ANOMALY DETECTION USING KERNEL DENSITY ESTIMATION WITH DENSITY MATRICES

UNIVERSIDAD NACIONAL DE COLOMBIA

Felipe Osorio, Sebastian Medina, Landneyker Betancourth

INTRODUCTION

Anomaly detection (AD) is a fundamental problem in a wide variety of fields including banking, cybersecurity, manufacturing, system management and medicine. Density estimation is at the heart of anomaly detection, regardless of whether the data is high dimensional or not.

Density estimation is a statistical task that involves constructing an estimate of the true probability density function (PDF) which generated a set of observations. Two possible use cases for density estimation are: assuming if the data comes from a proposed parametric model or employing a non-parametric model approach.

BACKGROUND AND RELATED WORK

> **Anomaly detection.**

> **Density estimation.**

> **Kernel density estimation (KDE).**

> **Random Fourier Features (RFF).**

> **Density Matrices.**

VIDEO LINK: https://youtu.be/AwpQhPBnqxQ

METHOD

PROCESS

The overall process is defined next:

- Input. An observation of a d-dimensional random sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$, number of random Fourier features $D \in \mathbb{N}$ and spread parameter $\gamma \in \mathbb{R}_{>0}$.
- Generate an observation $\omega_1, \ldots, \omega_N$ of a random sample of $\omega \sim N(\mathbf{0}, \mathbf{I}_D)$ and an observation $b_1, \ldots, b_N$ of a random sample of $b \sim \text{Uniform}(0, 2\pi)$ for building the map ϕ_{rff} from the random Fourier features method to approximate a Gaussian kernel with parameters γ and D.
- Apply ϕ_{rff} to each element $\mathbf{x}_i$:

$$\mathbf{z}_i = \phi_{\text{rff}}(\mathbf{x}_i) \tag{7}$$

- Density matrix estimation:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^T \tag{8}$$

The density estimation of a query point $\mathbf{x}$ is calculated using Born's rule

$$\hat{f}_\rho(\mathbf{x}) = \frac{\text{Tr}\left(\rho \phi_{\text{rff}}(\mathbf{x}) \phi_{\text{rff}}(\mathbf{x})^T\right)}{\mathcal{Z}} = \frac{\phi_{\text{rff}}(\mathbf{x})^T \rho \phi_{\text{rff}}(\mathbf{x})}{\mathcal{Z}} \tag{9}$$

where the normalizing constant is given by:

$$\mathcal{Z} = \left(\frac{\pi}{2\gamma}\right)^{\frac{d}{2}} \tag{10}$$

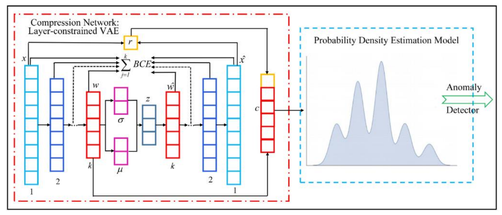
The overall DMKDE process.

We call f_p the DMKDE estimator. A threshold $\delta \geq 0$ is defined for the density of a new sample which will say if it has such a lower value in order to be classified as an outlier. Formally, a new sample $x \in \mathbb{R}^d$ will be an outlier if $f_p(x) \leq \delta$, otherwise it will be an inlier.

OC-SVM: Use the mathematical formulation of Support Vector Machines for the unsupervised imbalanced data classification problem. The main idea is to estimate a function that returns the value +1 in a region that captures all the inlier data points and −1 everywhere else by mapping the data on a kernel induced feature space and divide them with a maximum margin. For any new point that we want to classify, we simply return $f(x)$ as the evaluation of the side of the hyperplane the point falls on the feature space. The separation of the data points in the feature space follows the following quadratic optimization problem:

$$\min_{w \in F, \boldsymbol{\xi} \in \mathbb{R}^\ell, \rho \in \mathbb{R}} \frac{1}{2}\|w\|^2 + \frac{1}{\nu\ell} \sum_i \xi_i - \rho$$
$$\text{subject to } (w \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0$$

LAKE: Layer-constrained variational autoencoding kernel density estimation model for anomaly detection from high dimensional datasets, is a probability density-aware strategy that learns a probability density distribution of the high dimensional data in the training process that can effectively detect abnormal objects in the testing.



