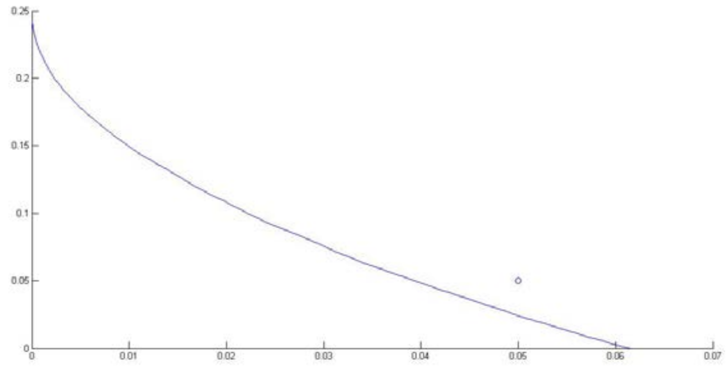

Figure 5. Bound line for the separation space for uniform distribution

Since $\dfrac{\left|A \bigcap convB\right| + \left|B \cap convA\right|}{n_A + n_B} = 0.06$, let $\varepsilon_A = 0.05; \varepsilon_B = 0.05$

According to the Fig. 5, $(0.05;0.05) \in D_l$

Let's build the ε-net for the set A by the following:

Algorithm of the ε-net building

Let's select $\varepsilon_A, \varepsilon_B \in D_{A,B}$.

Let set A contain the point with minimal y-coordinate. Let's denote the point of set A with minimal x-coordinate by $a_{\min}$, and the point with maximal x-coordinate by $a_{\max}$. Let's draw $k = \left[\dfrac{1}{\varepsilon_A}\right] + 1$ vertical lines from $a_{\min}$ to $a_{\max}$ in a manner that there are $\varepsilon_A n_A$ points in each of $\left[\dfrac{1}{\varepsilon_A}\right]$ bands. Vertical lines which separate the bands are described by the equations

$$x = C_i,\ i = \overline{1,k},$$

where constant C_i can be found from the equation

$$F(C_i) = i\varepsilon_A$$

For each i-th band, $1 \le i \le \left[\dfrac{1}{\varepsilon}\right]$, let's denote:

A^i is the set which contains points from the set A that are contained in the i-th band;

B^i is the set (may be empty) which contains points from the set B that are contained in the i-th band;

$ay^i_{\min}$, $ay^i_{\max}$ are points from the set A^i with minimal and maximal y-coordinates;

$by^i_{\min}$, $by^i_{\max}$ are points from the set B^i with minimal and maximal y-coordinates.

$N_A^{\varepsilon_A}$ is the set of points which we will select in the ε-net of the set A.

From the i-th band we will select two points in the set N_A. The first point is the point $ay^i_{\min}$. According to the assumption, set A is placed below the set B. The second point from the set A^i in the set $N_A^{\varepsilon_A}$ will be selected according to the following rule.

If $B^i = \varnothing$ (it means that i-th band does not contain points from the set B),
    add point $ay^i_{\max}$ to the set $N_A^{\varepsilon_A}$
else
    if $ay^i_{\max} < by^i_{\min}$ (it means that in i-th band convex hulls of sets A, B are not intersected),
        add point $ay^i_{\max}$ to the set $N_A^{\varepsilon_A}$
    else (some points of the set B exist in the i-th band and they are placed bellow some points of set A)

add point $a^i \in A$ to the set $N_A^{\varepsilon_A}$ such that point a^i is the nearest neighbor to the point $by^i_{\min}$. We will call point a^i the basis point of the set A.

In the same way we build horizontal bands and select two points from each band to the set $N_A^{\varepsilon_A}$.

Example 3.

