TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Detekce anomálií

František Kynych
14. 11. 2024 | MVD

Část I.:
Úvod do problematiky

Detekce anomálií

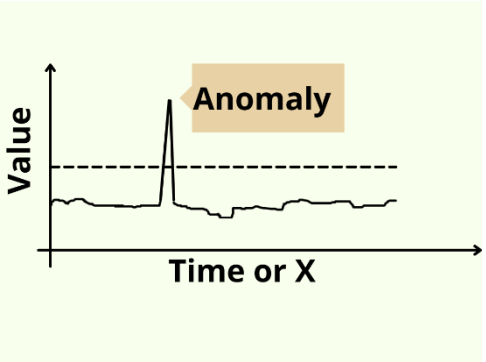
Definice

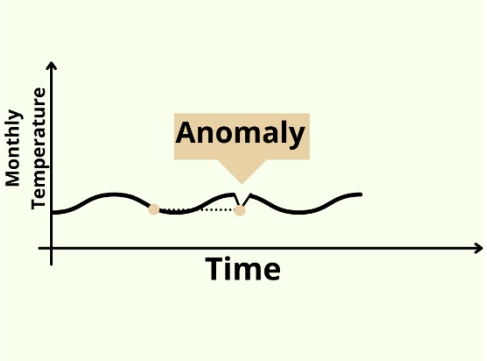
- Proces identifikace datových bodů (položek, událostí, …), které se výrazně odchylují od většiny dat
  - Předpokladem je dostatečná vzdálenost nebo odlišnost od normy
- Anomálie se také označuje např. jako outlier nebo novelty

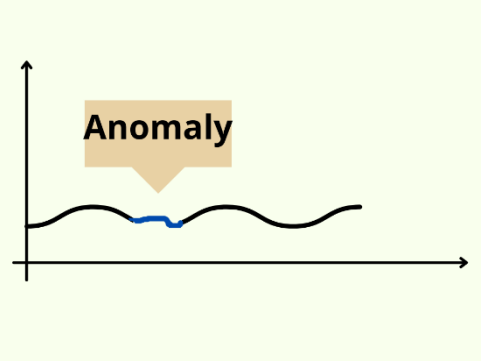
Druhy anomálií

Druhy anomálií

- Bodové anomálie (global outlier)
- Kontextuální anomálie (local outlier)
  - V rámci daného kontextu se jedná o anomálii
    - Např.: Naměřena vysoká teplota v prosinci
- Kolektivní anomálie
  - Anomálií je skupina datových instancí
    - Např.: Jednotlivé body nevypadají jako anomálie, ale jejich společný výskyt ano







https://medium.com/@jelkhoury880/introduction-to-anomaly-detection-methods-part-i-b1a2f389ffcb

Základní přístupy

Statistické profilování

- Detekce dat odchylujících se od statistických vlastností distribuce
- Histogram, Gaussovo rozdělení, Z-skóre, Boxplot

Metody založené na strojovém učení

- **Supervised**
  - Učení s popisky (labely) dat, kde model je trénován na rozpoznávání anomálií na základě označených příkladů
  - Logistická regrese, rozhodovací stromy, neuronové sítě

Základní přístupy

Metody založené na strojovém učení

- **Semi-supervised**
  - Popisky dat jsou dostupné pouze pro normální data, která model používá k učení reprezentace normálního chování
  - One-class SVM, autoenkodéry

- **Unsupervised**
  - Nepotřebujeme popisky dat, anomálie jsou identifikovány na základě inherentních struktur v datech
  - Např. metody založené na hustotě dat
    - Využití shlukovacích algoritmů pro nalezení anomálií
    - DBSCAN, Local Outlier Factor
  - K-means, Hierarchické shlukování

Praktické aplikace

- Odhalování podvodných transakcí a pojistných událostí
- Detekce kybernetických útoků
- Zjišťování neobvyklého chování zařízení
- Detekce anomálií v procesu výroby
- Hledání anomálií ve spotřebě energií
- Detekce havárie vody
- …

Proč je detekce anomálií složitá?

1. Neznámé chování anomálií - neznámé druhy a rozložení. Často je poznáme až v okamžik, kdy nastanou.

2. Heterogenní třídy – nepravidelné a většinou naprosto odlišné charakteristiky od jiné třídy.

3. Vzácnost a nevyváženost tříd – oproti normálním datům jsou vzácné a je složité (může být nemožné) vytvořit dataset s popisky anomálií. Chybný popisek u anomálie může mít velmi špatný vliv na algoritmus.

4. Různé druhy anomálií – bodové, kontextuální a kolektivní.

Část II.:
Přístupy k detekci anomálií

Histogram

- Neparametrická statistická technika
  - Nepředpokládá se žádné specifické rozdělení dat

Přístup rozdělen do dvou kroků:

1. Konstrukce histogramu
   - Určení šířky a počtu intervalů
2. Zkoumání, do jakého intervalu bod padne
   - Pokud bod nepadá do žádného intervalu, je pravděpodobně anomálií
   - Intervaly s nízkou četností mohou také indikovat anomálie

Nalezení šířky a počtu intervalů:
n … počet dat, r … rozsah dat (max – min), p … počet intervalů

a) $p = pevná\ hodnota$
b) $p = \lfloor \sqrt{n} \rfloor, nebo\ p = \lceil \sqrt{n} \rceil$
c) $p = 1 + log_2(n)$

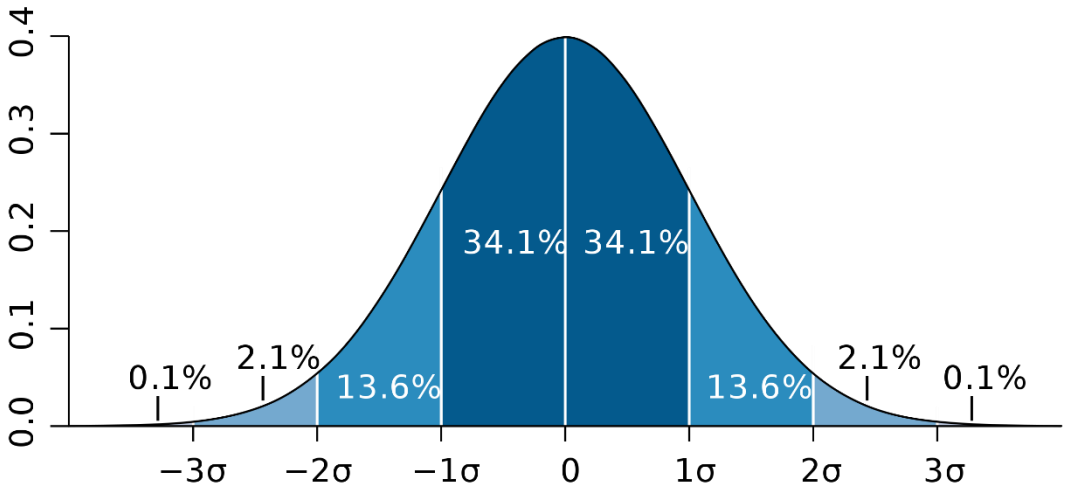
Šířka intervalu $s = \frac{r}{p}$

Gaussovo (normální) rozdělení

- Data X mají normální rozdělení pravděpodobnosti se střední hodnotou μ a rozptylem σ^2

$$X \approx N(\mu, \sigma^2)$$

- Data leží v rozsahu $\mu \pm 3\sigma$ s pravděpodobností 99,73 %
  - Anomálie leží mimo námi stanovený interval ε



https://en.wikipedia.org/wiki/Normal_distribution

Gaussovo (normální) rozdělení

Algoritmus:

