Machine Learning for Anomaly Detection

MLE SCHOOL'22
Machine Learning in Engineering Summer School @ TUHH

Antonin Sulc

Hamburg, September 13, 2022

CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

DESY.

TUHH
Technische Universität Hamburg

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

What's anomaly

- a person or thing that is different from what is usual, or not in agreement with something else and therefore not satisfactory [Cambridge Dictionary]

What's anomaly

- a person or thing that is different from what is usual, or not in agreement with something else and therefore not satisfactory [Cambridge Dictionary]
- a thing, situation, etc. that is different from what is normal or expected [Oxford Dictionary]

What's anomaly

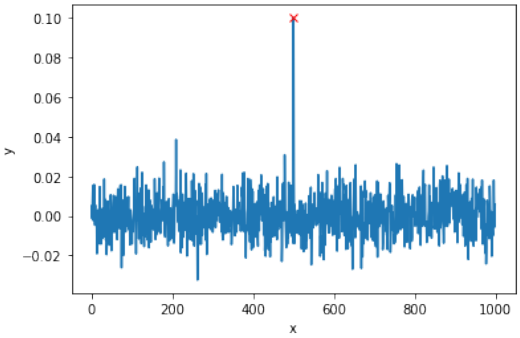
- a person or thing that is different from what is usual, or not in agreement with something else and therefore not satisfactory [Cambridge Dictionary]
- a thing, situation, etc. that is different from what is normal or expected [Oxford Dictionary]
- something different, abnormal, peculiar, or not easily classified [Merram-Webster]

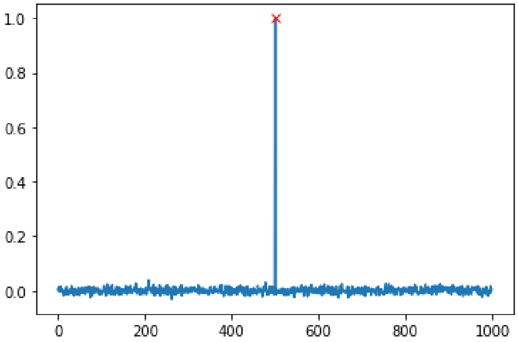
What's anomaly

- a person or thing that is different from what is usual, or not in agreement with something else and therefore not satisfactory [Cambridge Dictionary]
- a thing, situation, etc. that is different from what is normal or expected [Oxford Dictionary]
- something different, abnormal, peculiar, or not easily classified [Merram-Webster]
- …
- A person with no perceptible faults, who consistently fails at achieving matches thru all forms of social media [Urban Dictionary]

Point Anomaly

a single data sample that can be considered anomalous compared to the rest of the data [Wittkopp et al.(2022)Wittkopp, Wiesner, Scheinert, and Kao]





Point Anomaly - Cont'd

Point Anomaly - Cont'd

CNBC
Search quotes, news & videos
WATCHLIST | SIGN IN
CREATE FREE ACCOUNT
MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO
MAKE IT
USA • INTL

EUROPE ECONOMY

Euro zone inflation hits another record of 9.1% as food and energy prices soar

... Eurozone inflation hit a new record high in August of 9.1%, according to flash figures from Europe's statistics office Eurostat, with high energy prices the main driving force. Werknghwrjhb werbn re4rhn. The rate was above expectations, with a Reuters poll of economists anticipating a rate of 9%... [Ward-Glenton(2022)]

Point Anomaly - Cont'd

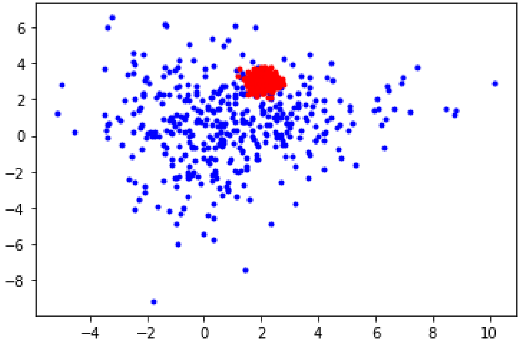
EUROPE ECONOMY

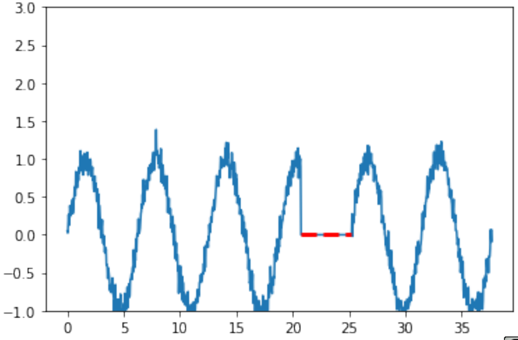
Euro zone inflation hits another record of 9.1% as food and energy prices soar

... Eurozone inflation hit a new record high in August of 9.1%, according to flash figures from Europe’s statistics office Eurostat, with high energy prices the main driving force. Werknghwrjhb werbn re4rhn. The rate was above expectations, with a Reuters poll of economists anticipating a rate of 9%... [Ward-Glenton(2022)]

Group Anomaly

a multiple data samples that can be considered anomalous compared to the rest of the data

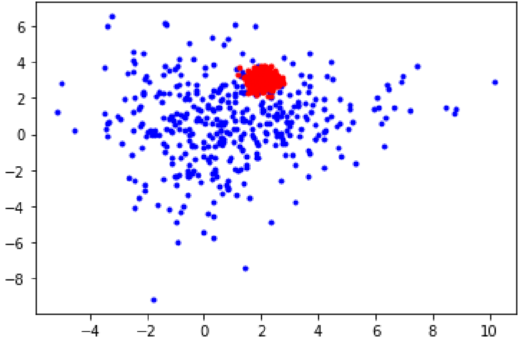


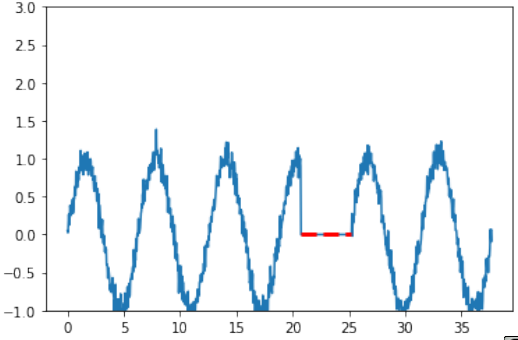


Group Anomaly

a multiple data samples that can be considered anomalous compared to the rest of the data

