Feature scaling in support vector data description

P. Juszczak[*] D.M.J. Tax[†] R.P.W. Duin[*]

[*]Pattern Recognition Group, Department of Applied Physics, Faculty of Applied Sciences,
Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

[†]Fraunhofer Institute FIRST.IDA
Kekuléstr.7, D-12489 Berlin, Germany

email: piotr@ph.tn.tudelft.nl

Keywords: one-class classification problems, feature scaling, support vector data description

Abstract

When in a classification problem only samples of one class are easily accessible, this problem is called a one-class classification problem. Many standard classifiers, like backpropagation neural networks, fail on this data. Some other classifiers, like k-means clustering or nearest neighbor classifier can be applied after some minor changes. In this paper we focus on the support vector data description classifier, which is especially constructed for one-class classification. But this method appears to be sensitive to scaling of the individual features of the dataset. We show that it is possible to improve its performance by adequate scaling of the feature space. Some results will be shown on artificial dataset and handwritten digits dataset.

1 Introduction

In the problem of one-class classification, one class of the data, called the target set, has to be distinguished from all the other possible objects, called outliers. This description should be constructed such that objects not originating from the target set are not accepted by the data description. It is assumed that almost no examples of the outlier class are available.

In general, the problem of one-class classification is harder than the problem of normal two-class classification. For normal classification the decision boundary is supported from both sides by examples of each of the classes. Because in the case of one-class classification only one set of data is available, only one side of the boundary is covered. On the basis of one class it is hard to decide how tight the boundary should fit around the data in each of the directions.

The absence of example outlier objects makes it also very hard to estimate the error that the classifier makes. The error of the first kind $\mathcal{E}_{\mathcal{I}}$, the target objects that are classified as outlier objects, can be estimated on the training set. The error of the second kind $\mathcal{E}_{\mathcal{II}}$, the outlier objects that will be classified as target objects, cannot be estimated without assumptions on the distribution of the outliers. As long as we do not have example outlier objects available, we assume therefore that the outliers are uniformly distributed in the feature space. This directly means, that when the chance of accepting an outlier object is minimized, the volume covered by the one-class classifier in the feature space should be minimized.

Using the uniform distribution for the outlier objects, implicitly assumes that the objects are represented by 'good' features. This means that outlier objects will be around the target class and not inside it. When it appears that there is still some overlap between the target objects and outlier objects, the representation of the objects should be changed such that the distinction becomes easier.

In this paper we show a few simple possibilities for rescaling the features in order to improve the distinction between the target and the outlier class for support vector data description (SVDD) classifier. The performance of the classifier depends on representation of the data. For some one-class classifiers this relation is strong: for instance in the mixture of Gaussians, SVDD, k-centers, k-means, nearest neighborhood for others like PCA, auto-encoder, diabolo network, normal density estimator this relation is weak [1]. The former classifiers are scale dependent and the second group of classifiers are scale independent. This means if we rescale the feature space the performance of the classifier for the first group significantly changes.

What do we mean by a good representation of the feature set in the case of the SVDD? It appears that the SVDD is written in terms of a few objects in the training set, the so-called support vectors (SVs). The number of SVs indicates how complicated the boundary is around the target set. For spherical representations of the data the number of support vectors is 3 or in some cases 2. This type of the boundary also minimizes the error of the second kind $\mathcal{E}_{\mathcal{II}}$, the outlier acceptance.

We applied three different types of scaling to obtain the most appropriate spherical shape of the target set:

1. scaling by variance - the features in each direction is divided by its variance

2. scaling by domain - all features are scaled to range [0, 1],

3. scaling to minmax - the minimum of the maximum value of feature in all directions is assigned as the radius of the sphere R, next the features are scaled to the range $[0, R]$.

We describe these methods in more detail and their mathematical description in the next section. The efficiency of appropriate rescaling the feature space is shown on a artificial datasets and a real world handwritten digits dataset. The results are presented in the section Experiments.

