# 1 Introduction

# 2 Basics

## 2.1 Unsupervised learning

### 2.1.1 Introduction of Unsupervised learning

<https://link.springer.com/article/10.1007/s11222-007-9033-z>

最广泛的应用技术，涉及领域包括统计学，计算机科学，生物，社会科学和心理学。

人们尝试用第一印象去处理数据：即把数据分类

Clustering is one of the most widely used techniques for exploratory data analysis, with applications ranging from statistics, computer science, biology to social sciences or psychology. In virtually every scientific field dealing with empirical data, people attempt to get a first impression on their data by trying to identify groups of “similar behavior” in their data.

<https://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=1763&context=ele_comeng_facwork>Data analysis plays an indispensable role for understanding various phenomena. Cluster analysis, primitive exploration with little or no prior knowledge, consists of research developed across a wide variety of communities. The diversity, on one hand, equips us with many tools. On the other hand, the profusion of options causes confusion.

Introduction:

第一段：clustering与人的关系

第二段：无监督学习

第四段：介绍Clustering算法：Hard partitional clustering和Hierarchical clustering

<http://eprints.iisc.ernet.in/273/1/p264-jain.pdf>

非常详细的一篇总结

<https://www.researchgate.net/publication/307667810_MACHINE_LEARNING_TECHNIQUES_USED_IN_BIG_DATA>

\*Big data和machine learning

<https://arxiv.org/pdf/1112.6209.pdf&amp>

unsupervised learning 第二页

<http://mlg.eng.cam.ac.uk/zoubin/papers/ul.pdf>

unsupervised learning，k-means

<https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>

<https://medium.com/machine-learning-for-humans/unsupervised-learning-f45587588294>

<https://www.guru99.com/unsupervised-machine-learning.html>

<https://pythonistaplanet.com/applications-of-unsupervised-learning/>

Supervised and unsupervised learning

### 2.1.2 Clustering

<https://apandre.wordpress.com/visible-data/cluster-analysis/>

图片信息来源

<https://medium.com/machine-learning-for-humans/unsupervised-learning-f45587588294>

<https://www3.nd.edu/~rjohns15/cse40647.sp14/www/content/lectures/12%20-%20k-means%20Clustering.pdf>

Clustering &Association

<https://en.wikipedia.org/wiki/Cluster_analysis>

<https://www.geeksforgeeks.org/clustering-in-machine-learning/>

<https://www.ims.uni-stuttgart.de/institut/mitarbeiter/schulte/theses/phd/algorithm.pdf>

可参考单元描述和章节描述

<http://infolab.stanford.edu/~ullman/mining/pdf/cs345-cl.pdf>

<https://www.ims.uni-stuttgart.de/institut/mitarbeiter/schulte/theses/phd/algorithm.pdf>

<http://www.schulteimwalde.de/research/phd.html>

博士论文distance

<http://www.mit.edu/~9.54/fall14/slides/Class13.pdf>

technique 图源

<https://www2.stat.duke.edu/~kheller/thesis.pdf>

bayesian methods (p.22-23)

<https://projecteuclid.org/download/pdfview_1/euclid.ba/1508378464>

Bayesian nonparametric

### 2.1.3 k-means algorithm

<http://www.mit.edu/~9.54/fall14/slides/Class13.pdf>

MIT PPT讲义：介绍了clustering定义 分类

<https://chlyy2005.wordpress.com/2012/08/20/latex-algorithm%E7%9A%84%E5%86%99%E6%B3%95/>

Latex 算法教程

<http://www.sohu.com/a/215996636_163476>

中文解释Clustering

<https://users.ics.aalto.fi/sami/thesis/node9.html>

## 2.2 Autoencoder

<https://dbs.uni-leipzig.de/file/Saalmann_Ausarbeitung.pdf>

<http://www.deeplearningbook.org/contents/autoencoders.html>

各种各样的autoencoder

<https://en.wikipedia.org/wiki/Autoencoder#cite_note-:0-3>

<https://www.deeplearningbook.org/contents/autoencoders.html>

<https://arxiv.org/pdf/1606.05908.pdf>

VAE 解释了VAE与autoencoder的关系

<https://pdfs.semanticscholar.org/0611/46b1d7938d7a8dae70e3531a00fceb3c78e8.pdf>

<https://en.wikipedia.org/wiki/Autoencoder#cite_note-:2-18>

<http://www.fanyeong.com/2018/06/01/vae/>

page 7: VAE和autoencoder的区别

The idea of autoencoders has been popular in the field of neural networks for decades, and the first applications date back to the '80s.[[6]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-6)[[3]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-:0-3)[[7]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-7) Their most traditional application was [dimensionality reduction](https://en.wikipedia.org/wiki/Dimensionality_reduction) or [feature learning](https://en.wikipedia.org/wiki/Feature_learning), but more recently the autoencoder concept has become more widely used for learning [generative models](https://en.wikipedia.org/wiki/Generative_model) of data.[[8]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-VAE-8)[[9]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-gan_faces-9) Some of the most powerful [AIs](https://en.wikipedia.org/wiki/Artificial_intelligence) in the 2010s involved sparse autoencoders stacked inside of [deep](https://en.wikipedia.org/wiki/Deep_learning) neural networks.[[10]](https://en.wikipedia.org/wiki/Autoencoder#cite_note-domingos-10)

