Assignment 5: Evaluating one-class classifiers for anomaly detection

A.C. Miranda, R. Quintero, and C. Pérez

November 3, 2021

1 Assignment Description

The students have evaluated 60 databases on different algorithms to conclude on their performance according to AUC. The evaluated algorithms are Bagging-Random Miner (BRM) with different, Gaussian MIxture Model (GMM), Isolation Forest (ISOF), and One-Class Support Vector Machine (OCSVM). Particularly BRM was evaluated using three different dissimilarity measures. These results were compared in cases where data was processed without normalization, using MinMax scaling, and using Standard normalizing. At the end the students provide box plots, Friedman tests with posthoc test, and CD diagrams that explain these patterns.

2 Methodology

2.1 Data preparation

The 60 databases were given in .dat format that led to processing in order to work with .csv files from a containing directory. Each one of these databases had a training and testing separate file with its respective attributes and class column, from which testing and training information where obtained. Subsequently, two different functions allowed for the creation of two different sets of normalized variables using MinMax scaling and Standard scaling.

2.2 Model evaluation

Each database was trained using BRM, GMM, ISOF, OCSVM. However, BRM was additionally evaluated using three different dissimilarity metrics: Manhattan distance, Cosine distance, and Chebyshev distance. The resulting AUC of each model evaluation was stored for each database without normalizing and for each scaled database.

3 Results

3.1 Algorithm evaluation

As previously mentioned, each algorithm was tested for all databases to contrast their results under different data transformations. For example, table 1 shows a comparison of AUC performance in the glass0 database when scaling.

Table 1. AUC results for the glass0 database under different scaling conditions.

Database	Model	AUC Normal	AUC MinMax	AUC Standard
glass0	BRM	0.756158	0.76601	0.763547
glass0	GMM	0.310345	0.310345	0.310345
glass0	ISOF	0.586207	0.586207	0.586207
glass0	OCSVM	0.703202	0.703202	0.703202
glass0	manhattanBRM	0.731527	0.738916	0.738916
glass0	cosineBRM	0.748768	0.753695	0.753695
glass0	chebyshevBRM	0.763547	0.761084	0.763547

Box-plots of AUC distribution were generated in order to better appreciate the performance of the different models under these conditions of data transformation. Figure **??** summarizes the results conveyed when evaluating the databases in its normal state, that is without any normalizing performed. On the other hand, Figure 2 and Fig 3 show respectively the distributed AUC results of the databases when MinMax and Standard scaling is performed.

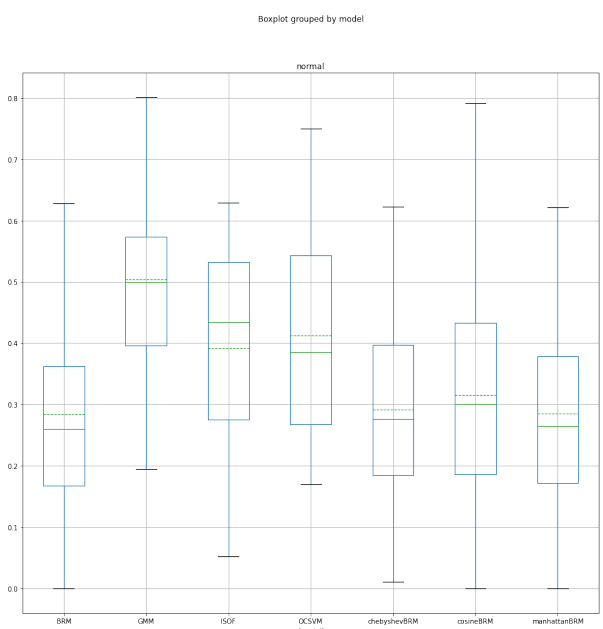

Fig. 1. Box-plot of AUC distribution in databases without scaling.