2 Theory

Now we would like to give a short description of the SVDD. For more information we refer to [2, 1]. In the SVDD the data is enclosed by a hypersphere with minimum volume. By minimizing the volume of the feature space, or equivalently minimizing the radius R we hope to minimize the chance of accepting outlier objects. So in analogy to the support vector classifier [4] we can define the structural error:

$$\mathcal{E}_{struct}(R, a) = R^2 \tag{1}$$

where a is the center of the sphere and equation (1) has to be minimized with the constraint:

$$|x_i - a|^2 \leq R^2, \qquad \forall i \tag{2}$$

To allow the possibility of outliers in the training set, we can introduce slack variables ξ, and minimize the following error function:

$$\mathcal{E}(R, a, \xi) = R^2 + C \sum_i \xi_i \tag{3}$$

where C gives the tradeoff between the volume of the data description and the errors it is making on the target data.

So we constrain the solution such that almost all objects are in the sphere:

$$|x_i - a|^2 \leq R^2 + \xi \qquad \xi \leq 0, \qquad \forall i \tag{4}$$

By introducing the Lagrange multipliers α, γ and constructing the Lagrangian from equation (3) according to constraints (4) one obtains:

$$\begin{aligned} L(R, a, \xi, \alpha, \gamma) &= R^2 + C \sum_i \xi_i \\ &- \sum_i \alpha_i (R^2 + \xi i - (x_i \cdot x_i - 2a \cdot x_i + a \cdot a)) \\ &- \sum_i \gamma_i \xi_i \end{aligned} \tag{5}$$

Setting partial derivatives R, a, ξ of $L(R, a, \xi, \alpha, \gamma)$ to 0 gives the constraints:

$$\sum_i \alpha_i = 1 \tag{6}$$

$$a = \sum_i \alpha_i x_i \tag{7}$$

$$0 \leq \alpha_i \leq C \tag{8}$$

Applying equations (6-8) to equation (5) we obtain an equation for the error L:

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \tag{9}$$

The minimization of the error function (9) with the constraint (8) is a well-known problem called the quadratic programming problem and standard algorithms exist.

Finally the function that describes the boundary decision for one-class classification problem it can be state as:

$$\begin{aligned} f_{SVDD}(z, \alpha, R) &= I(|z - a|^2 \leq R^2) \\ &= I((z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) \\ &+ \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2) \end{aligned} \tag{10}$$

where z is a new test object and I is a function defined as:

$$I(A) = \begin{cases} 1 & \text{if A is true} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Support vectors are those elements of the dataset for which $\alpha > 0$. For all others $\alpha = 0$. When the function f_{SVDD} is calculated for a new object z only support vectors are non-zero elements in the sums.

The hypersphere is a very rigid boundary around the data and often does not give a good description of it. The idea of support vector data description is to map the training data nonlinearly into a higher-dimensional feature space and construct a separating hyperplane with maximum margin there. This yields a nonlinear decision boundary in the input space. By the use of a kernel function, it is possible to compute the separating hyperplane without explicitly carrying out the map into the feature space. Introducing a kernel function

$$k(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \tag{12}$$

one can avoid to compute the dot product $x_i \cdot x_j$. A kernel function is any kind of a function that obeys Mercer's Theorem [3]. The most often used kernels are:

1. the polynomial kernel:
$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^d \tag{13}$$
where d is the degree of polynomial

2. the radial basis function (RBF) e.g.:
$$K_\gamma(|x_j - x_i|) = exp\left(\frac{|x_j - x_i|^2}{s^2}\right) \tag{14}$$

3. the sigmoid kernel, a combination of sigmoid functions $S(u)$:
$$K(x_i, x_j) = \tanh(k(x_i \cdot x_j) + c) \tag{15}$$

satisfies Mercer conditions only for some values of parameters k, c [4].

