An Exploration of Normalizing Flows for Outlier Generation

Till Bungert

July 12, 2021

Problem Statement

- Outliers and outlier detection are important topics
- Trend to use generative models to improve classifiers
- INNs/Normalizing Flows not yet used

Outline

Methods

- Normalizing Flows
- Archetypal Analysis
- Discriminator and Classifier Training

Experiments

- Toy Data
- Qualitative Comparison
- Archetypal Analysis
- Analysis of the Latent Space
- Discriminator Performance
- Conclusion
- Future Work

Theoretical Background

Normalizing Flows

$$p_X(x_i) = p_Z(\mathrm{INN}(x_i))|J_X|^{-1} \tag{1}$$

$$\mathrm{NLL} = \frac{1}{N}\sum_i \|\mathrm{INN}(x_i)\|^2 - \log J_X \tag{2}$$

Theoretical Background

Normalizing Flows

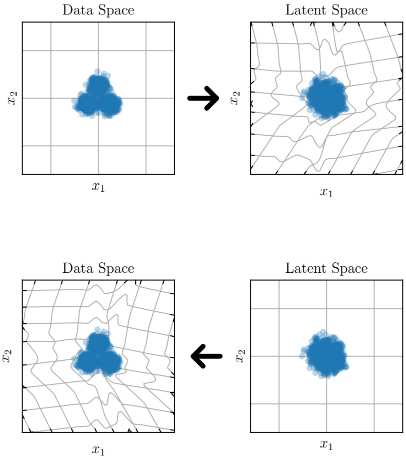


Figure: Mapping between data and latent space

Theoretical Background

Extremal Sampling

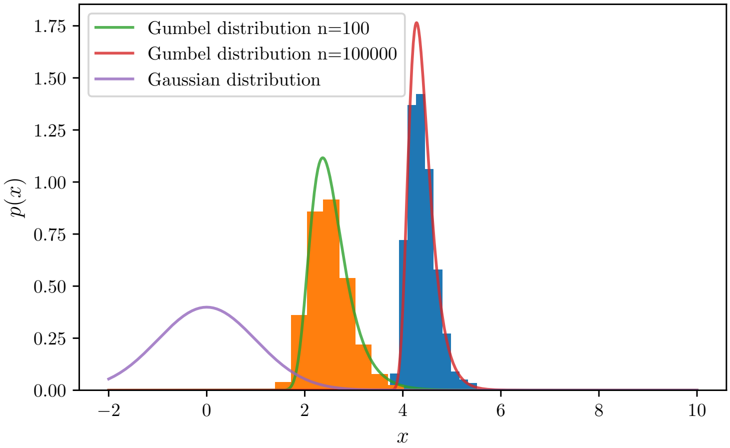


Figure: Gumbel distribution

Theoretical Background

Extremal Sampling

$$\|\mathsf{X}\|_2 \sim \chi_\mathsf{d} \tag{3}$$

$$\begin{aligned}\mathbb{E}\|\mathsf{X}\|_2 &= \sqrt{2}\frac{\Gamma(\frac{\mathsf{d}+1}{2})}{\Gamma(\frac{\mathsf{d}}{2})} \\ \mathrm{Var}\|\mathsf{X}\|_2 &= \mathsf{d} - (\mathbb{E}\|\mathsf{X}\|_2)^2\end{aligned} \tag{4}$$