1. Normal distribution

Vertical and horizontal bands for the set A are illustrated in the Fig. 6

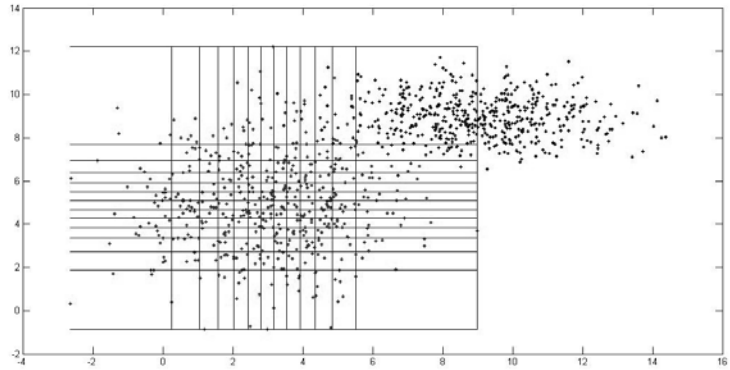

Figure 6. Vertical and horizontal bands for the set A which is generated by the normal distribution

2. Uniform distribution

Vertical and horizontal bands for the set A are illustrated in the fig.7

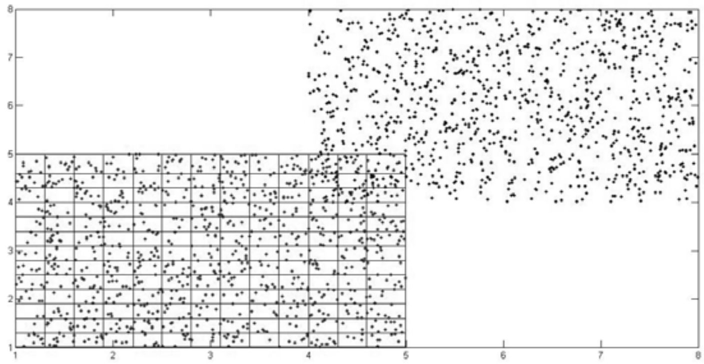

Figure 7. Vertical and horizontal bands for the set A which is generated by the uniform distribution

Lemma 3. *The set* $N_A^{\varepsilon_A}$ *is* ε*-net for the set* A.

Proof. Let's make an indirect proof. Assume $N_A^{\varepsilon_A}$ is not an ε-net of the set A. It means that there exists a halfspace $H \subset R^2$ that contains at least $\varepsilon_A n_A$ sets of point A, but each point does not belong to the set $N_A^{\varepsilon_A}$. Let's denote Z as the set of points from the set A that belong to the halfspace H and $|Z| \geq \varepsilon_A n_A$. Consider a point $z \in Z$. This point belongs to one horizontal and one vertical band. Together with point z one of extreme points of horizontal or vertical band or basis point belongs to the halfspace H. According to the building process, set $N_A^{\varepsilon_A}$ consists of extreme and basis points of the set A, so $Z \cap N_A^{\varepsilon_A} \neq \varnothing$. This contradicts the assumption.

Lemma is proved.

According to the algorithm, ε-net $N_A^{\varepsilon_A}$ consists of $\left[\dfrac{4}{\varepsilon_A}\right]$ points. ε-net $N_B^{\varepsilon_B}$ is built using the same algorithm.

Example 4.

1. Normal distribution

ε-nets of the sets A, B are illustrated in the Fig. 8

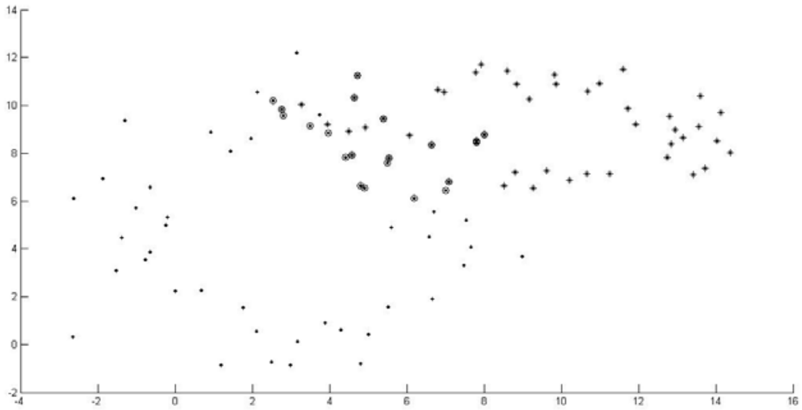

Figure 8. ε-nets of the sets A, B which are generated by the normal distribution.

