Deep Clustering based on a Mixture of Autoencoders

Advisor: Jia-Ling, Koh

Speaker: Zi-Xin Chen

Source: IEEE'19

Date: 2022/09/13

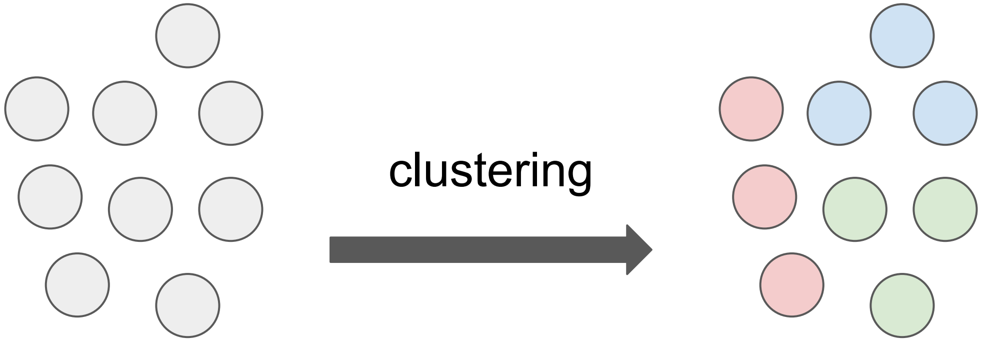
Outline

- Introduction
- Method
- Experiment
- Conclusion

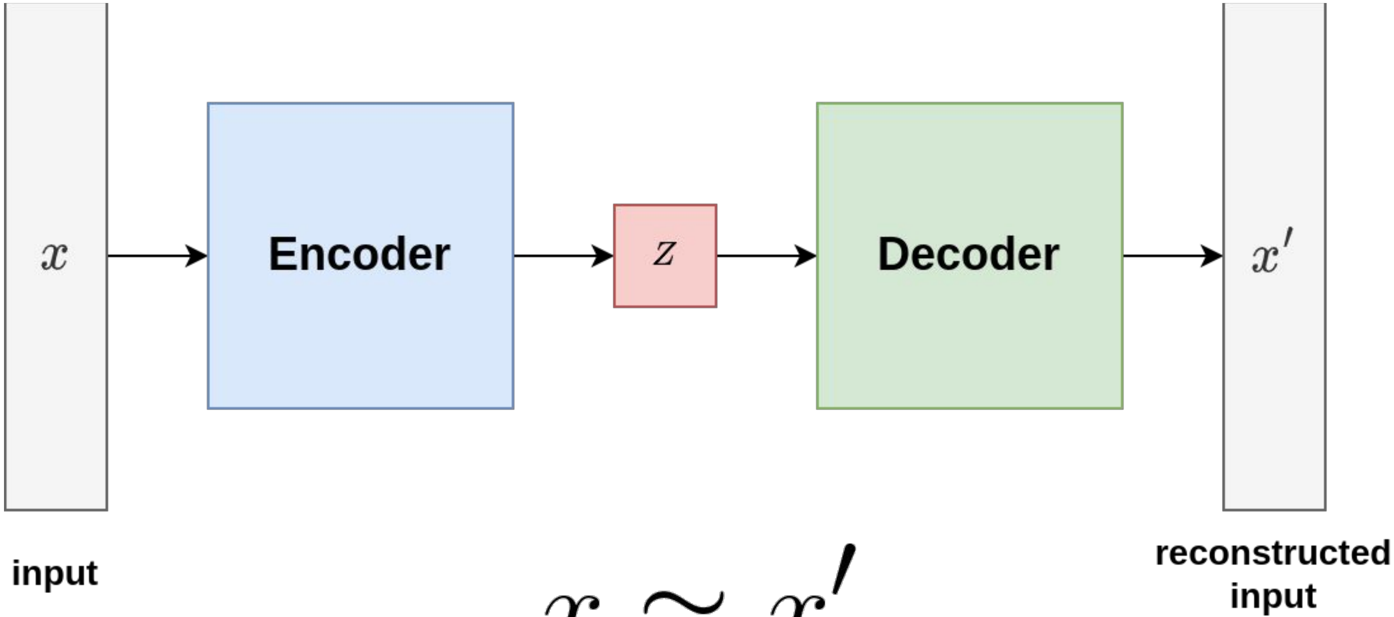
Introduction

Clustering

- During clustering process, similar items are grouped together and distinct samples are separated