In our experiments we are using the Gaussian kernel of the form (14). This kernel is independent of the position of the data set with respect to the origin, it only uses the distance $|x_i - x_j|$ between objects. For the Gaussian kernel no finite mapping $\Phi(x_i)$ of an object x can be given. The fact that $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_i) = 1$ means that the mapped object $x_i^* = \Phi(x_i)$ has norm equal to 1. Because of that, both the Lagrangian (9) and the discriminant function (10) simplify. Now we have to minimize the Lagrangian:

$$L = -\sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \tag{16}$$

to evaluate if a new object z is a target or an outlier, formula (10) can be rewritten as:

$$\begin{aligned} &f_{SVDD}(z, \alpha, R) = \\ &I\left(\sum_i \alpha_i exp\left(\frac{-|z - x_i|^2}{s^2}\right) > \right. \\ &\left.\frac{1}{2}(B - R)^2\right) \end{aligned} \tag{17}$$

where $B = 1 + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j)$ only depends on the support vectors x_i and not on the object z.

We applied three different types of rescaling the feature space to obtain the most appropriate spherical shape of the target set:

1. scaling by variance - the features in each direction l are divided by their variance σ_l in this direction.
$$\widehat{x}_{l,m} = \frac{x_{l,m}}{\sigma_l}, \quad \underset{m=1\dots n}{\forall} \wedge \underset{l=1\dots k}{\forall} \tag{18}$$
where $\widehat{x}_{l,m}$ is a scaled feature, n is a number of samples in the training set and k is a number of dimensions of a feature space.

2. scaling by domain - all feature are scaled to range [0 1]
$$\widehat{x}_{l,m} = \left|\frac{x_{l,m}}{\max\limits_m(x_{l,m})}\right| \quad \underset{m=1\dots n}{\forall} \wedge \underset{l=1\dots k}{\forall} \tag{19}$$
where $\max\limits_m(x_{l,m})$ is the maximum value of the data in the l direction.

3. scaling to minmax - minimum of the maximum value of features in all k directions is assigned as the radius of the sphere R. All other features are then rescaled to the range $[0R]$.

$$\widehat{x}_{l,m} = \left|\frac{x_{l,m}}{\min\limits_l(\max\limits_m(x_{l,m}))}\right| \quad \underset{m=1\dots n}{\forall} \wedge \underset{l=1\dots k}{\forall} \tag{20}$$

Now we rewrite equation (17) for the rescaled data

$$\begin{aligned} &f_{SVDD}(\widehat{z}, \alpha', R') = \\ &I\left(\sum_i \alpha'_i exp\left(\frac{-|\widehat{z} - \widehat{x}_i|^2}{s^2}\right) > \right. \\ &\left.\frac{1}{2}(B' - R')^2\right) \end{aligned} \tag{21}$$

Where ˆ - indicates values directly rescaled by equations (18 - 20) and ′ - indicates changed values after rescaling. By minimizing the volume of the target set we minimize the structural error $\mathcal{E}_{struct}$ in equation (1).

3 Experiments

In this section we want to evaluate how well the rescaling procedures work on an artificial and a real-world data. We start with describing artificial data.
Higleyman (a normally distributed two dimension data with different covariance matrices Highleyman classes are defined by N([1 1],[1 0; 0 0.25]) for class A and N([2 0],[0.01 0; 0 4]) for class B) , difficult (a

$\mathcal{E}_{\mathcal{II}}[0.01]$	$\mathcal{E}_{\mathcal{II}}[0.1]$
Higleyman	
0.51 (0.06)	0.38 (0.06)
difficult 2D	
0.80 (0.21)	0.18 (0.07)
difficult 5D	
0.93 (0.12)	0.30 (0.07)
difficult 10D	
0.93 (0.11)	0.42 (0.07)
banana	
0.38 (0.05)	0.04 (0.04)

Table 1: The error of the second kind $\mathcal{E}_{\mathcal{II}}$ for SVDD with the error of the first kind $\mathcal{E}_{\mathcal{I}}$ set to 0.01 in first column and 0.1 in the second. In brackets the standard deviation.

