Unsupervised Deep Embedding for Clustering Analysis

Advisor: Jia-Ling, Koh

Speaker: Zi-Xin Chen

Source: ICML’16

Date: 2022/10/11

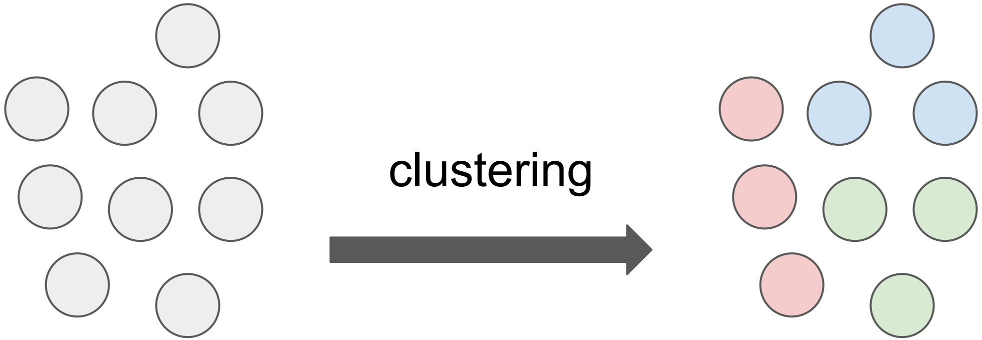
Outline

- Introduction
- Method
- Experiment
- Conclusion

Introduction

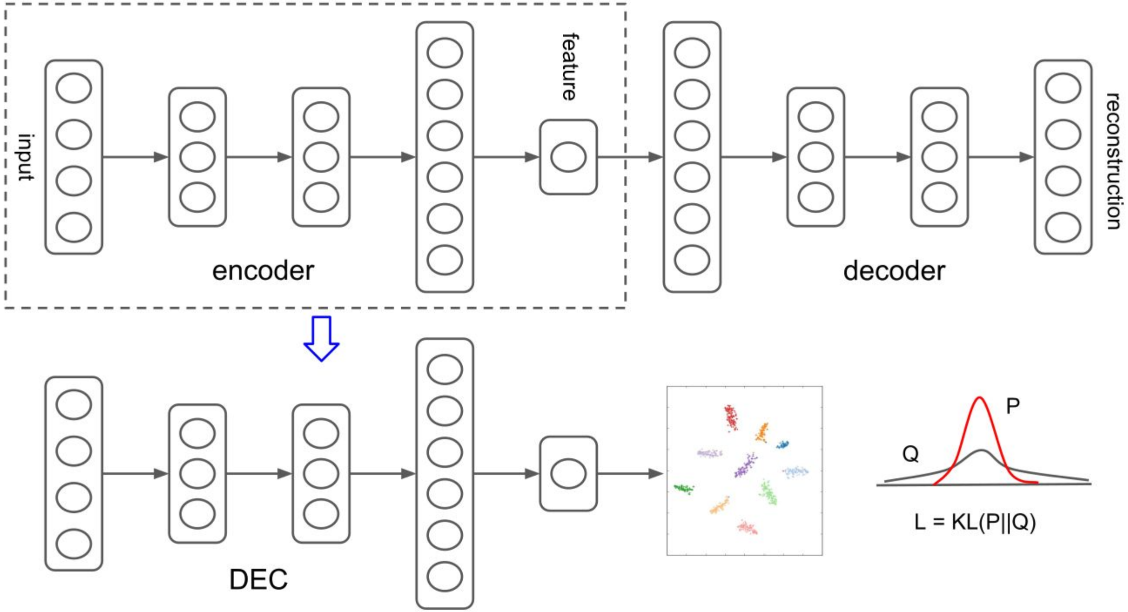
Clustering

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



Deep Clustering - DEC

- Unsupervised learning of the **feature space** where to perform clustering



Method

Problem Formulation

X : data space

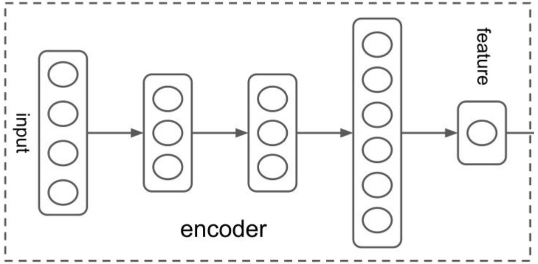
n points : $\{x_i \in X\}_{i=1}^{n}$

Z : feature space

n points : $\{z_i \in Z\}_{i=1}^{n}$

k centroids : $\{\mu_j \in Z\}_{j=1}^{k}$

