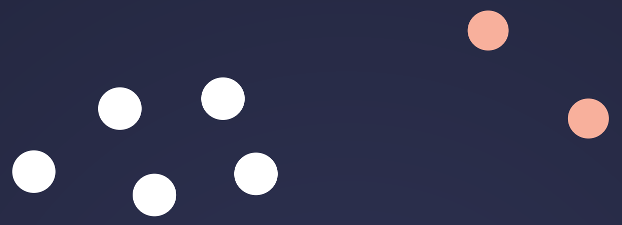


# Anomaly Detection with Variational Autoencoders

João Pereira

Deep Learning Sessions Lisbon

www.joao-pereira.pt mail@joao-pereira.pt

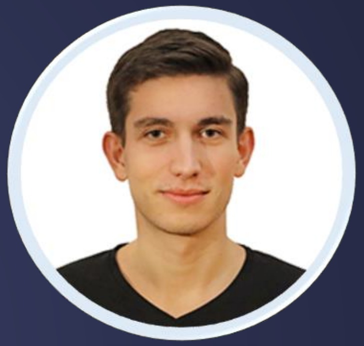
May 6, 2020

# About me

PDEng Data Science trainee @ TU/e

Based in Eindhoven, The Netherlands

Electrical and Computer Engineer (IST)



JADS Jheronimus Academy of Data Science

Founded by TU Eindhoven and Tilburg Univ.

BSc, MSc, PhD, PDEng, Professional edu.

300 partnerships

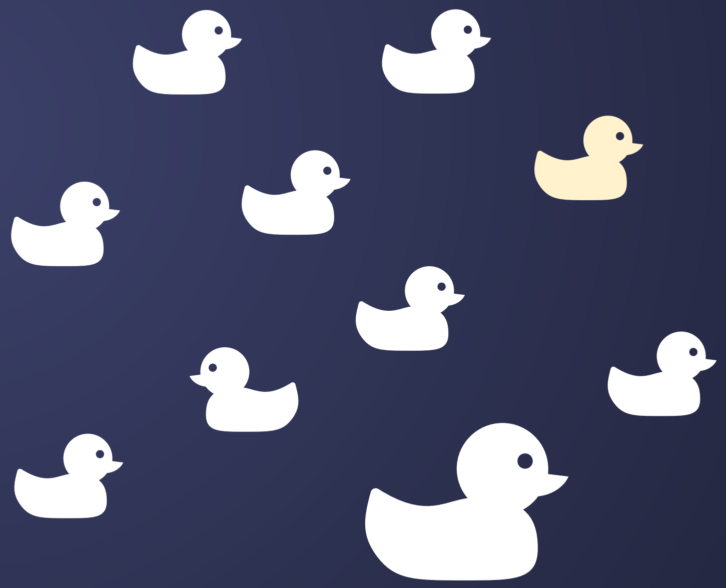
1. Introduction
2. VAEs
3. Applications

# What is anomaly detection?

- Anomalies are deviations from **normal** behaviour.

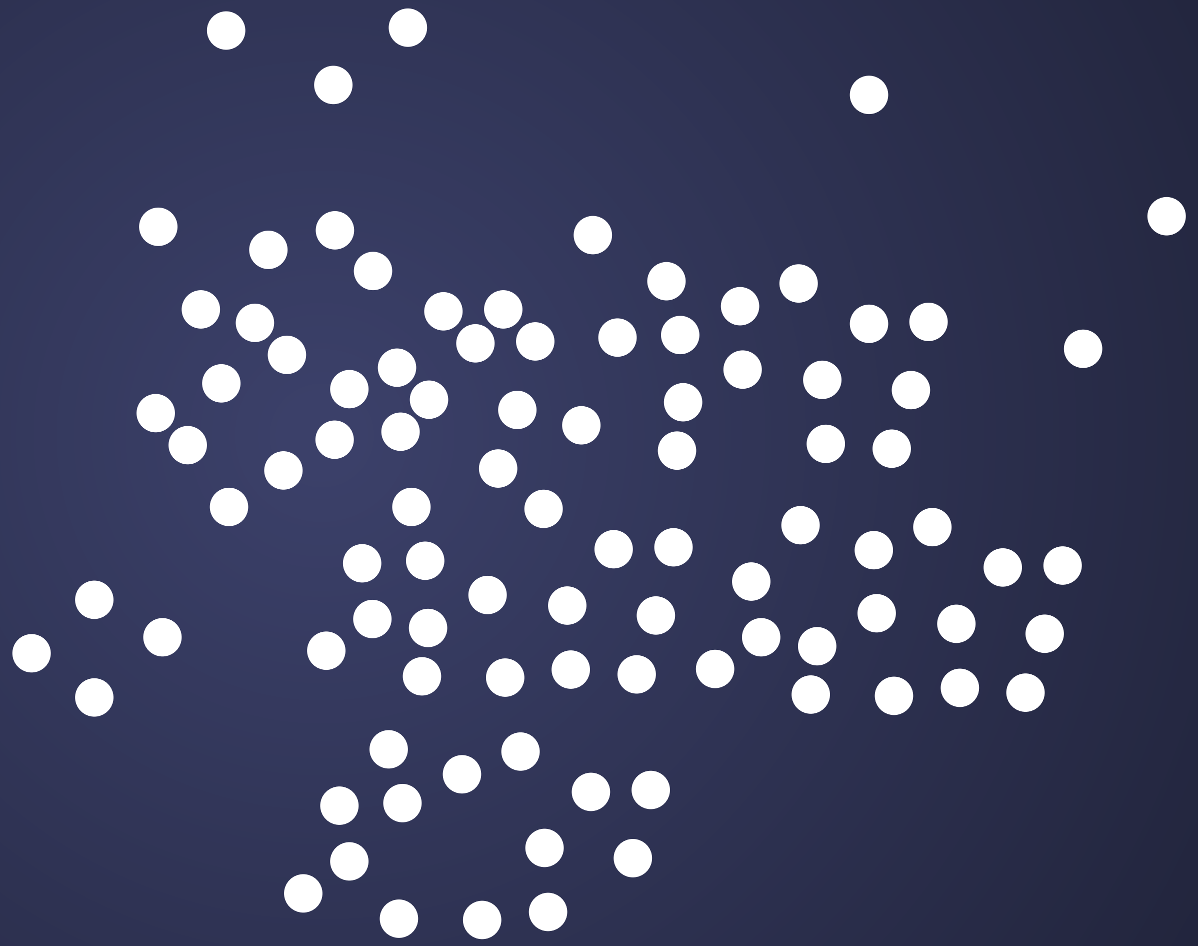
- **Applications**:

    Fault detection
    Fraud detection
    Cyber intrusion detection
    Video surveillance
    …



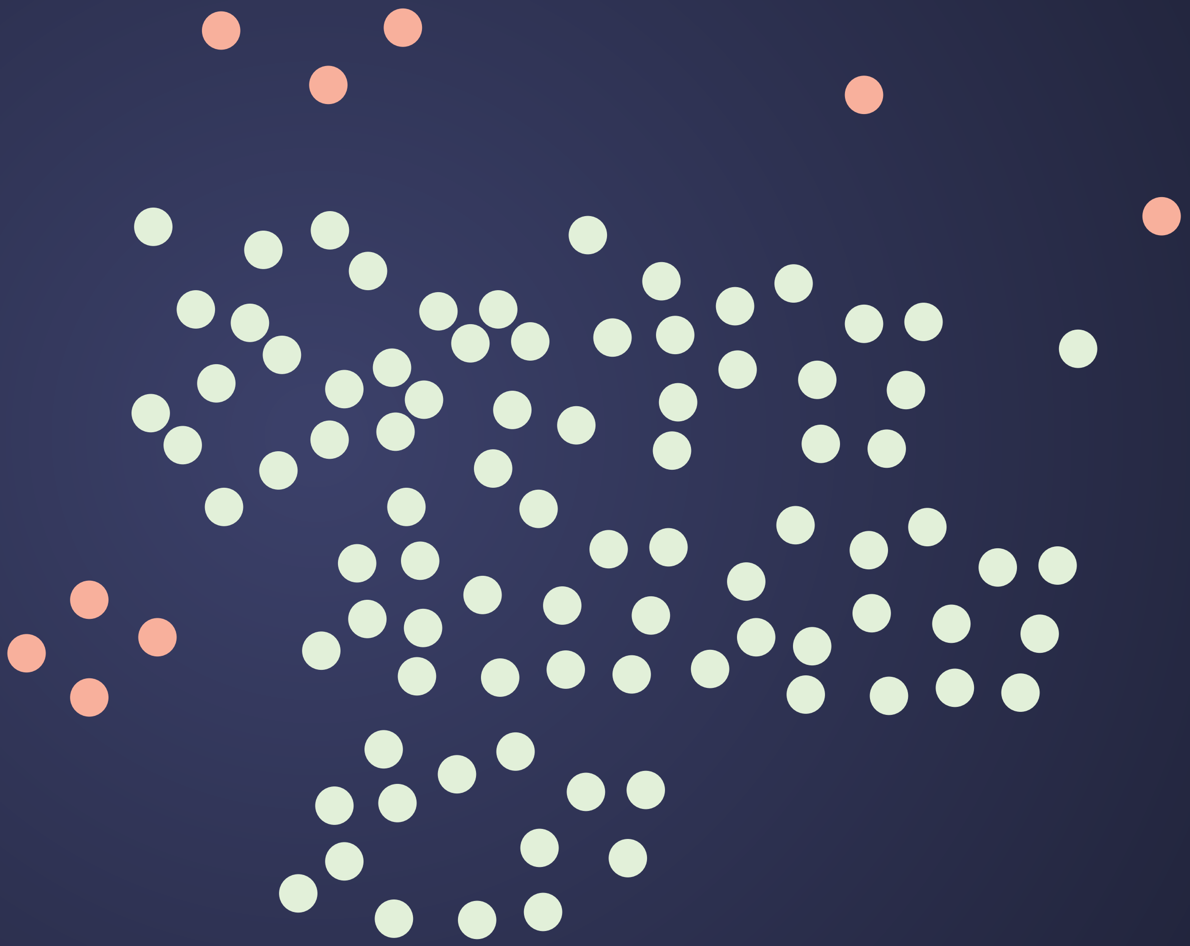
# Problem

$$\mathcal{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{N}$$



# Problem

$$\mathcal{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{N}$$



- “Normal”
- “Anomalous”