> individual samples are often not interesting,

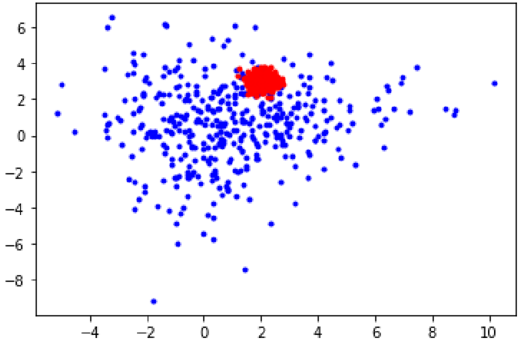


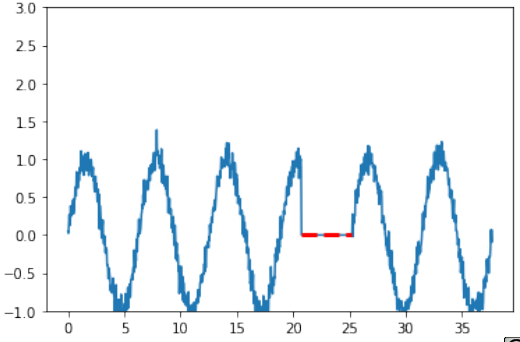


Group Anomaly

a multiple data samples that can be considered anomalous compared to the rest of the data

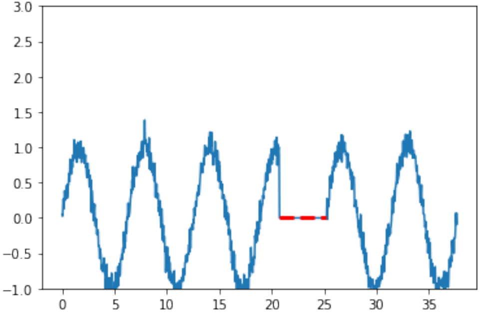
- individual samples are often not interesting,
- anomalous group of signals is an noticeably dense with respect to entire dataset.





Contextual Anomaly

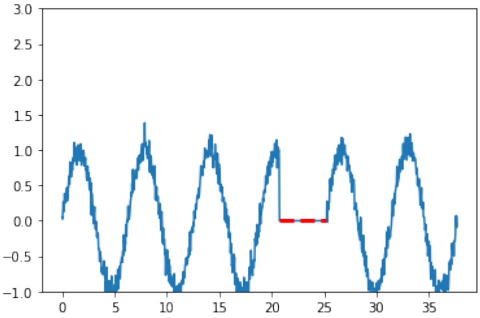
samples that are anomalous in a specific context only (but not otherwise) are called contextual anomalies[Wittkopp et al.(2022)Wittkopp, Wiesner, Scheinert, and Kao]



Contextual Anomaly

samples that are anomalous in a specific context only (but not otherwise) are called contextual anomalies[Wittkopp et al.(2022)Wittkopp, Wiesner, Scheinert, and Kao]

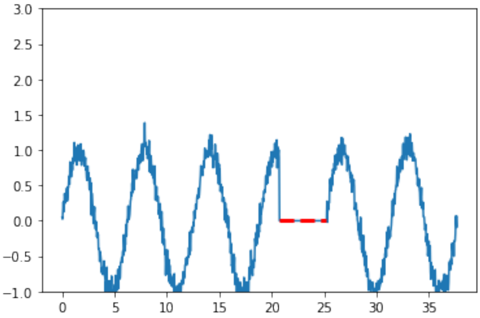
> can have the same feature-set (behavioral properties) as normal samples,



Contextual Anomaly

samples that are anomalous in a specific context only (but not otherwise) are called contextual anomalies[Wittkopp et al.(2022)Wittkopp, Wiesner, Scheinert, and Kao]

- can have the same feature-set (behavioral properties) as normal samples,
- but are still anomalous within a specific context defined by their contextual properties.



Contextual Anomaly - Cont'd

CNBC

Search quotes, news & videos | WATCHLIST | SIGN IN | CREATE FREE ACCOUNT

MARKETS | BUSINESS | INVESTING | TECH | POLITICS | CNBC TV | INVESTING CLUB | PRO | MAKE IT | USA • INTL

EUROPE ECONOMY

Euro zone inflation hits another record of 9.1% as food and energy prices soar

Contextual Anomaly - Cont'd

... Eurozone inflation hit a new record high in August of 9.1%, according to flash figures from Europe's statistics office Eurostat, with high energy prices the main driving force. We enjoyed delicious pizza with a pineapple. The rate was above expectations, with a Reuters poll of economists anticipating ...

Contextual Anomaly - Cont'd

CNBC

Search quotes, news & videos | WATCHLIST | SIGN IN | CREATE FREE ACCOUNT

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE IT USA · INTL

EUROPE ECONOMY

Euro zone inflation hits another record of 9.1% as food and energy prices soar

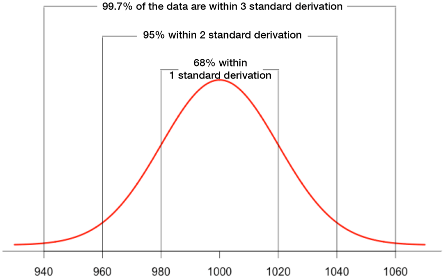
... Eurozone inflation hit a new record high in August of 9.1%, according to flash figures from Europe's statistics office Eurostat, with high energy prices the main driving force. We enjoyed delicious pizza with a pineapple. The rate was above expectations, with a Reuters poll of economists anticipating ...

Roll your sleeves!

https://github.com/sulcantonin/MLE2022

Anomaly Detection - Basics - 3-Sigma Rule

> Mean μ defines an average value ($\mu = 1000$).

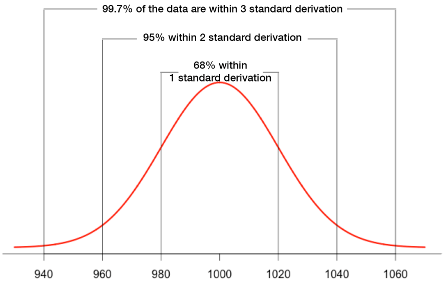


[Magakian(2018)]

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Basics.ipynb

Anomaly Detection - Basics - 3-Sigma Rule

- Mean μ defines an average value ($\mu = 1000$).
- Standard deviation σ, defines how far the normal distribution is spread around the mean ($\sigma = 20$).

