INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD
VI Semester B.Tech in Information Technology
Report - Group Assignment 2

Data Mining and Warehousing

Deep One-Class Classification

By :-

Palak Mittal (IIT2018117)

Nehal Singh (IIT2018119)

Abhishek Kumar Gupta (IIT2018187)

Puja Kumari (IIT2018191)

Prabha Kumari (IIT2018195)

1. INTRODUCTION

Anomaly detection (AD) is identifying rare items ,observations or events which raise suspicion by differing significantly from the majority of the data.Basically, It is the task of discerning unusual samples in data . This is treated as an unsupervised learning problem where the anomalous samples are not known a priori and it is assumed that the majority of the training dataset consists of "normal" data ((The term "normal" means not anomalous and is unrelated to the Gaussian distribution).This is known as one-class classification. AD algorithms are often trained on data collected during the normal operating state of a machine.

2. PROPOSED PROBLEM :

In this problem we have to detect anomalies in our data . We have proposed a novel approach to deep AD inspired by kernel-based one-class classification and minimum volume estimation. Our method, Deep Support Vector Data Description (Deep SVDD), trains a neural network while minimizing the volume of a hypersphere that encloses the network representations of the data .

3. ALGORITHM

- First of all , We build on the kernel-based SVDD and minimum volume estimation by finding a data-enclosing hypersphere of smallest size.
- Then , we employ a neural network that is jointly trained to map the data into a hypersphere of minimum volume.
- Then Deep SVDD optimization and selection of the hypersphere is done .

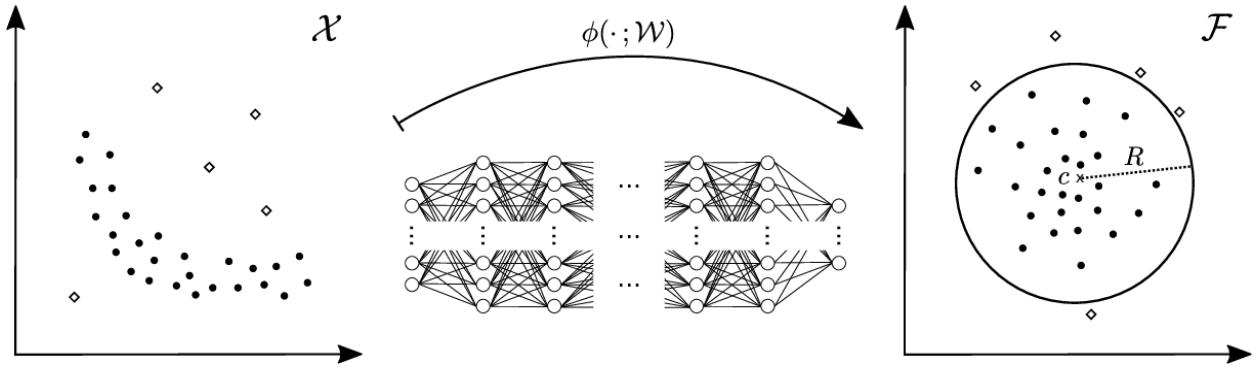


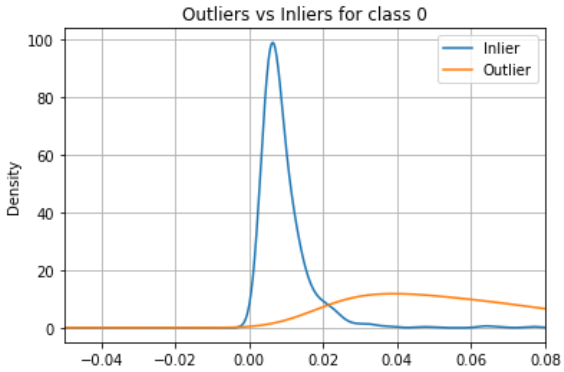
Fig : Deep SVDD learns Neural Network Transformation