# Two ways to go...

**Classification**
supervised

$$p(\mathbf{y}|\mathbf{x})$$

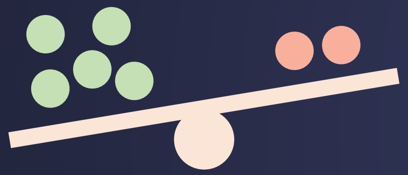
**Density Estimation**
unsupervised

$$p(\mathbf{x})$$

Anomaly Score

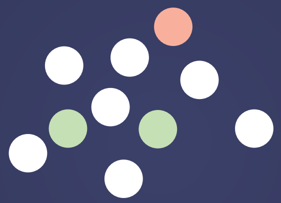
$$\mathbf{y} \in \{\bullet, \bullet\}$$

# Challenges



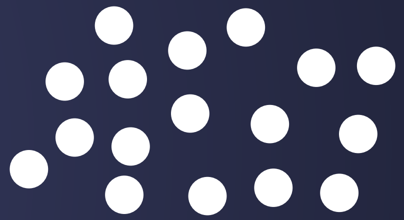
**High Imbalance**

# anomalies << # normal



**Scarce Labels**

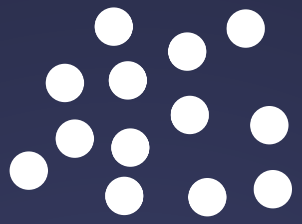
expensive, time

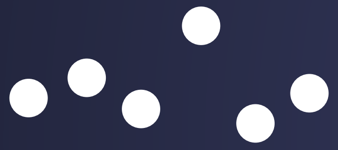


**Data Dimension and Size**

curse of dimensionality

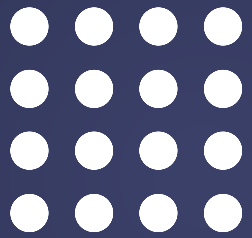
# Data is not i.i.d.





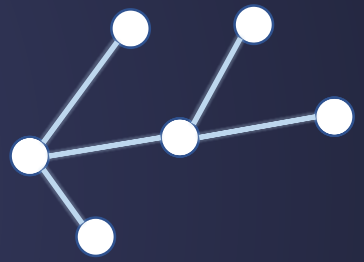
**Sequences**
e.g., time series, text

Temporal



**Images**

Spatial



**Graphs**
e.g., {social, transaction} networks

Relational

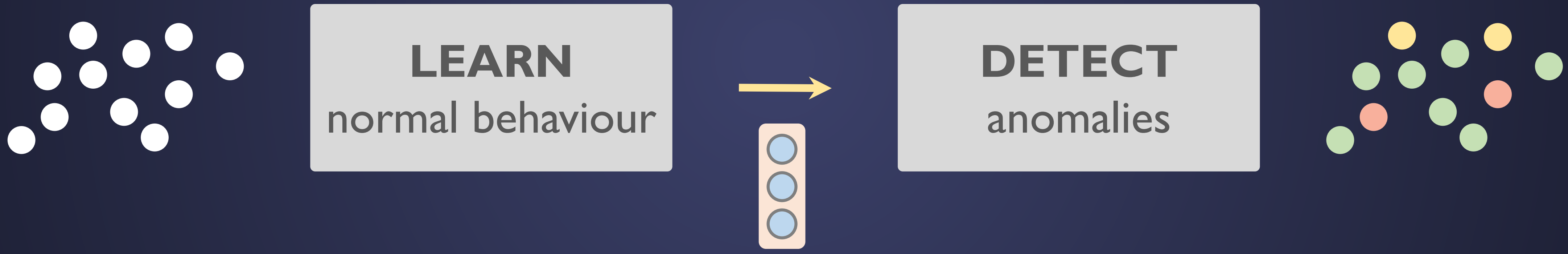
1. Introduction
2. VAEs
3. Applications

# We learn a representation!

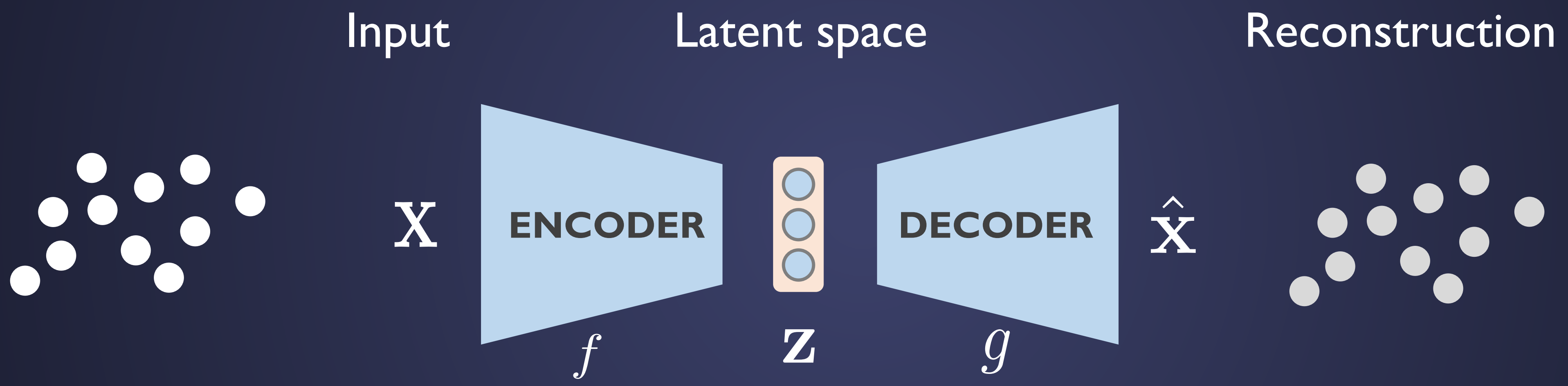
- Low-dimensional
- Structured
- Expressive

$$\mathcal{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{N} \qquad \mathcal{Z} = \{\mathbf{z}^{(i)}\}_{i=1}^{N}$$

