Denis Derkach

Anomaly Detection

Advanced Techniques

2021

MLHEP 2021

NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

LAMBDA • HSE

Yandex

SCHOOL OF DATA ANALYSIS

EPFL

SIT Schaffhausen Institute of Technology

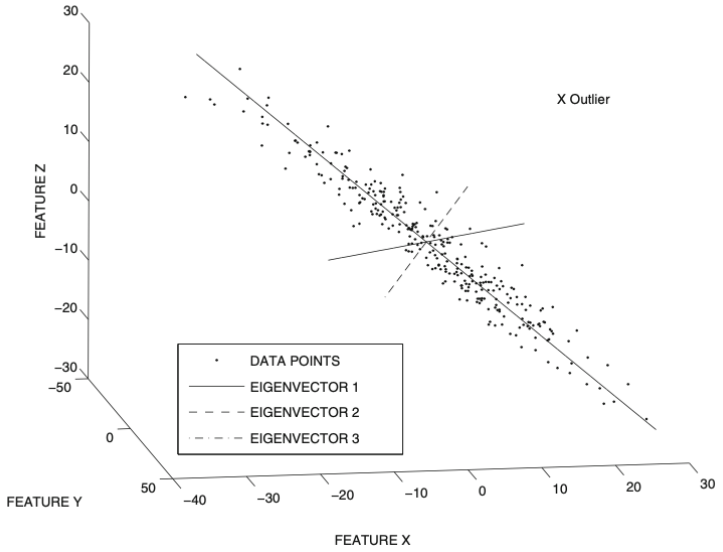
Distance scores

Introduction

- we know that anomalies are rare and deviate from the populous ”normal” class;
- ”normal” class is usually concentrated in some area of feature space;
- can we use this property?

Principle component analysis

- selects the k-dimensional hyperplane that minimizes the squared projection error over the remaining dimensions;
- all points can be projected to this hyperplane;
- a data point, which is far away from its projection is deemed as anomalous.
- anomaly score: normalized distance of the data point to the centroid of the sample along main components.

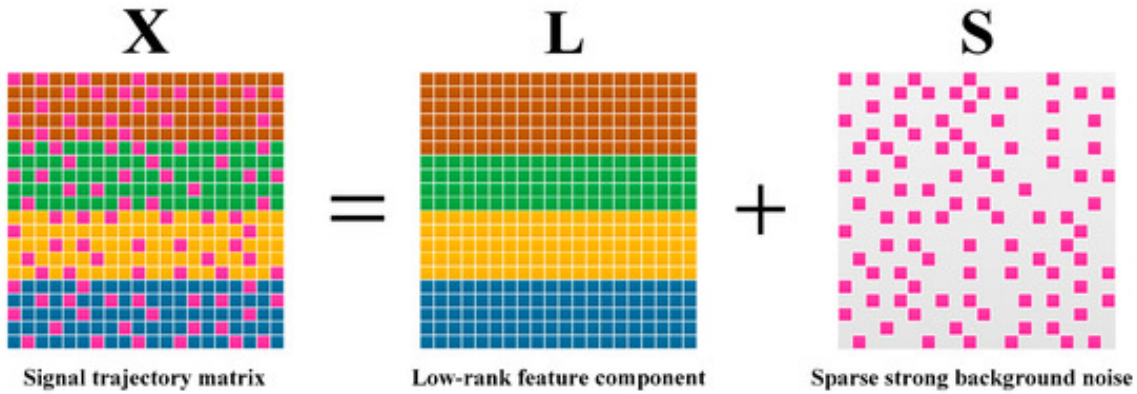


PCA for anomaly detection: issues

- sensitivity to noise
  - in presence of multiple outliers PCA can have difficulties in determining the main component.
- normalization issues
  - in case of very different feature scale, the variation of one components can eclipse other variations.
- regularization Issues
  - not really stable for small datasets.

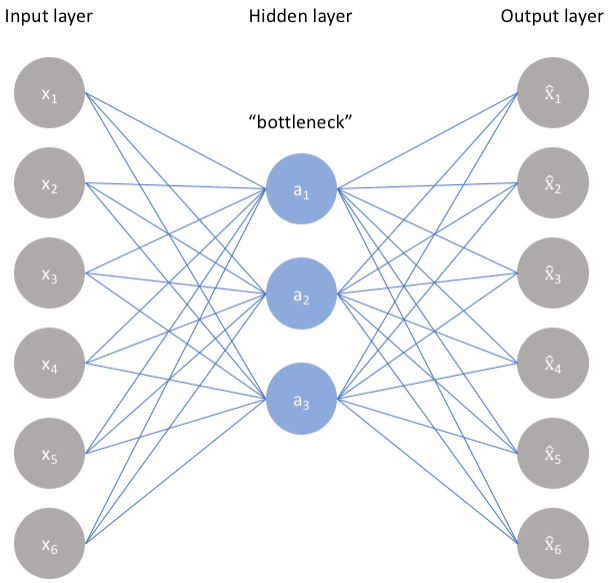
Robust PCA

The presence of many outliers can be overcome by using Robust PCA analysis. The analysis seeks to separate low-rank trends from sparse outliers within a data matrix:



Several methods of finding the decomposition exist.