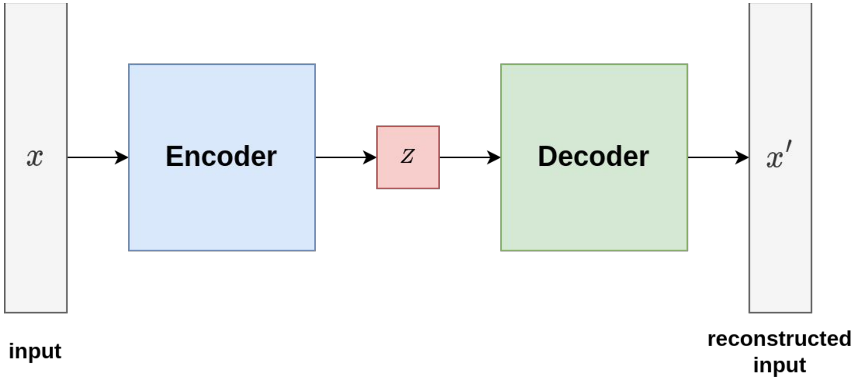
$$f_\theta : X \to Z$$



Initialization

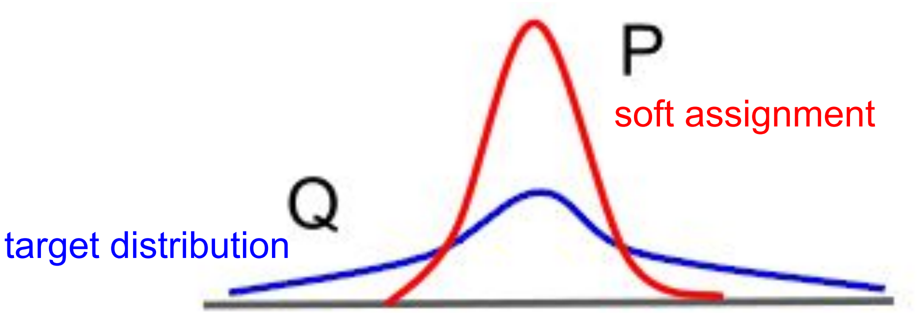
1. Train an autoencoder
2. Discard the decoder
3. Get embedded data
4. Perform k-means to get initial centroids $\{\mu_j \in Z\}_{j=1}^{k}$

*training data: 100000 unlabeled images from ***STL-10***



Objective function

- Model is trained by matching the **soft assignment** to the **target distribution**



Soft assignment

- The probability of assigning sample ***i*** to cluster ***j***

embedded data point centroid

$$q_{ij} = \frac{(1 + ||z_i - \mu_j||^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'}(1 + ||z_i - \mu_{j'}||^2/\alpha)^{-\frac{\alpha+1}{2}}}$$