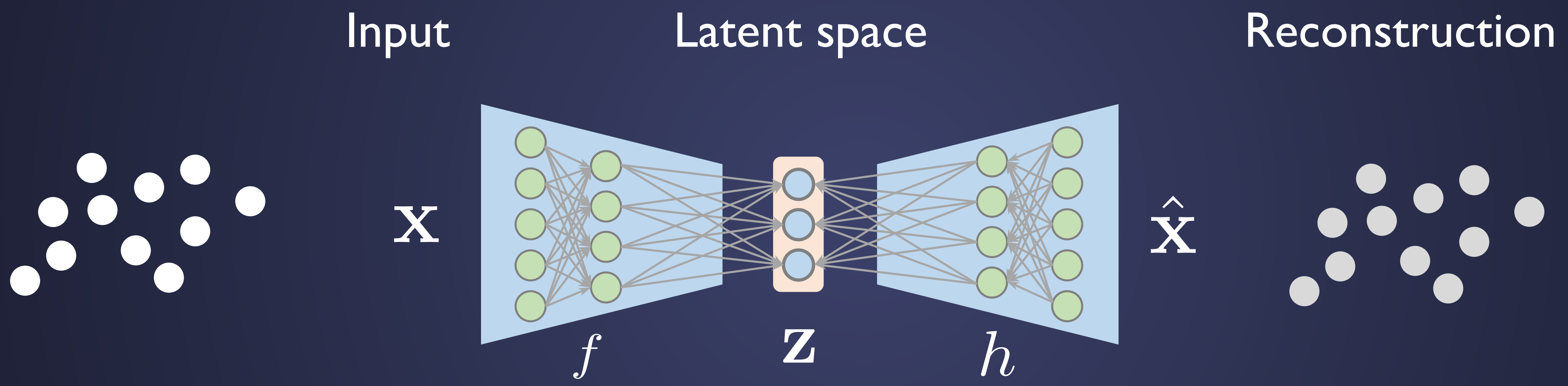
# Anomaly Detection Strategy



# Autoencoders

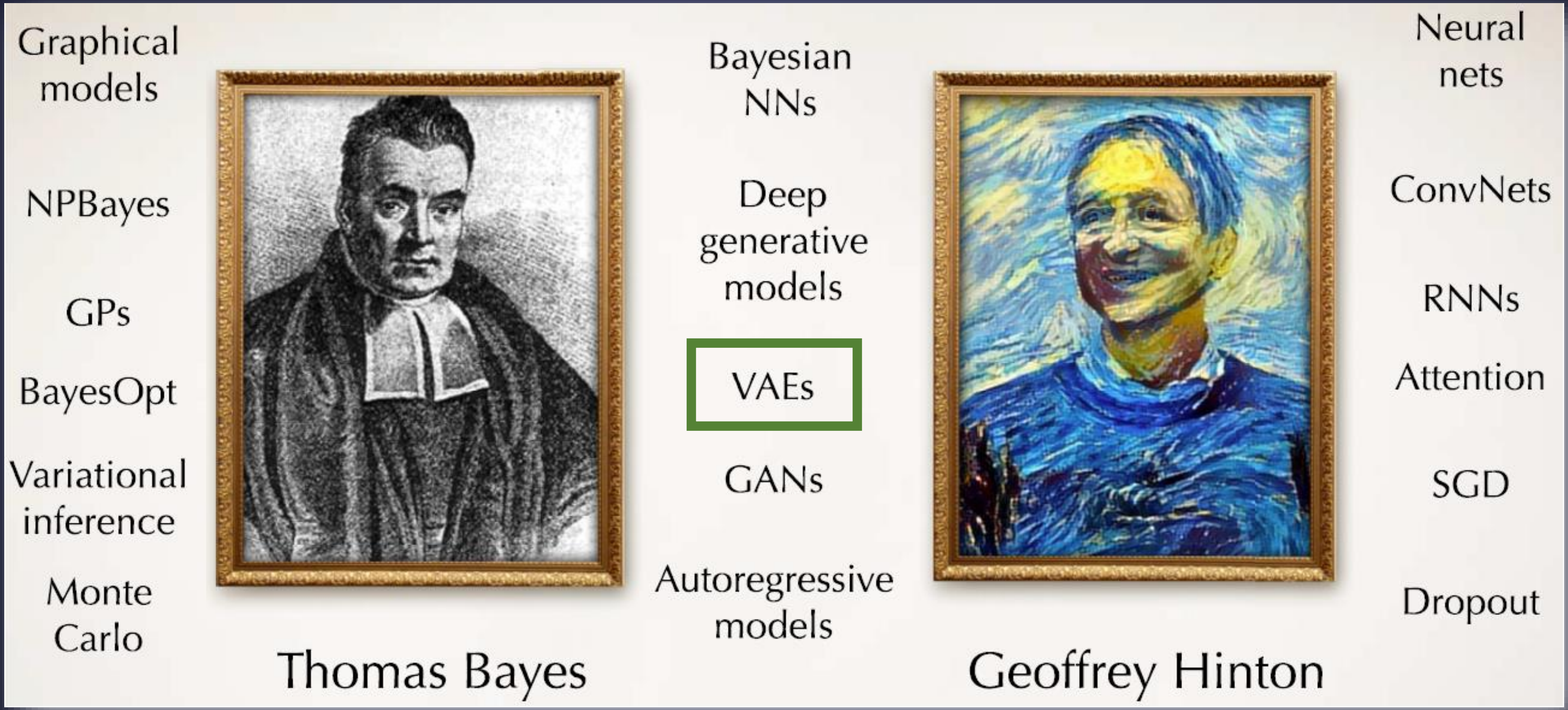


# Autoencoders




Loss function: $$\mathcal{L}\left(\mathbf{x}, \hat{\mathbf{x}}\right) = \left|\left|\mathbf{x} - \hat{\mathbf{x}}\right|\right|_2^2$$

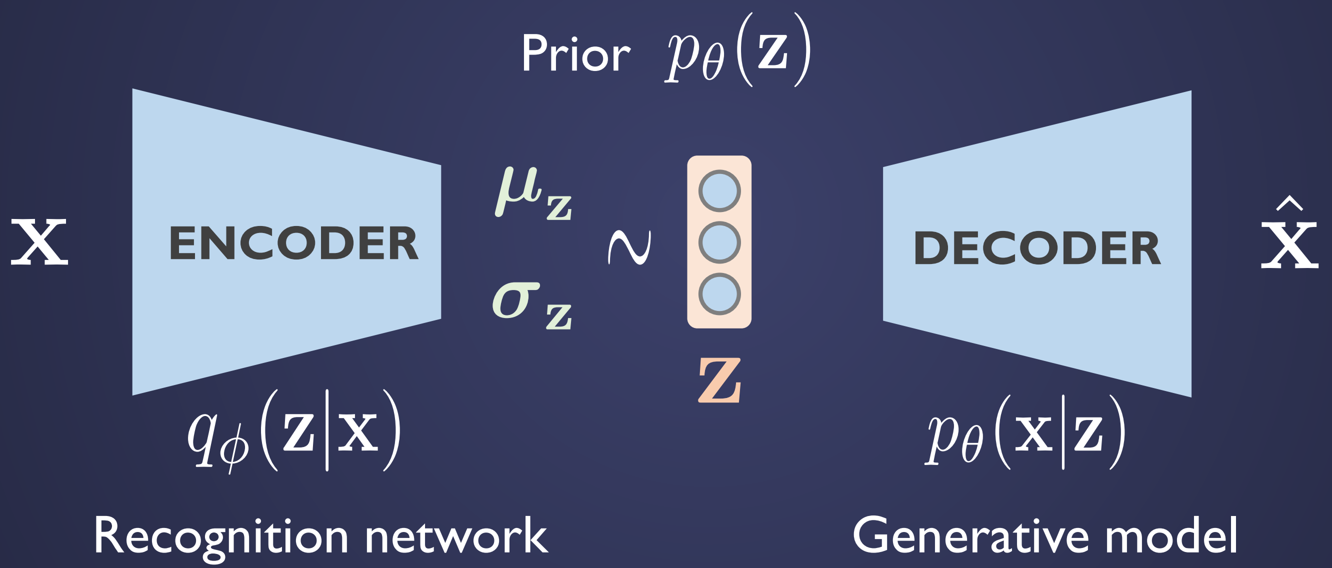
# Bayesian Deep Learning




Source: Keynote by Yee Whye Teh @ NIPS'17

# Variational Autoencoders

Kingma and Welling, 2014

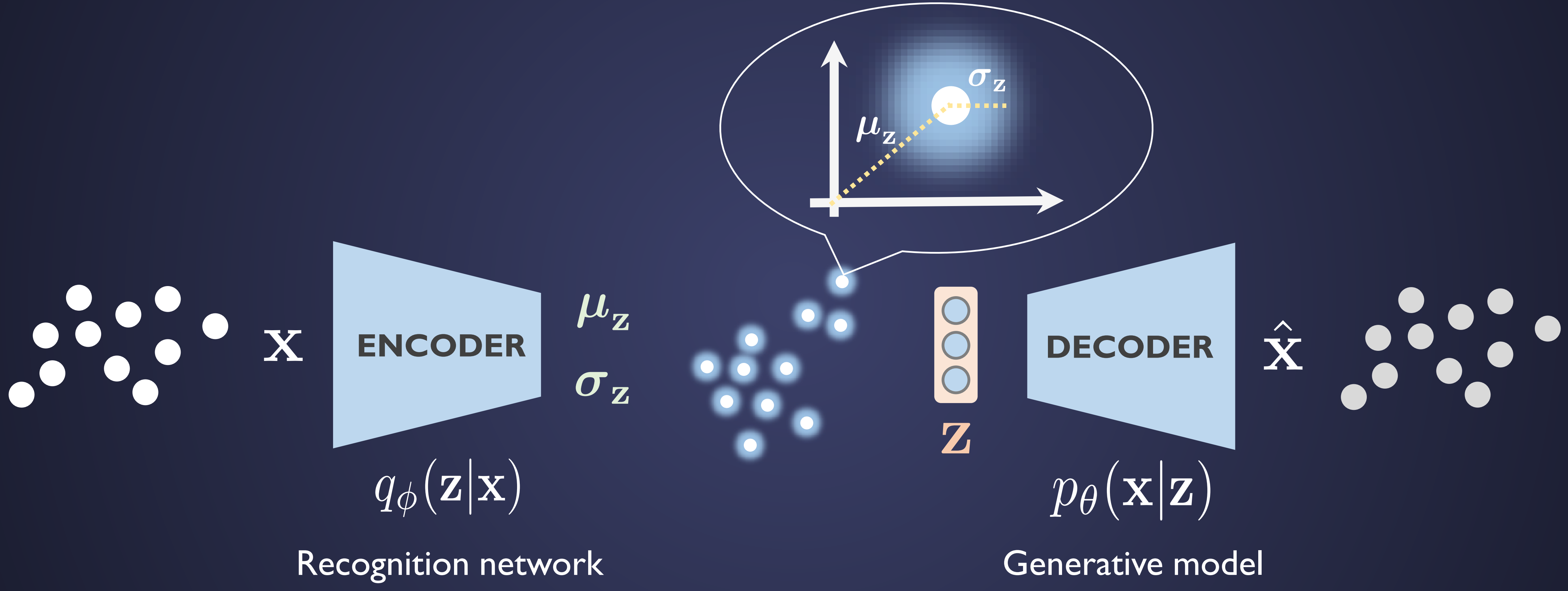



$$\mathbf{z} = \boldsymbol{\mu}_\mathbf{z} + \boldsymbol{\sigma}_\mathbf{z}\epsilon$$

$$\epsilon \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I})$$

Reparameterization trick

# Variational Autoencoders



# Variational Autoencoders

Kingma and Welling, 2014

We would like:

$$p_\theta(\mathbf{x}) = \int_{\mathbf{z}} \underbrace{p_\theta(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{z})}_{p(\mathbf{x},\mathbf{z})}\, d\mathbf{z} \longrightarrow \text{Intractable} ☹$$

Build a tractable lower bound using *amortized variational inference:*

$$\log p_\theta(\mathbf{x}) = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathcal{D}_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})\right)}_{=\mathcal{L}_{\mathrm{ELBO}}(\theta,\phi;\mathbf{x})} + \underbrace{\mathcal{D}_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})\right)}_{\geq 0}$$

# Variational Autoencoder

Kingma and Welling, 2014

**Objective**: Maximize the Evidence Lower Bound (ELBO)

$$\log p_\theta(\mathbf{x}) \geq \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathcal{D}_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})\right)}_{\mathcal{L}_{\mathrm{ELBO}}(\theta, \phi;\ \mathbf{x})}$$