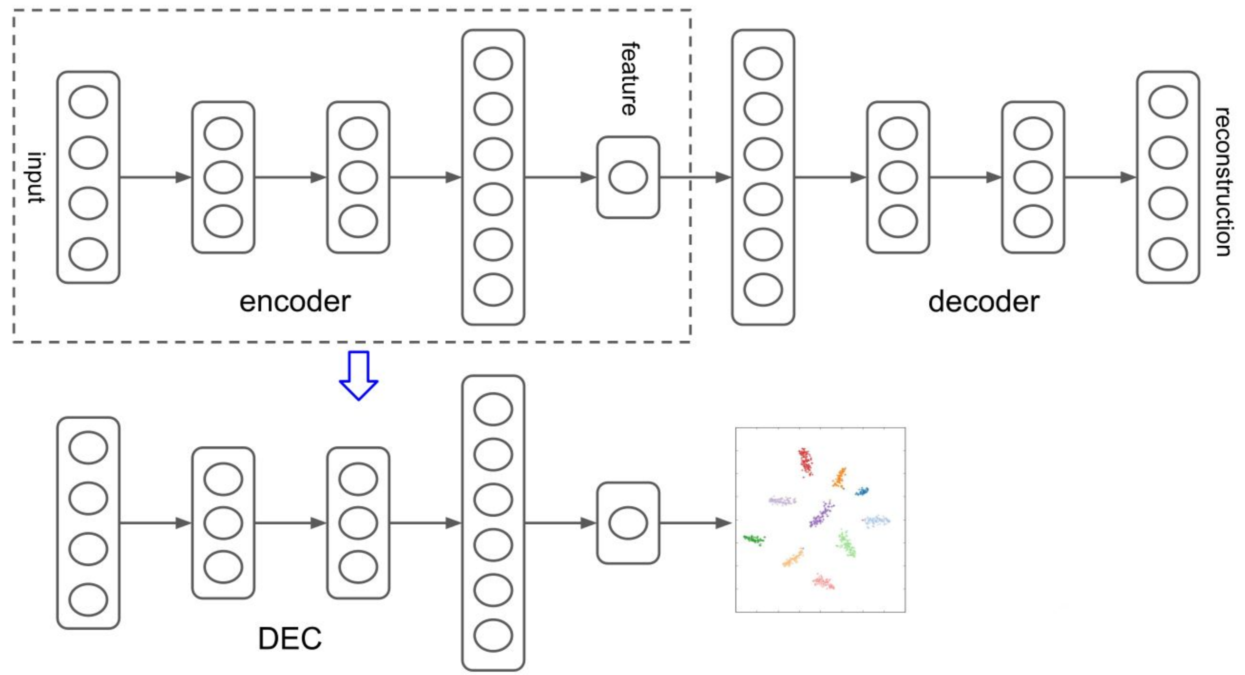


Autoencoder

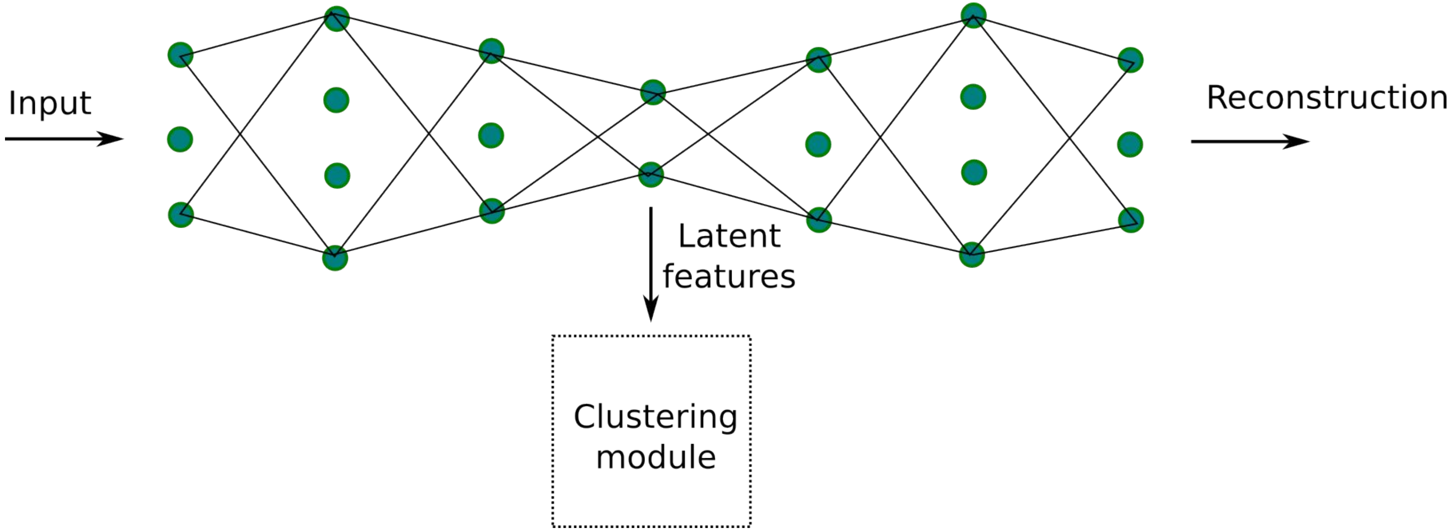


$$x \approx x'$$

DEC - Deep Embedded Clustering (2016)