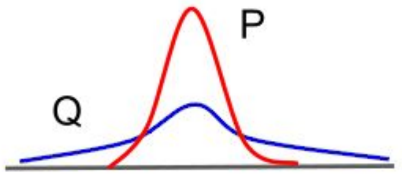
Auxiliary target distribution

$$p_{ij} = \frac{q_{ij}^2 / f_j}{\sum_{j'} q_{ij'}^2 / f_{j'}}$$

$$f_j = \sum_i q_{ij}$$

Soft cluster frequencies

KL divergence



L = KL(P||Q)

$$L = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

p_{ij}: Target distribution

q_{ij}: Soft assignment

$$= \sum_i \sum_j p_{ij} (\log(p_{ij}) - \log(q_{ij}))$$

Experiment

Dataset

dataset	size	classes	type
MNIST	70000	10	image
STL-10	1300	10	image
REUTERS	685071	4	text
REUTERS-10K	10000	4	text

Evaluation Metric

- Accuracy for classification

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n} \mathbf{1}\{l_i = c_i\}}{n}$$

- Accuracy for clustering

Permutes clustering labels to match the ground truth labels

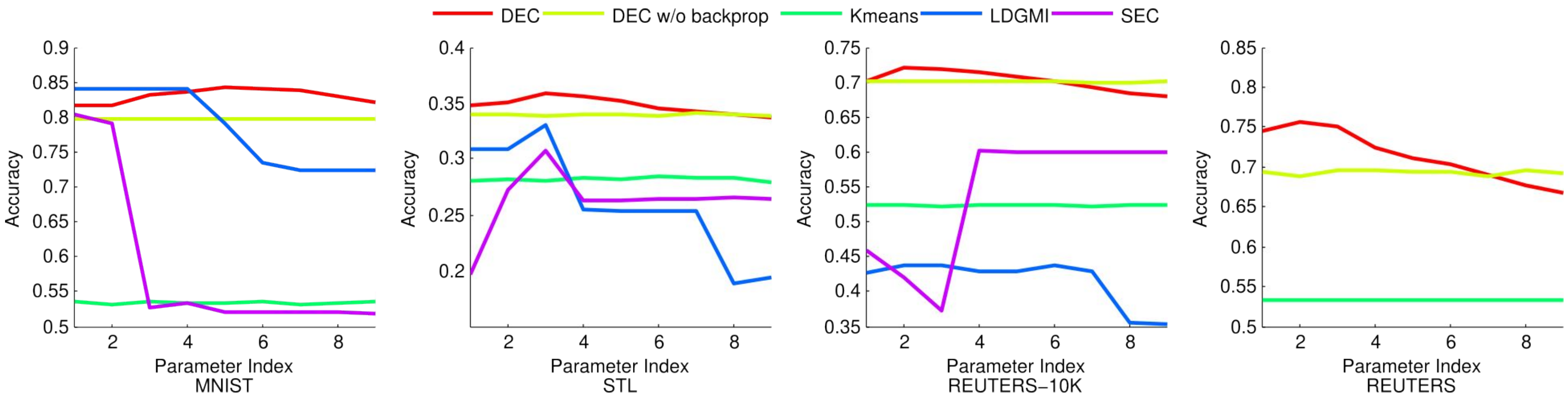
$$\mathrm{ACC} = \max_{m} \frac{\sum_{i=1}^{n} \mathbf{1}\{l_i = m(c_i)\}}{n}$$

Table 2. Comparison of clustering accuracy (Eq. 10) on four datasets.

Method	MNIST	STL-HOG	REUTERS-10k	REUTERS
k-means	53.49%	28.39%	52.42%	53.29%
LDMGI	84.09%	33.08%	43.84%	N/A
SEC	80.37%	30.75%	60.08%	N/A
DEC w/o backprop	79.82%	34.06%	70.05%	69.62%
DEC (ours)	**84.30%**	**35.90%**	**72.17%**	**75.63%**