Autoencoders



Two parts of the network:

- encoder $h = f(x)$;
- decoder $r = g(h)$

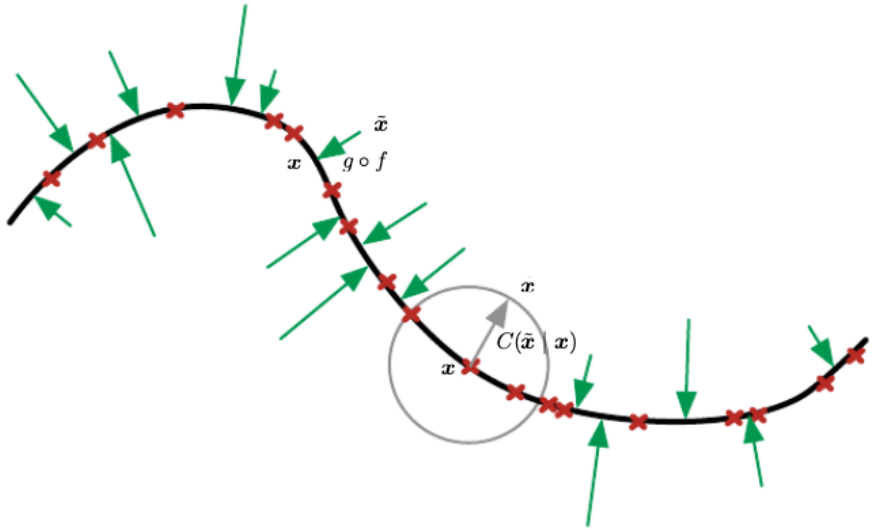
Generally, we want to find a transformation

$$g(f(x)) = x$$

The approach can be made more flexible than PCA transform.

AE: learning manifold

In fact, we learn a manifold, where normal class is situated:

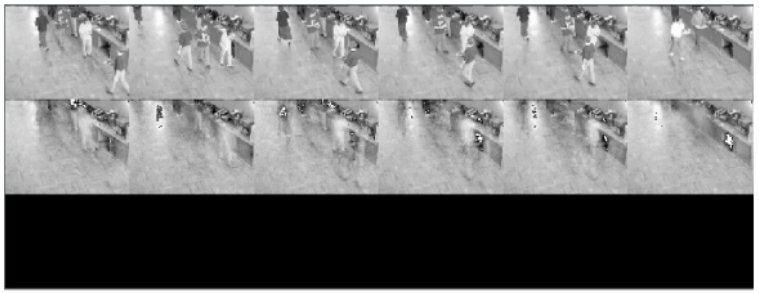


We can keep the same anomaly score as in PCA case.

Robust Autoencoder

- same problem as in the PCA case;
- same regularisation using sparse matrix S;
- can be learned iteratively;
- different architectures possible.

(a) RCAE.



(b) RPCA.

Variational Autoencoders

- ”normal” manifold can be created with probabilistic model;
- anomaly score remains distance based but we can sample from ”normal” distribution several events and average the distance.

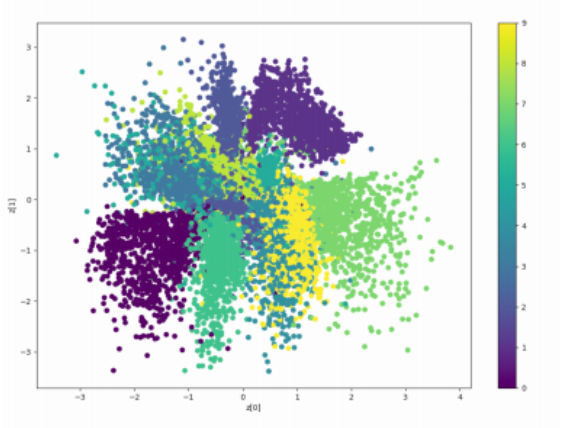


Figure 2.11: 2D plot of (variationally)autoencoded digits.