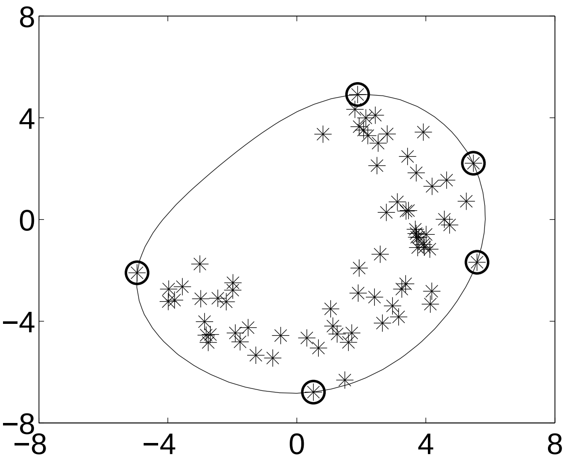


Figure 1: A scatterplot of the banana shape data distribution

normally distributed data overlapping on one of the side with equal covariance matrices of their distribution, two, five or ten dimensional) and banana shape (a scatterplot is shown in figure 1) data sets [1] are used in the experiments. Those data sets were chosen as representations of different distribution of the target set to verify performance of our rescaling methods on various shapes and dimensions of the target data.

Each set contains 500 target and 500 outliers objects. For the training set 250 of target objects are drawn and the remainder (including outliers) are used for testing. Two SVDD's were trained, with two different thresholds on the fraction rejection: 0.01 and 0.1. It means that the error of the first kind $\mathcal{E}_{\mathcal{I}}$ should be 0.01 and 0.1 respectively. This fraction of the data is related to the number of support vectors. Experiments were repeated 100 times. The means and standard deviations (in brackets) for the error of the second kind $\mathcal{E}_{\mathcal{II}}$ are shown in table 2. Table 1 shows as a reference the results for the SVDD classifiers for the same parameters and data without rescaling the feature set.

The value of the error of the first kind $\mathcal{E}_{\mathcal{I}}$ for the data before and after rescaling was obtained with small deviation (0.1 of its value) for all classes and for clarity were not included in the tables.

Fixing the error of the first kind, we observe the influence of rescaling the feature set on the classification performance by the change in the error of the second kind - outliers accepted. The table 2 shows results for rescaled distributions of the data mentioned above for three scaling methods. Numbers in bolds indicate in what cases the error $\mathcal{E}_{\mathcal{II}}$ was decreased after rescaling for the particular class. The fact that we have chosen a particular value for $\mathcal{E}_{\mathcal{I}}$ means that the total error is decreased.

method	$\mathcal{E}_{\mathcal{II}}[0.01]$	$\mathcal{E}_{\mathcal{II}}$ [0.1]
Higleyman		
variance	**0.45** (0.05)	0.41 (0.05)
domain	0.77 (0.04)	0.63 (0.07)
sphere	**0.46** (0.05)	0.40 (0.06)
difficult 2D		
variance	**0.42** (0.14)	0.25 (0.10)
domain	**0.46** (0.12)	0.30 (0.08)
sphere	**0.40** (0.14)	0.27 (0.08)
difficult 5D		
variance	**0.44** (0.11)	**0.25** (0.07)
domain	**0.41** (0.11)	0.30 (0.07)
sphere	**0.42** (0.12)	**0.28** (0.08)
difficult 10D		
variance	**0.50** (0.12)	0.42 (0.10)
domain	**0.45** (0.12)	**0.37** (0.09)
sphere	**0.50** (0.13)	**0.40** (0.10)
banana		
variance	0.38 (0.05)	0.04 (0.05)
domain	**0.35** (0.06)	0.28 (0.05)
sphere	0.38 (0.05)	0.04 (0.02)

Table 2: The error of the second kind $\mathcal{E}_{\mathcal{II}}$ for rescaled data, the error of the first kind $\mathcal{E}_{\mathcal{I}}$ set to 0.01 in first column and 0.1 in the second. In brackets the standard deviation. The numbers in bold point show an improvement of the performance after scaling the data.

Summarizing, we checked the influence of scaling on the performance on SVDD classifier with: different data distributions, with different number of dimensions and different number of SV which is related to $\mathcal{E}_{\mathcal{I}}$.