Theoretical Background

Extremal Sampling

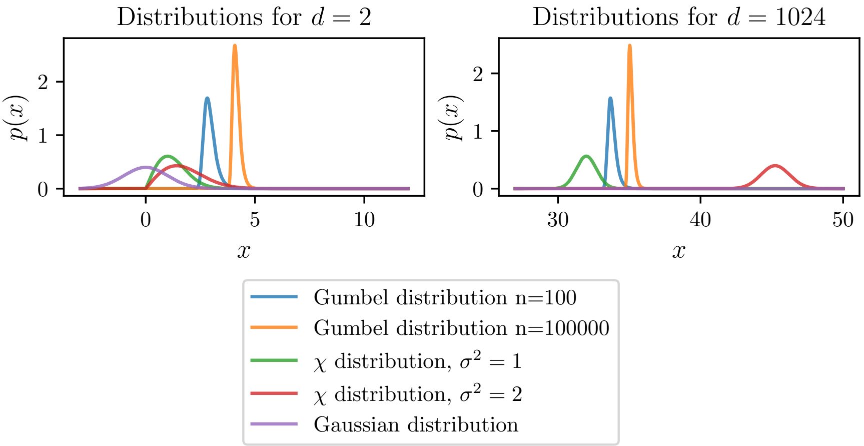


Figure: Gumbel distribution

Theoretical Background

Archetypal Analysis

$$x_i = \sum_j a_{ij} z_j \tag{5}$$

$$z_j = \sum_i b_{ji} x_i \tag{6}$$

$$\begin{aligned} a_{ij} \geq 0 \text{ and } \sum_{j=1}^{m} a_{ij} = 1 \\ b_{ji} \geq 0 \text{ and } \sum_{i=1}^{n} b_{ji} = 1 \end{aligned} \tag{7}$$

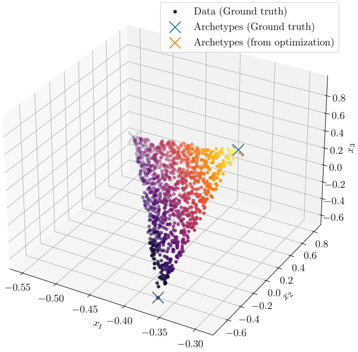
Theoretical Background

Archetypal Analysis

$$
\begin{aligned}
&\min_{a,b} \sum_i \|x_i - \sum_j a_{ij} z_j\|^2 \\
&= \min_{a,b} \sum_l \|x_l - \sum_j a_{lj} \sum_i b_{ji} x_i\|^2
\end{aligned} \tag{8}
$$

Theoretical Background

Archetypal Analysis



Theoretical Background

DeepAA

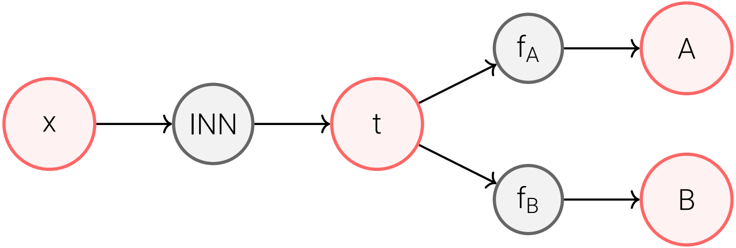


Figure: Forward pass of the normalizing flow with additional layers for the mapping to archetype coefficients.

Theoretical Background

Sampling

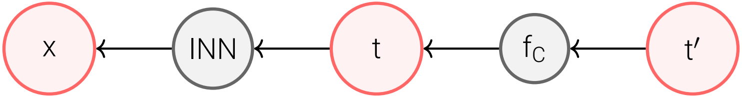


Figure: Inverse computation from lower dimensional sample created from archetype coefficients to data space.

$$A \sim \mathrm{Dirichlet}(c_1, \ldots, c_k)$$
$$t' = Az_{\text{fixed}}$$

Theoretical Background

Nullspace Sampling

$$A = f_A(t) = \mathrm{softmax}(M_A t)$$
$$t = f_C(Az_{\mathrm{fixed}}) = M_C A z_{\mathrm{fixed}}$$