DCN - Deep Clustering Network (2017)



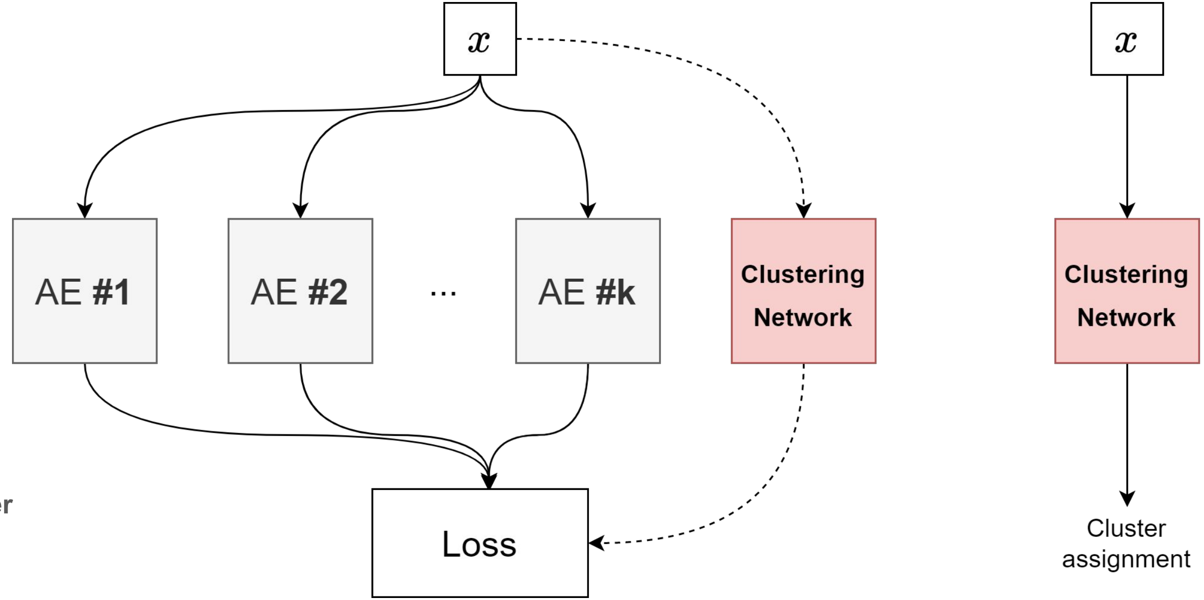
Loss = reconstruction error + clustering error

Problem

- Features are collapsed to a single point in the embedded space

Deep Autoencoder MIxture Clustering (DAMIC)

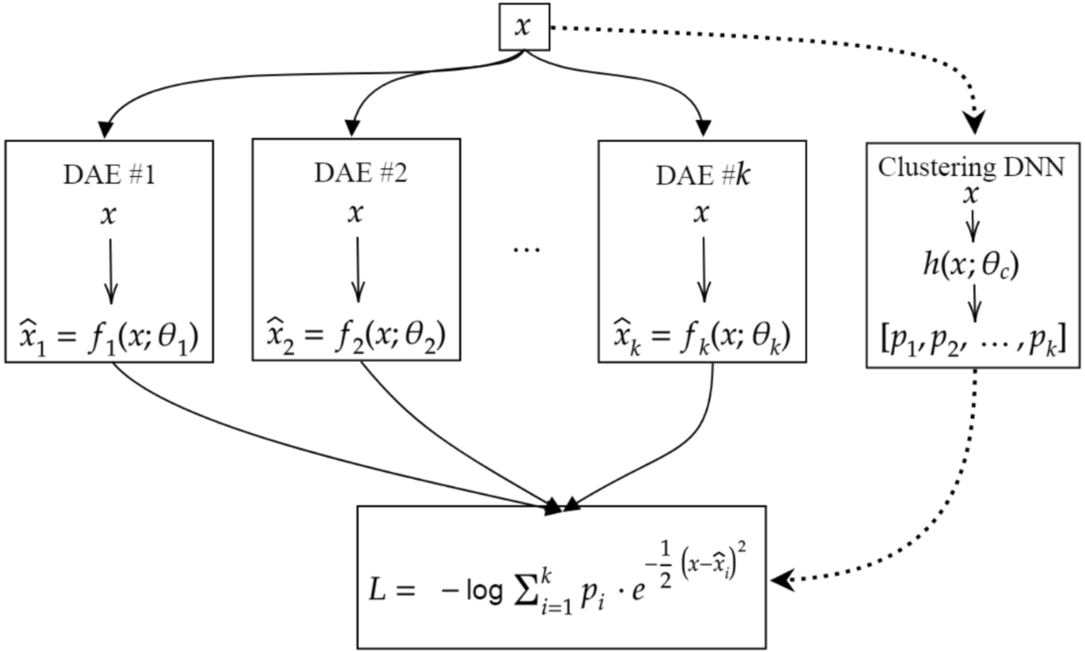
- Each cluster is represented by an autoencoder neural network

