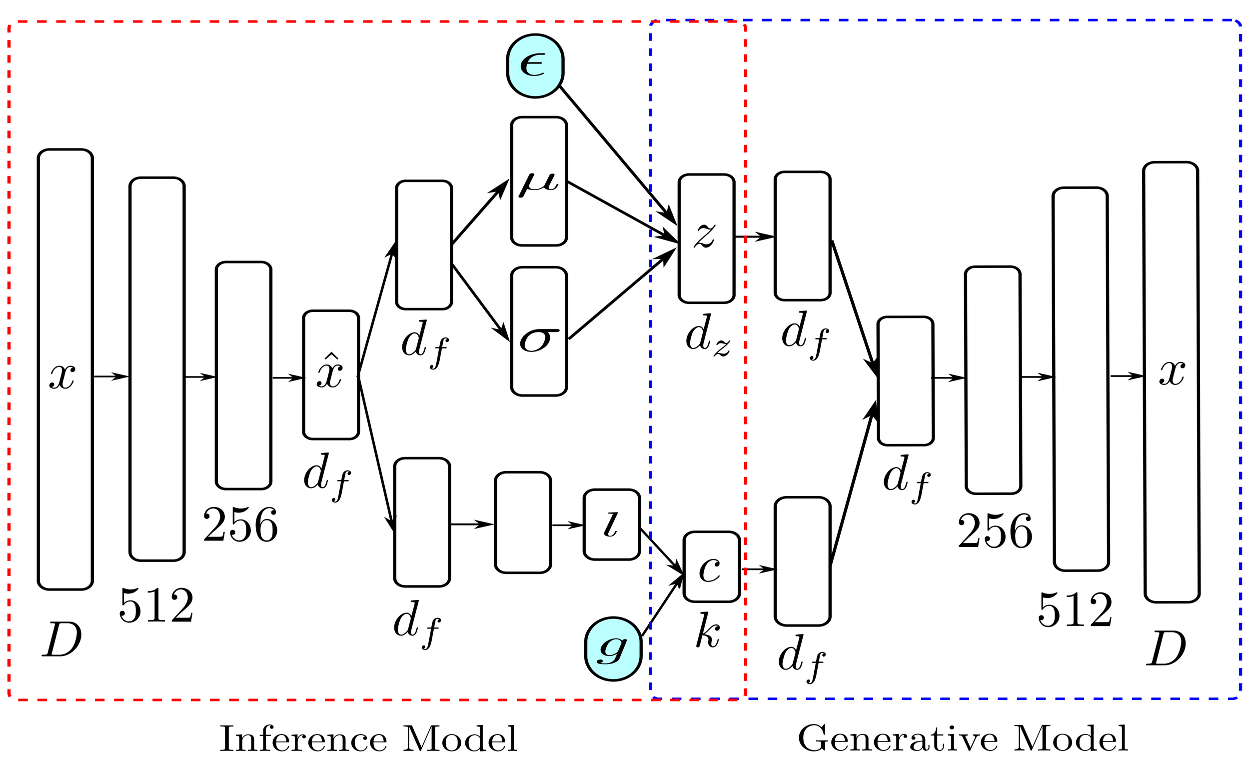
$$p(c) = Cat(c|\pi),$$
$$p(z) = \mathcal{N}(z|0, I),$$
$$p_\theta(x|z,c) = f(x; z, c, \theta)$$

Evidence Lower Bound (ELBO)

$$\log p_\phi(x) \geq \mathbb{E}_{c,z\sim q_\phi(c,z|\hat{x})}\left[\log \frac{p_\theta(x,c,z)}{q_\phi(c,z|\hat{x})}\right]$$
$$\geq \mathbb{E}_{c,z\sim q_\phi(c,z|\hat{x})}\left[\log p_\theta(x|c,z)\right] - KL\left(q_\phi(c|\hat{x})\,||\,p(c)\right) - KL\left(q_\phi(z|\hat{x})\,||\,p(z)\right)$$



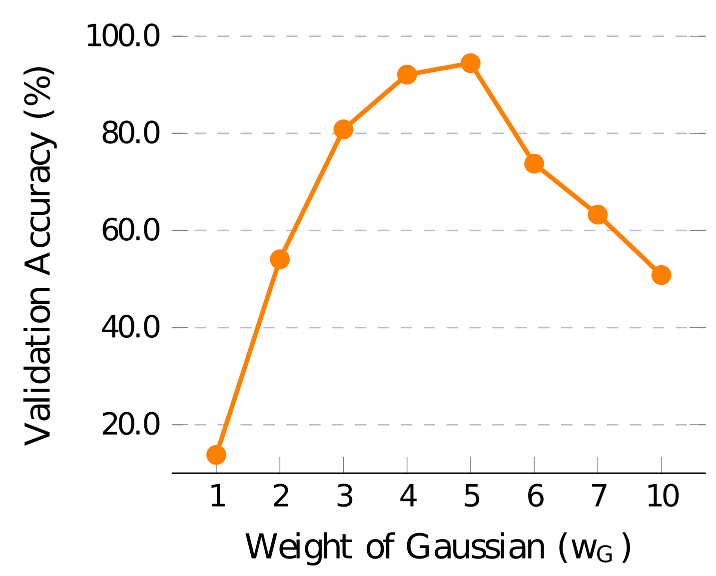
Network Architecture



Quantitative Evaluation

MNIST test error-rate (%) for kNN.