2. Uniform distribution

ε-nets of the sets A, B are illustrated in the Fig. 9

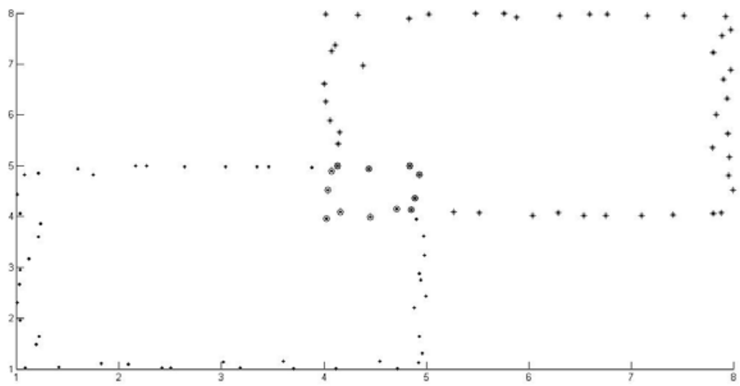

Figure 9. ε-nets of the sets A, B which are generated by the uniform distribution.

3 ε-Nets' Separating

Let's separate sets $N_A^{\varepsilon_A}$ and $N_B^{\varepsilon_B}$ using the separation algorithm of the convex hulls, which is described in [9].

Separation algorithm of the convex hulls

1. Build convex hulls $convN_A^{\varepsilon_A}$ and $convN_B^{\varepsilon_B}$.
2. Find outlier points. In order to minimize the algorithm's time complexity, we find outliers only among the basis points. Point $x \in NB_A$, where $x \in convN_B^{\varepsilon_B}$, is the outlier point of the set $N_A^{\varepsilon_A}$.

The set of outlier points of the set $N_A^{\varepsilon_A}$ is denoted by $P_A^{\varepsilon_A}$.

3. Reject the outliers from the set $N_A^{\varepsilon_A}$ and build the convex hull of the set $N'_A = N_A^{\varepsilon_A} \setminus P_A^{\varepsilon_A}$.
4. Among the edges of the polygons $convN'_A$ and $convN_B^{\varepsilon_B}$ find the edge so that points of the set N'_A and points of the set $N_B^{\varepsilon_B}$ are placed in different halfspaces which are generated by the line l containing this edge.
5. The line l is the separating line for the sets $N_A^{\varepsilon_A}$ and $N_B^{\varepsilon_B}$. In order to minimize the algorithm's time complexity, we find separating line among the edges containing basis points.

If ε-nets are not linear separable, we propose to use Voronoi diagram [8] .

Example 5.

1. Normal distribution

The separating line for ε-nets is illustrated in the Fig. 10

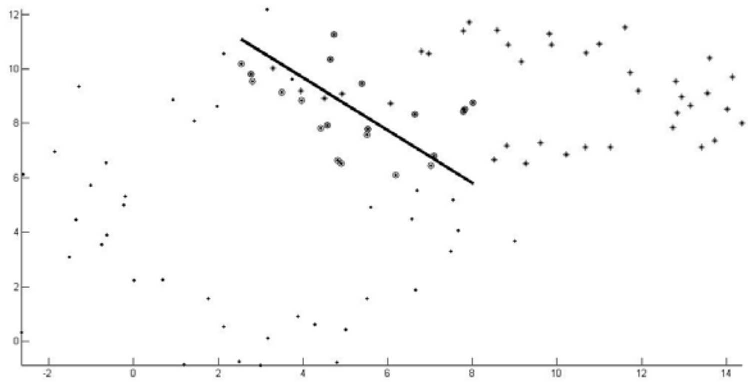

Figure 10. The separating line for ε-nets for normal distribution.