1. Vypočteme $\mu_1, \dots, \mu_n$ a $\sigma_1^2, \dots, \sigma_n^2$ pro data X o dimenzi n, m = počet dat

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_{i,j}$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (x_{i,j} - \mu_j)^2$$

2. Vypočteme $p(x)$ pro instanci x

$$p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right)$$

3. Anomálie, pokud $p(x) < \varepsilon$

- Za předpokladu, že jednotlivé příznaky jsou nezávislé

Vícerozměrné normální rozdělení

- Není potřeba počítat jednotlivé pravděpodobnosti $p(x_j; \mu_j, \sigma_j^2)$
- **Vektor** středních hodnot μ a kovarianční matice Σ

$$p(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

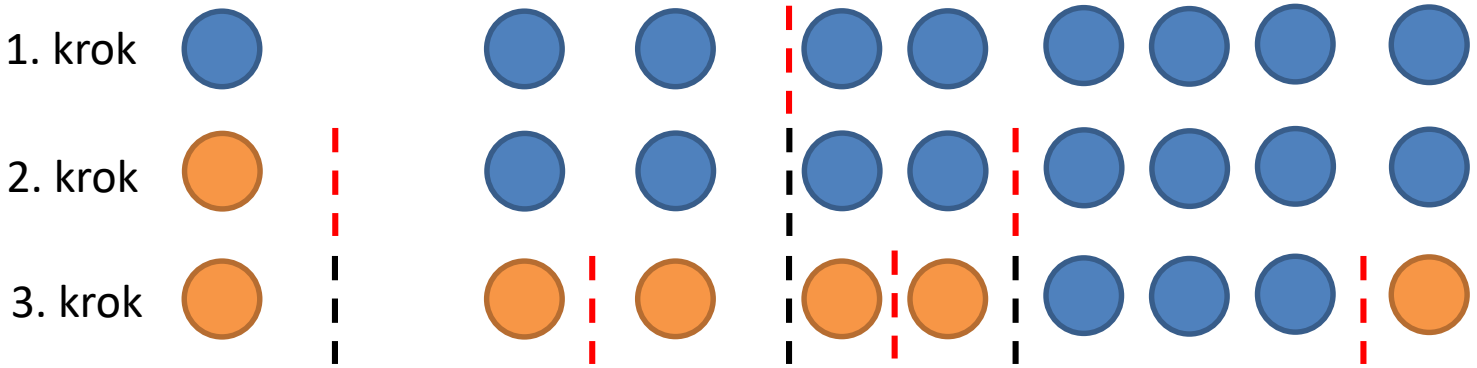
$|\Sigma|$ … determinant kovarianční matice

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)(x_i - \mu)^T$$

- Potřeba splnit podmínku $m > n$
- Postup algoritmu obdobný předchozímu snímku

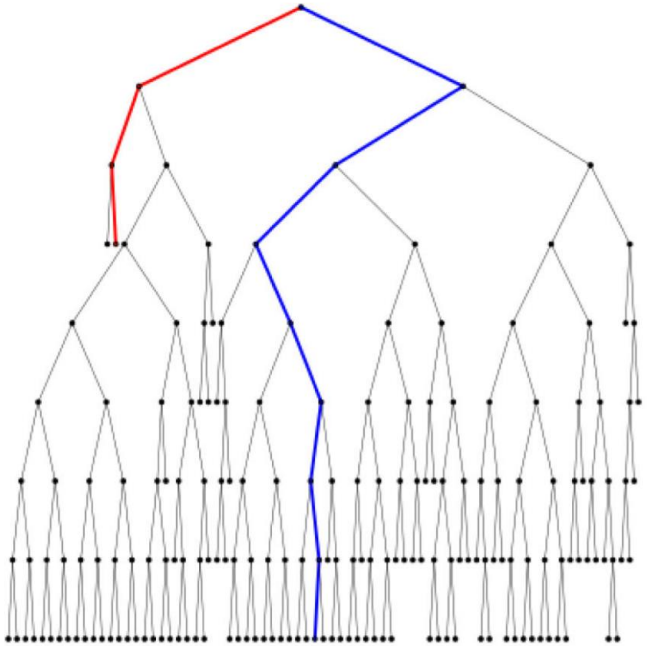
Isolation forest

- Vytvoření stromu pomocí výběru náhodného příznaku dat a náhodného rozdělení na dvě části
  - Data dělíme rekurzivně dál, dokud nemáme pouze izolované instance (nebo duplicitní data) nebo je dosažen limit hloubky stromu
  - Před dělením datasetu použijeme pouze podmnožinu dat při tvorbě každého stromu