Reconstruction term

$$\propto -||\mathbf{x} - \mu_{\mathbf{x}}||^2$$

Regularization term

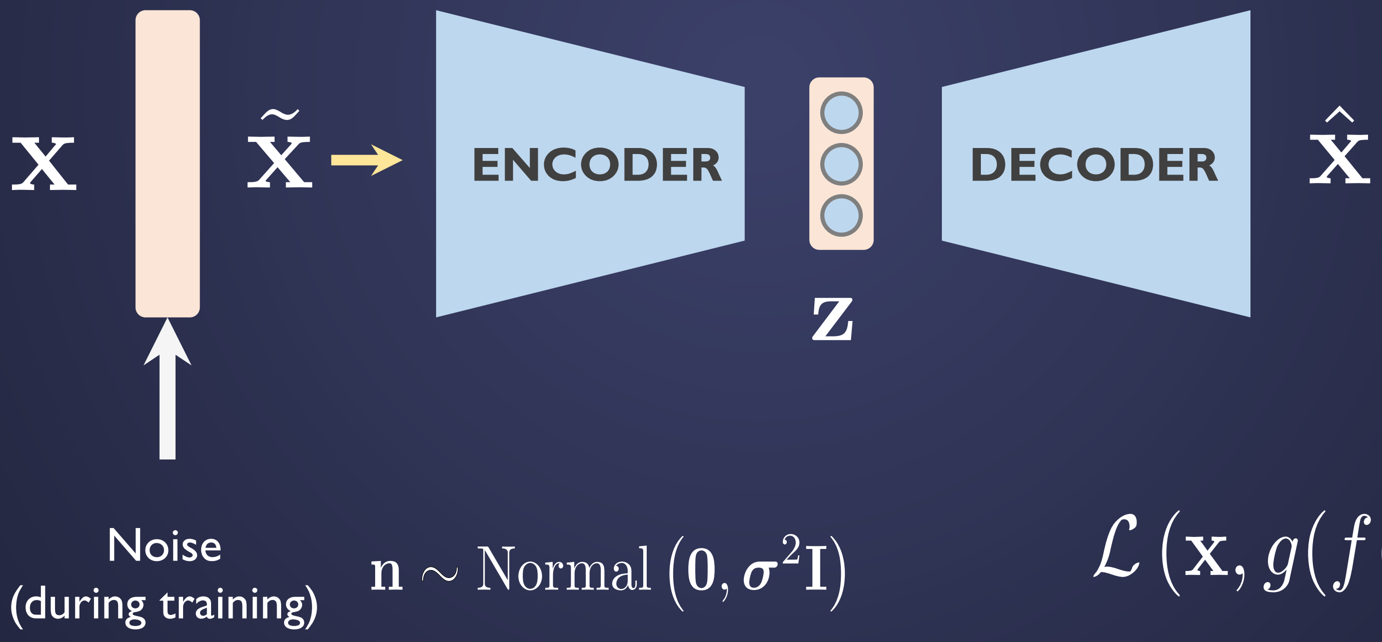
# Which encoder/decoder?

| | | |
|---|---|---|
| **~ iid** | ⟶ | Feed-forward NN |
| **Sequences** | ⟶ | Recurrent NN (e.g., LSTM, GRU) |
| **Images** | ⟶ | Convolutional NN (e.g., ResNet, VGG16) |
| **Graphs** | ⟶ | Graph NN (e.g., GCN) |

# Regularization (I)

**Denoising criterion**: learn to reconstruct $\mathbf{x}$ from a corrupted version $\tilde{\mathbf{x}}$.

Bengio *et al.*, 2015

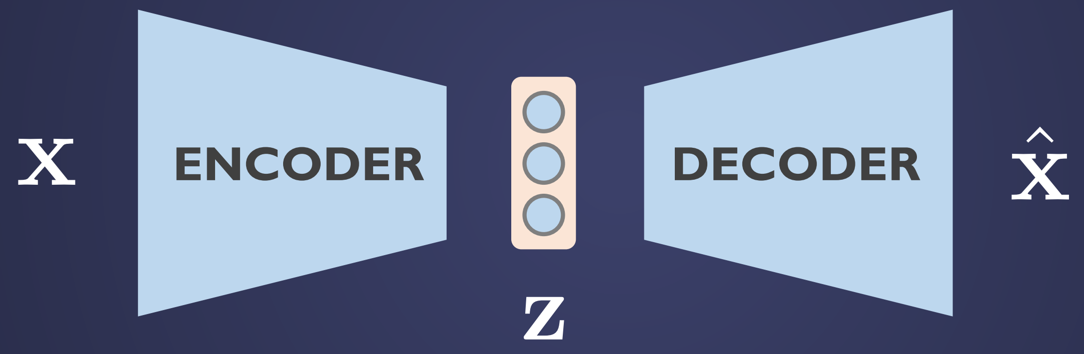



$\mathbf{n} \sim \mathrm{Normal}\left(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}\right)$

$\mathcal{L}\left(\mathbf{x}, g\left(f\left(\tilde{\mathbf{x}}\right)\right)\right)$

# Regularization (2)

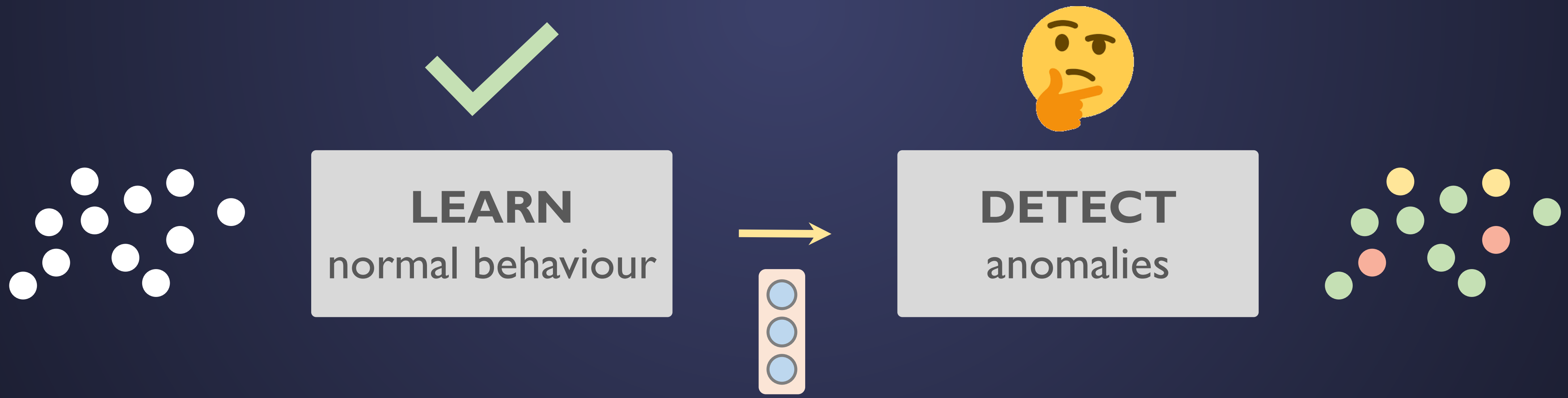
**Representation sparsity:** promote a sparse $\mathbf{z}$.




$$\mathcal{L}\left(\mathbf{x}, g(f(\tilde{\mathbf{x}}))\right) + \Omega(\mathbf{z}) \qquad \text{e.g., } \Omega\left(\mathbf{z}\right) = \lambda \|\mathbf{z}\|_1$$

# Now, we have a data representation (z)...

# How do we detect anomalies?



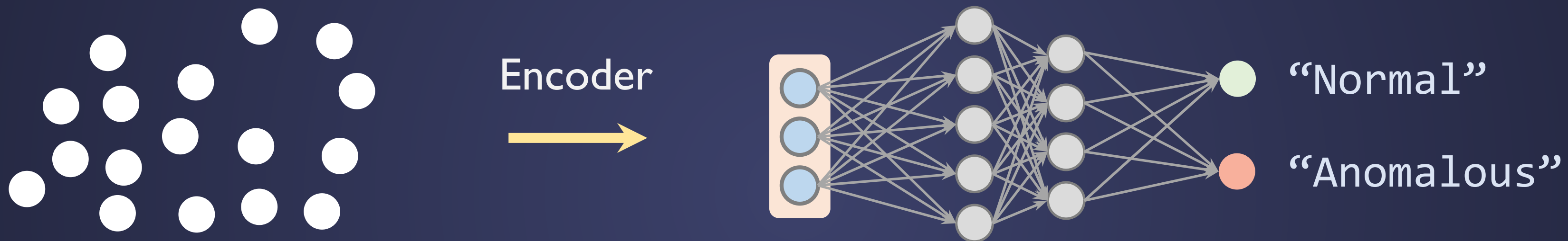
# Detection Strategy

**Availability of labels**

Scarce — Abundant

Unsupervised — Semi-supervised — Supervised

# Supervised Detection

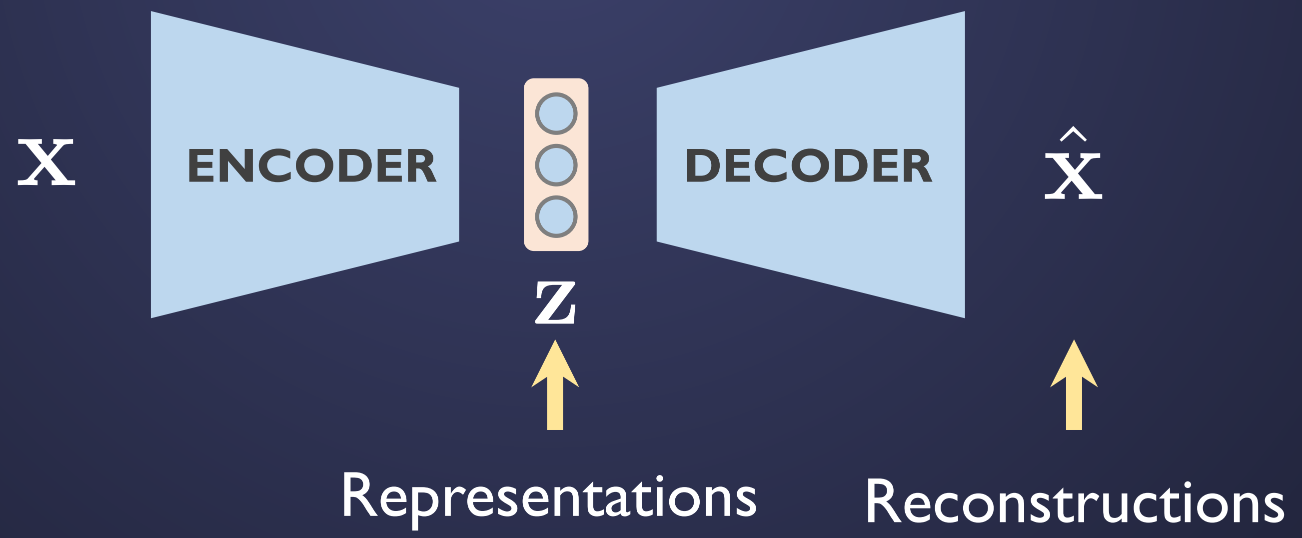



$\mathcal{X}$ $\mathcal{Z}$

# Unsupervised Detection

**Philosofy**:

- VAE trained on mostly normal data
- **Reconstruction quality** for anomalies is **worst** → **M**ethod 1
- Anomalies are **represented differently** in $\mathcal{Z}$ → **M**ethod 2