The input is \emph{I}. The output of a encoder is \emph{E}(\emph{I}) and the reconstructed input i.e. the output of a decoder is \emph{D}(\emph{E}(\emph{I})). Let $f$ be the computations of the encoder and $g$ be the computations of the decoder, then the encoded representation is computed as in Eq.\ref{Eq\_encoder} and the reconstructed input is computed as in Eq.\ref{Eq\_decoder}. The loss function is represented by the reconstruction error given in Eq.\ref{AE\_loss}.

\begin{align}

\label{Eq\_encoder}

\emph{E}(\emph{I}) = f(\emph{I})

\end{align}

\begin{align}

\label{Eq\_decoder}

\emph{D}(\emph{E}(\emph{I})) = g(\emph{E}(\emph{I}))

\end{align}

\begin{align}

\label{AE\_loss}

\emph{L} = \frac{1}{\lvert\emph{J}\rvert} \sum\limits\_{\emph{j}\in\emph{J}} (\emph{$I\_{j}$} - \emph{D}(\emph{E}(\emph{$I\_{j}$})))^{2}

\end{align}

\begin{figure}

\centering

\includegraphics[scale=.3]{basic\_AE.eps}

\caption{The structure of a basic Autoencoder}

\label{fig:basic\_AE}

\end{figure}

### 2.2.1 Neural Network

### 2.2.2 Introduction of Autoencoder

<https://zhuanlan.zhihu.com/p/34998569>

<https://jaan.io/what-is-variational-autoencoder-vae-tutorial/>

<https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc>

<https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc>

<https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/>

<https://www.researchgate.net/publication/228394623_A_brief_review_of_feed-forward_neural_networks>

activation function

VAE

<https://spaces.ac.cn/archives/5716>

变分推断理解VAE等模型

### 2.2.3 l2 normalized autoencoder

## 2.3 Adversarial Autoencoder(AAE)

### 2.3.1 GAN

<https://medium.com/ai-society/gans-from-scratch-1-a-deep-introduction-with-code-in-pytorch-and-tensorflow-cb03cdcdba0f>

<https://www.geeksforgeeks.org/generative-adversarial-network-gan/>

<https://www.analyticsvidhya.com/blog/2017/06/introductory-generative-adversarial-networks-gans/>

<https://blog.usejournal.com/introduction-to-gans-and-its-common-applications-29c8f22bc2c6>

<https://www.cnblogs.com/fxjwind/p/9275744.html>

GAN详细讲解

<https://medium.com/@jonathan_hui/gan-whats-generative-adversarial-networks-and-its-application-f39ed278ef09>

可参考写论文 GAN

### 2.3.2 basic AAE

<https://towardsdatascience.com/paper-summary-adversarial-autoencoders-f89bfa221e48>

<https://duvenaud.github.io/learn-discrete/slides/AdversarialAutoencoders.pdf>

AAE

<https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/Adversarial-Autoencoders.pdf>

GAN，AAE，VAE

<http://openaccess.thecvf.com/content_cvpr_2017/papers/Zhang_Age_ProgressionRegression_by_CVPR_2017_paper.pdf>

GAN

<https://www.aclweb.org/anthology/P17-1179.pdf>

<https://www.jiqizhixin.com/articles/2017-10-1-1>

推导KL散度和JS散度

可参考结果分析

<https://towardsdatascience.com/paper-summary-adversarial-autoencoders-f89bfa221e48>

参考自编码器选择

<https://arxiv.org/pdf/1811.07605.pdf>

VAE和AAE的区别， AAE弥补了目标函数中的KL，可引用

<http://kissg.me/2017/12/17/papernotes03/>

AAE 与 GAN 的不同在于: GAN 通常利用神经网络输出层的像素级的数据分布; 而 AAE 倚赖自编码训练来捕获数据分布. 在 AAE 的训练过程中, 可以令 q(z) 去拟合更低维度的更简单的分布. (这就起到了简化和降维的作用)

<https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/Adversarial-Autoencoders.pdf>

AAE具有的缺点：论文不多

## 2.4 Continual Learning

判别模型和生成模型的方法通常不通用，因此判别模型的CL方法不能直接英语与生成模型，CL生成模型的方法还有很大的探索空间。已存在的方法要不用于VAE要不用于GAN。但有一种方法可同时应用于对抗和变分框架，即DGR。

Paper: <https://arxiv.org/pdf/1812.09111.pdf>

### 2.4.1 Catastrophic Forgetting

1994\_04\_Catastrophic forgetting in connectionist networks

介绍灾难性遗忘的原因结果和解决方法

## 2.5 Deep Generative Replay

.

### 2.5.2 Deep Generative Model

### 2.5.3 Generative Replay

# 3 Implementation

## 3.1 Dataset

### 3.1.1 MNIST

### 3.1.2 Split MNIST

### 3.1.3 Fashion MNIST

### 3.1.4 Split Fashion MNIST

## 3.2 Proposed Model

### 3.2.1 l2 Normalized Deep Dense Autoencoder

### 3.2.2 Adversarial Autoencoder (AAE)

# 4 Experiments and Results

<https://papers.nips.cc/paper/7225-gradient-episodic-memory-for-continual-learning.pdf>

测试方法

<https://www.cs.uic.edu/~liub/lifelong-learning/continual-learning.pdf>

介绍各种方法

4.1 Unsupervised Clustering

## 4.2 Continual Unsupervised Clustering without DGR

## 4.3 Continual Unsupervied Clustering with DGR

# 5 Summary and Discussion

References

[1]Continual Lifelong Learning with Neural Networks: A Review

[2]Continual learning with deep generative deplay

[3] A. Robins. Catastrophic forgetting, rehearsal and pseudorehearsal. Connection Science,

7(2):123–146, 1995.

[4]Re-evaluating

[5] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 2017.