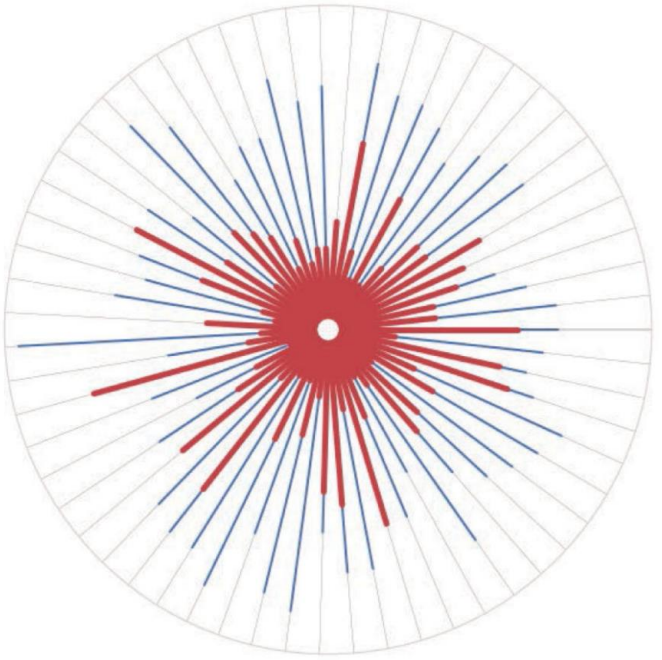


- Náhodný proces opakujeme $n \times$ pro vytvoření lesu

Isolation forest



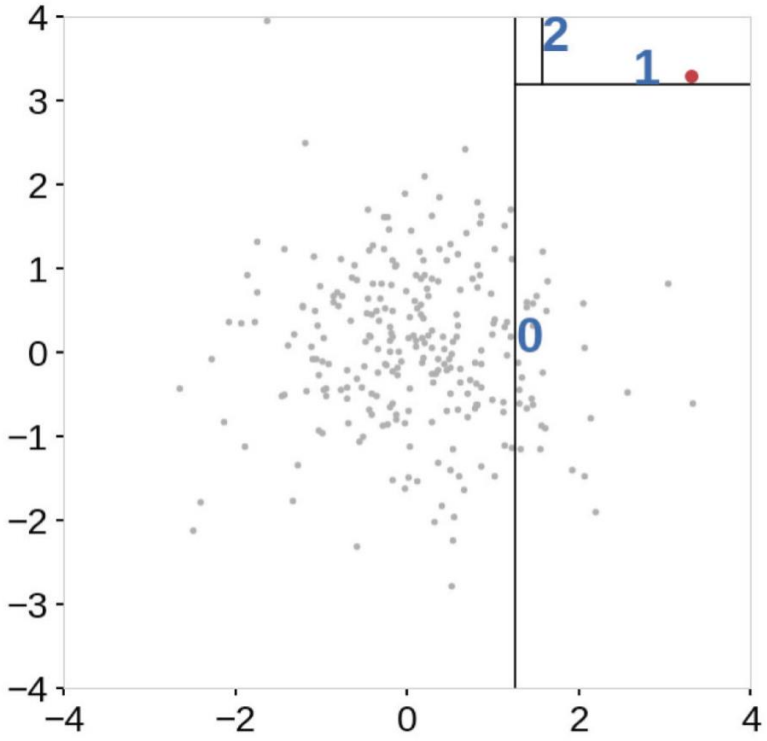
(a) Representation of a single tree in a forest.



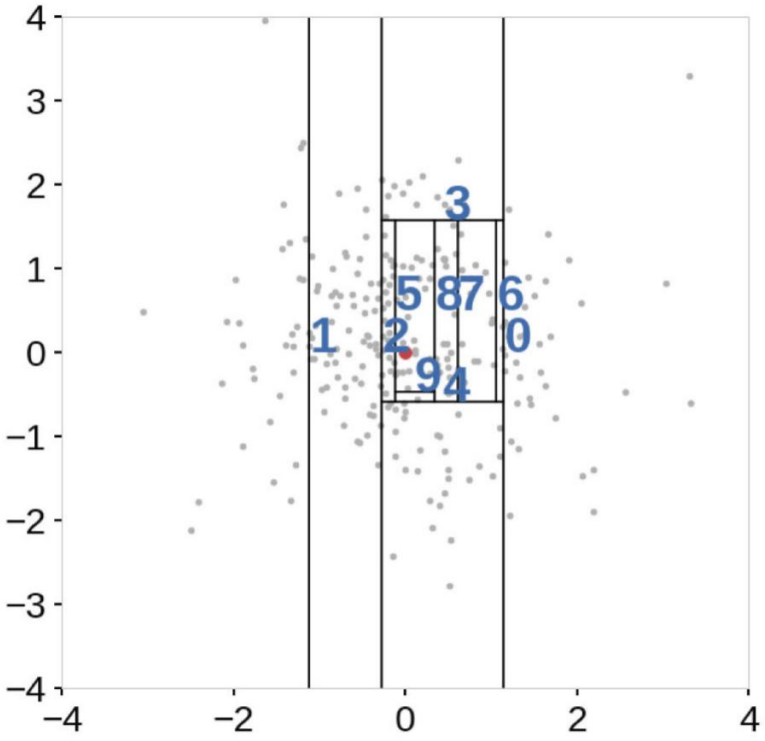
(b) Representation of a full forest where each radial line corresponds to a tree.

https://ieeexplore.ieee.org/abstract/document/8888179

Isolation forest



(a) Anomaly point



(b) Nominal point

https://ieeexplore.ieee.org/abstract/document/8888179

Isolation forest

- Po vytvoření lesu lze získat skóre pro vybraný datový bod x a datech n

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

$E(h(x))$ … průměrná délka cesty (přes všechny stromy) k bodu x

$c(n)$ … průměrná délka cesty ke každému listu

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

$$H(i) = \ln(i) + 0.5772156649 \ldots$$

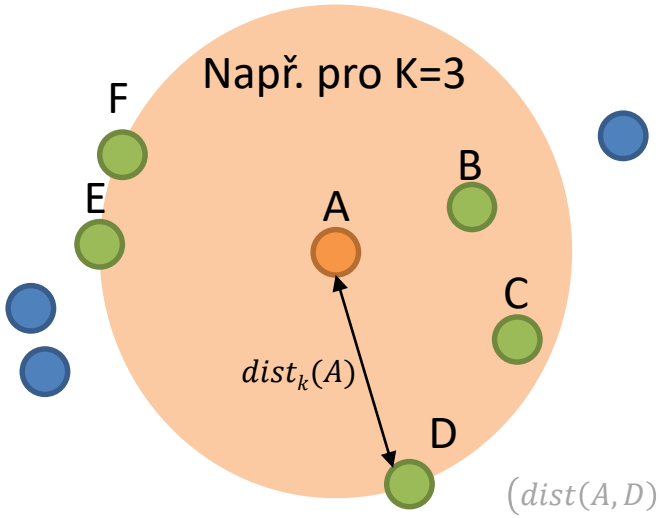
(Harmonické číslo) (Eulerova konstanta)

Local Outlier Factor

- Přístup vhodný pro hledání lokálních (kontextuálních) anomálií
- Založeno na hustotě bodů, odvozeno z DBSCAN algoritmu

Postup:

1. K-vzdálenost bodu A -> $dist_k(A)$ = vzdálenost mezi A a k-tým nejbližším sousedem



$(dist(A,D) = dist(A,E) = dist(A,F))$

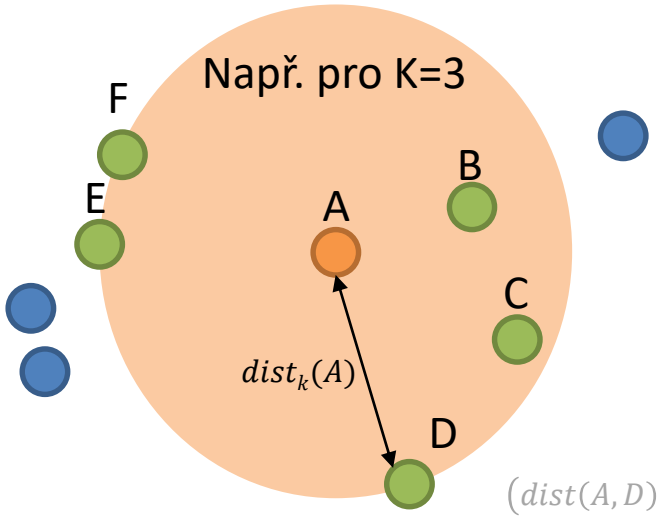
Local Outlier Factor

2. K-vzdálenostní okolí bodu A

$$N_k(A) = \{p \mid p \in D \backslash \{A\}, dist(A,p) \leq dist_k(A)\}$$

- Všechny body p, jejichž vzdálenost od bodu A je menší než $dist_k(A)$

$$N_k(A) = \{B, C, D, E, F\}$$

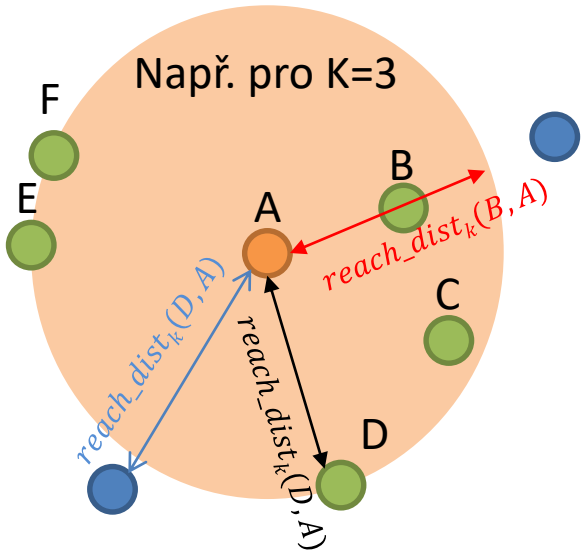


$$\left(dist(A,D) = dist(A,E) = dist(A,F)\right)$$

Local Outlier Factor

3. Dosažitelná vzdálenost (reachability distance)

$$reach\_distk(A,p) = \max\{dist_k(p), dist(A,p)\}$$



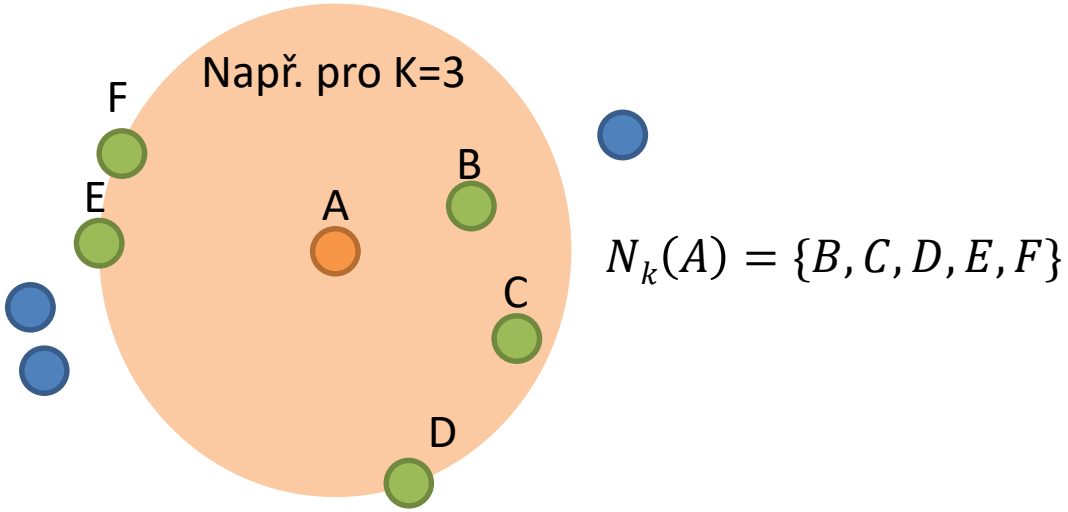
- Vypočítat pro všechny body p v k-okolí bodu A

Local Outlier Factor

4. Hustota lokální dosažitelnosti bodu A (local reachability density)

$$lrd_k(A) = \frac{|N_k(A)|}{\sum_{p \epsilon Nk(A)} reach\_distk(A,p)}$$

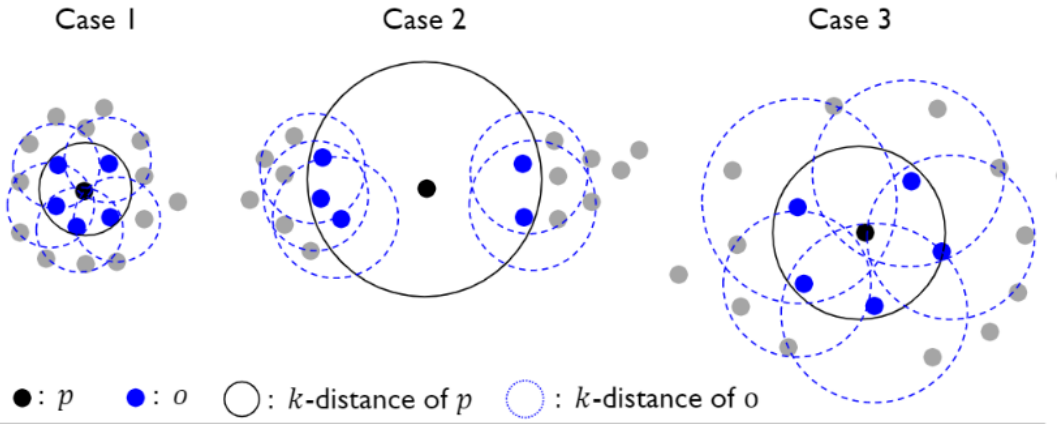
Pokud je bod A v husté oblasti, tak hodnota jmenovatele bude nízká (=$reach\_distk(A,p)$ budou malé hodnoty) -> velká hodnota $lrd_k(A)$