# Unsupervised Detection

## **Method 1 – Reconstruction Quality**

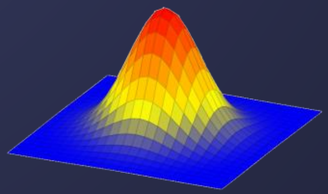
Reconstruction Error

$$\frac{1}{L}\sum_{l=1}^{L}\left\|\mathbf{x}-\underbrace{\mathbb{E}\left[p_\theta\left(\mathbf{x}|\mathbf{z}_l\right)\right]}_{\boldsymbol{\mu}_\mathbf{x}}\right\|_1$$

"Reconstruction Probability"

$$\frac{1}{L}\sum_{l=1}^{L}\log p(\mathbf{x}|\mathbf{z}_l)$$

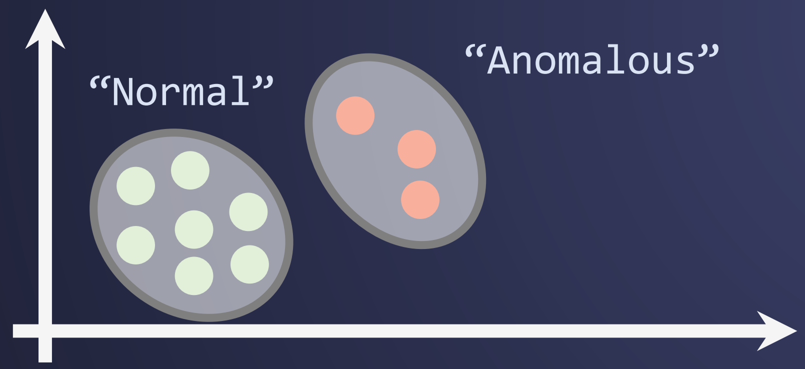
L Monte Carlo samples $\mathbf{z}_l \sim q_\phi(\mathbf{z}|\mathbf{x})$