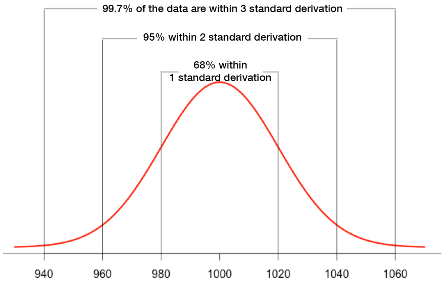


[Magakian(2018)]

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Basics.ipynb

Anomaly Detection - Basics - 3-Sigma Rule

- Mean μ defines an average value ($\mu = 1000$).
- Standard deviation σ, defines how far the normal distribution is spread around the mean ($\sigma = 20$).
- 68% of all values fall between $[\mu - \sigma, \mu + \sigma]$, i.e. [980, 1020].

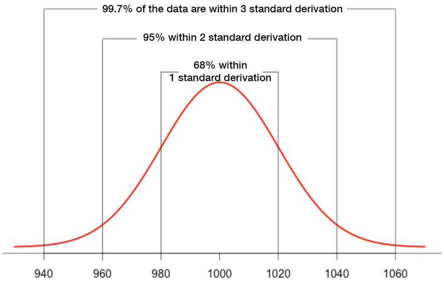


[Magakian(2018)]

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Basics.ipynb

Anomaly Detection - Basics - 3-Sigma Rule

- Mean μ defines an average value ($\mu = 1000$).
- Standard deviation σ, defines how far the normal distribution is spread around the mean ($\sigma = 20$).
- 68% of all values fall between $[\mu - \sigma, \mu + \sigma]$, i.e. [980, 1020].
- 95% of all values fall between $[\mu - 2\sigma, \mu + 2\sigma]$, i.e. [960, 1040].

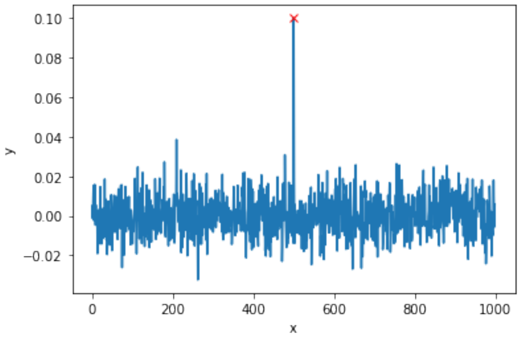


[Magakian(2018)]

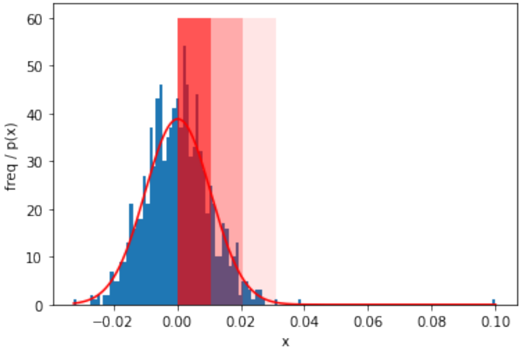
See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Basics.ipynb

Anomaly Detection - 3-Sigma Example

Plot



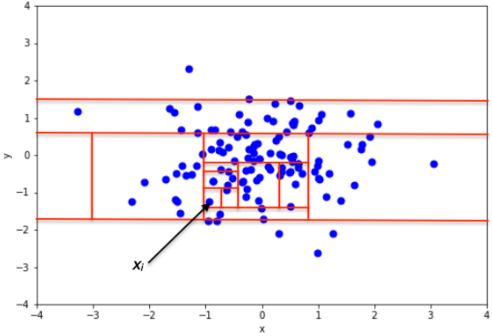
Histogram

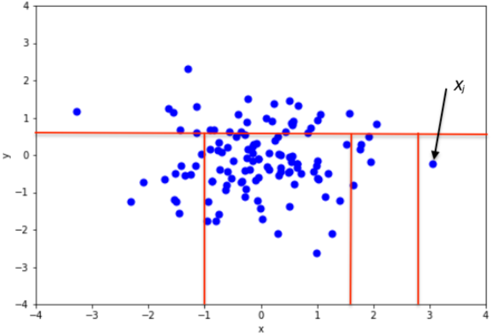


See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Basics.ipynb

Anomaly Detection - Isolation Forest

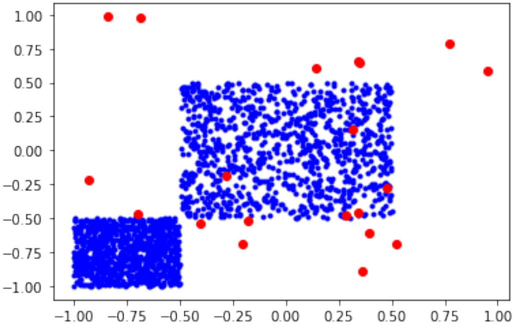
Isolation forest split the data space using lines that are orthogonal to the origin, and assigns higher anomaly scores to data points that need few splits to be isolated. [wik(2022)]

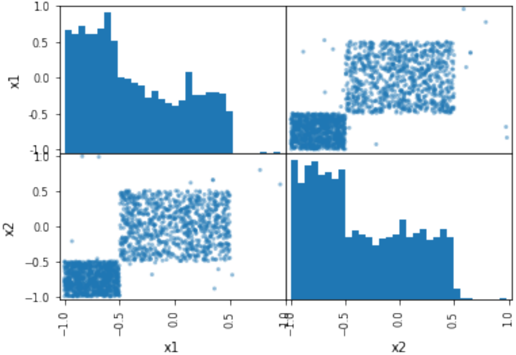




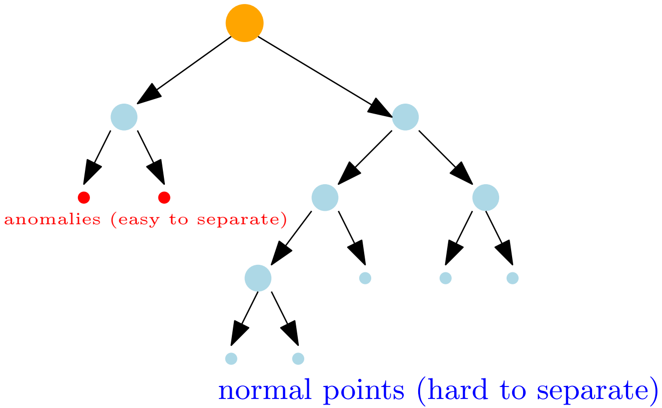
Anomaly Detection - Isolation Forest

What if the points are not normally distributed?