The same procedure was applied to real-word dataset handwritten digit images. The data used in experiments was taken from the Special Database 3 distributed on CD-ROM by the U.S. National Institute for Standards and Technology (NIST) [5] and preprocessed to 16x16 images NIST16 database. The

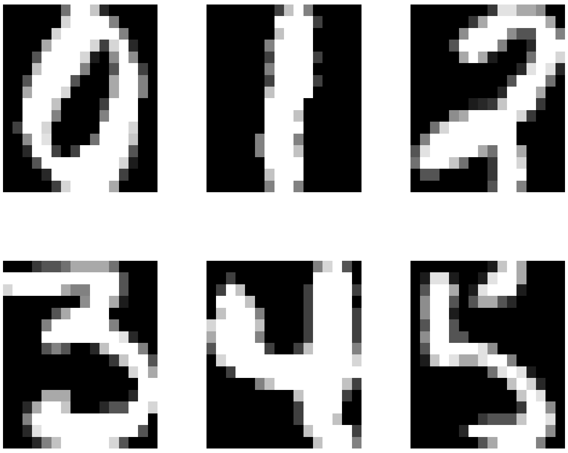


Figure 2: Examples of handwritten digits from NIST16 database.

NIST16 database contains 2000 images of 10 classes, each class-digit is represented by 200 images, figure 3 shows examples of the data. SVDD was trained on every 10 situations when one of the digits is a target class and others are outliers. We use images directly as the 256 dimensional data set.

class	$\mathcal{E}_{\mathcal{I}}$	$\mathcal{E}_{\mathcal{II}}$
0	0.03 (0.02)	0.81 (0.14)
1	0.04 (0.01)	0.97 (0.14)
2	0.02 (0.02)	0.91 (0.05)
3	0.03 (0.02)	0.59 (0.15)
4	0.02 (0.02)	0.84 (0.06)
5	0.03 (0.02)	0.95 (0.03)
6	0.03 (0.02)	0.70 (0.10)
7	0.03 (0.02)	0.76 (0.10)
8	0.02 (0.02)	0.90 (0.06)
9	0.03 (0.02)	0.90 (0.06)

Table 3: Classification results for the NIST16 digits, where each digit is considered the target class once. $\mathcal{E}_{\mathcal{I}}$ - error of the first kind, $\mathcal{E}_{\mathcal{II}}$ - error of the second kind.