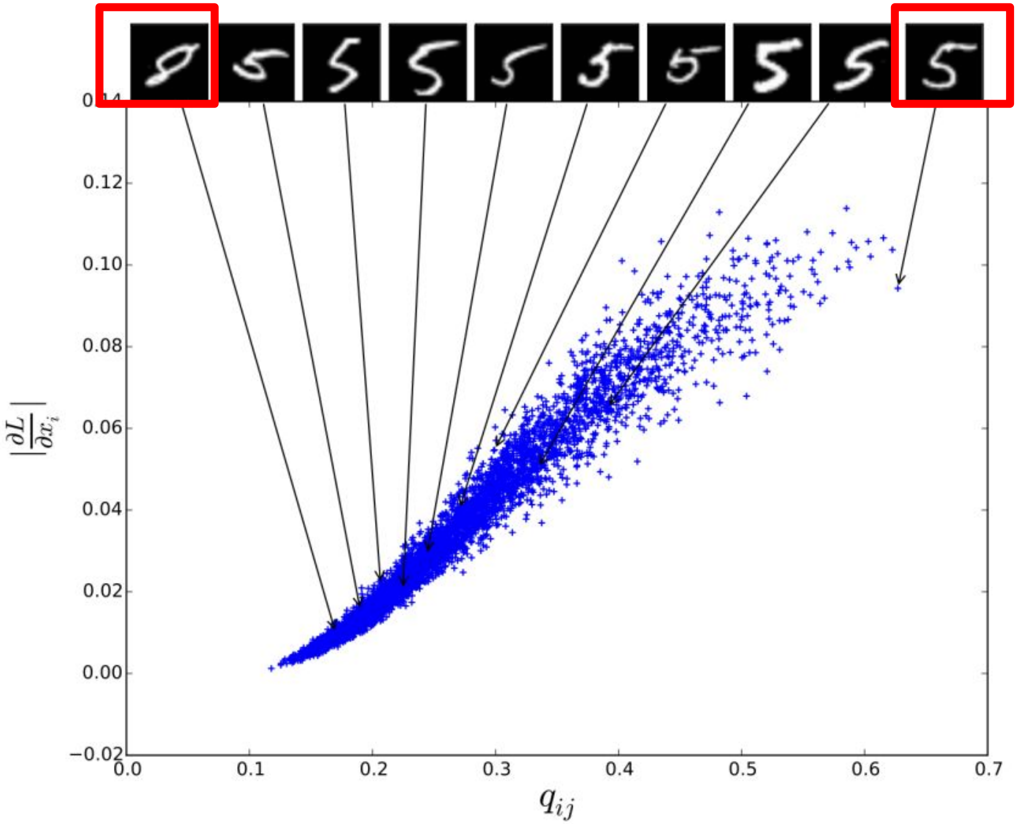
*LDMGI, SEC : spectral clustering based

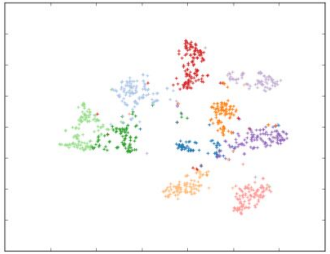
*DEC w/o backprop : freeze the non-linear mapping f_θ


DEC
DEC w/o backprop
Kmeans
LDGMI
SEC
Accuracy
Parameter Index
MNIST
STL
REUTERS−10K
REUTERS
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.4
0.35
0.3
0.25
0.2
0.45
2
4
6
8

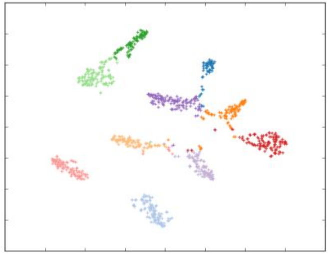
(a) MNIST

(b) STL-10

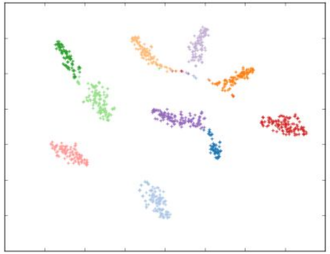

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−0.02
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
q_{ij}
$\left|\frac{\partial L}{\partial x_i}\right|$