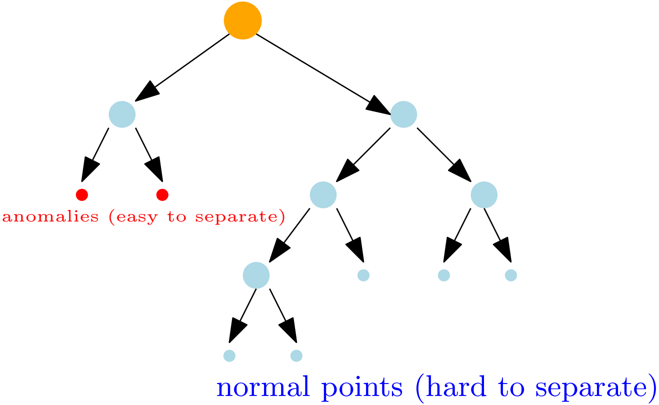


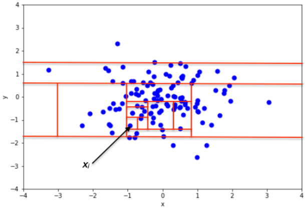


Anomaly Detection - Isolation Forest

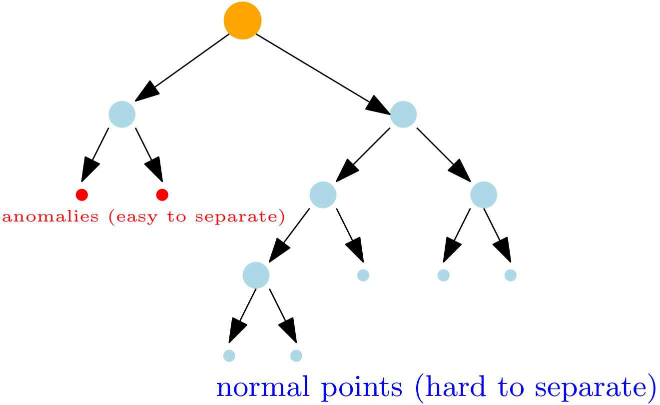


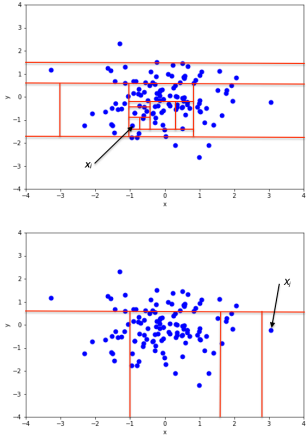
Anomaly Detection - Isolation Forest





Anomaly Detection - Isolation Forest



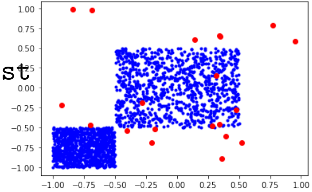


Anomaly Detection - Isolation Forest Example

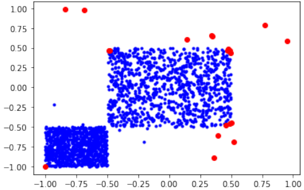
```python
from sklearn.ensemble import IsolationForest
import numpy.random as r
r.seed(42)
Xn1 = r.rand(1000,2) - 0.5
Xn2 = (r.rand(1000,2) - 2) * 0.5
Xn = np.concatenate((Xn1,Xn2))
Xa = 2 * (r.rand(20,2) - 0.5)
X = np.concatenate((Xn,Xa))

cont= 20.0 / 2020.0 # ratio of anomalies
if = IsolationForest(contamination = cont, random_state = 42)
l = if.fit_predict(X)
```

Generated Samples



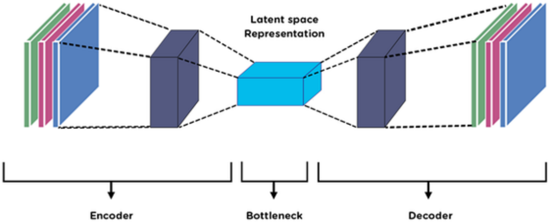
Results (Isolation Forest)



Anomaly Detection - Auto-encoder

What if we have data and want to train a model of a set of data points $\{x_1 \ldots x_N\}$?

$\mathbf{x}$ $\quad\quad$ $\mathbf{z} = enc(\mathbf{x})$ $\quad\quad$ $\hat{\mathbf{x}} = dec(\mathbf{z})$



[Birla(2019)]

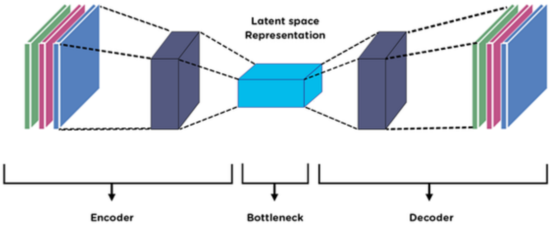
$$s = \|\mathbf{x} - dec(enc(\mathbf{x}))\|_2$$

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Autoenc.ipynb

Anomaly Detection - Auto-encoder

What if we have data and want to train a model of a set of data points $\{x_1 \ldots x_N\}$?

$\mathbf{x}$ $\quad$ $\mathbf{z} = enc(\mathbf{x})$ $\quad$ $\hat{\mathbf{x}} = dec(\mathbf{z})$



[Birla(2019)]

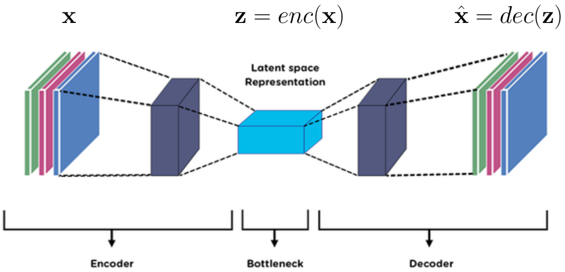
$$s = \|\mathbf{x} - dec(enc(\mathbf{x}))\|_2$$

+/- Trains the network ”generatively”.

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Autoenc.ipynb

Anomaly Detection - Auto-encoder

What if we have data and want to train a model of a set of data points $\{x_1 \dots x_N\}$?