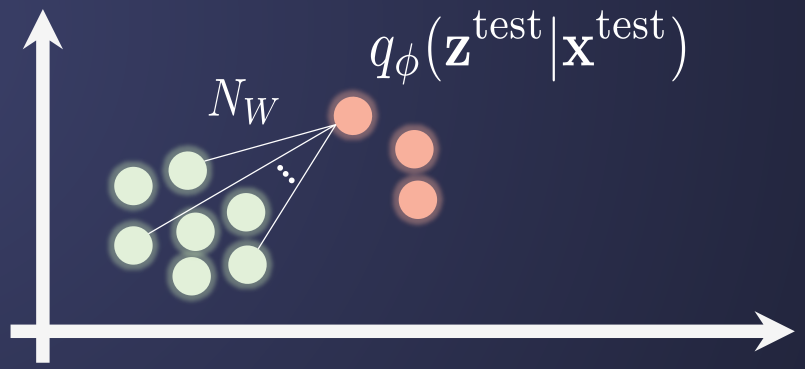
# Unsupervised Detection

## Method 2 – Latent Space

Clustering

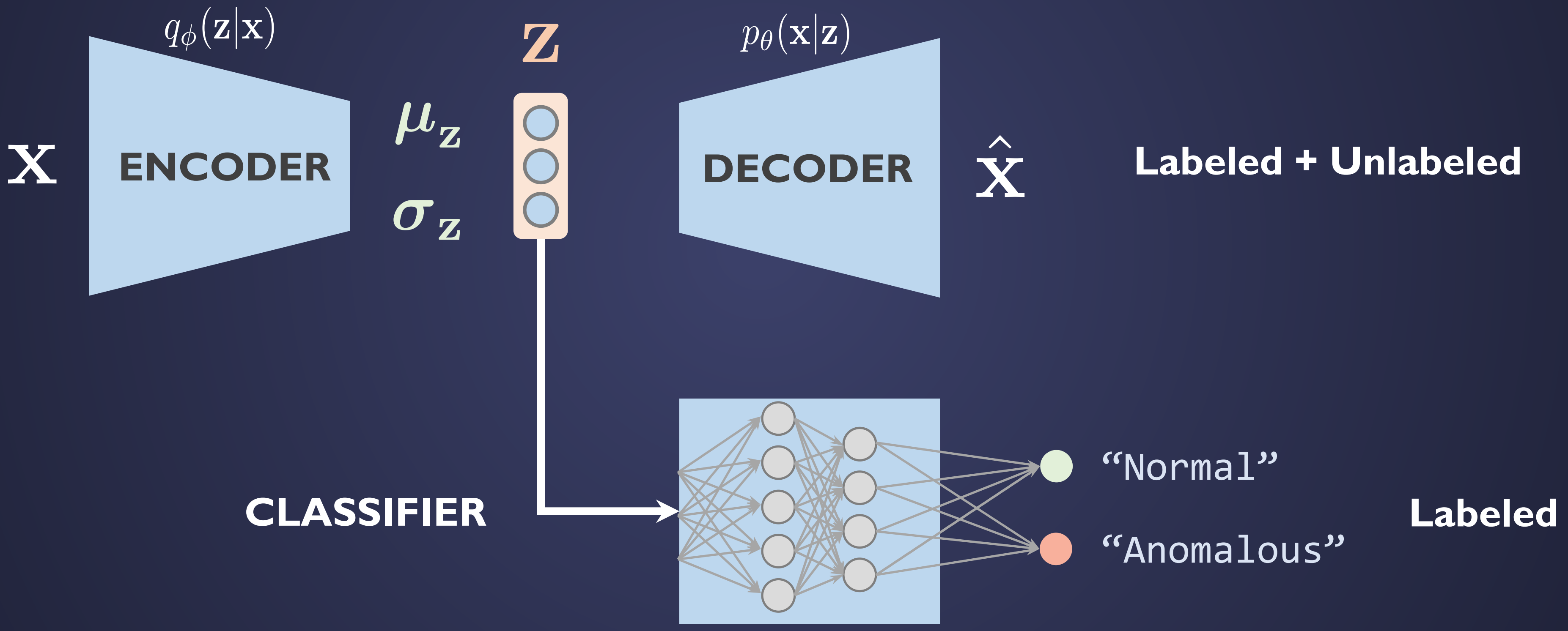



Wasserstein distance




$$\text{median}\{W(\mathbf{z}^{\text{test}}, \mathbf{z}^{i})^2\}_{i=1}^{N_W}$$

# Semi-supervised learning with VAEs



1. Introduction
2. VAEs
3. **Applications**

# Applications

Sensor time series
Brain images
Network graphs

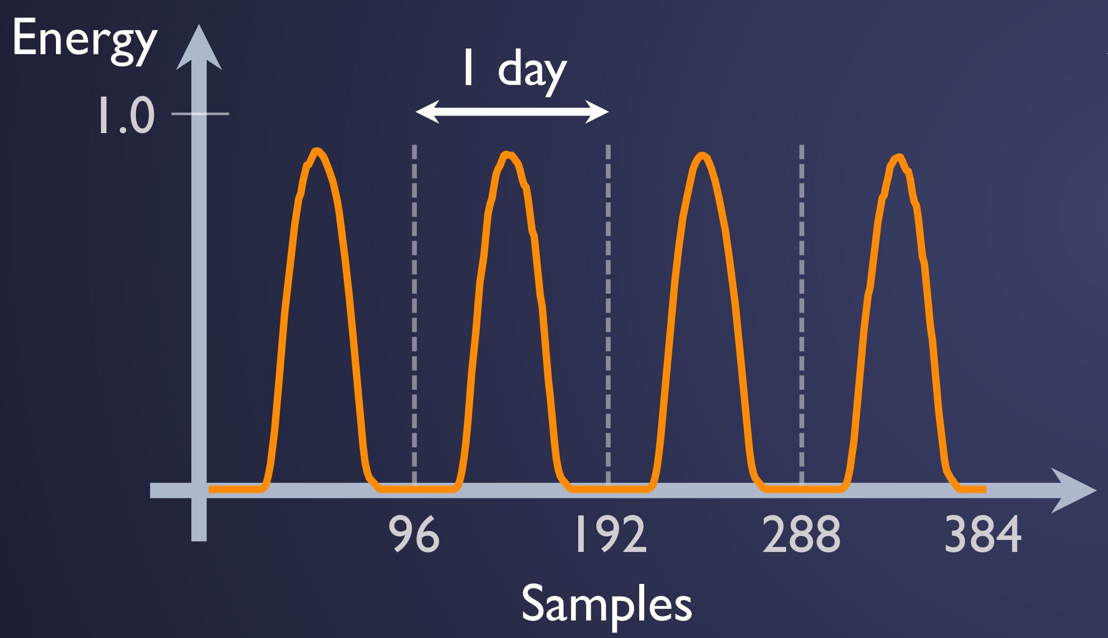
# Example 1

## Sensor Time Series

# Example 1 – Sensor Time Series

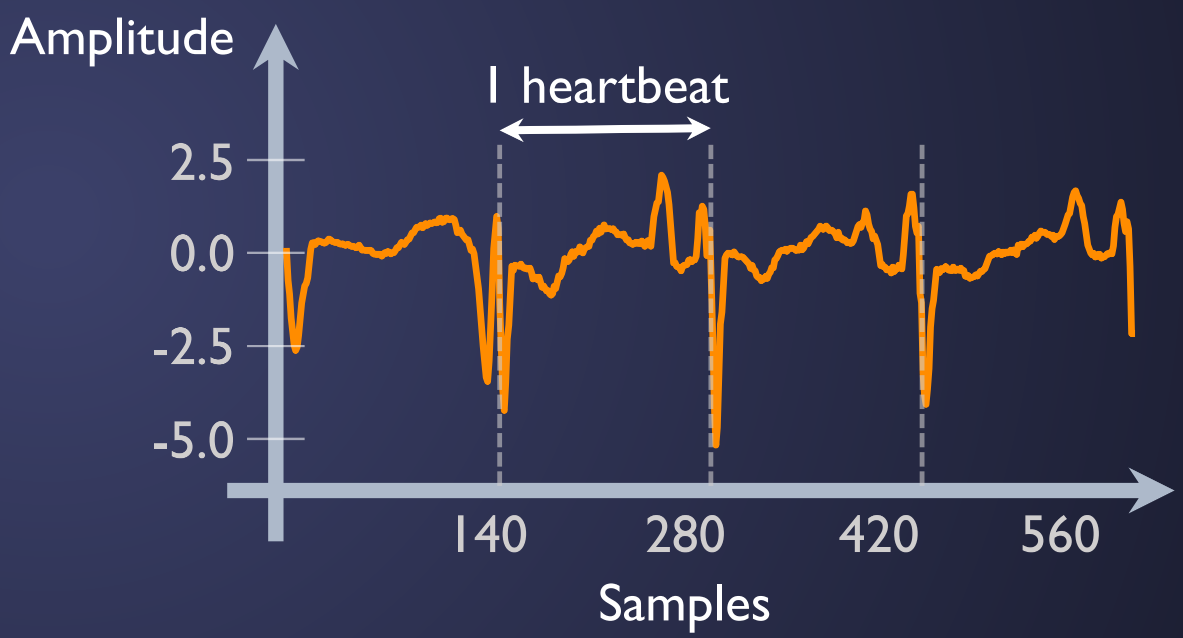
Pereira & Silveira, 2018

**Solar PV energy generation**




Proprietary dataset
Unlabeled

**Electrocardiogram**



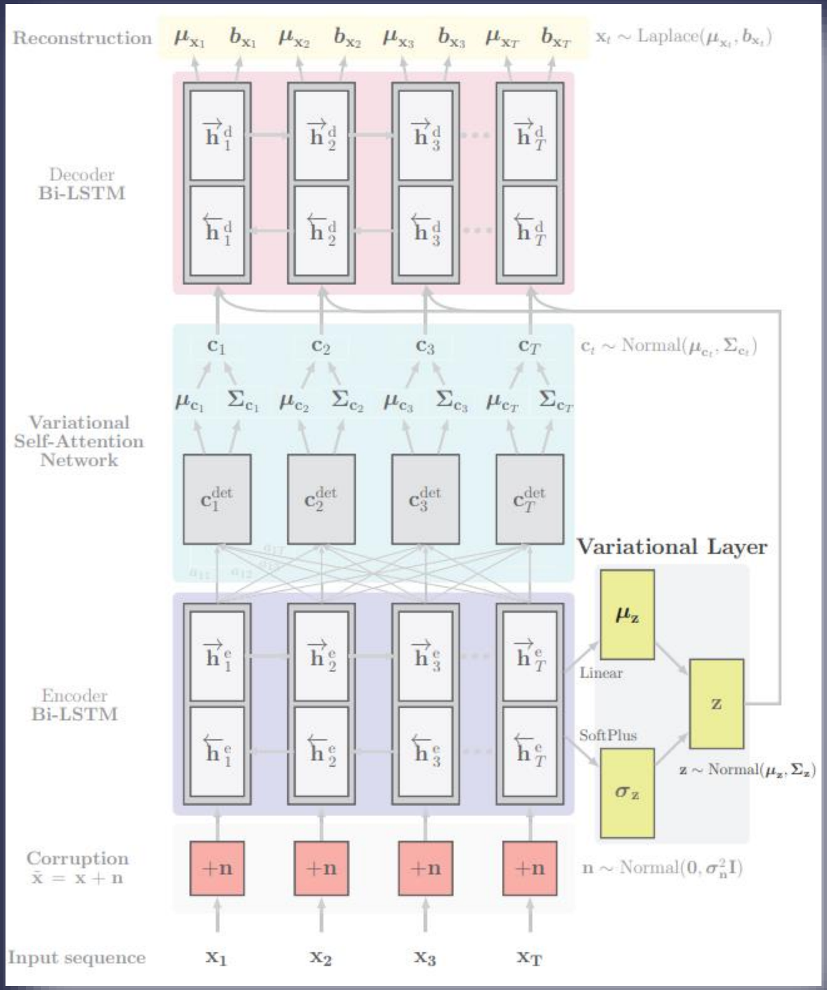

Public ECG5000
Labeled

# Example 1 – Sensor Time Series

Pereira & Silveira, 2018

What does this reminds you of?!

Seq2Seq + attention



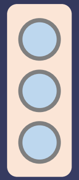
# Example 1 - Sensor Time Series

Pereira & Silveira, 2018

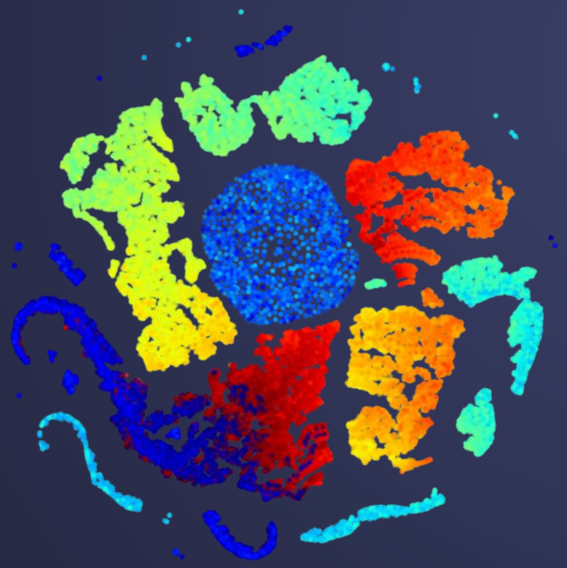
**Solar Energy,** Method 1 – Reconstruction Quality