Theoretical Background

Nullspace Sampling

$$A = M_A t$$

$$t = M_A^+ A + [I - M_A^+ M_A] w$$

Theoretical Background

Nullspace Sampling

$$t = f_C(Az_{\text{fixed}}) + [I - M_A^+ M_A]w$$

Theoretical Background

Discriminator Training

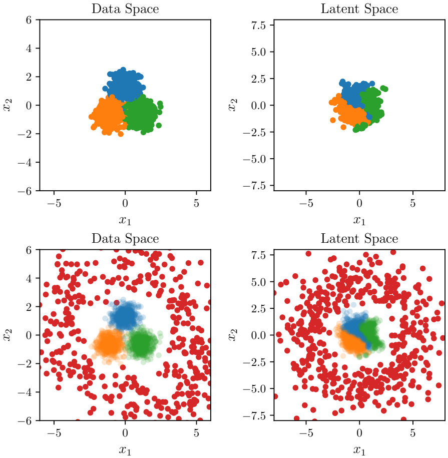
$$\mathbb{E}_{\hat{x},\hat{y}\sim p_{in}} \log D(x, y) + \mathbb{E}_{z\sim p_{out}} \log(1 - D(z, y)) \tag{9}$$

Theoretical Background

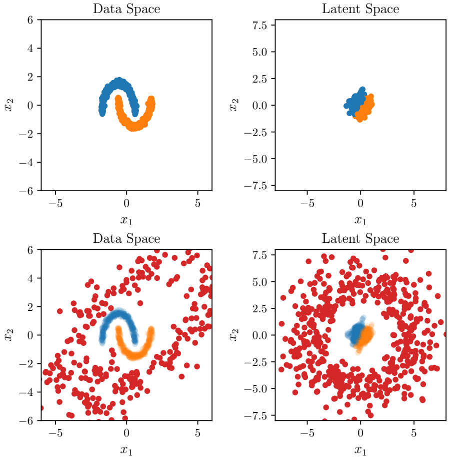
Classifier Training

$$\mathbb{E}_{\hat{x},\hat{y}\sim p_{in}} \log p(y=\hat{y}|\hat{x}) - \beta\mathbb{E}_{z\sim p_{out}} \mathrm{KL}(\mathcal{U}(y)|p(y|z)) \tag{10}$$

