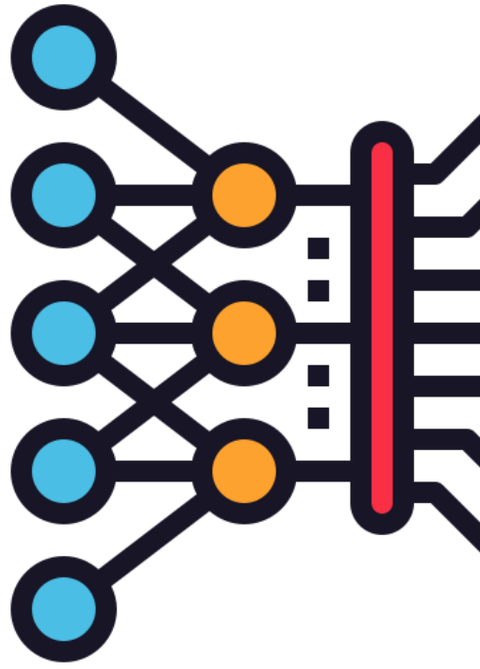


Autoencoder

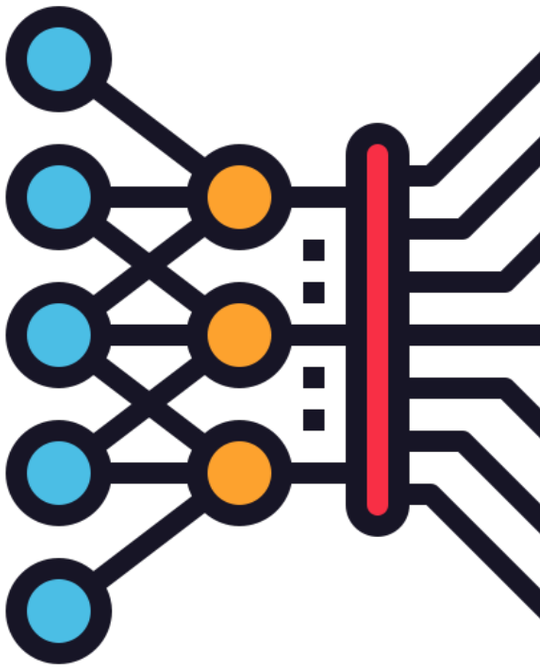
Subject: **Data Mining**

Lecturers: **PHAUK Sokkhey** & **CHAN Sophal**

Presenter: **CHOENG Veyseng**

Academic Year: 2021 - 2022

Table of Content



What is Autoencoder?

Autoencoder (Bhuvaneshwari et al., 2021) is type of neutral network concept (A deep learning concept) and known as unsupervised learning model.

It is consisted of 2 main phases: Encoding process and Decoding process.

For the encode phase, the input data are converted to low-dimensionality that represent to filter the most important features.

For the decode phase is contradicted to the encode phase that it works to expand all available features for outputting the high-dimensionality.

Autoencoder Applications

We can use this deep learning Autoencoder concepts for some applications (Arul, 2021):

- Dimensional Reduction
- Image Denoising
- Feature Extraction
- Recommendation System
- Image Compression …

Advantages of Autoencoder

- Help us to reduce the dimensionality that could eventually reduce the resource consuming and time consuming for data training process
- A high functional feature to make the data transition better than CNN
- It is the unsupervised learning model that could benefit to the conceptual design that can have the ability to train the data in the unlabeled way.

Disadvantages of Autoencoder

- It try to capture as much information as possible rather than capture the most optimal information from the features
- In order build a real model, we need to collect a huge amount of data, consider on process time, hyper-parameters, model validation at first.
- Autoencoder might eventually represent manifold lies because the training dataset are not the representative of testing dataset

Demonstration of Autoencoder

Using Keras library to detect the anomaly data with financial fraud detection.

References

[1] Bhuvaneshwari, M., Kanaga, E. G. M., Anitha, J., Raimond, K., & George, S. T. (2021). A comprehensive review on deep learning techniques for a BCI-based communication system. In *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics* (pp. 131–157). Elsevier. https://doi.org/10.1016/B978-0-12-821633-0.00013-1

[2] Arul, V. H. (2021). Deep learning methods for data classification. In *Artificial Intelligence in Data Mining* (pp. 87–108). Elsevier. https://doi.org/10.1016/B978-0-12-820601-0.00001-X

[3] Quora - **Conner Davis**, Data Scientist at Microsoft Answered March 11, 2017·Upvoted by Drew Levin, PhD Computer Science & Complex Adaptive Systems, University of New Mexico and Akshay Khatri, MS Computer Science, Columbia University (2018)