Local Outlier Factor

5. Local Outlier Factor

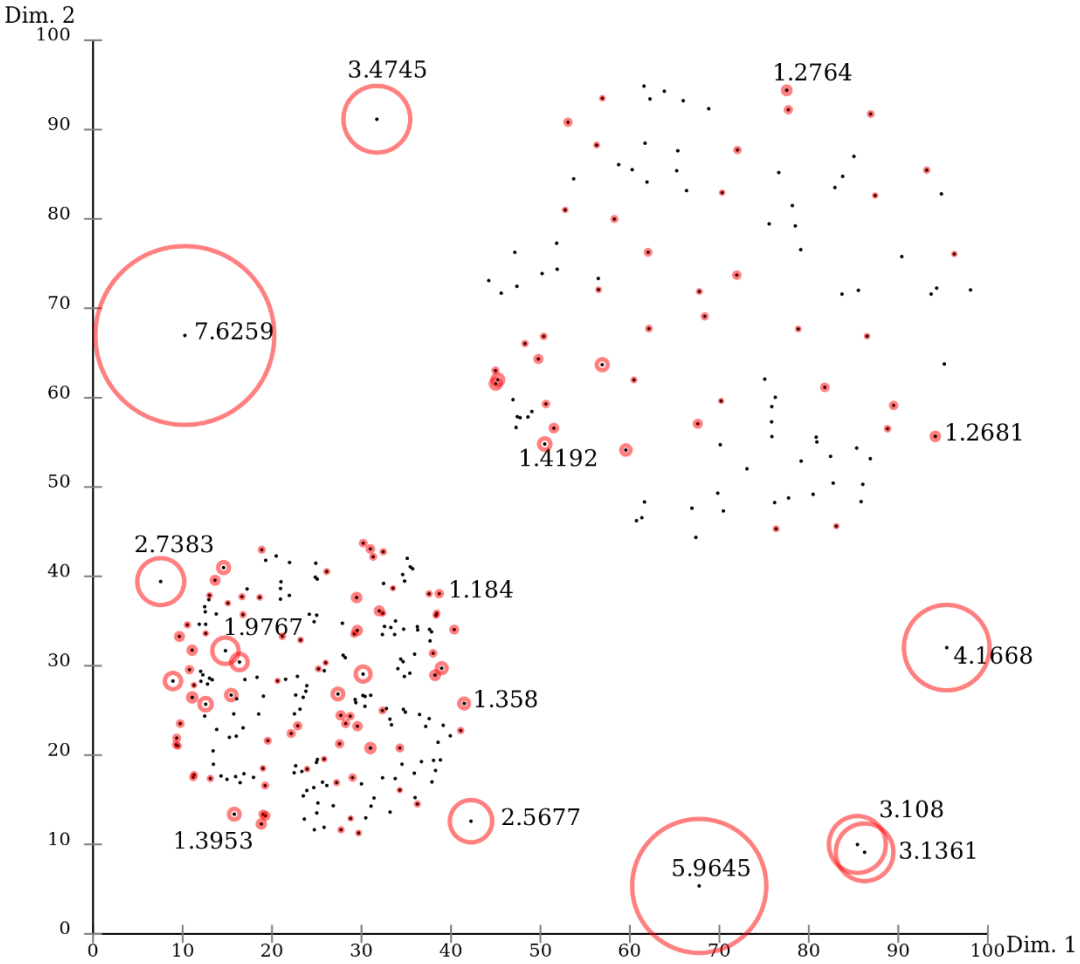
$$LOF_k(A) = \frac{\sum_{p \epsilon Nk(A)} \frac{lrd_k(p)}{lrd_k(A)}}{|N_k(A)|}$$



Case	$lrd_k(p)$	$lrd_k(o)$	$LOF_k(p)$
Case 1	Large	Large	Small
Case 2	Small	Large	Large
Case 3	Small	Small	Small

https://github.com/pilsung-kang/Business-Analytics-IME654-

Local Outlier Factor



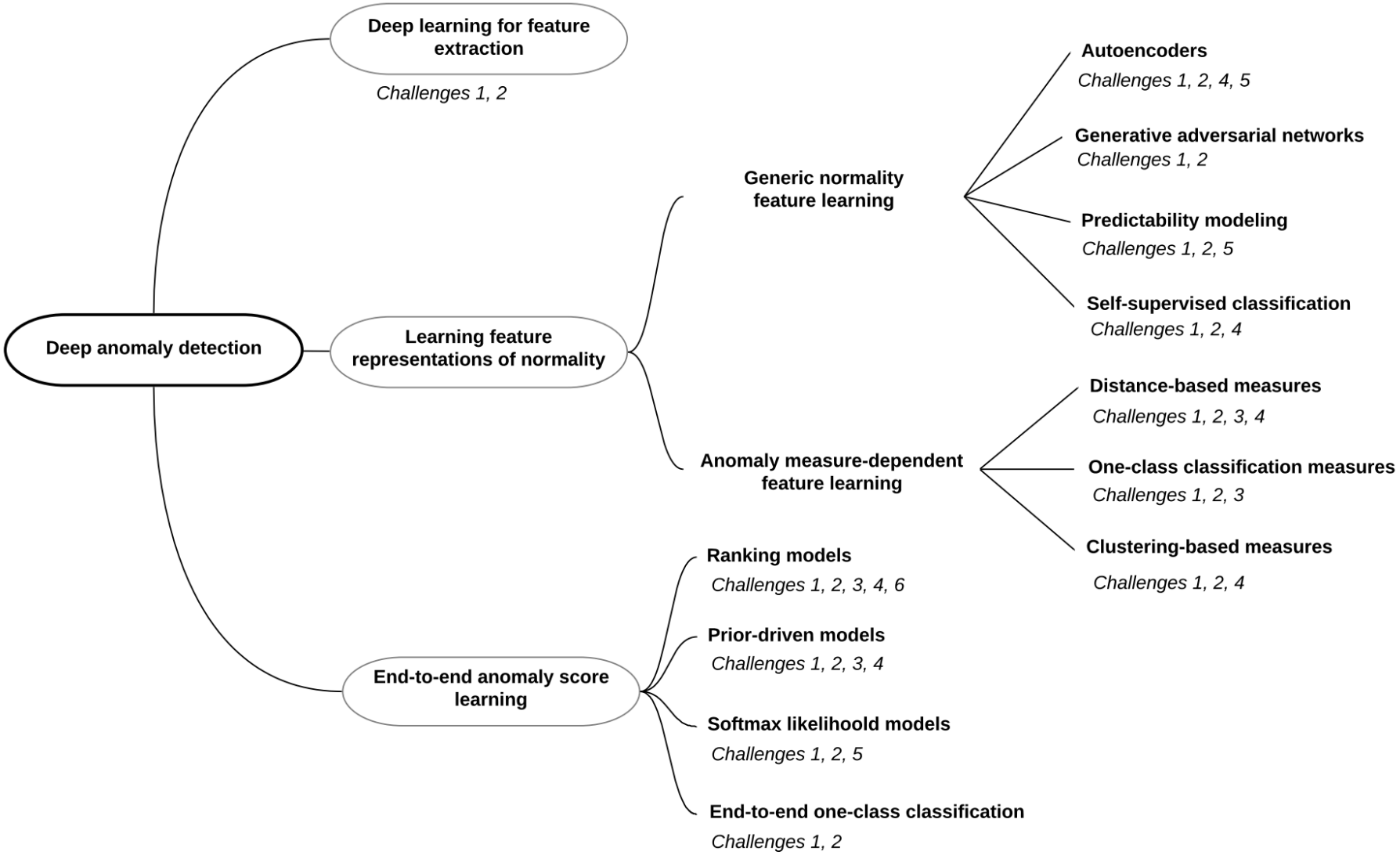
https://en.wikipedia.org/wiki/Local_outlier_factor

TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Část III.:
Přístupy využívající hlubokých neuronových sítí