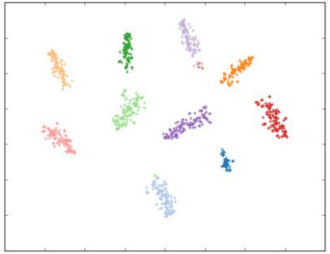
(a) Epoch 0



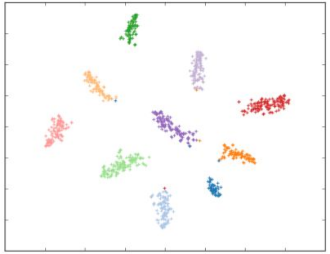
(b) Epoch 3



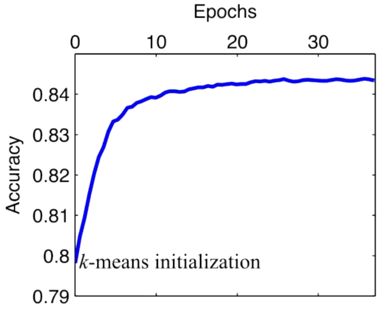
(c) Epoch 6



(d) Epoch 9



(e) Epoch 12



(f) Accuracy vs. epochs

Contribution of Autoencoder Initialization

Table 3. Comparison of clustering accuracy (Eq. 10) on autoencoder (AE) feature.

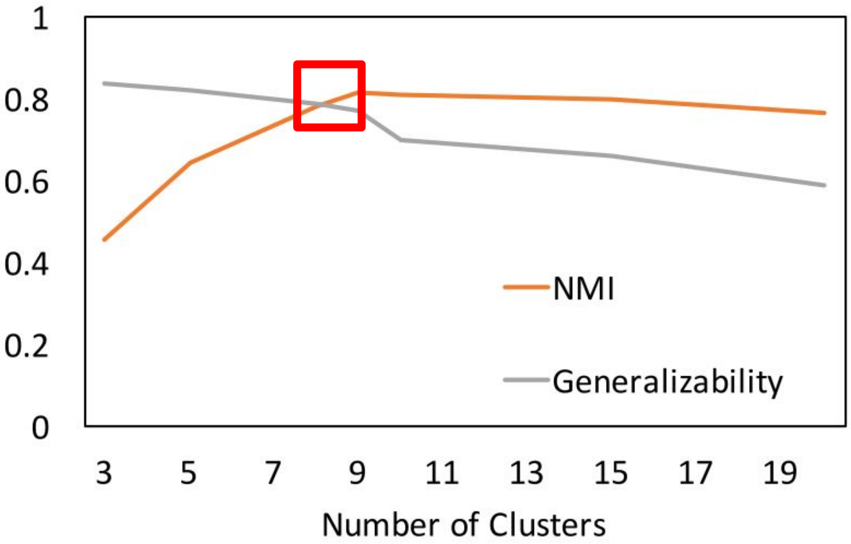
Method	MNIST	STL-HOG	REUTERS-10k	REUTERS
AE+k-means	81.84%	33.92%	66.59%	71.97%
AE+LDMGI	83.98%	32.04%	42.92%	N/A
AE+SEC	81.56%	32.29%	61.86%	N/A
DEC (ours)	**84.30%**	**35.90%**	**72.17%**	**75.63%**

Performance on Imbalanced Data

Table 4. Clustering accuracy (Eq. 10) on imbalanced subsample of MNIST.

Method \ r_{min}	0.1	0.3	0.5	0.7	0.9
k-means	47.14%	49.93%	53.65%	54.16%	54.39%
AE+k-means	66.82%	74.91%	77.93%	80.04%	81.31%
DEC	70.10%	80.92%	82.68%	84.69%	85.41%

Number of clusters



$$G = \frac{L_{\text{train}}}{L_{\text{validation}}}$$

Conclusion

Conclusion

- Present an algorithm that clusters a set of data points in a jointly optimized feature space
- A way to learn a representation specialized for clustering without ground truth cluster membership labels