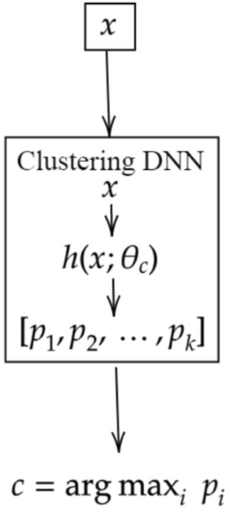


*AE: Autoencoder

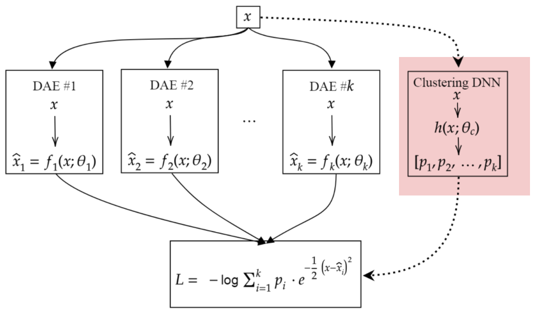
Method

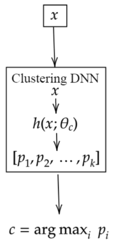
Architecture





Clustering Network



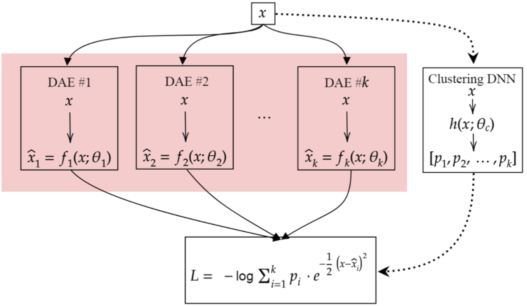


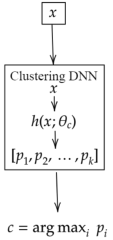
Soft cluster assignment:

Clustering network

$$p(c = i | x; \theta_c) = \frac{\exp(w_i \boxed{h(x)} + b_i)}{\sum_{j=1}^{k} \exp(w_j h(x) + b_j)}, \quad i = 1, \ldots, k$$

Autoencoder Network



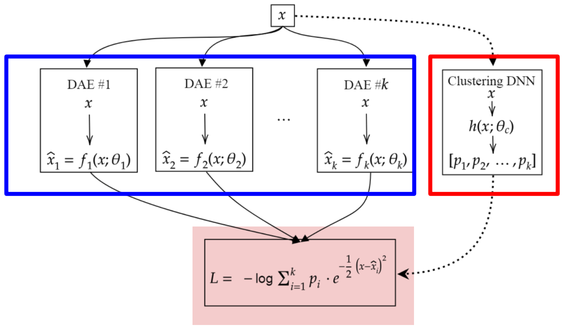


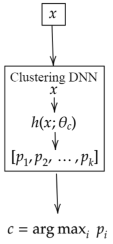
Reconstruction error:

$$d(x_t, f_i(x_t;\theta_i)) = \frac{1}{2}\|x_t - \boxed{f_i}(x_t;\theta_i)\|^2$$

Autoencoder of i-th cluster

Loss function





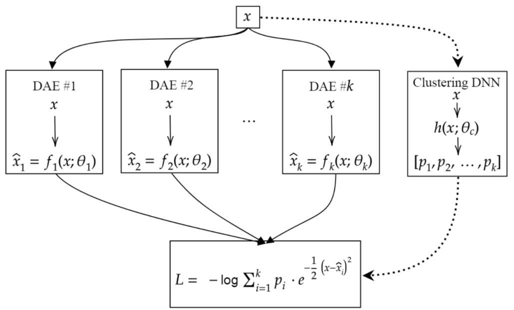
$$L(\theta_1, \dots, \theta_k, \theta_c)$$

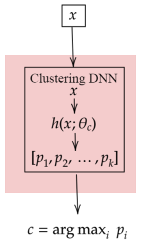
$$= -\sum_{t=1}^{n} \log \left(\sum_{i=1}^{k} p(c_t = i | x_t; \theta_c) \exp(-d(x_t, f_i(x_t; \theta_i))) \right)$$