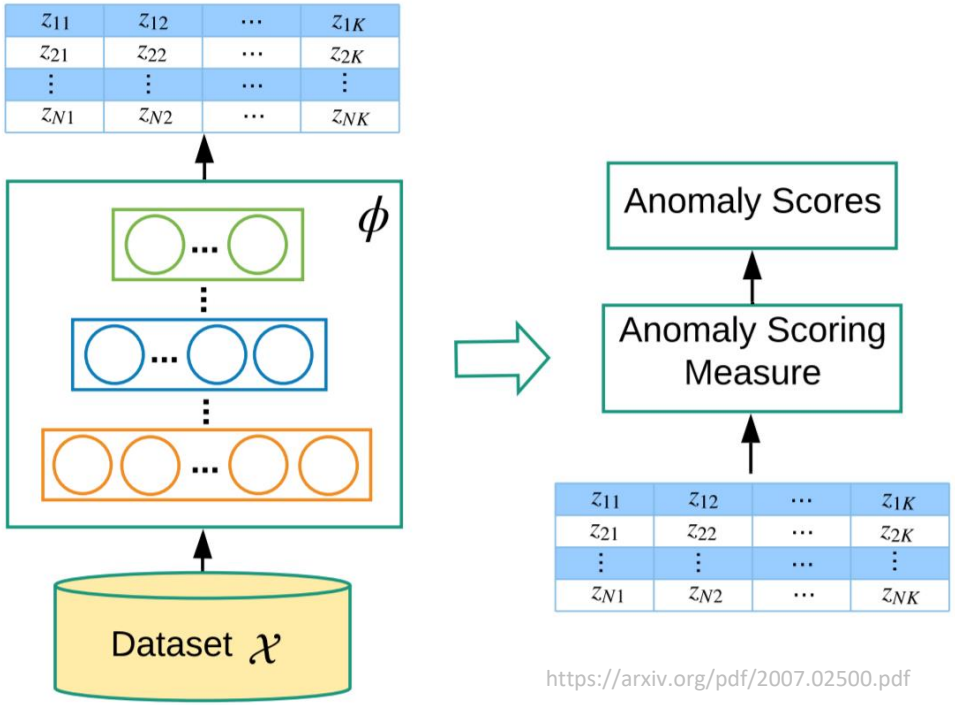
Výzvy detekce anomálií

1. Nízký recall – je obtížné detekovat větší množství anomálií, aniž bychom nezvýšili chybnou detekci (false positives)
2. Detekce anomálií ve vysoce dimenzionálních datech a/nebo v datech se závislostmi
3. Efektivní učení normality/abnormality
4. Odolnost vůči šumu
5. Detekce komplexních anomálií
6. Možnost vysvětlení anomálie