[Birla(2019)]

$$s = \|\mathbf{x} - dec(enc(\mathbf{x}))\|_2$$

+/- Trains the network "generatively".

- Technically you are not training anomaly detection, but training a model of data

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Autoenc.ipynb

Anomaly Detection - Auto-encoder Example

```python
enc = nn.Sequential(
    nn.Linear(2, 8),
    nn.ReLU(),
    nn.Linear(8, 16))
dec = nn.Sequential(
    nn.Linear(16, 8),
    nn.ReLU(),
    nn.Linear(8, 2))
def score_reconstruction(x):
    return vector_norm(dec(enc(x)) - x, dim=-1)
```

Reconstr. loss - score

$$\downarrow \|\mathbf{x} - dec(enc(\mathbf{x}))\|_2$$

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Autoenc.ipynb

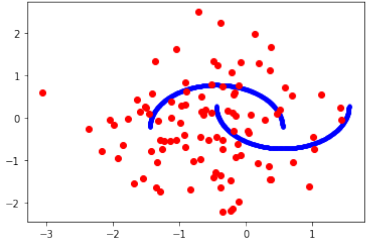
Anomaly Detection - Auto-encoder Example Cont'd

```python
optimizer = optim.Adam( list(enc.parameters()) +
                        list(dec.parameters()))
# reconstruction (X_hat)
X_hat = model(X)
# criterion ||X - X_hat||_2
loss = F.mse_loss(X_hat, X)
optimizer.zero_grad()
loss.backward()
optimizer.step()
```

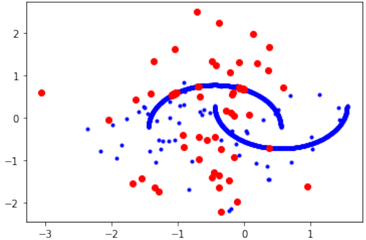
See https://github.com/sulcanton-in/MLE2022/blob/main/MLE_Autoenc.ipynbg

Anomaly Detection - Auto-encoder Example Cont'd

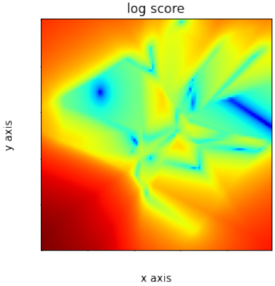
Input Points



Trained Classification



$$s = \|\mathbf{x} - dec(enc(\mathbf{x}))\|_2$$

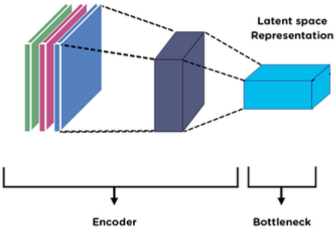


See https://github.com/sulcantonin/MLE2022/blob/main/MLE_Autoenc.ipynb

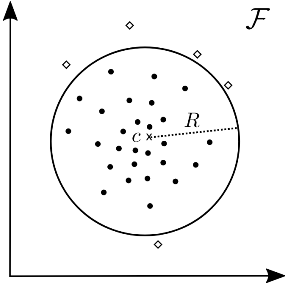
Anomaly Detection - One Class Loss (OCL)

What if we have data and want to train a model of a set of data points $\{x_1 \ldots x_N\}$?

$\mathbf{x}$ $\quad\quad$ $\mathbf{z} = enc(\mathbf{x})$



$$\begin{aligned} s &= \|enc(\mathbf{x}) - \mathbf{c}\|_2 \\ &= \|\mathbf{z} - \mathbf{c}\|_2 \end{aligned}$$



$$s = \|enc(\mathbf{x}) - \mathbf{c}\|_2$$

Networks trains to project **x** to fit the hyphersphere center **c**

[Ruff(2019)]

+ Trains the network ”discriminatively”,
- One has to be careful with trivial solutions

  i. e. $\mathbf{c} \neq 0$, no biases and unbounded non-linearity

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_OCL.ipynb

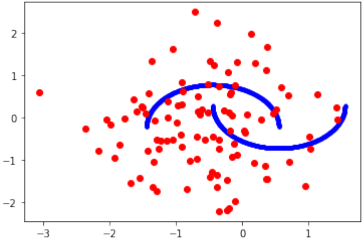
Anomaly Detection - One Class Loss

```
enc = nn.Sequential(nn.Linear(2,8, bias = False),
    nn.Linear(8,16, bias = False),
    nn.Linear(16,4, bias = False))
# random center of 4D hypersphere
c = torch.randn((1,4), requires_grad = False)
optimizer = optim.Adam(enc.parameters())
# reconstruction (X_hat)
X_hat = enc(X)
loss = vector_norm(X_hat - c, dim = -1).mean()
# Optimisation enc s.t. min ||c - enc(X_hat)||_2
optimizer.zero_grad()
loss.backward()
optimizer.step()
```

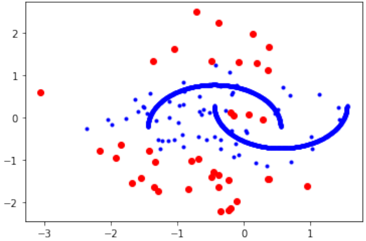
See https://github.com/sulcantonin/MLE2022/blob/main/MLE_OCL.ipynb

Anomaly Detection - One Class Loss Example Cont'd

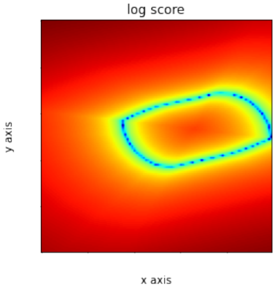
Input Points



Trained Classification



$s = \|\mathbf{c} - enc(\mathbf{x})\|_2$



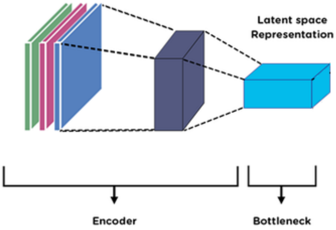
See https://github.com/sulcantonin/MLE2022/blob/main/MLE_OCL.ipynb

Anomaly Detection - Semi-Supervised

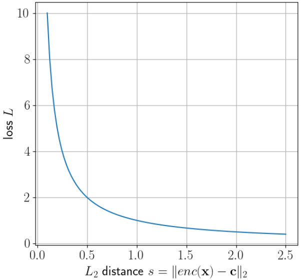
What if we have data and want to train a model of a set of data points $\{x_1 \dots x_N\}$ with very few **known** anomalies?