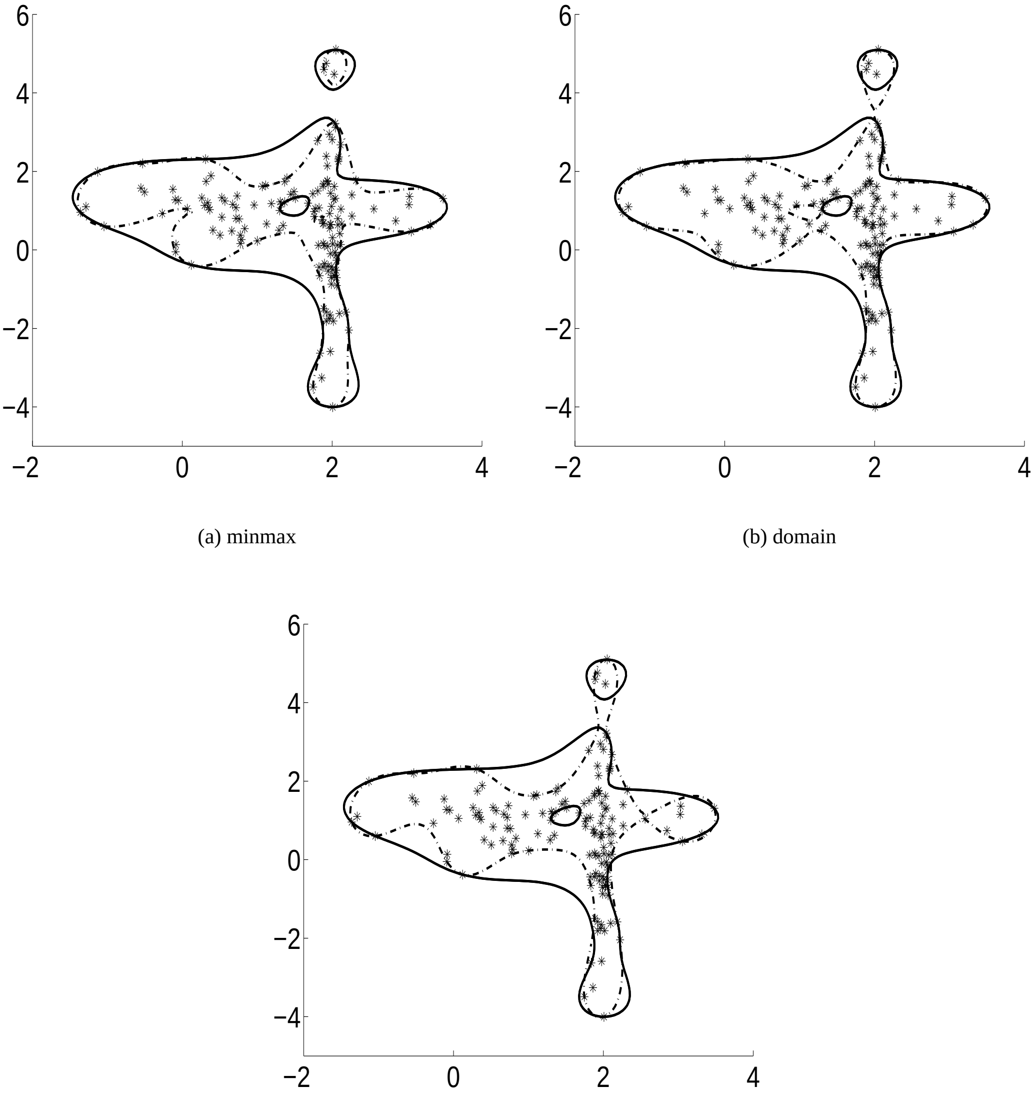
variance		
class	$\mathcal{E}_{\mathcal{I}}$	$\mathcal{E}_{\mathcal{II}}$
0	0.11 (0.04)	**0.05** (0.03)
1	0.06 (0.04)	**0.08** (0.04)
2	0.16 (0.05)	**0.13** (0.06)
3	0.20 (0.05)	**0.08** (0.04)
4	0.14 (0.05)	**0.17** (0.04)
5	0.21 (0.06)	**0.08** (0.03)
6	0.12 (0.04)	**0.07** (0.04)
7	0.09 (0.03)	**0.09** (0.04)
8	0.17 (0.05)	**0.21** (0.04)
9	0.14 (0.04)	**0.05** (0.02)

domain		
class	$\mathcal{E}_{\mathcal{I}}$	$\mathcal{E}_{\mathcal{II}}$
0	0.11 (0.04)	**0.09** (0.05)
1	0.08 (0.04)	**0.08** (0.04)
2	0.15 (0.06)	**0.32** (0.09)
3	0.17 (0.06)	**0.20** (0.07)
4	0.13 (0.04)	**0.26** (0.05)
5	0.18 (0.05)	**0.20** (0.06)
6	0.14 (0.04)	**0.12** (0.04)
7	0.11 (0.04)	**0.13** (0.04)
8	0.16 (0.05)	**0.35** (0.06)
9	0.13 (0.04)	**0.12** (0.04)
minmax		
class	$\mathcal{E}_{\mathcal{I}}$	$\mathcal{E}_{\mathcal{II}}$
0	0.10 (0.04)	**0.03** (0.02)
1	0.07 (0.04)	**0.10** (0.06)
2	0.12 (0.05)	**0.14** (0.06)
3	0.14 (0.04)	**0.08** (0.04)
4	0.12 (0.04)	**0.21** (0.03)
5	0.16 (0.05)	**0.10** (0.03)
6	0.12 (0.04)	**0.05** (0.03)
7	0.08 (0.04)	**0.08** (0.03)
8	0.15 (0.05)	**0.22** (0.06)
9	0.13 (0.04)	**0.06** (0.03)

Table 4: Results for one-class digits classification problems with rescaled by three linear methods feature space, $\mathcal{E}_{\mathcal{I}}$ - error of the first kind, $\mathcal{E}_{\mathcal{II}}$ - error of the second kind. The numbers in bold point show an improvement of the performance after scaling the data.