Toy Data

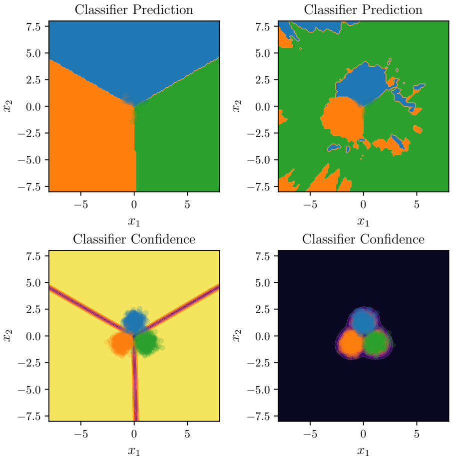


Toy Data



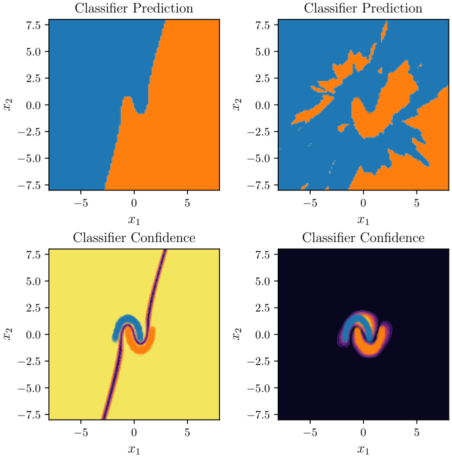
Toy Data

Classifier



Toy Data

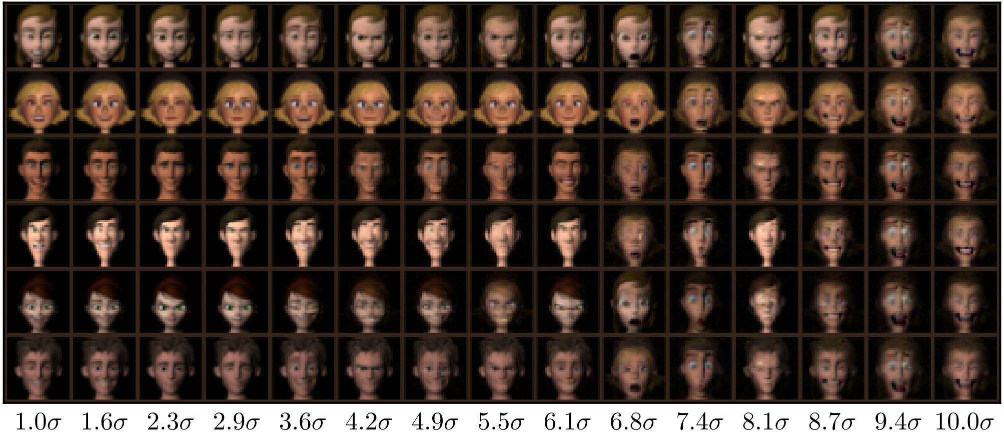
Classifier



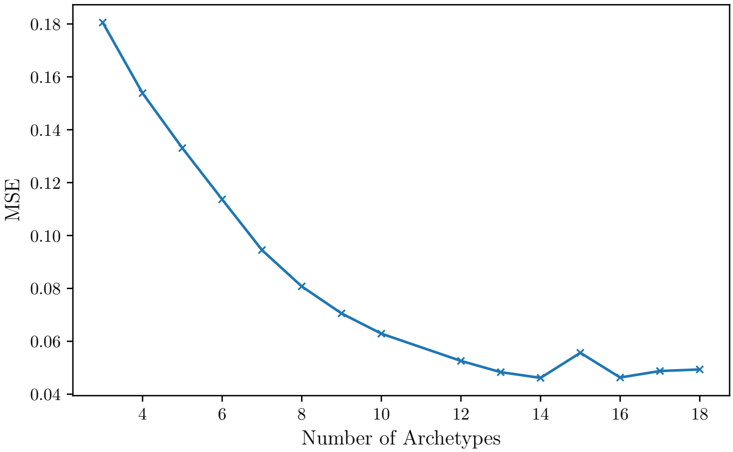
EMNIST

0.5σ 0.6σ 0.8σ 0.9σ 1.1σ 1.2σ 1.4σ 1.5σ 1.6σ 1.8σ 1.9σ 2.1σ 2.2σ 2.4σ 2.5σ

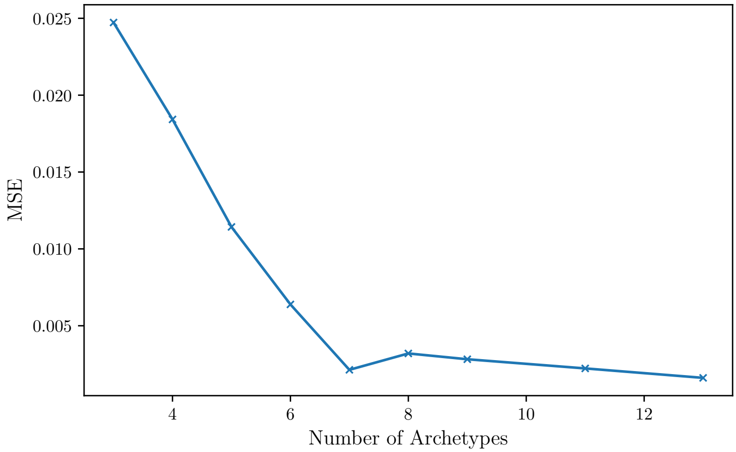
FERG


1.0σ 1.6σ 2.3σ 2.9σ 3.6σ 4.2σ 4.9σ 5.5σ 6.1σ 6.8σ 7.4σ 8.1σ 8.7σ 9.4σ 10.0σ

Number of Archetypes for Digits

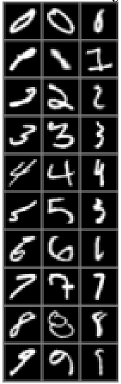


Number of Archetypes for People



Archetypes (Digits)