Anomaly $\rightarrow \|enc(\mathbf{x}) - \mathbf{c}\|^{-1}$

$\mathbf{x}$ $\qquad$ $\mathbf{z} = enc(\mathbf{x})$



$$\begin{aligned} s &= \|enc(\mathbf{x}) - \mathbf{c}\|_2 \\ &= \|\mathbf{z} - \mathbf{c}\|_2 \end{aligned}$$



$$s = \|enc(\mathbf{x}) - \mathbf{c}\|^l \text{ where } l \in \{-1, 1\}$$

Known anomalies ($l = -1$) encourage enc and $\mathbf{c}$ to move away.

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_SAD.ipynb

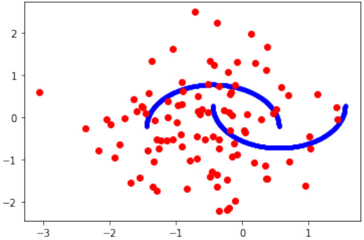
Anomaly Detection - Semi-Supervised Example

```
enc = nn.Sequential(nn.Linear(2,8),
    nn.Linear(8,16),
    nn.Linear(16,4))
# random center of 4D hypersphere
enc.c = nn.Parameter(torch.randn((1,4), requires_grad = True))
optimizer = optim.Adam(enc.parameters())
# reconstruction (X_hat)
X_hat = enc(X)
loss = (vector_norm((X_hat - c), dim = -1)**1).mean()
# Optimisation enc s.t. min ||c - enc(X_hat)||_2
optimizer.zero_grad()
loss.backward()
optimizer.step()
```

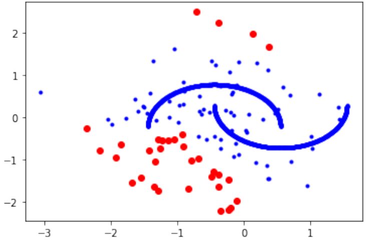
Erratum: the loss should be correct now,
sorry for confusion.

Anomaly Detection - Semi-supervised Example Cont'd

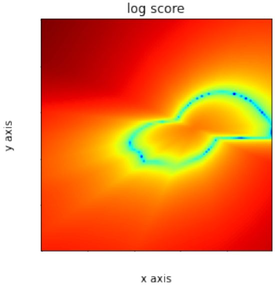
Input Points



Trained Classification

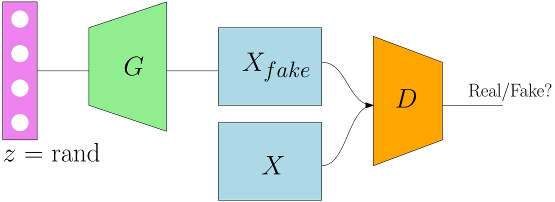


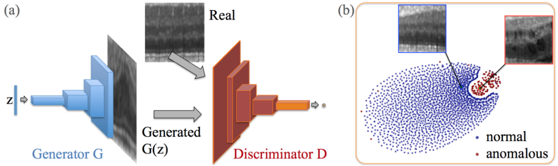
$$s = \|\mathbf{c} - enc(\mathbf{x})\|_2$$



See https://github.com/sulcantonin/MLE2022/blob/main/MLE_SAD.ipynb

Anomaly Detection - Generative Adversarial Networks





- > G generates fake samples and tries to fool discriminator D.
- > D receives fake and real samples and tries to distinguish if an input is fake or real.

[Goodfellow et al.(2020)Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio, Wiggers(2019),

Schlegl et al.(2017)Schlegl, Seeböck, Waldstein, Schmidt-Erfurth, and Langs]

$$\arg\min_{z_\gamma}\{(1-\lambda)\|x - G(z_\gamma)\| + \lambda CrossEntropyLoss(D(G(z_\gamma)), 1)\}$$

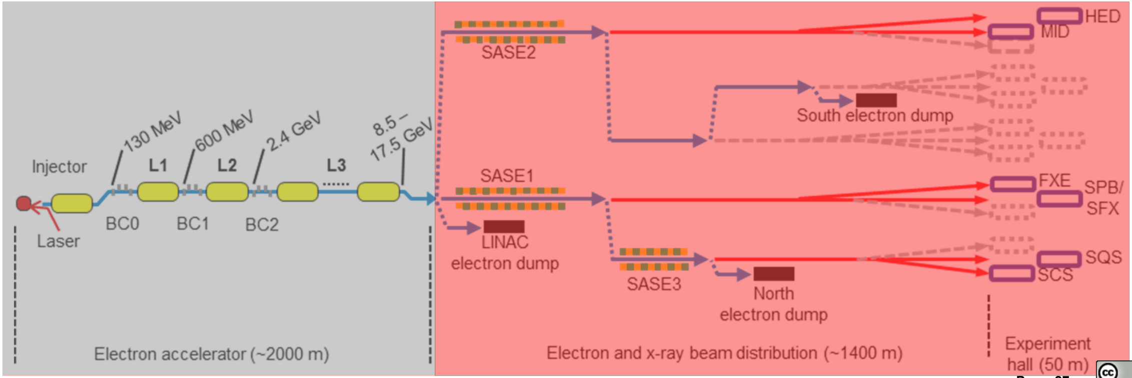
See https://github.com/sulcantonin/MLE2022/blob/main/MLE_GAN.ipynb

Anomaly Detection at European XFEL

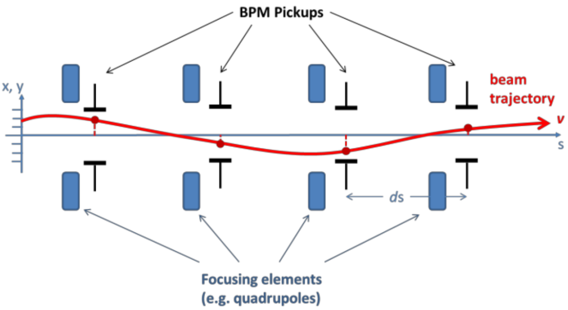
Orbit Monitoring

Assumption

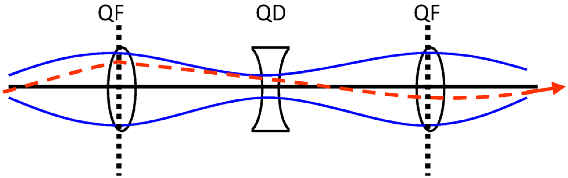
There is a systematic pattern shown in orbits given by the physical construction of EuXFEL.