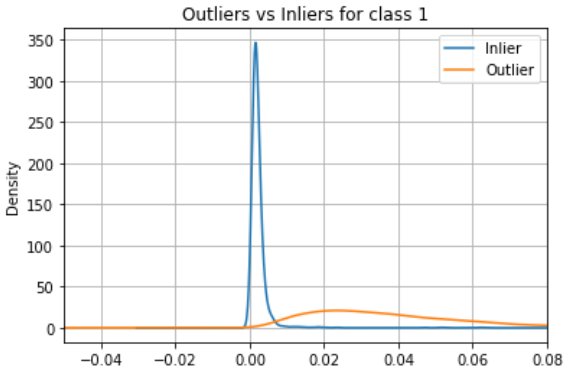
4. RESULT :

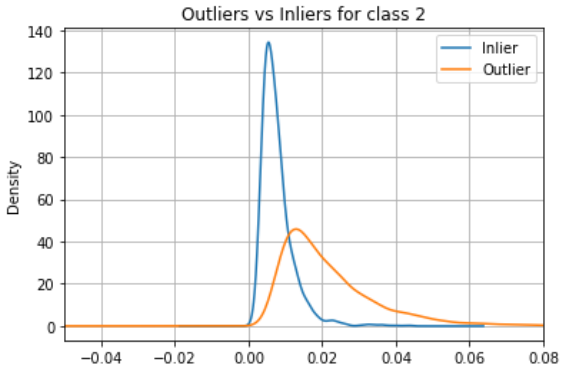
ROC SCORES FOR EACH CLASSES WERE:

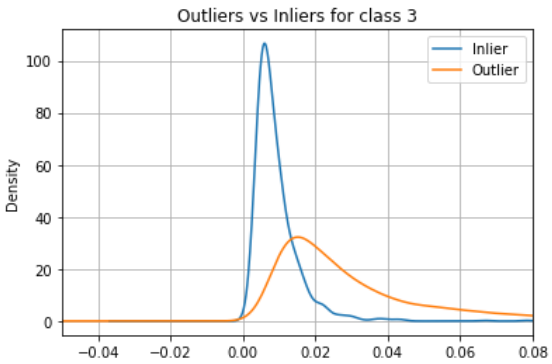
```
ROC scores for class  0  is:  98.79004706095299
ROC scores for class  1  is:  99.60754439450295
ROC scores for class  2  is:  91.3406906727797
ROC scores for class  3  is:  90.24187491051664
ROC scores for class  4  is:  92.38666816626986
ROC scores for class  5  is:  85.27766453775423
ROC scores for class  6  is:  97.23715100812308
ROC scores for class  7  is:  95.31597221619877
ROC scores for class  8  is:  94.77695282303326
ROC scores for class  9  is:  95.8508558112126
```

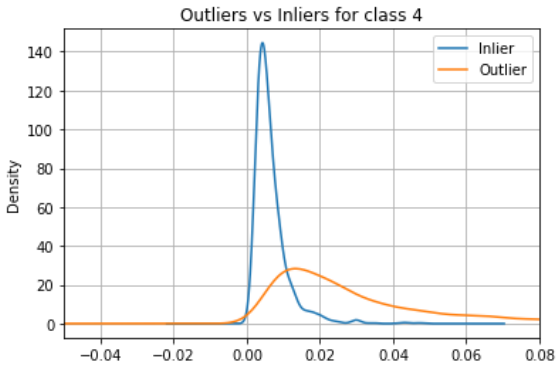
Outliers V/S inliers for each class :