Method	k=3	k=5	k=10
VAE	18.43	15.69	14.19
DLGMM	9.14	8.38	8.42
VaDE	2.20	2.14	2.22
Proposed	3.46	3.30	3.44

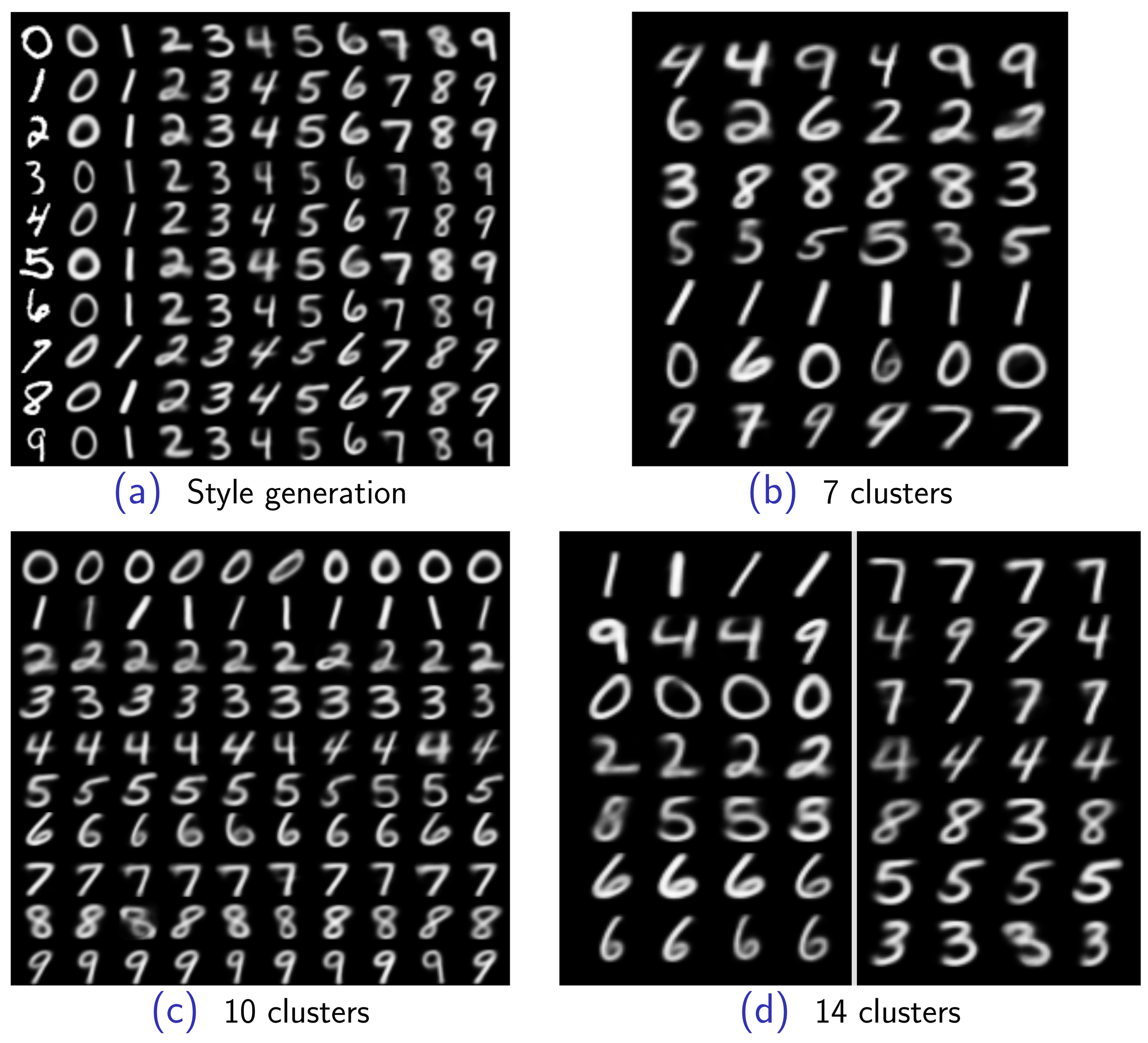


Hyperparameter setup on Gaussian weight regularizer for MNIST

Clustering performance, ACC (%) and NMI (%), on all datasets

Method	MNIST		USPS		REUTERS-10K	
	ACC	NMI	ACC	NMI	ACC	NMI
k-means	53.24	-	66.82	-	51.62	-
GMM	53.73	-	-	-	54.72	-
AE+k-means	81.82	74.73	69.31	66.20	70.52	39.79
AE+GMM	82.18	-	-	-	70.13	-
GMVAE	82.31 (± 4)	-	-	-	-	-
DCN	83.00	81.00	-	-	-	-
DEC	86.55	83.72	74.08	75.29	73.68	49.76
IDEC	88.06	86.72	76.05	78.46	75.64	49.81
VaDE	94.46	-	-	-	79.83	-
Proposed	85.75 (± 8)	82.13 (± 5)	72.58 (± 3)	65.48 (± 1)	76.74 (± 6)	52.42 (± 5)