Orbit Monitoring - FODO Lattice



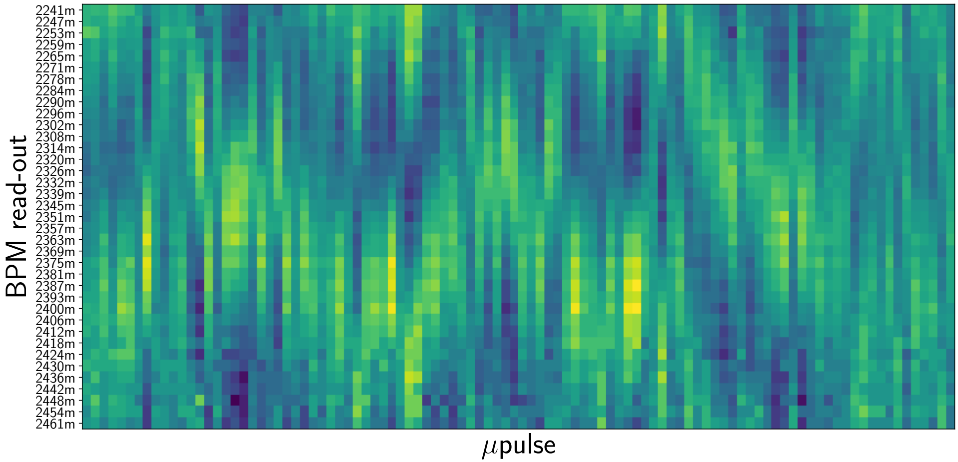
[Wendt(2011)]



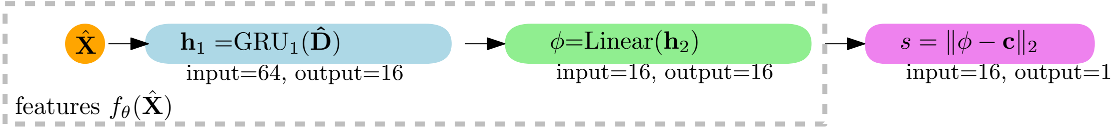
[Holzer(2006)]

Orbit Monitoring

Orbit Monitoring



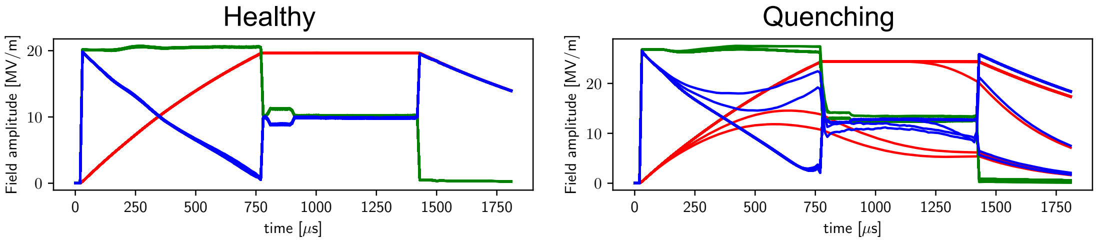
Orbit Monitoring




```
rnn = nn.GRU(64,16)
lin = nn.Linear(16,16)
c = nn.randn((16))
s = vector_rnom(c - lin(rnn(x)[0], dim = -1)
```

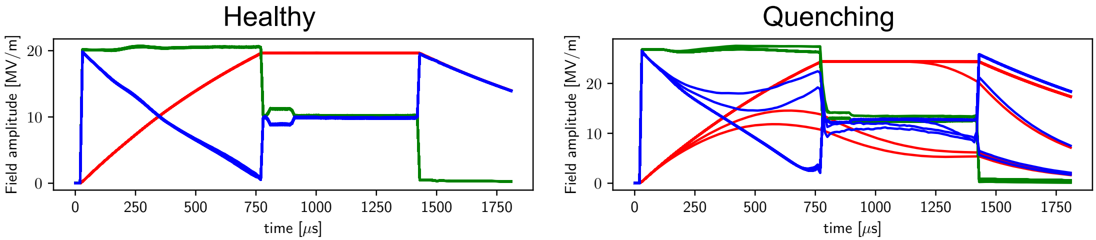
See https://github.com/sulcantonin/MLE2022/blob/main/MLE_orbit.ipynb

Monitoring Superconducting LLRF Cavities



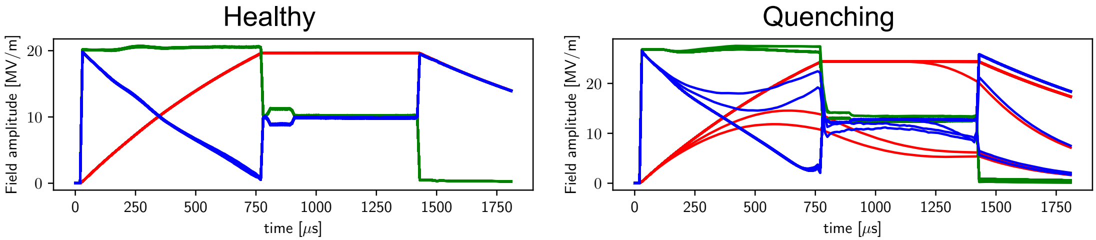
- We record an envelope (phase, amplitude) of three signals - probe, forward and reflected signals.

Monitoring Superconducting LLRF Cavities



- We record an envelope (phase, amplitude) of three signals - probe, forward and reflected signals.
- Each signal consists of a pair of 1820 values (amplitude, phase) per pulse.

Monitoring Superconducting LLRF Cavities



- We record an envelope (phase, amplitude) of three signals - probe, forward and reflected signals.
- Each signal consists of a pair of 1820 values (amplitude, phase) per pulse.
- Quench (right) means a loss of superconductivity in a cavity, which has a significant effect on the quality factor.

Data-Driven Monitoring of Superconducting LLRF Cavities

$$\underbrace{\mathbf{h}_1 = \mathrm{LSTM}(\mathbf{x}) \longrightarrow \mathbf{h}_2 = \mathrm{LSTM}(\mathbf{h}_1) \longrightarrow \phi = \mathrm{Linear}(\mathbf{h}_2)}_{f_\theta(\mathbf{x})} \longrightarrow s = \|\phi(\mathbf{h}_2) - \mathbf{c}\|_2^2 \longrightarrow$$

input=1092, output=256; input=256, output=256; input=256, output=64; input=64, output=1

> A RNN is assigning a score to series of cavity pulses.