T=12 (<96)
dim(z)=3

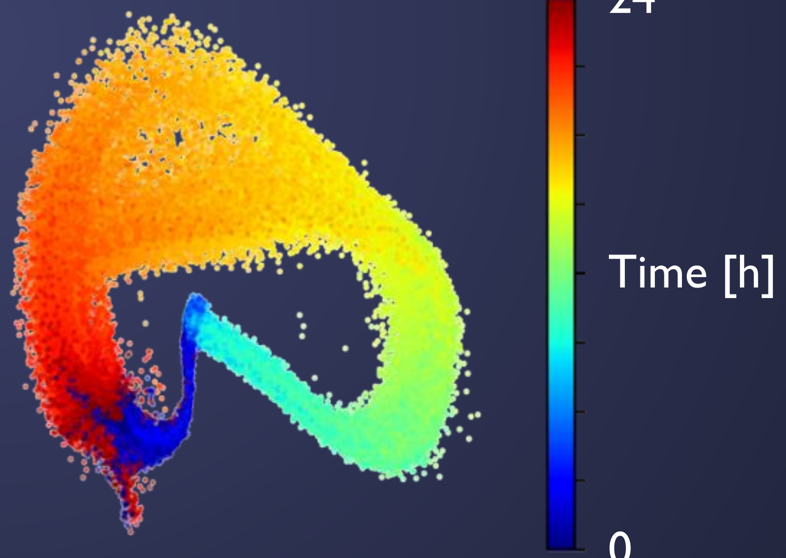
**Variational Latent Space**



PCA → 2D
t-SNE



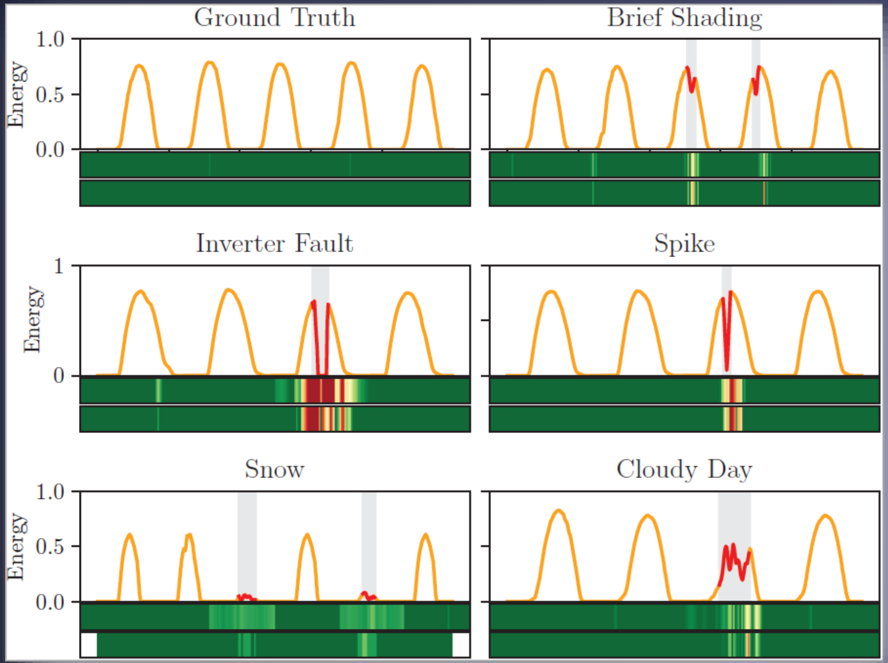
t-SNE

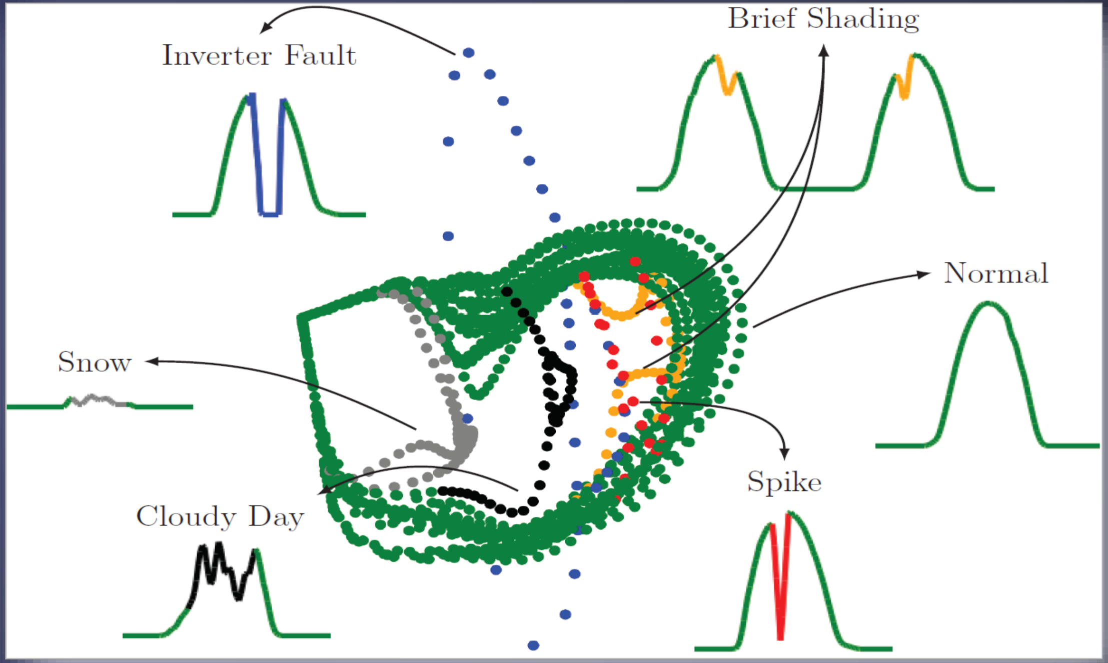



PCA

# Example 1 - Sensor Time Series

**Solar Energy,** Method 1 – Reconstruction Quality




Top bar: reconstruction error
Bottom bar: reconstruction probability

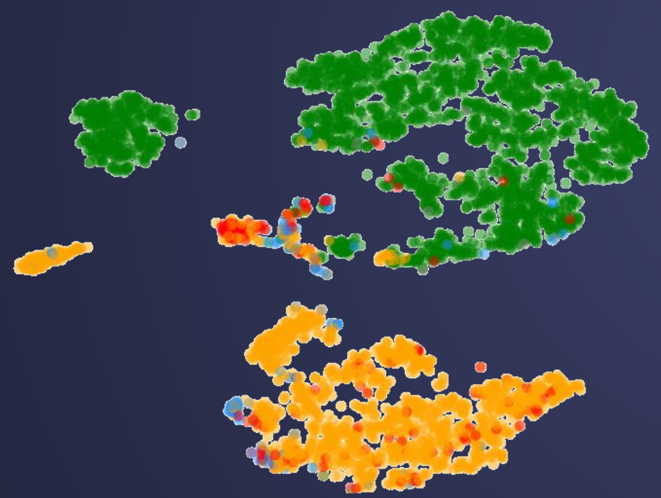
| Reconstruction Error | "Reconstruction Probability" |
|---|---|
| $\frac{1}{L}\sum_{l=1}^{L}\left\Vert \mathbf{x} - \underbrace{\mathbb{E}\left[p_\theta\left(\mathbf{x}\vert\mathbf{z}_l\right)\right]}_{\mu_\mathbf{x}}\right\Vert_1$ | $\frac{1}{L}\sum_{l=1}^{L}\log p(\mathbf{x}\vert\mathbf{z}_l)$ |

# Example 1 - Sensor Time Series

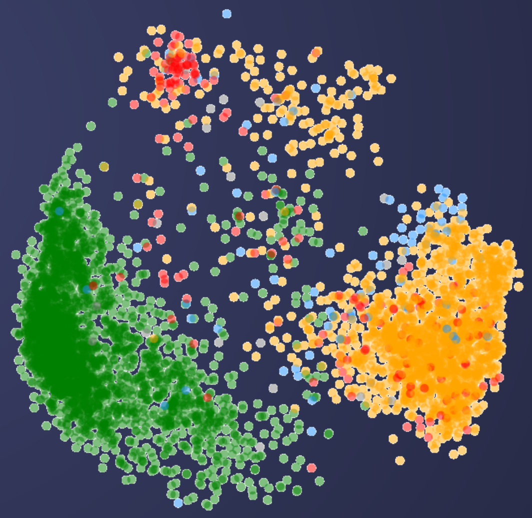
**ECG5000,** Method 2 – Latent Space

● "Normal"

T=140
dim(z)=5



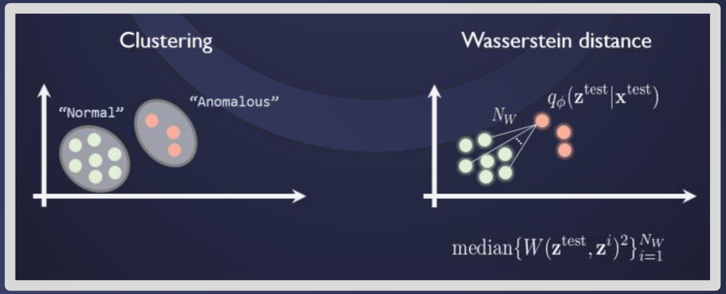
t-SNE



PCA

# Example 1 - Sensor Time Series

**ECG5000,** Method 2 – Latent Space



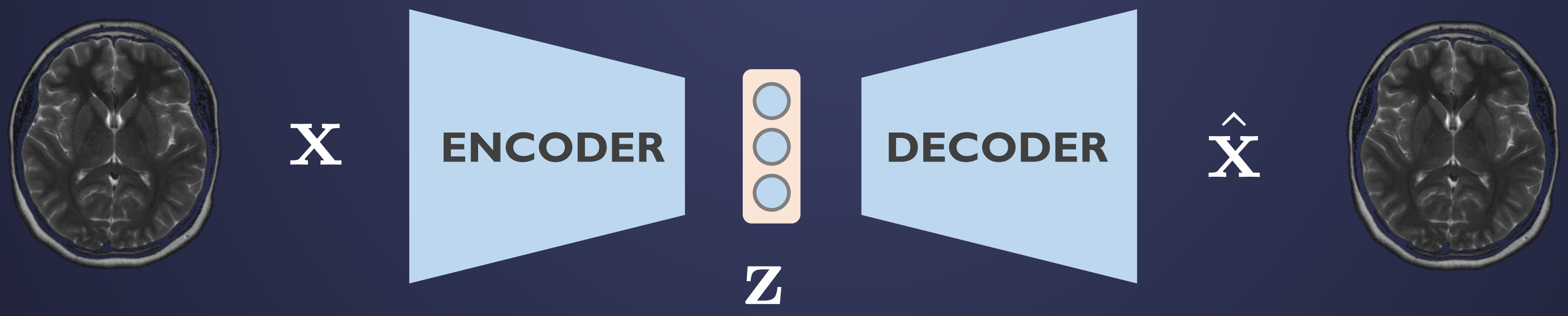

| Source | S/U | Model | AUC | Acc | F1 |
|---|---|---|---|---|---|
| Proposed | S | VRAE+SVM | 0.9836 | 0.9843 | 0.9844 |
| | **U** | **VRAE+Clust/W** | **0.9819** | **0.9596** | **0.9522** |
| Lei *et al.*, 2017 | S | SPIRAL-XGB | 0.9100 | - | - |
| Karim *et al.*, 2017 | S | F-t ALSTM-FCN | - | 0.9496 | - |
| Malhotra *et al.*, 2017 | S | SAE-C | - | 0.9340 | - |
| Liu *et al.*, 2018 | U | oFCMdd | - | - | 0.8084 |

# Example 2

Brain Images

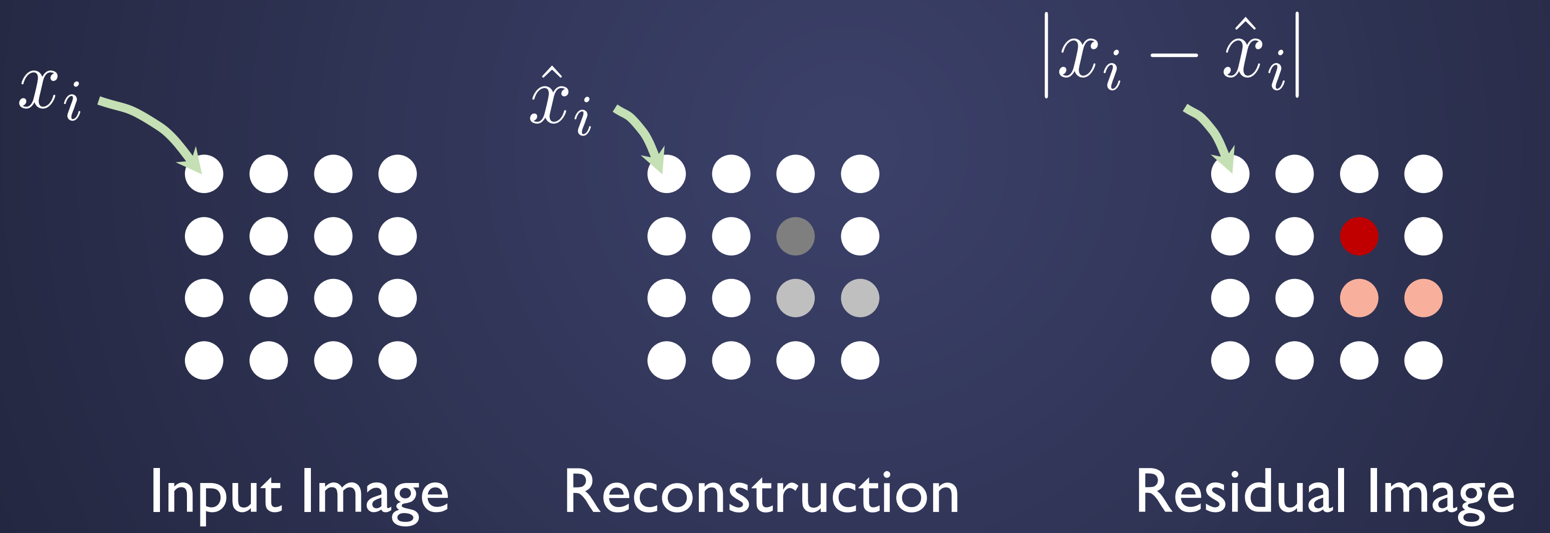
# Example 2 – Brain Images

- Detect **brain lesions**: trauma, infection, cancer...
- Early detection is crucial.
- Magnetic Ressonance Images (MRI)