https://arxiv.org/pdf/2007.02500.pdf

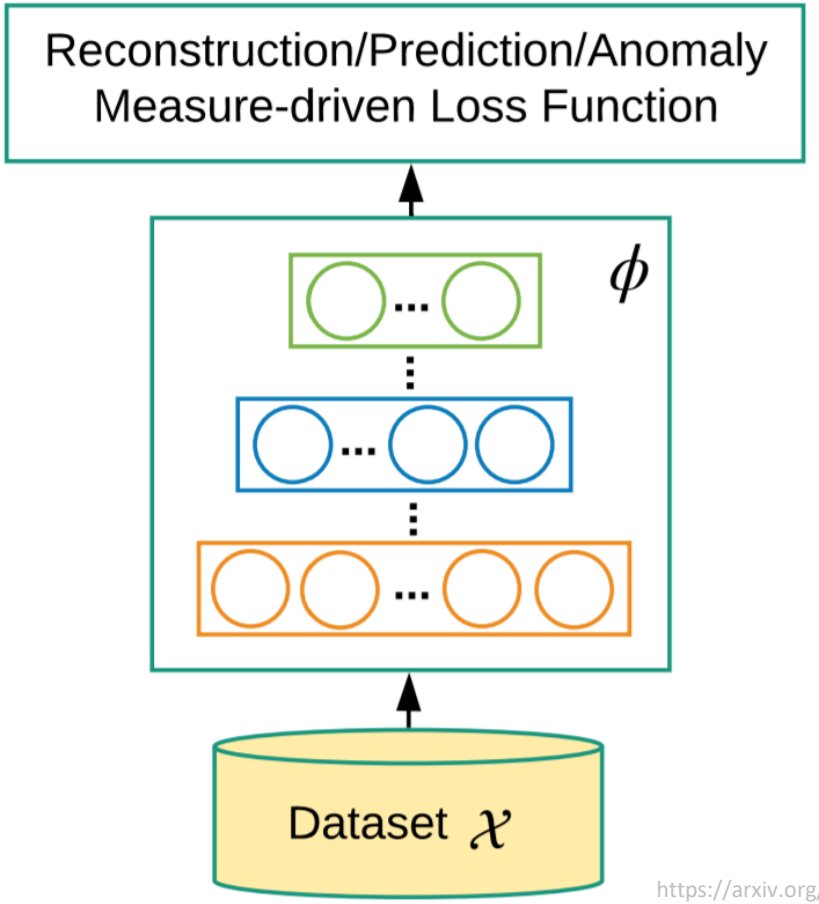
Deep Learning for Feature Extraction



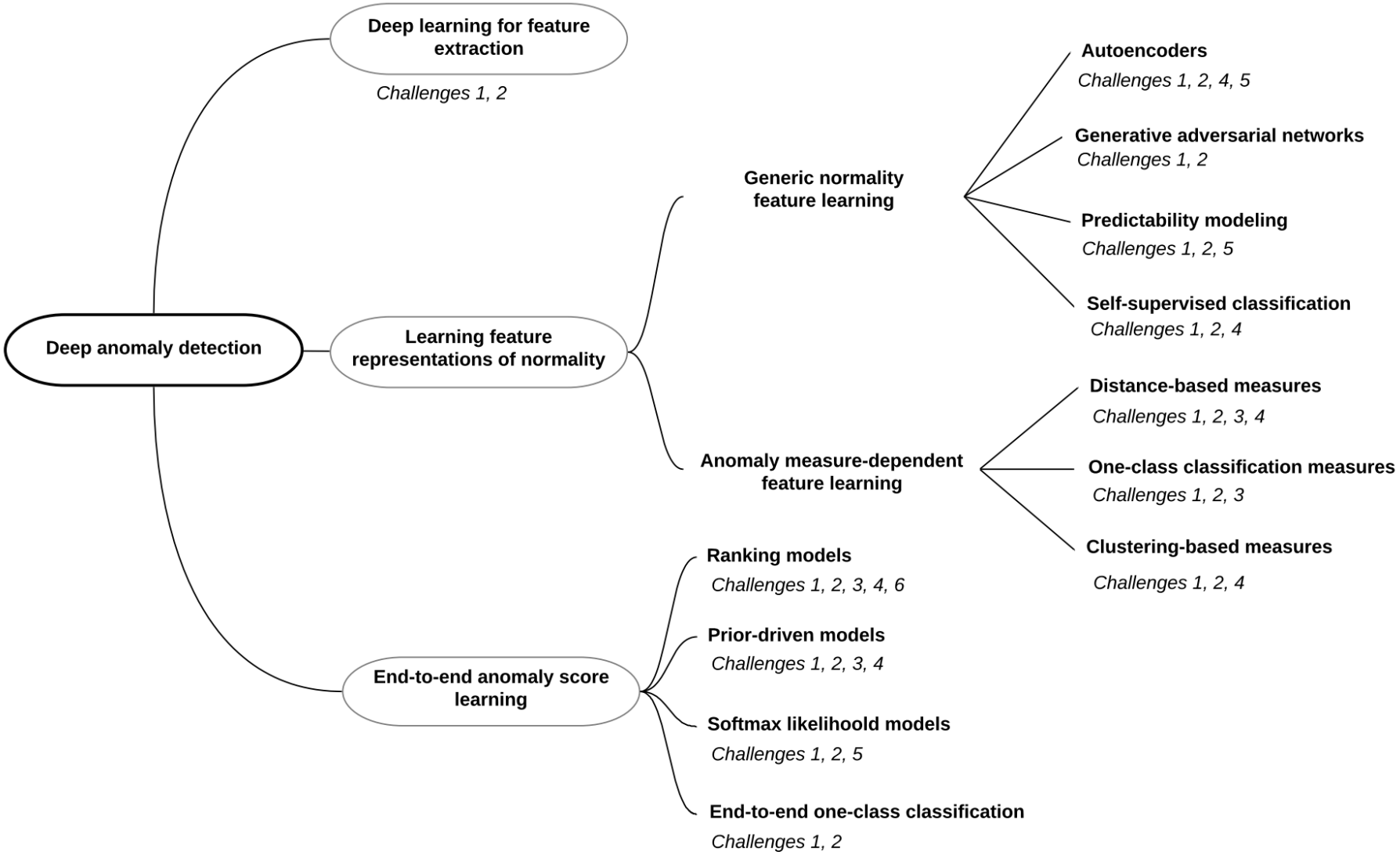
https://arxiv.org/pdf/2007.02500.pdf

$$z = \phi(\chi;\ \theta)$$
$$\phi\colon \chi \longmapsto z$$
$$\chi \in \mathbb{R}^D, z \in \mathbb{R}^K$$
$$D \gg K$$