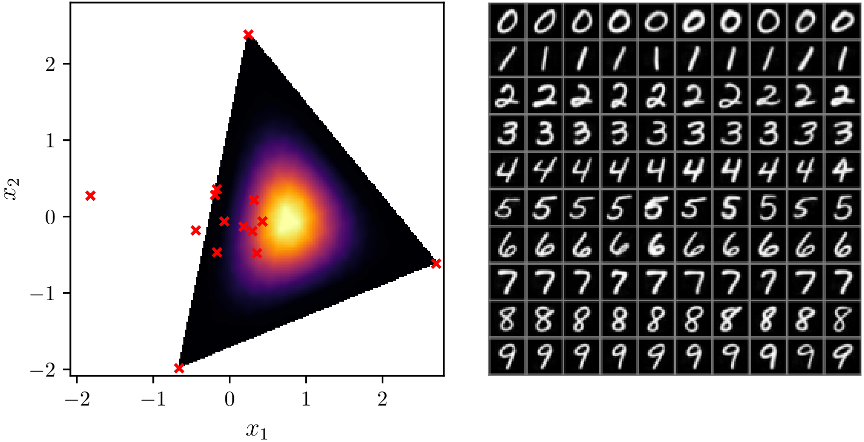
Model with
$k = 3$ Archetypes



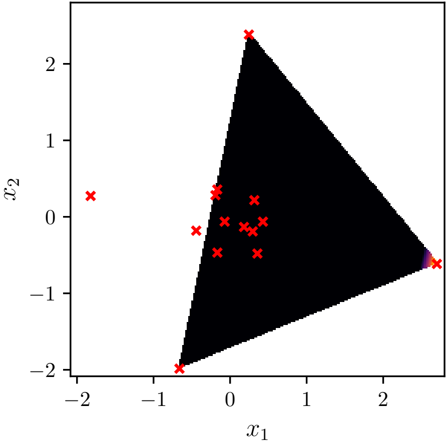
Model with
$k = 14$ Archetypes

Archetypes (People)

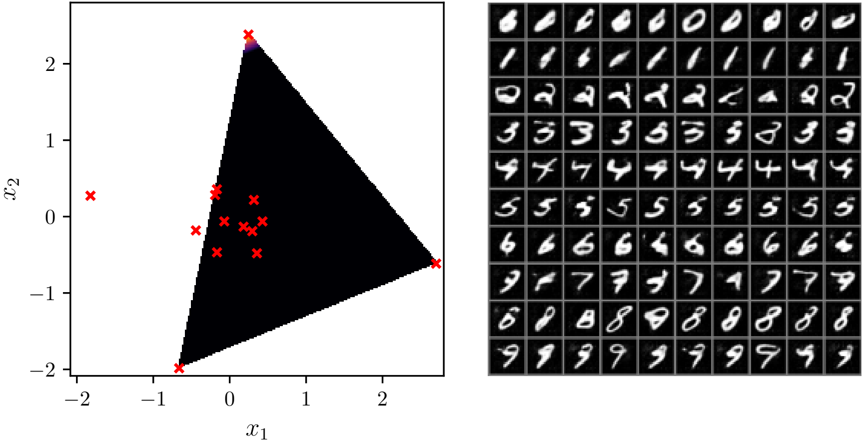
Sampling Inliers (Digits)



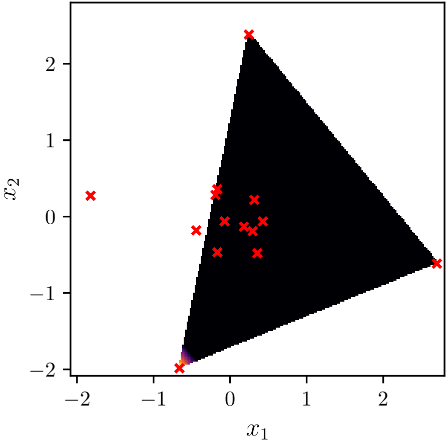
Sampling Outliers (Digits)