# Example 2 – Brain Images

**Anomaly score**: pixel-wise reconstruction error



# Example 2 – Brain Images

Unsupervised Detection of Lesions in Brain MRI Using Constrained Adversarial Auto-encoders, Chen & Konukoglu, 2018

**Models**

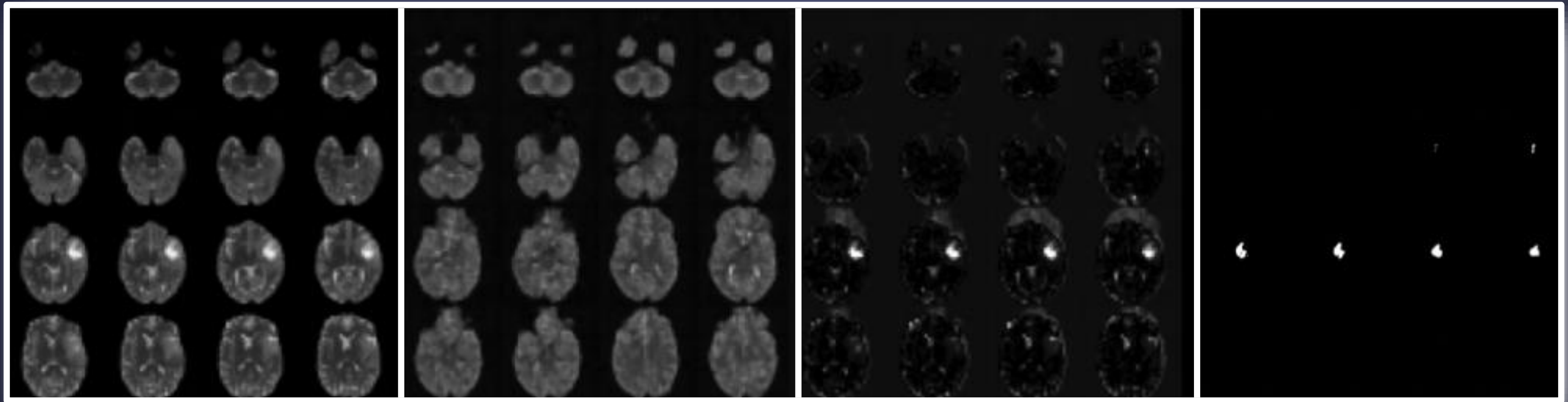
- Variational Autoencoder
- Adversarial Autoencoder

**Regularization**

- "Representation consistency" $\lambda \|\mathbf{z} - \hat{\mathbf{z}}\|^2$

# Example 2 – Brain Images

$\mathbf{x}$ $\hat{\mathbf{x}}$ $|\mathbf{x} - \hat{\mathbf{x}}|$ Ground truth



Unsupervised Detection of Lesions in Brain MRI Using Constrained Adversarial Auto-encoders, Chen & Konukoglu, 2018

# Example 3

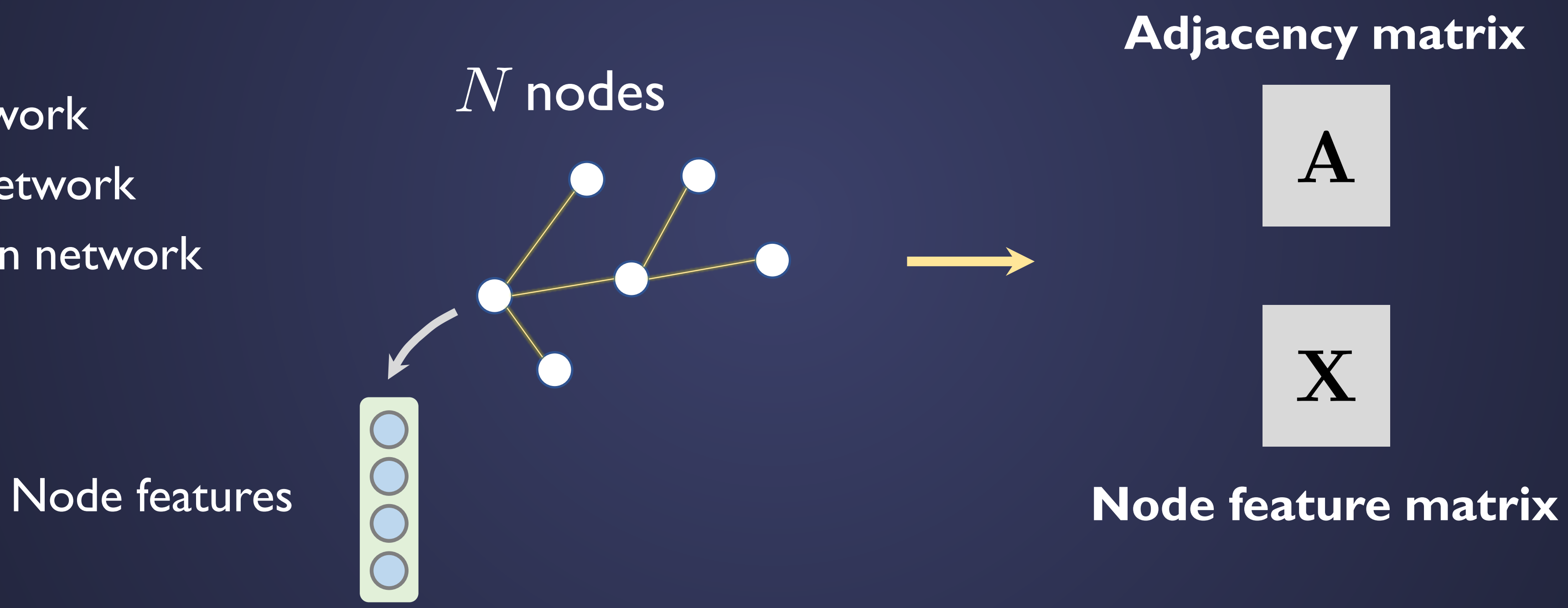
## Network Graphs

# Example 3 – Network Graphs

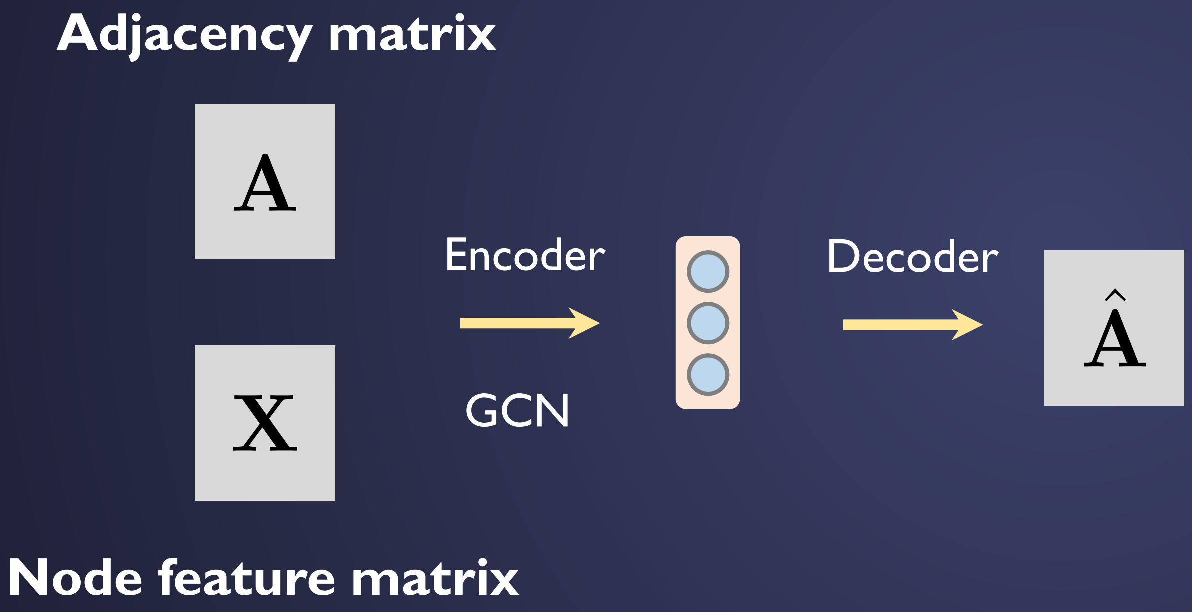
Social network

Citation network

Transaction network



# Example 3 – Network Graphs




Anomaly score: $NLL(\mathbf{A}, \hat{\mathbf{A}})$

Semi-supervised classification using GCN



Variational Graph Autoencoder, Kipf & Welling, 2016

# Take home messages

- Deep learning is about representation learning
- Anomaly detection is not solved
- VAEs are flexible
- Scale well to big data
- Deal with class imbalance

# Take home messages

Anomaly Detection ❤ Deep Learning

# References

## Variational Autoencoder


- Auto-Encoding Variational Bayes, Kingma & Welling, 2014 (Link)
- Stochastic Backpropagation and Approximate Inference in Deep Generative Models, Rezende et al., 2014 (Link)
- Denoising Criterion for Variational Autoencoding Framework, Bengio et al., 2015 (Link)


## Semi-supervised Learning


- Semi-Supervised Learning with Deep Generative Models, Mohamed et al., 2014 (Link)
- Adversarial Autoencoders, Goodfellow et al., 2015 (Link)

# References

## Anomaly Detection in Time Series


- LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection, Malhotra et al., 2015 (Link)
- Variational Inference for On-line Anomaly Detection in High-Dimensional Time Series, Bayer et al., 2016 (Link)


## Graph Convolutional Networks and VGAE


- Deep Learning with Graph-structured Representations, Kipf, 2020 (Link)
- Variational Graph Auto-Encoders, Kipf & Welling, 2016 (Link)

# References

## Anomaly Detection in Images


- Unsupervised Detection of Lesions in Brain MRI Using Constrained Adversarial Auto-encoders, Chen & Konukoglu, 2018 (Link)


## Anomaly Detection in Graphs


- Deep Anomaly Detection on Attributed Networks, Ding et al., 2019 (Link)


## My works (Link)

# Thank you for your attention!

mail@joao-pereira.pt
www.joao-pereira.pt