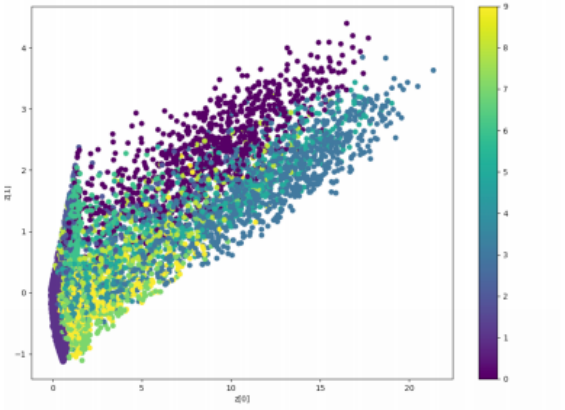


Figure 2.12: 2D plot of autoencoded digits.

Recap

- Linear methods are quite powerful for anomaly detection.
- Most of the analysis is done in the latent space.
- Issues:
  - data need to be correlated and not heavily clustered;
  - might be overfit;
  - lacks interpretability.

Probability Scores

Generative modeling

- Some generative modeling produce expilcit estimate of probability of sample:
  - Variational autoencoders.
  - Flow-based models.
- Can we use it to find anomaly?

Figure: N. Schucher

Constructing Score Function

- direct probability is overly optimistic for anomalous samples (tail problem!);
- one can try to construct a different probability-based measure:
  - Watanabe-Akaike Information Criterion;
  - use in-batch dependencies.
- empirically these approaches work better.

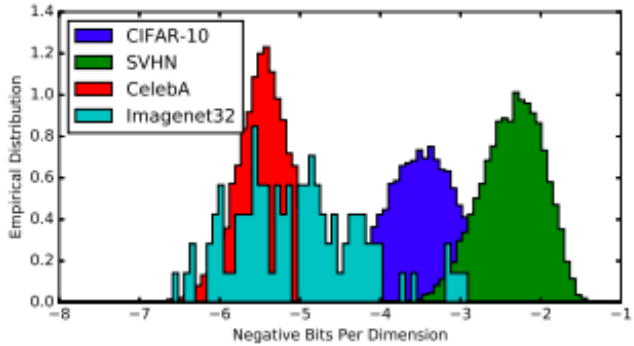


Figure 1. Density estimation models are not robust to OoD inputs. A GLOW model (Kingma & Dhariwal, 2018) trained on CIFAR-10 assigns much higher likelihoods to samples from SVHN than samples from CIFAR-10. .

H. Choi, WAIC, but Why. Generative Ensembles for Robust Anomaly Detection

Advanced Ideas

$(1+\varepsilon)$-class classification

Two-class classification:

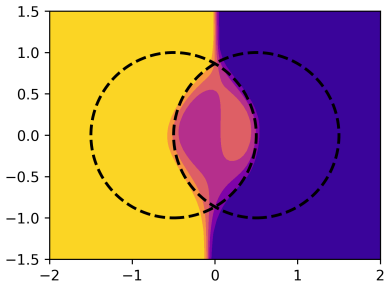
- undefined in empty regions;
- recovers proper probabilities;

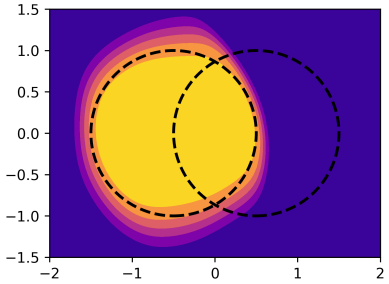
One-class classification:

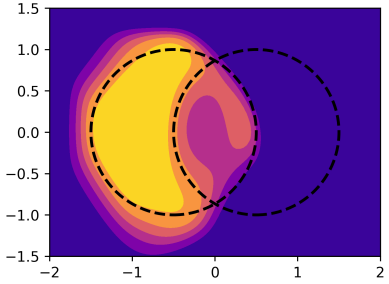
- defined everywhere;
- ignores negative class;

$(1+\varepsilon)$-class classification:

- shifts two-class solution towards a one-class solution;







Approach Classification

Numerous approaches have been developed:

- Extreme value analysis (Z-score).
- Probabilistic and statistical models (Generative models).
- Linear models (Principle Component Analysis)/
- Proximity-based models (Clustering)
- Information theoretic models (Minimal Description Analysis).
- High-dimensional outlier detection (isolation forest).

Methods can be combined into sequential and independent ensembles.

C. Aggarwal, Outlier Analysis

Summary

- Anomaly detection problem attracts a lot attention both from researchers and practitioners communities.
- Method should be selected based on the problem to be analysed.
- Many recent development in this area.