The training and the test set are constructed as follows: for the training set a half of target data was taken (100 images) and for the test set all other classes plus 100 images from target class. The fraction rejection for SVDD was set to 0.05 what means that we expect that $0.05 * 100 = 5$ data points will be lying on the boundary. PCA with a set of variance thresholds was applied (0.8, 0.9, 0.95, 0.99) to reduce the dimension of the feature space. Results for the threshold equals 0.95 are included in table 3. (The SVDD performed the best for the variance threshold of 0.95.) Because the PCA was taken on nine different target sets, digits in this case, different number of dimensions were obtained when 0.05 of the variance was retained. On average, the dimensionality was 72 so we reduce our space from 256 to about 72.

The results for the handwritten digit recognition for the same threshold for the fraction rejection and PCA with rescaled data are shown in table 4. This data was used to test the behavior of the scaling methods for unknown distributions of the target data, with many feature dimensions. We included $\mathcal{E}_{\mathcal{I}}$ in the tables, because, unlike for the previous experiments on the artificial data, this time $\mathcal{E}_{\mathcal{I}}$ considerably changes its value after rescaling.



(a) minmax

(b) domain

(c) variance

Figure 3: Differences in shapes of the SVDD classifier trained on no-rescaled and rescaled (minimax, domain, variance) data: the continuous line - SVDD trained on no-rescaled data, dashed line - SVDD trained on rescaled data

From the results in tables 1 - 2 we cannot distinguish which scaling method is the best, they are performing quite similar. But we can say that in almost every example the error $\mathcal{E}_{\mathcal{II}}$ is smaller after scaling then before. Some of $\mathcal{E}_{\mathcal{II}}$ errors rise, this phenomena can be explained as follows: imagine a sinusoidal shape boundary separating two classes. If we divide it along longer axis (frequency of the sinusoid arise) more support vectors are needed to describe this boundary. When we fix the number of support vectors by specifying $\mathcal{E}_{\mathcal{I}}$, we lack of the sufficient number of support vectors to describe the more complicated boundary. In that case the boundary cannot be followed and the error increases.

From results in tables 3 - 4 some regularity can be observed. In general, the minmax method causes the smallest rise in the error $\mathcal{E}_{\mathcal{I}}$ while scaling by variance decreases $\mathcal{E}_{\mathcal{II}}$. This regularity was obtained for a particular distribution of the data and can not be generalized for other distributions.

All rescaling procedures used in this paper are based on the same principles. Rescale the feature space such that the target distribution is more spherically distributed and the areas which do not contain target data are easily separated from the target class.

The rescaling methods perform well for data with ellipsoidal shapes or with directions with a different length from the means (the Mahalanobis distance)