2. Uniform distribution

The separating line for ε-nets is illustrated in the Fig. 11

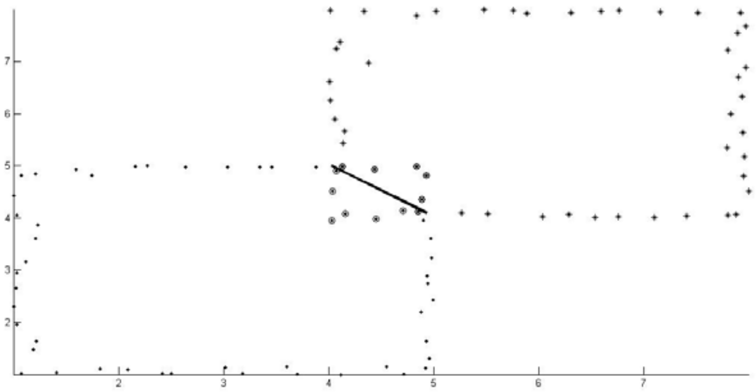

Figure 11. The separating line for ε-nets for uniform distribution

According to the theorem 1, separating line for the ε-nets $N_A^{\varepsilon_A}$ and $N_B^{\varepsilon_B}$ is ε-separating for the sets A and B.

Example 6.

1. Normal distribution

The ε-separating line for sets A and B is illustrated in the Fig. 12

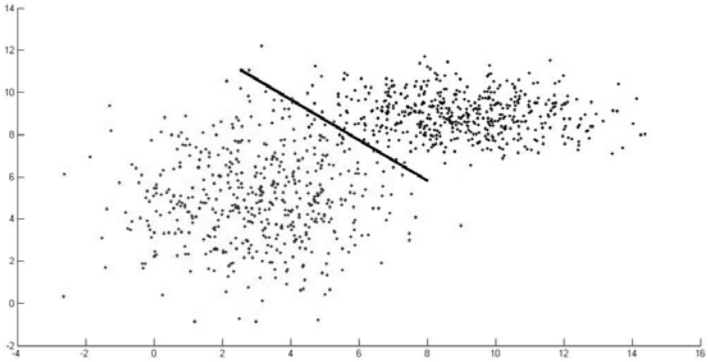

Figure 12. The ε-separating line for the sets A, B which are generated by the normal distribution.

2. Uniform distribution

The ε-separating line for sets A and B is illustrated in the Fig. 13

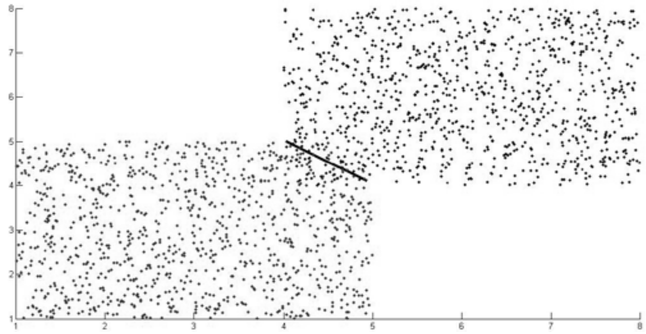

Figure 13. The ε-separating line for the sets A, B which are generated by the uniform distribution.

Let's compare the classification using the algorithm described above and the classification using the Support Vector Machine (SVM) [2].

Example 7.

1. Normal distribution

Classification using ε-nets gives 3.0% errors; classification by SVM 2.9% errors (Fig. 14)

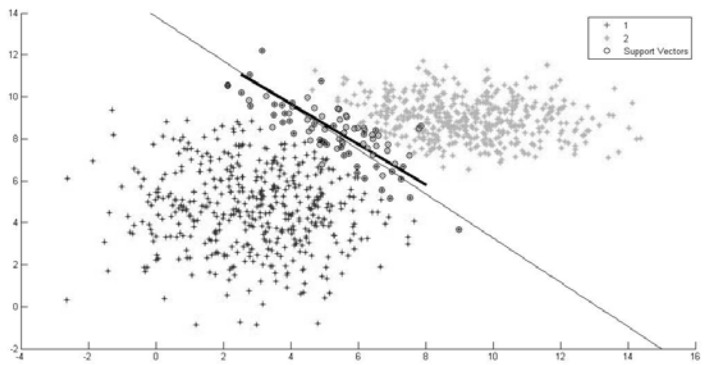

Figure 14. Comparing classification for the normal distribution

2. Uniform distribution

Classification using ε-nets gives 5.9% errors; classification by SVM 5.9% errors (Fig. 15). The separating lines coincide.