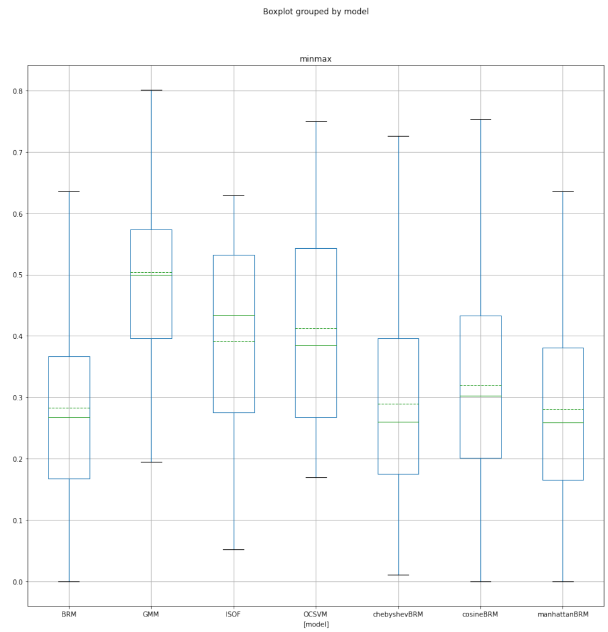

Fig. 2. Box-plot of AUC distribution in databases with MinMax scaling.

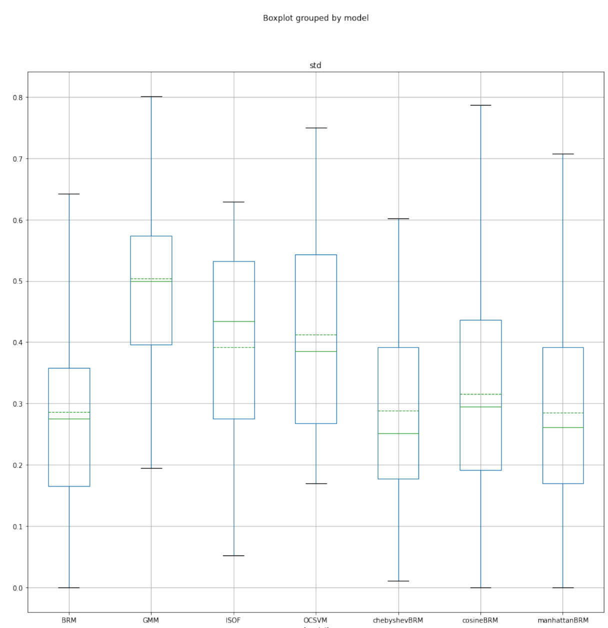

Fig. 3. Box-plot of AUC distribution in databases with Standard scaling.

3.2 Statistical tests

In order to draw conclusions from the evaluated models, Friedman tests with posthoc tests were conducted for the three different data base scenarios.

Firstly, the statistical analysis was conducted for 7 populations with 60 paired samples without scaling. The family-wise significance level of the tests is alpha=0.050. We rejected the null hypothesis that the population is normal for the populations ISOF (p=0.004) and cosineBRM (p=0.001). Therefore, we assume that not all populations are normal.

Because we have more than two populations and the populations and some of them are not normal, we use the non-parametric Friedman test as omnibus test to determine if there are any significant differences between the median values of the populations. We use the post-hoc Nemenyi test to infer which differences are significant. We report the median (MD), the median absolute deviation (MAD) and the mean rank (MR) among all populations over the samples. Differences between populations are significant, if the difference of the mean rank is greater than the critical distance CD=1.163 of the Nemenyi test.

We reject the null hypothesis (p=0.000) of the Friedman test that there is no difference in the central tendency of the populations GMM (MD=0.500+-0.095, MAD=0.134, MR=2.758), OCSVM (MD=0.385+-0.141, MAD=0.202, MR=2.833), ISOF (MD=0.435+-0.133, MAD=0.194, MR=3.350), cosineBRM (MD=0.300+-0.126, MAD=0.195, MR=4.475), chebyshevBRM (MD=0.277+-0.111, MAD=0.155, MR=4.842), manhattanBRM (MD=0.264+-0.119, MAD=0.156, MR=4.867), and BRM (MD=0.260+-0.108, MAD=0.140, MR=4.875). Therefore, we assume that there is a statistically significant difference between the median values of the populations. Based on the post-hoc Nemenyi test, we assume that there are no significant differences within the following groups: GMM, OCSVM, and ISOF; ISOF and cosineBRM; cosineBRM, chebyshevBRM, manhattanBRM, and BRM. All other differences are significant.