Clustering network

Autoencoder Network

Cluster assignment

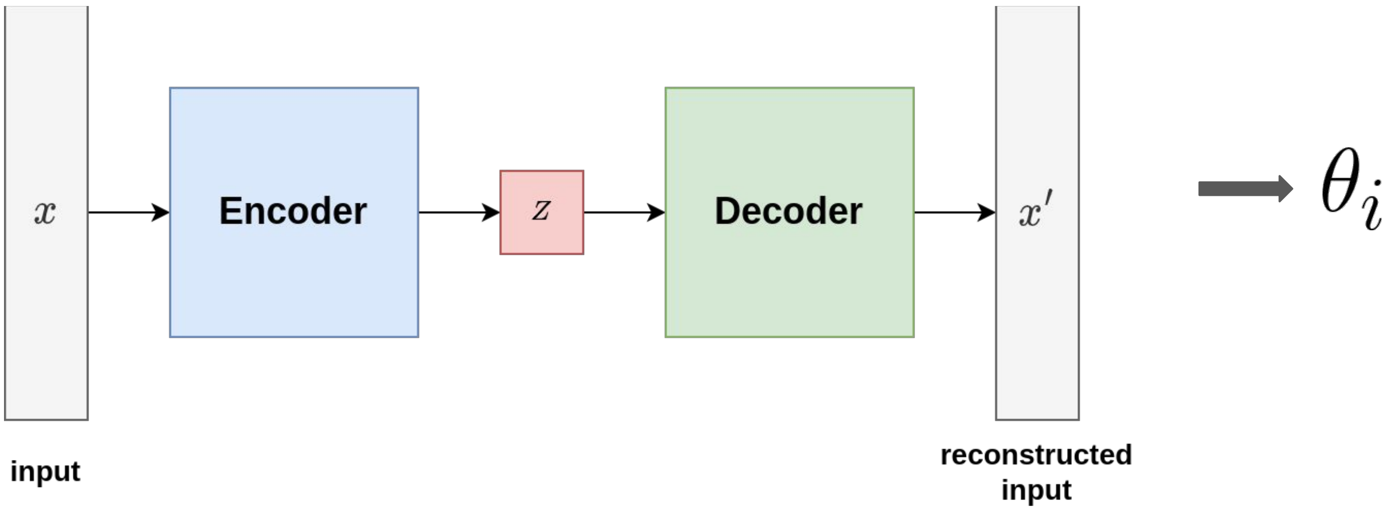




$$\hat{c} = \arg\max_{i=1}^{k} p(c = i | x; \theta_c) = \arg\max_{i=1}^{k} (w_i h(x) + b_i)$$

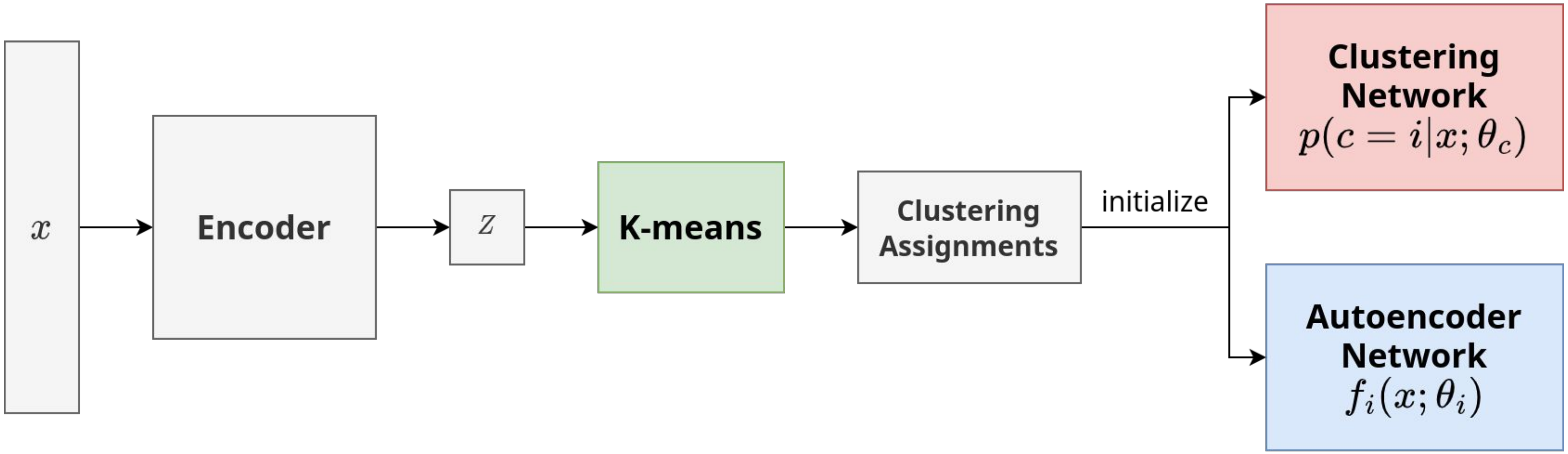
Pre-training

- Train a single autoencoder for the **entire dataset**



Pre-training

- Apply **k-means** algorithm in the embedded space
- Use the k-means clustering to **initialize** the network parameters