Learning Feature Representations of Normality

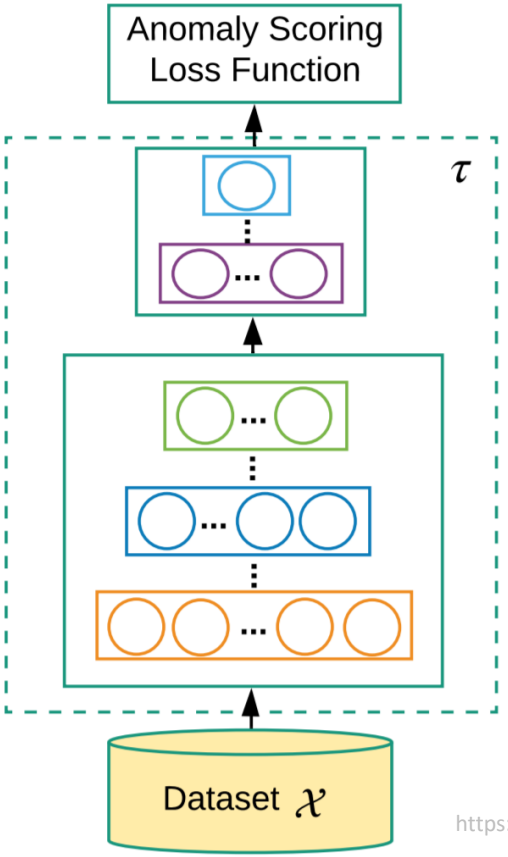


https://arxiv.org/pdf/2007.02500.pdf

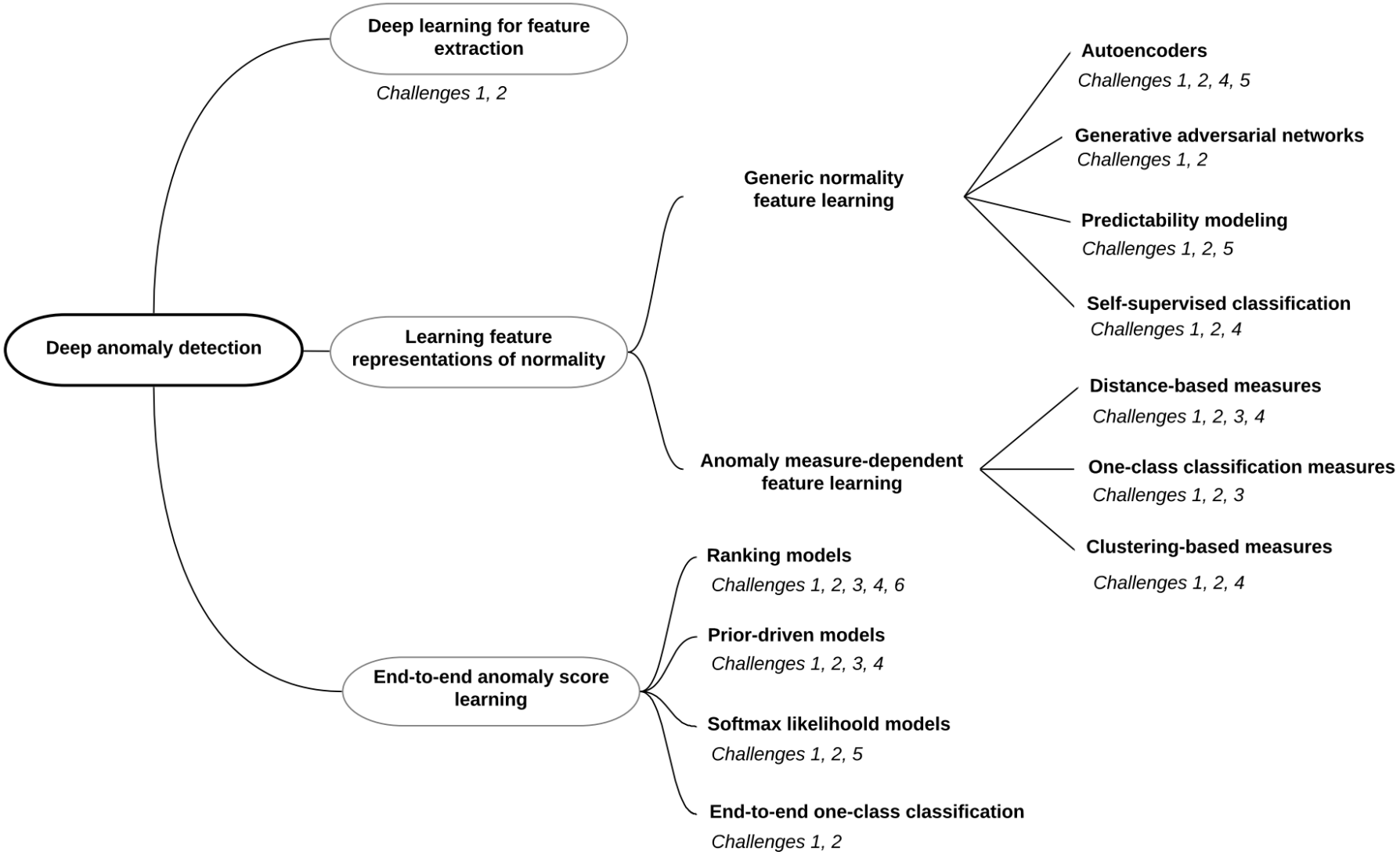


https://arxiv.org/pdf/2007.02500.pdf

End-to-end Anomaly Score Learning



https://arxiv.org/pdf/2007.02500.pdf



https://arxiv.org/pdf/2007.02500.pdf

Charakteristiky modelů

Method	Ref.	Sup.	Objective	DA	DP	PT	Archit.	Activation	# layers	Loss	Data
OADA	[65] (4)	Semi	Reconstruction	Yes	No	No	AE	ReLU	3	MSE	Video
Replicator	[57] (5.1.1)	Unsup.	Reconstruction	No	No	No	AE	Tanh	2	MSE	Tabular
RandNet	[29] (5.1.1)	Unsup.	Reconstruction	No	Yes	Yes	AE	ReLU	3	MSE	Tabular
RDA	[175] (5.1.1)	Semi	Reconstruction	No	No	No	AE	Sigmoid	2	MSE	Tabular
UODA	[91] (5.1.1)	Semi	Reconstruction	No	No	Yes	AE & RNN	Sigmoid	4	MSE	Sequence
AnoGAN	[138] (5.1.2)	Semi	Generative	No	No	No	Conv.	ReLU	4	MAE	Image
EBGAN	[170] (5.1.2)	Semi	Generative	No	No	No	Conv. & MLP	ReLU/lReLU	3-4	GAN	Image & Tabular
FFP	[86] (5.1.3)	Semi	Predictive	Yes	No	Yes	Conv.	ReLU	10	MAE/MSE	Video
LSA	[1] (5.1.3)	Semi	Predictive	No	No	No	Conv.	lReLU	4-7	MSE & KL	video
GT	[48] (5.1.4)	Semi	Classification	Yes	Yes	No	Conv.	ReLU	10-16	CE	Image
E^3Outlier	[157] (5.1.4)	Semi	Classification	Yes	Yes	No	Conv.	ReLU	10	CE	Image
REPEN	[112] (5.2.1)	Unsup.	Distance	No	No	No	MLP	ReLU	1	Hinge	Tabular
RDP	[155] (5.2.1)	Unsup.	Distance	No	No	No	MLP	lReLU	1	MSE	Tabular
AE-1SVM	[104] (5.2.2)	Unsup.	One-class	No	No	No	AE & Conv.	Sigmoid	2-5	Hinge	Tabular & image
DeepOC	[161](5.2.2)	Semi	One-class	No	No	No	3D Conv.	ReLU	5	Hinge	Video
Deep SVDD	[132] (5.2.2)	Semi	One-class	No	No	Yes	Conv.	lReLU	3-4	Hinge	Image
Deep SAD	[133] (5.2.2)	Semi	One-class	No	No	Yes	Conv. & MLP	lReLU	3-4	Hinge	Image & Tabular
DEC	[162] (5.2.3)	Unsup.	Clustering	No	Yes	Yes	MLP	ReLU	4	KL	Image & Tabular
DAGMM	[179] (5.2.3)	Unsup.	Clustering	No	Yes	No	AE & MLP	Tanh	4-6	Likelihood	Tabular
SDOR	[117] (6.1)	Unsup.	Anomaly scores	No	No	Yes	ResNet & MLP	ReLU	50 + 2	MAE	Video
PReNet	[114] (6.1)	Weak	Anomaly scores	Yes	No	No	MLP	ReLU	2-4	MAE	Tabular
MIL	[145] (6.1)	Weak	Anomaly scores	No	Yes	Yes	3DConv. & MLP	ReLU	18/34 + 3	Hinge	Video
PUP	[107] (6.2)	Unsup.	Anomaly scores	No	No	No	MLP	ReLU	3	Likelihood	Sequence
DevNet	[115] (6.2)	Weak	Anomaly scores	No	No	No	MLP	ReLU	2-4	Deviation	Tabular
APE	[30] (6.3)	Unsup.	Anomaly scores	No	No	No	MLP	Sigmoid	3	Softmax	Tabular
AEHE	[45] (6.3)	Unsup.	Anomaly scores	No	No	No	AE & MLP	ReLU	4	Softmax	Graph
ALOCC	[135] (6.4)	Semi	Anomaly scores	Yes	No	No	AE & CNN	lReLU	5	GANs	Image
OCAN	[174] (6.4)	Semi	Anomaly scores	No	No	Yes	LSTM-AE & MLP	ReLU	4	GANs	Sequence
Fence GAN	[103] (6.4)	Semi	Anomaly scores	No	Yes	No	Conv. & MLP	lReLU/Sigmoid	4-5	GANs	Image & Tabular
OCGAN	[120] (6.4)	Semi	Anomaly scores	No	No	No	Conv.	ReLU/Tanh	3	GANs	Image

https://arxiv.org/pdf/2007.02500.pdf

Vyhodnocení

True Positive, True Negative, False Positive, False Negative

Recall, Precision

F1 skóre

False Positive Rate, False Negative Rate

Dostupné datasety

Domain	Data	Size	Dimension	Anomaly (%)	Type
Intrusion detection	KDD Cup 99 [13]	4,091-567,497	41	0.30%-7.70%	Tabular
Intrusion detection	UNSW-NB15 [100]	257,673	49	≤9.71%	Streaming
Excitement prediction	KDD Cup 14	619,326	10	6.00%	Tabular
Dropout prediction	KDD Cup 15	35,091	27	0.10%-0.40%	Sequence
Malicious URLs detection	URL [93]	2.4m	3.2m	33.04%	Streaming
Spam detection	Webspam [160]	350,000	16.6m	39.61%	Tabular/text
Fraud detection	Credit-card-fraud [34]	284,807	30	0.17%	Streaming
Vandal detection	UMDWikipedia [76]	34,210	N/A	50.00%	Sequence
Mutant activity detection	p53 Mutants [13]	16,772	5,408	0.48%	Tabular
Internet ads detection	AD [13]	3,279	1,555	14.00%	Tabular
Disease detection	Thyroid [13]	7,200	21	7.40%	Tabular
Disease detection	Arrhythmia [13]	452	279	14.60%	Tabular
Defect detection	MVTec AD	5,354	N/A	35.26%	Image
Video surveillance	UCSD Ped 1 [81]	14,000 frames	N/A	28.6%	Video
Video surveillance	UCSD Ped 2 [81]	4,560 frames	N/A	35.9%	Video
Video surveillance	UMN [106]	7,739 frames	N/A	15.5%- 18.1%	Video
Video surveillance	Avenue [90]	30,652 frames	N/A	12.46%	Video
Video surveillance	ShanghaiTech Campus	317,398 frames	N/A	5.38%	Video
Video surveillance	UCF-Crime	1,900 videos (13.8m frames)	N/A	13 crimes	Video
System log analysis	HDFS Log [164]	11.2m	N/A	2.90%	Sequence
System log analysis	OpenStack log	1.3m	N/A	7.00%	Sequence

https://arxiv.org/pdf/2007.02500.pdf

Užitečná literatura / odkazy

- Scikit-learn - Novelty and Outlier Detection
- Deep Learning for Anomaly Detection: A Review
- Implementation of SOTA Deep Anomaly Detection Methods