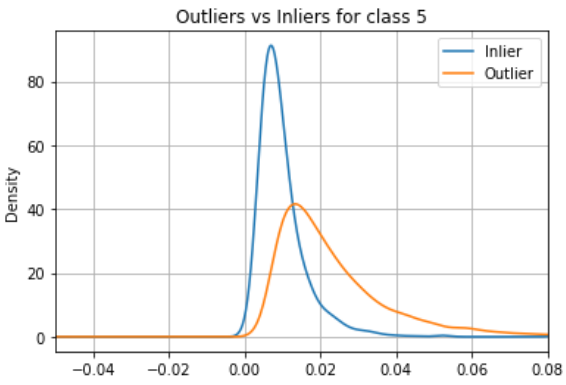












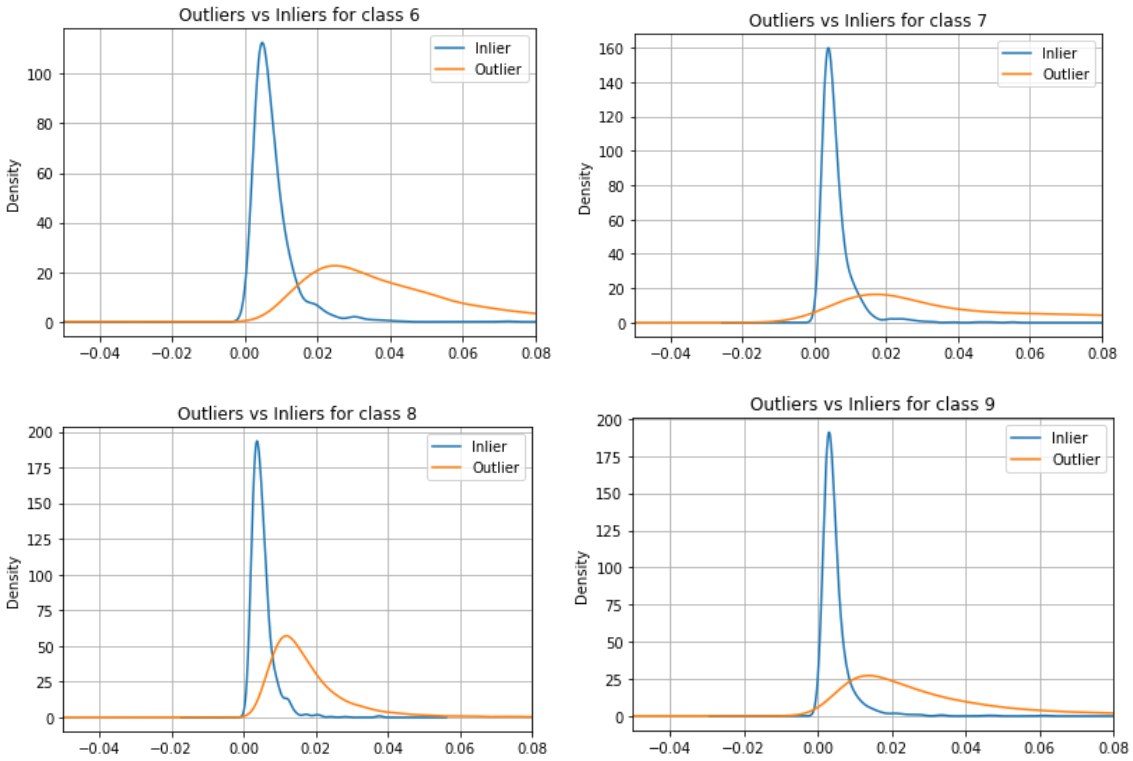


Table 1 : Average AUCs in % with StdDevs (over 10 seeds) per one-class experiment on MNIST .

According to this table , The One-Class Deep SVDD (IMPLEMENTED) showed better/similar performance as compared to One-Class Deep SVDD (In Paper).

Normal Class	DEEP ONE-CLASS (IN PAPER)	DEEP ONE-CLASS (IMPLEMENTED)
0	98.0±0.7	98.79004706095299
1	99.7±0.1	99.60754439450295
2	91.7±0.8	91.3406906727797
3	91.9±1.5	90.24187491051664
4	94.9±0.8	92.38666816626986
5	88.5±0.9	85.27766453775423
6	98.3±0.5	97.23715100812308

7	94.6±0.9	95.31597221619877
8	93.9±1.6	94.77695282303326
9	96.5±0.3	95.8508558112126

So, here we can see that we had implemented the algorithm and achieved the accuracy (almost better) as mentioned in the research paper.

5. CONCLUSION

We introduced a deep one-class classification objective for unsupervised AD in this work. Our method, Deep SVDD trains a deep neural network while optimiz- ing a data-enclosing hypersphere in output space. Through this Deep SVDD extracts common factors of variation from the data. It demonstrates quantitatively as well as qualitatively the sound performance of Deep SVDD.