This can be appreciated in Table 2, and seen in Figure 4.

model	MR	MED	MAD	CI	γ	Magnitude
GMM	2.758	0.500	0.134	[0.399, 0.589]	0.000	negligible
OCSVM	2.833	0.385	0.202	[0.270, 0.553]	0.671	medium
ISOF	3.350	0.435	0.194	[0.284, 0.550]	0.392	small
cosineBRM	4.475	0.300	0.195	[0.192, 0.444]	1.195	large
chebyshevBRM	4.842	0.277	0.155	[0.187, 0.410]	1.543	large
manhattanBRM	4.867	0.264	0.156	[0.173, 0.411]	1.620	large
BRM	4.875	0.260	0.140	[0.169, 0.385]	1.745	large

Table 2. Summary of populations without scaling.

Secondly, the statistical analysis was conducted for 7 populations with 60 paired samples scaled using MinMax. The family-wise significance level of the tests is alpha=0.050. We rejected the null hypothesis that the population is nor-

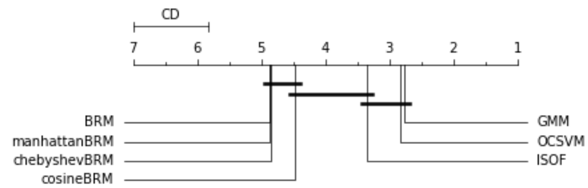

Fig. 4. CD diagram in databases without scaling.

mal for the populations ISOF (p=0.004) and cosineBRM (p=0.001). Therefore, we assume that not all populations are normal.

Because we have more than two populations and the populations and some of them are not normal, we use the non-parametric Friedman test as omnibus test to determine if there are any significant differences between the median values of the populations. We use the post-hoc Nemenyi test to infer which differences are significant. We report the median (MD), the median absolute deviation (MAD) and the mean rank (MR) among all populations over the samples. Differences between populations are significant, if the difference of the mean rank is greater than the critical distance CD=1.163 of the Nemenyi test.

We reject the null hypothesis (p=0.000) of the Friedman test that there is no difference in the central tendency of the populations GMM (MD=0.500+-0.095, MAD=0.134, MR=2.750), OCSVM (MD=0.385+-0.141, MAD=0.202, MR=2.817), ISOF (MD=0.435+-0.133, MAD=0.194, MR=3.342), cosineBRM (MD=0.295+-0.128, MAD=0.182, MR=4.483), chebyshevBRM (MD=0.251+-0.115, MAD=0.130, MR=4.817), BRM (MD=0.275+-0.119, MAD=0.162, MR=4.867), and manhattanBRM (MD=0.261+-0.118, MAD=0.167, MR=4.925). Therefore, we assume that there is a statistically significant difference between the median values of the populations. Based on the post-hoc Nemenyi test, we assume that there are no significant differences within the following groups: GMM, OCSVM, and ISOF; ISOF and cosineBRM; cosineBRM, chebyshevBRM, BRM, and manhattanBRM. All other differences are significant.

This can be appreciated in Table 3, and seen in Figure 5.

model	MR	MED	MAD	CI	γ	Magnitude
GMM	2.750	0.500	0.134	[0.399, 0.589]	0.000	negligible
OCSVM	2.817	0.385	0.202	[0.270, 0.553]	0.671	medium
ISOF	3.342	0.435	0.194	[0.284, 0.550]	0.392	small
cosineBRM	4.483	0.295	0.182	[0.193, 0.449]	1.282	large
chebyshevBRM	4.817	0.251	0.130	[0.180, 0.410]	1.883	large
BRM	4.867	0.275	0.162	[0.165, 0.403]	1.512	large
manhattanBRM	4.925	0.261	0.167	[0.173, 0.408]	1.581	large

Table 3. Summary of populations with MinMax scaling.