Experiment

Dataset

Image		
MNIST	70000 images	10 classes
Fashion	70000 images	10 classes

Text		
20NEWS	18846 documents (only use 2000 words)	20 groups
RCV1	365968 documents (only use 2000 words)	20 topics

Evaluation measures

- NMI
- ARI
- ACC

NMI

Entropy before clustering — Entropy after clustering

$$I(C,T) = \boxed{H(T)} - \boxed{H(T|C)}$$

$$\mathrm{NMI} = \frac{I(C,T)}{\sqrt{H(T) \cdot H(C)}}$$

ARI

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}$$

$X \backslash Y$	Y_1	Y_2	$\cdots$	Y_s	sums
X_1	n_{11}	n_{12}	$\cdots$	n_{1s}	a_1
X_2	n_{21}	n_{22}	$\cdots$	n_{2s}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_r	n_{r1}	n_{r2}	$\cdots$	n_{rs}	a_r
sums	b_1	b_2	$\cdots$	b_s	

$$\overbrace{ARI}^{\text{Adjusted Index}} = \frac{\overbrace{\sum_{ij}\binom{n_{ij}}{2}}^{\text{Index}} - \overbrace{[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}^{\text{Expected Index}}}{\underbrace{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}]}_{\text{Max Index}} - \underbrace{[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}_{\text{Expected Index}}}$$

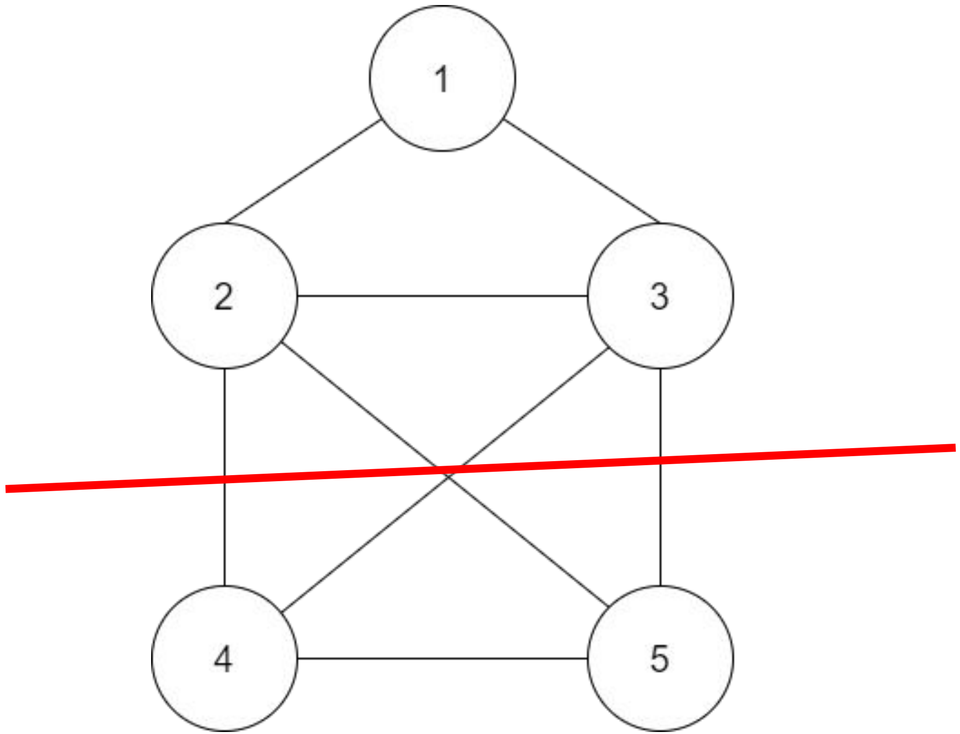
ACC

$$\mathrm{AC} = \frac{\delta(s_i, \mathrm{map}(r_i))}{N}$$

Baseline Methods

- K-means
- Spectral Clustering
- DAE + KM
- DCN
- DEC

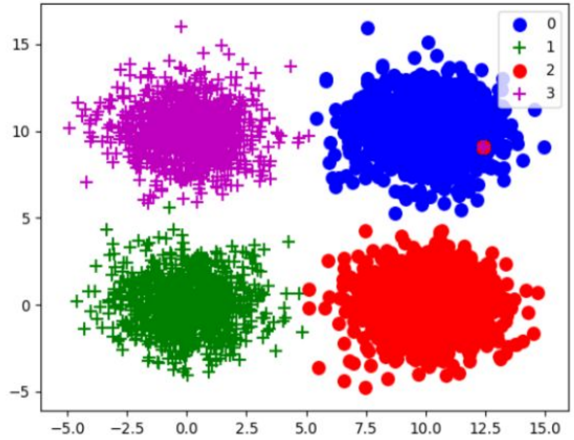
Spectral Clustering



Synthetic dataset

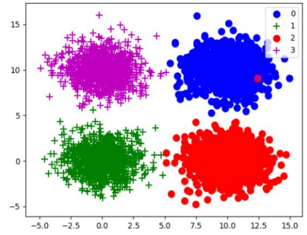
$$x_t = (\sigma(W \cdot v_t))^2 \quad t = 1, \ldots, n$$