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_cavity.ipynb

Data-Driven Monitoring of Superconducting LLRF Cavities

$$\underbrace{\mathbf{h}_1 = \mathrm{LSTM}(\mathbf{x}) \xrightarrow{} \mathbf{h}_2 = \mathrm{LSTM}(\mathbf{h}_1) \xrightarrow{} \phi = \mathrm{Linear}(\mathbf{h}_2)}_{f_\theta(\mathbf{x})} \xrightarrow{} s = \|\phi(\mathbf{h}_2) - \mathbf{c}\|_2^2 \xrightarrow{}$$

input=1092, output=256 | input=256, output=256 | input=256, output=64 | input=64, output=1

- A RNN is assigning a score to series of cavity pulses.
- Each datum **x** (pulse) consist of (probe, forward and reflected signals).

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_cavity.ipynb

Data-Driven Monitoring of Superconducting LLRF Cavities

$$\mathbf{h}_1 = \mathrm{LSTM}(\mathbf{x}) \longrightarrow \mathbf{h}_2 = \mathrm{LSTM}(\mathbf{h}_1) \longrightarrow \phi = \mathrm{Linear}(\mathbf{h}_2) \longrightarrow s = \|\phi(\mathbf{h}_2) - \mathbf{c}\|_2^2 \longrightarrow$$

input=1092, output=256; input=256, output=256; input=256, output=64; input=64, output=1

$$\underbrace{\mathbf{h}_1 \rightarrow \mathbf{h}_2 \rightarrow \phi}_{f_\theta(\mathbf{x})}$$

- A RNN is assigning a score to series of cavity pulses.
- Each datum **x** (pulse) consist of (probe, forward and reflected signals).
- We have a disproportionately smaller dataset with faults (1300 faults) and (almost) unlimited access to healthy data.

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_cavity.ipynb

Data-Driven Monitoring of Superconducting LLRF Cavities

$$\underbrace{\mathbf{h}_1=\mathrm{LSTM}(\mathbf{x}) \longrightarrow \mathbf{h}_2=\mathrm{LSTM}(\mathbf{h}_1) \longrightarrow \phi=\mathrm{Linear}(\mathbf{h}_2)}_{f_\theta(\mathbf{x})} \longrightarrow s = \|\phi\left(\mathbf{h}_2\right) - \mathbf{c}\|_2^2 \longrightarrow$$

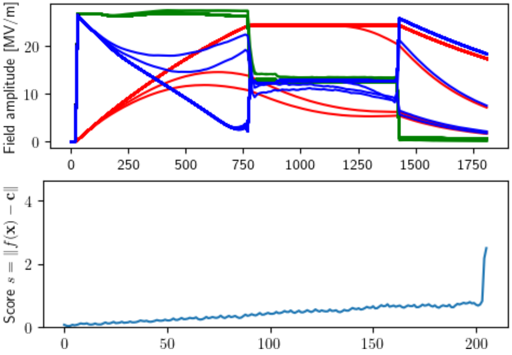
input=1092, output=256 — input=256, output=256 — input=256, output=64 — input=64, output=1

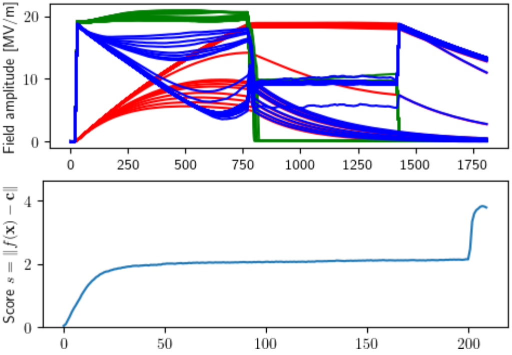
- A RNN is assigning a score to series of cavity pulses.
- Each datum **x** (pulse) consist of (probe, forward and reflected signals).
- We have a disproportionately smaller dataset with faults (1300 faults) and (almost) unlimited access to healthy data.
- Semi-supervised anomaly loss [Ruff(2019)]

$$L\left(\theta\right) = \|f_\theta(\mathbf{x}) - \mathbf{c}\|_2^y + \|f_\theta(\mathbf{x}) - \mathbf{c}\|_2 \text{ where } y \in \{-1, 1\}.$$

See https://github.com/sulcantonin/MLE2022/blob/main/MLE_cavity.ipynb

Results - Quenches





Thank you!

This is the joint work of A. Eichler and Raimund Kammering!

Contact

DESY. Deutsches Elektronen-Synchrotron

www.desy.de

Antonin Sulc
0000-0001-7767-778X
MCS DESY
antonin.sulc@desy.de

Bibliography I

Isolation forest, Aug 2022.
URL https://en.wikipedia.org/wiki/Isolation_forest.

Deepak Birla.
Autoencoders, Mar 2019.
URL https://medium.com/@birla.deepak26/autoencoders-76bb49ae6a8f.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.
Generative adversarial networks.
Communications of the ACM, 63(11):139–144, 2020.

BJ Holzer.
Lattice design in high-energy particle accelerators.
2006.

Bibliography II

Martin Magakian.
Anomaly detection with the normal distribution, Dec 2018.
URL
https://anomaly.io/anomaly-detection-normal-distribution/index.html.

L Ruff.
Deep semi-supervised anomaly detection.
arXiv preprint arXiv:1906.02694, 2019.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs.
Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, 2017.
URL https://arxiv.org/abs/1703.05921.

Bibliography III

Hannah Ward-Glenton.
Euro zone inflation hits another record of 9.1% as food and energy prices soar, Aug 2022.
URL https://www.cnbc.com/2022/08/31/euro-zone-inflation-hits-another-record-of-9point1percent-as-food-and-en html.

Manfred Wendt.
Overview of recent trends and developments for bpm systems.
Technical report, Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2011.

Bibliography IV

Kyle Wiggers.
Generative adversarial networks: What gans are and how they've evolved, Dec 2019.
URL https://venturebeat.com/ai/gan-generative-adversarial-network-explainer-ai-machine-learning/.

Thorsten Wittkopp, Philipp Wiesner, Dominik Scheinert, and Odej Kao.
A taxonomy of anomalies in log data.
In International Conference on Service-Oriented Computing, pages 153–164. Springer, 2022.