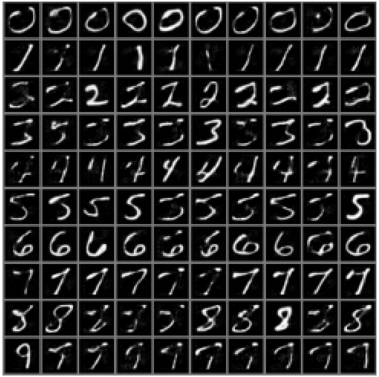


Sampling Outliers (Digits)

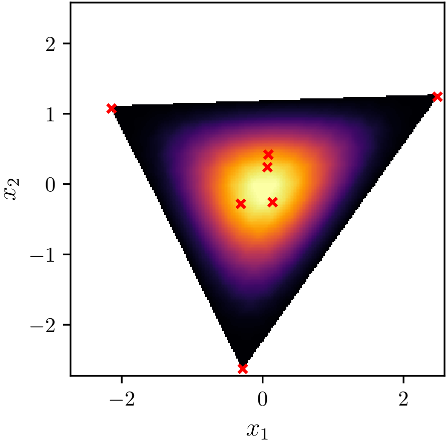


Sampling Outliers (Digits)

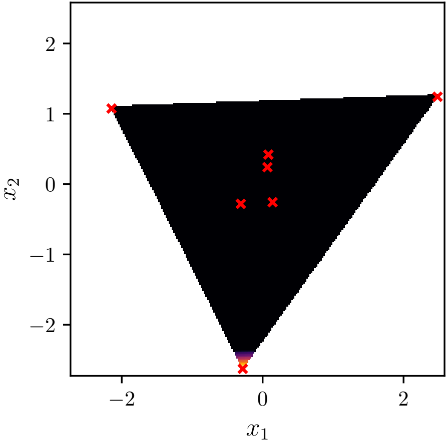




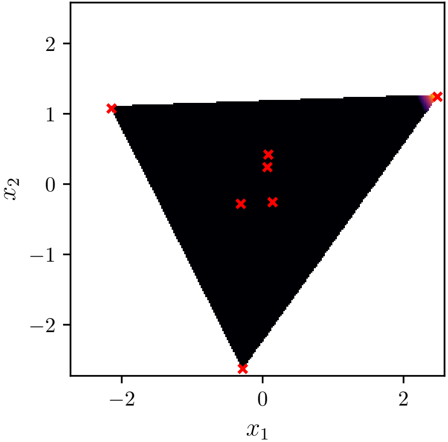
Sampling Inliers (People)



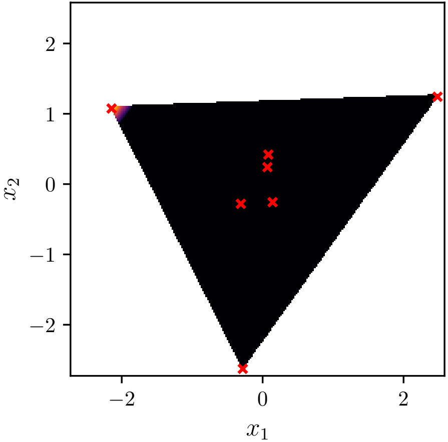
Sampling Outliers (People)



Sampling Outliers (People)



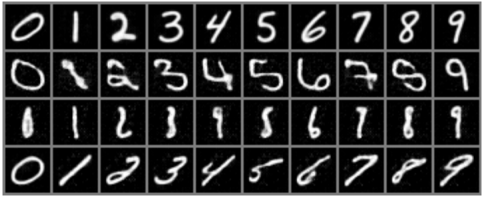
Sampling Outliers (People)



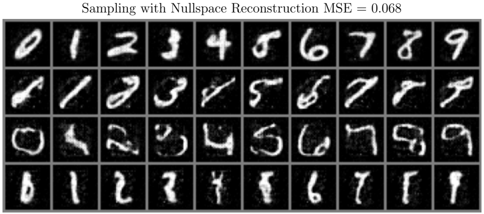
Sampling with Nullspace

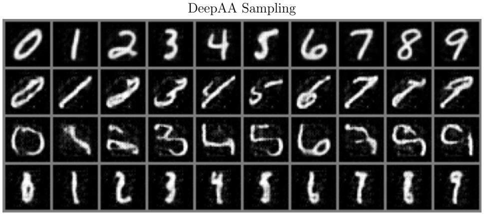
Sampling with Nullspace Reconstruction MSE = 0.022