v_t is the t-th point in the latent domain

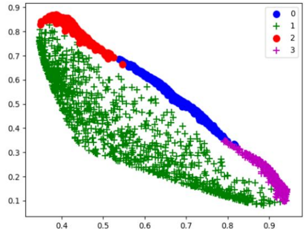


(a) Latent domain, v.

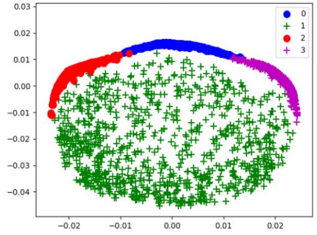
Synthetic dataset



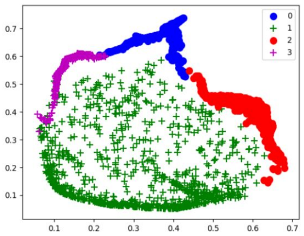
(a) Latent domain, v.



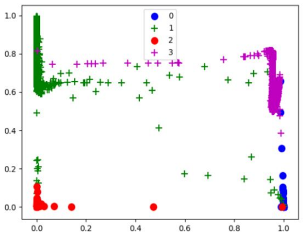
(b) NMF



(c) SVD



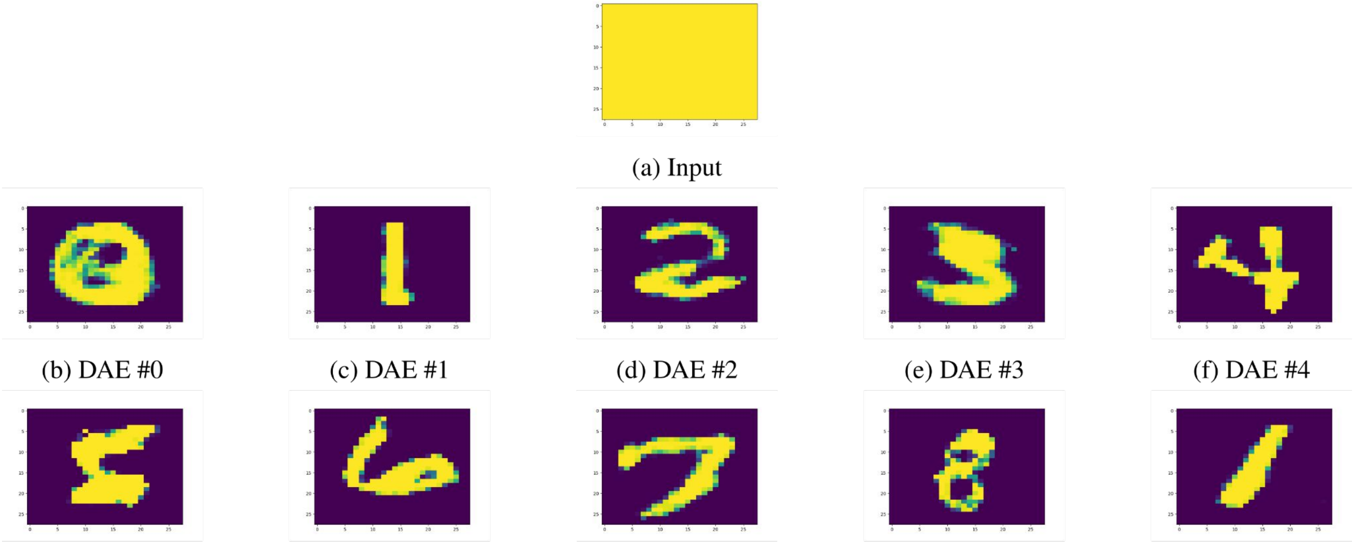
(d) DAE+KM



(e) DAMIC

Table 2: Objective measures for the synthetic database.

Method	DAMIC	DAE+KM	SC	KM
NMI	**0.94**	0.83	0.82	0.80
ARI	**0.96**	0.84	0.83	0.81



(a) Input

(b) DAE #0 (c) DAE #1 (d) DAE #2 (e) DAE #3 (f) DAE #4

(g) DAE #5 (h) DAE #6 (i) DAE #7 (j) DAE #8 (k) DAE #9

Figure 3: The outputs of the different DAEs with a vector of all-ones input.