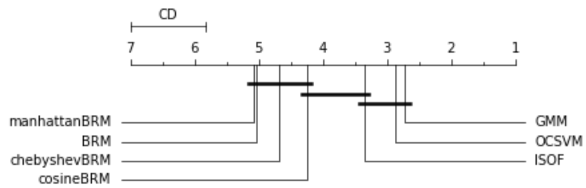

Fig. 5. CD diagram in databases with MinMax scaling.

Lastly, the statistical analysis was conducted for 7 populations with 60 paired samples using Standard scaling. The family-wise significance level of the tests is alpha=0.050. We rejected the null hypothesis that the population is normal for the populations ISOF (p=0.004) and cosineBRM (p=0.001). Therefore, we assume that not all populations are normal.

Because we have more than two populations and the populations and some of them are not normal, we use the non-parametric Friedman test as omnibus test to determine if there are any significant differences between the median values of the populations. We use the post-hoc Nemenyi test to infer which differences are significant. We report the median (MD), the median absolute deviation (MAD) and the mean rank (MR) among all populations over the samples. Differences between populations are significant, if the difference of the mean rank is greater than the critical distance CD=1.163 of the Nemenyi test.

We reject the null hypothesis (p=0.000) of the Friedman test that there is no difference in the central tendency of the populations GMM (MD=0.500+-0.095, MAD=0.134, MR=2.750), OCSVM (MD=0.385+-0.141, MAD=0.202, MR=2.817), ISOF (MD=0.435+-0.133, MAD=0.194, MR=3.342), cosineBRM (MD=0.295+-0.128, MAD=0.182, MR=4.483), chebyshevBRM (MD=0.251+-0.115, MAD=0.130, MR=4.817), BRM (MD=0.275+-0.119, MAD=0.162, MR=4.867), and manhattanBRM (MD=0.261+-0.118, MAD=0.167, MR=4.925). Therefore, we assume that there is a statistically significant difference between the median values of the populations. Based on the post-hoc Nemenyi test, we assume that there are no significant differences within the following groups: GMM, OCSVM, and ISOF; ISOF and cosineBRM; cosineBRM, chebyshevBRM, BRM, and manhattanBRM. All other differences are significant. significant.

This can be appreciated in Table 4, and seen in Figure 6.

4 Conclusions

This assignment let us gather conclusions of the performance that different algorithms have on one class anomaly detection. A directory comprised by 60 different datasets allowed for a robust comparison of algorithm performance using different statistical tests as well as the generation of box-plot diagrams that aided the interpretation of such results.

model	MR	MED	MAD	CI	γ	Magnitude
GMM	2.750	0.500	0.134	[0.399, 0.589]	0.000	negligible
OCSVM	2.817	0.385	0.202	[0.270, 0.553]	0.671	medium
ISOF	3.342	0.435	0.194	[0.284, 0.550]	0.392	small
cosineBRM	4.483	0.295	0.182	[0.193, 0.449]	1.282	large
chebyshevBRM	4.817	0.251	0.130	[0.180, 0.410]	1.883	large
BRM	4.867	0.275	0.162	[0.165, 0.403]	1.512	large
manhattanBRM	4.925	0.261	0.167	[0.173, 0.408]	1.581	large

Table 4. Summary of populations with Standard scaling.

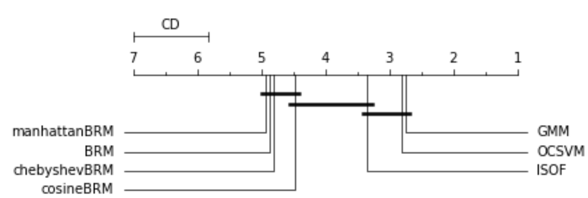

Fig. 6. CD diagram in databases with Standard scaling.