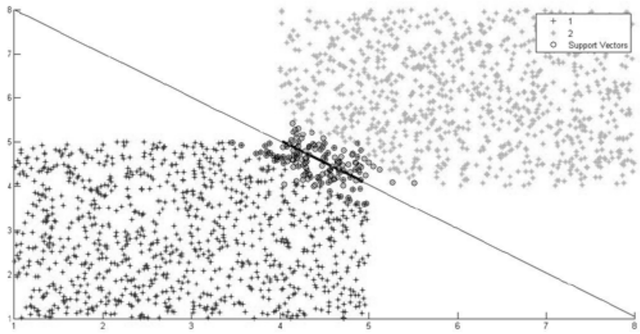

Figure 15. Comparing classification for the uniform distribution

4 Conclusions

The algorithm of building ε-nets for two sets is described in the paper. The ε-nets, constructed according to this algorithm have size $\left\lceil \frac{4}{\varepsilon_A} \right\rceil$, which does not depend on the size of the set. It is shown in the paper that for separating two sets one can use their ε-nets, which considerably reduce the complexity of the separating algorithm for large sets. Two examples in the paper illustrate the algorithm's effectiveness.

References

1. Aronov B., Ezra E., Sharir M. "Small-size epsilon-nets for axis-parallel rectangles and boxes", Symposium on Theory of Computing, 2009, P.P. 639–648
2. Christopher J.C. Burges "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2(2), 1998, P.121–167.
3. Embrechts P., Hofert M. A note on generalized inverses Mathematical Methods of Operations Research , 2013, 77(3), 423-432
4. Gärtner B., Hoffmann M. Computational Geometry, http://www.ti.inf.ethz.ch/ew/lehre/CG12/lecture/CG%20lecture %20notes.pdf
5. Hausler S. VC Dimension. A Tutorial for the Course Computational Intelligence, http://www.igi.tugraz.at/lehre/CI
6. Haussler D. and Welzl E. "Epsilon-nets and simplex range queries", Discrete Comput. Geom., 1987, №2, P.P. 127–151
7. Ivanchuk M. A. , Malyk I. V. "Using ε-Nets for Linear Separation of Two Sets in a Euclidean Space Rd", Cybernetics and Systems Analysis, Vol.51, Issue 6 (2015), P. 965-968.
8. Ivanchuk Maria A., Malyk Igor V. "Building expert medical prognostic systems using Voronoi diagram", Hindawi Publishing Corporation. - International Journal of Computational Mathematics, Volume 2015, Article ID 415146, 4 pages. – DOI 10.1155/2015/415146
9. Ivanchuk Mariya A., Malyk Igor V. "Separation of convex hulls as a way for modeling of systems of prediction of complications in patients", Journal of Automation and Information Sciences, 2015, Vol.47, Issue 4, P.78-84, DOI: 10.1615/JAutomatInfScien.v47.i4.80
10. Kulkarni J., Govindarajan S. "New ε-Net Constructions", Canadian Conference on Computational Geometry (CCCG), 2010, P.P.159-162

11. Matousek J., Seidel R., Welzl E. "How to Net a Lot with Little:Small ε-Nets for Disks and Halfspaces" In Proc. sixth annual symposium on Computational geometry, P.P. 16–22, 1990
12. Tucker H.G. "A Generalization of the Glivenko-Cantelli Theorem", The Annals of Mathematical Statistics, Vol. 30, No. 3, Sep., 1959, pp. 828-830