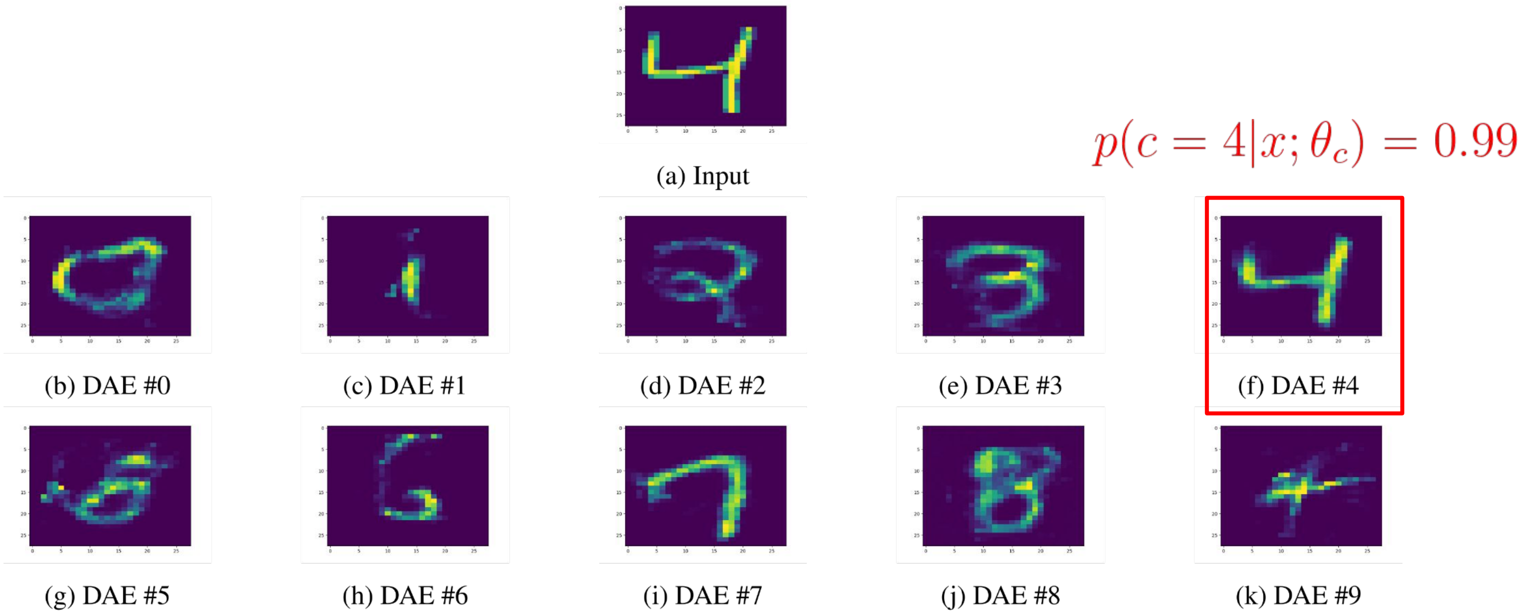


Figure 4: An example of the outputs of the different DAEs with the digit ‘4’ as the input.

Table 3: Objective measures of the MNIST database.

Method	DAMIC	DCN	DAE+KM	DEC	KM
NMI	**0.87**	0.81	0.74	0.80	0.50
ARI	**0.81**	0.75	0.67	0.75	0.37
ACC	**0.89**	0.83	0.80	0.84	0.53

Table 5: Objective measures of the 20NEWS database.

Method	DAMIC	DCN	DAE+KM	SC	KM
NMI	**0.57**	0.48	0.47	0.40	0.41
ARI	**0.42**	0.34	0.28	0.17	0.15
ACC	**0.56**	0.44	0.42	0.34	0.30

Table 4: Objective measures of the Fashion database.

Method	DAMIC	DCN	DAE+KM	DEC	KM
NMI	**0.65**	0.55	0.60	0.54	0.51
ARI	**0.49**	0.42	0.45	0.40	0.37
ACC	**0.60**	0.50	0.57	0.51	0.47

Table 6: Objective measures of the RCV1 database.

Method	DAMIC	DCN	DAE+KM	DEC	KM
NMI	**0.62**	0.61	0.59	0.08	0.58
ARI	**0.41**	0.33	0.33	0.01	0.29
ACC	**0.47**	**0.47**	0.46	0.14	0.47

Table 7: Ablation study on the MNIST database.

Method	DAMIC	Pre-training only	Joint-training only	KM
NMI	**0.87**	0.74	0.71	0.50
ARI	**0.81**	0.67	0.53	0.37
ACC	**0.89**	0.80	0.60	0.53

Conclusion

Conclusion

- Proposed a clustering technique that use autoencoder to represent cluster instead of single centroid vector, which enables a much richer representation of each cluster
- The proposed clustering technique can avoid data collapsing problem