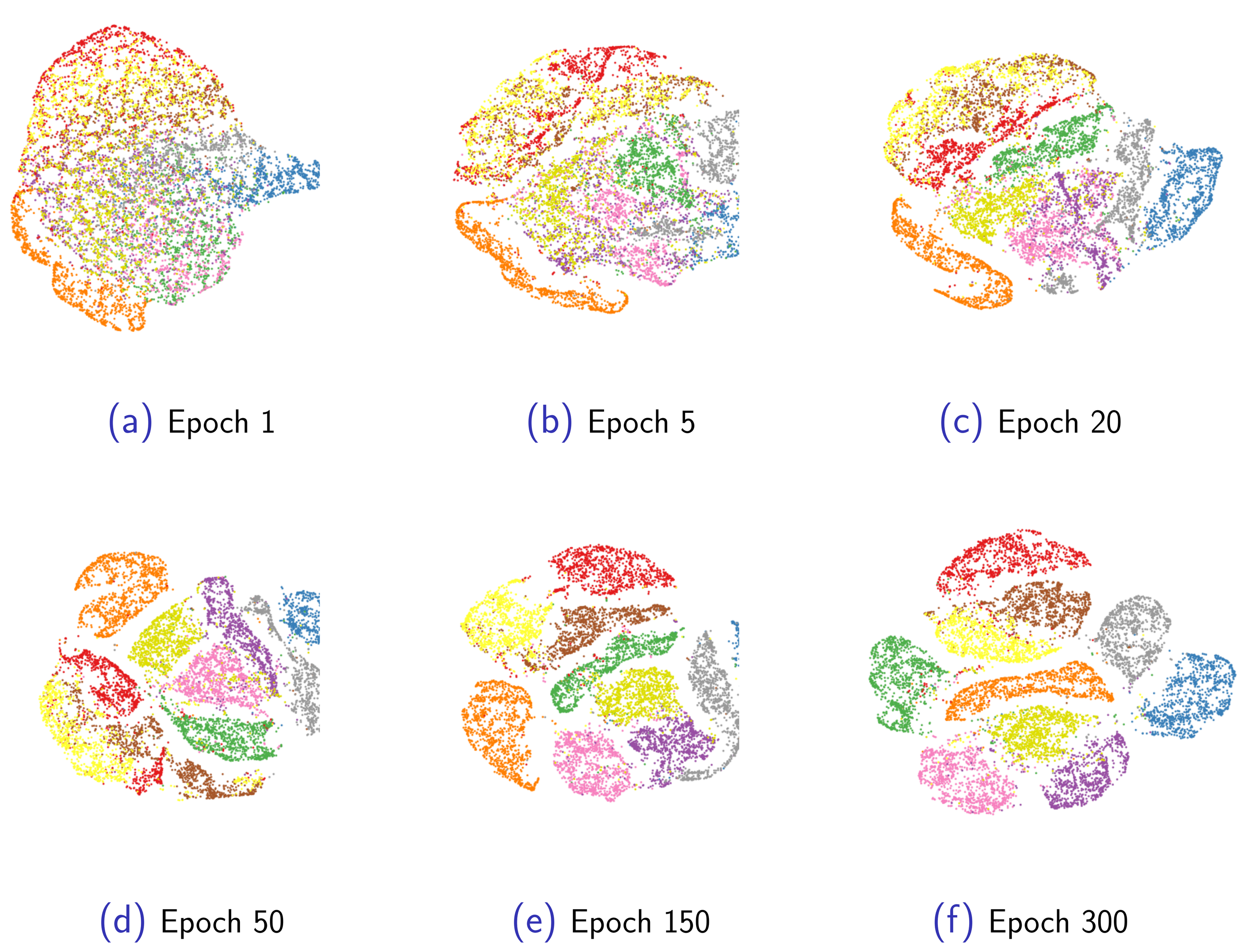
Qualitative Evaluation



(a) Style generation (b) 7 clusters (c) 10 clusters (d) 14 clusters

Generated images of our proposed model trained on MNIST.

Clustering Visualization



(a) Epoch 1 (b) Epoch 5 (c) Epoch 20 (d) Epoch 50 (e) Epoch 150 (f) Epoch 300

Second workshop on Bayesian Deep Learning (NIPS 2017), Long Beach, CA, USA. https://gitlab.com/mipl/simple-vae-clustering