DeepAA Sampling

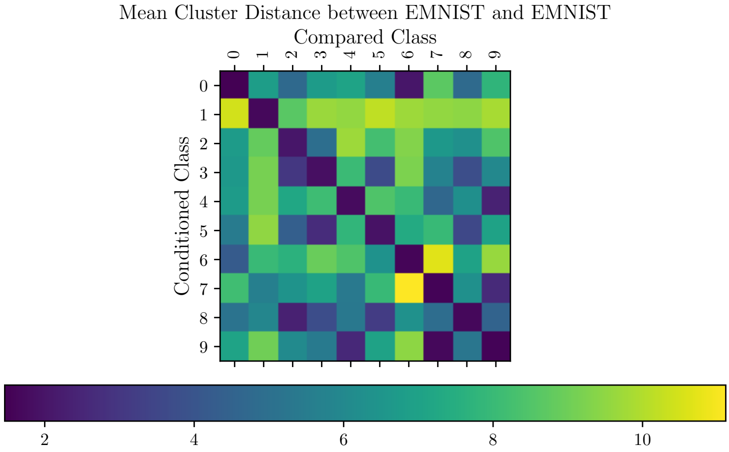


Sampling with Nullspace (biased)

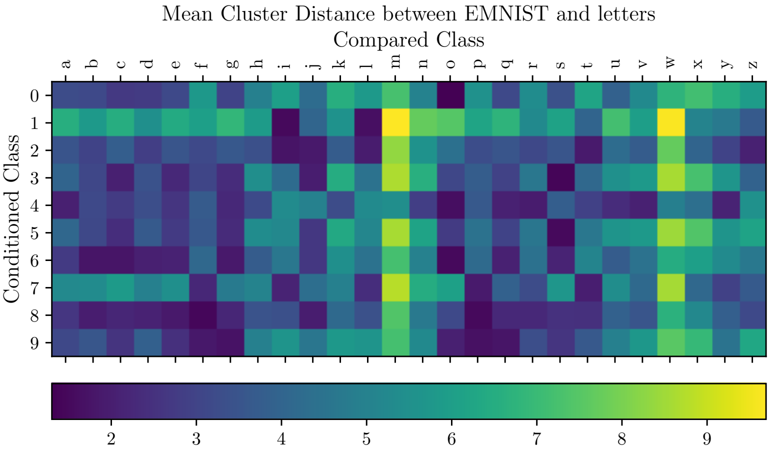




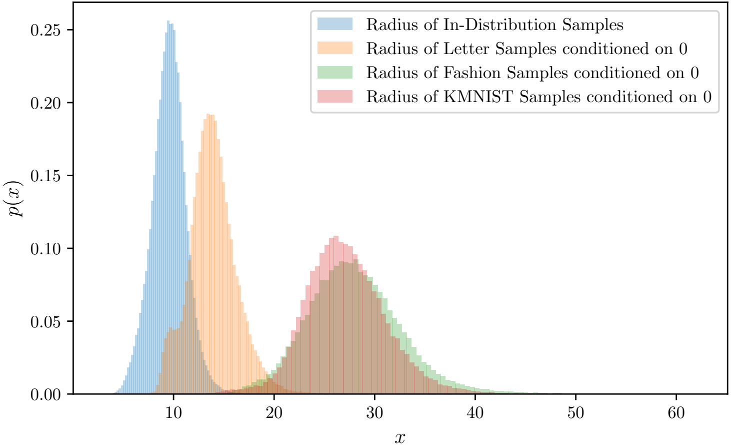
Cluster Distance between Classes



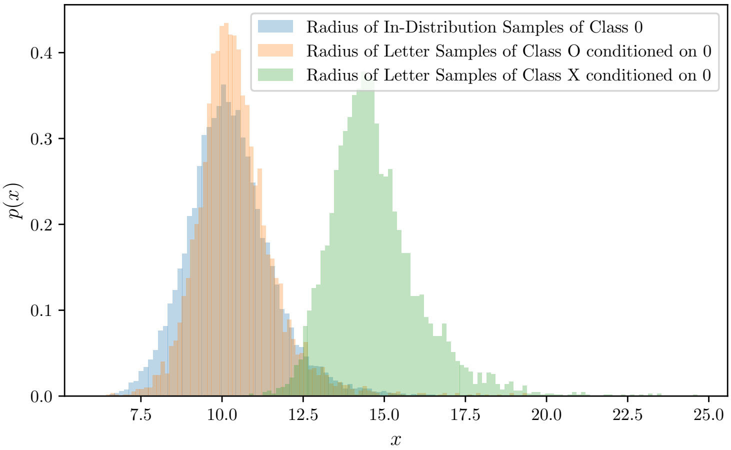
Cluster Distance between Classes



Distribution of Latent Space Radii



Distribution of Latent Space Radii



Discriminator Performance

outliers	letters		fashion		kmnist	
metric	AUC	AP	AUC	AP	AUC	AP
DeepAA, p = 0.1	0.64	0.33	0.88	0.77	0.67	0.47
DeepAA, p = 0.3	0.57	0.27	0.52	0.37	0.46	0.34
DeepAA, p = 0.5	0.48	0.22	0.38	0.31	0.42	0.33
DeepAA, p = 0.6	0.47	0.22	0.39	0.32	0.37	0.31
DeepAA, p = 0.8	0.43	0.20	0.48	0.35	0.39	0.32
DeepAA, p = 1.2	0.45	0.21	0.74	0.51	0.63	0.42
INN, Gumbel, p = 2.5	0.68	0.35	0.45	0.34	0.47	0.35
INN, Normal, p = 1.5	0.58	0.27	0.12	0.25	0.38	0.32
INN, Normal, p = 2.0	0.65	0.32	0.20	0.27	0.48	0.35
INN, Normal, p = 2.5	0.69	0.39	0.80	0.61	0.72	0.50
INN, Normal, p = 3.0	0.52	0.24	0.42	0.33	0.15	0.25
fashion	0.73	0.44	1.00	1.00	0.97	0.95