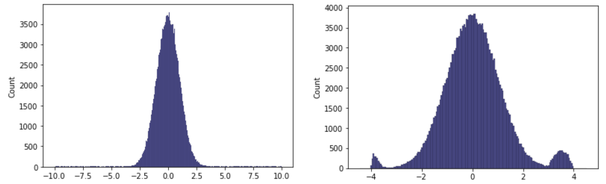
An overview on layer-constrained variational autoencoding kernel density estimation model.

DATASETS

KDDCUP: The KDDCUP data set from UCI Machine Learning Repositor is a network intrusion data set. We use one-hot representation to encode them and obtain a 122-dimensional data set. As 20% of them are marked as normal and others are labeled as attack, and normal samples make up a small percentage of the total. As a result, we classify normal samples as anomalies in our experiment.

NORMAL: We generate a classical example in statistics which consists of a one-dimensional normal standard sample and generate a portion of anomalies outside 3σ region using an uniform distribution.

BETA: In this case we generate inlier cases using a one-dimensional normal standard distribution, and outliers in this case represent two samples of a displaced Beta distribution.



NORMAL DATA SET BETA DATA SET

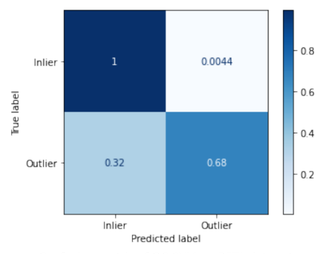
The purpose is to see how effective Kernel Density Estimation using Density Matrices is at recognizing anomalies by using as a binary classifier. We compare it against other algorithms which are used in practice for detecting anomalies as a binary classification problem.

RESULTS

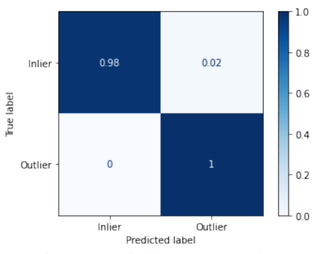
Results of the methods with accuracy, $F1$, recall and precision:

Method	KDD			
	Accuracy	F_1	Precision	Recall
OC-SVM	0.208	0.045	0.684	0.023
LAKE	**0.600**	**0.750**	**0.750**	**0.750**
DMKDE	0.199	0.0	0.0	0.004
Method	NORMAL			
	Accuracy	F_1	Precision	Recall
OC-SVM	0.979	0.494	0.328	**1**
LAKE	0.980	0	0	0
DMKDE	**0.998**	**0.929**	**0.978**	0.885
Method	BETA			
	Accuracy	F_1	Precision	Recall
OC-SVM	**0.983**	0.757	**0.859**	0.677
LAKE	0.921	0.0002	0.0002	0.0002
DMKDE	0.981	**0.806**	0.675	**1**

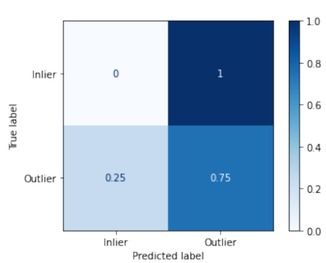
Average precision, recall and $F1$ from DMKDE and all baselines. For each metric, the best result is shown in bold.



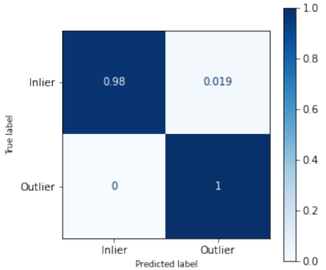
Confusion matrix of OC-SVM on BETA data set.



Confusion matrix of OC-SVM on NORMAL data set.



Confusion matrix of LAKE on KDD data set.



Confusion matrix of DMKDE on BETA data set showing a perfect score on detecting outliers.