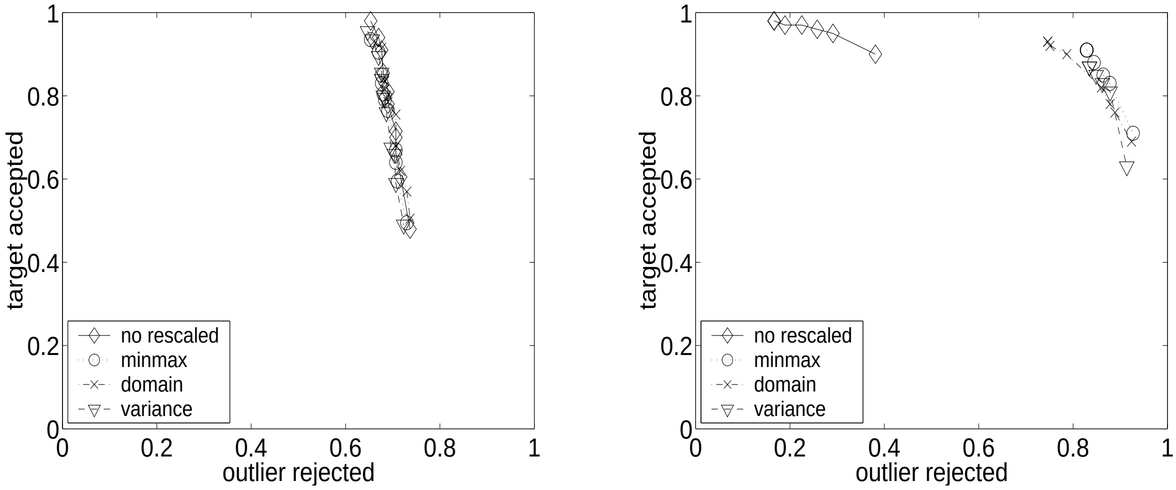


Figure 4: Receiver-operating characteristic curves for different scaling methods for banana shape data (on the right) and handwritten digits (on the left) when a digit 2 is a target set

to the decision boundary circumscribes data in the feature space. An example of classifiers trained on two-dimensional rescaled and no rescaled Highleman classes, connected into one target class, are shown in the figure 3. The tighter boundary around the target set was obtained after rescaling features. On these examples we can explain results of our experiments on hand-written digits. By finding the more narrow boundary around a target set we can minimize the error of outlier acceptance and only if we don't have a good representation of a target set in a training set some target elements in a test set will be classified as outliers and $\mathcal{E}_{\mathcal{I}}$ will be larger after rescaling. For some data with the derivation equals in each direction from its mean and areas do not contain target data inside the boundary describes it, see figure 1, our three rescaling methods do not perform well on this example (figure 4) because all of them will produce the same shape of the target distribution in the rescaled feature space only with the smaller volume. For this kind of distribution some non-linear scaling should be applied to obtain a more uniform data distribution.

The figure 4 shows ROC curves [6] for SVDD trained on banana shape dataset and handwritten digit with '2' as a target set. 'Target accepted' and 'outlier rejected' can be obtained respectively: $1-\mathcal{E}_{\mathcal{I}}$, $1-\mathcal{E}_{\mathcal{II}}$. From the first figure no improvement can be observed between presented rescaling methods and no-rescaled SVDD. From the second figure ROC curves for classifiers trained on rescaled data are shows a large decrease of $\mathcal{E}_{\mathcal{II}}$ error with a small increase of $\mathcal{E}_{\mathcal{I}}$ at the same time.

4 Conclusions

In this paper we presented some simple methods of rescaling the feature space to optimize the boundary around a training set of objects in one-class classification problem. We introduced simple linear methods to obtain more spherical distribution of the target set to minimize chance of a target rejection and an outlier acceptance. We verified those methods on the one-class classifier SVDD. Experiments on artificial and some real-world data suggest that scaling to domain, variance and minmax method decrease the error on the outlier data in most cases. In the future we will study more sophisticated methods of rescaling the data, we will also verify the performance of other one-class classifiers.

References

[1] D.M.J. Tax, 'One-class classification' , PhD Thesis, Delft University of Technology, http://www.ph.tn.tudelft.nl/~davidt/thesis.pdf ISBN: 90-75691-05-x, 2001

[2] D.M.J. Tax, R.P.W. Duin, 'Support Vector Data Description' , Pattern Recognition Letters, December 1999, vol. 20(11-13), pg. 1191-1199

[3] B. Schölkopf, C. J. C. Burges, A. J. Smola, 'Advances in kernel methods', The MIT Press 1999

[4] V. N. Vapnik, 'The nature of statistical learning theory', 1995 Springer-Verlog, New York,Inc.

[5] C.L. Wilson and M.D.Garris, Handprinted character database 3, February 1992, http://www.nist.gov/srd/nistsd19.htm, National Institute for Standards and Technology, Advanced Systems Division pg.43

[6] Metz, C. 'Basic principles of ROC analysis' Seminars in Nuclear Medicine,VIII(4) 1978