letters	0.81	0.56	0.50	0.36	0.48	0.35

Classifier Performance

outliers	letters			fashion			kmnist		
metric	AUC	AP	ACC	AUC	AP	ACC	AUC	AP	ACC
DeepAA, p = 0.1	0.89	0.64	1.00	1.00	1.00	1.00	0.99	0.98	1.00
DeepAA, p = 0.3	0.90	0.66	1.00	1.00	1.00	1.00	0.99	0.99	1.00
DeepAA, p = 0.6	0.89	0.63	1.00	1.00	1.00	1.00	0.98	0.97	1.00
DeepAA, p = 0.8	0.90	0.68	1.00	1.00	1.00	1.00	0.98	0.98	1.00
DeepAA, p = 1.0	0.89	0.70	0.99	1.00	1.00	0.99	0.97	0.96	0.99
DeepAA, p = 1.2	0.89	0.68	1.00	1.00	1.00	1.00	0.98	0.96	1.00
INN, Gumbel, p = 2.5	0.90	0.72	1.00	1.00	1.00	1.00	0.99	0.99	1.00
fashion	0.89	0.67	1.00	1.00	1.00	1.00	0.99	0.99	1.00
letters	0.98	0.95	1.00	1.00	1.00	1.00	0.99	0.99	1.00
none	0.89	0.68	0.99	1.00	1.00	0.99	0.97	0.95	0.99

Conclusion

- INNs/Normalizing Flows well suited for outlier generation
- Archetypal Analysis interesting approach for interpretable latent space

Future Work

- Use bigger models to test on ImageNet
- Work unconditionally
- Goal: Completely unsupervised model of outliers