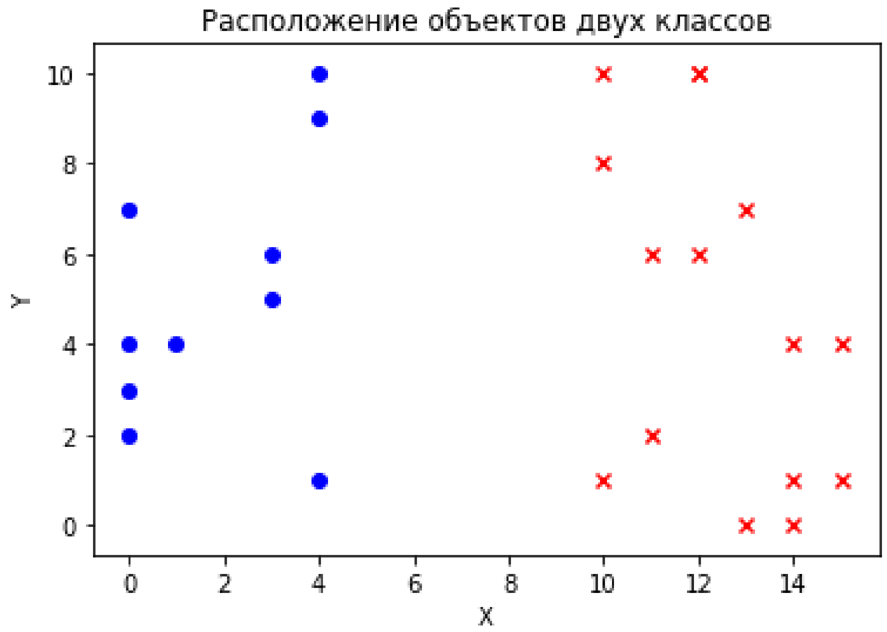
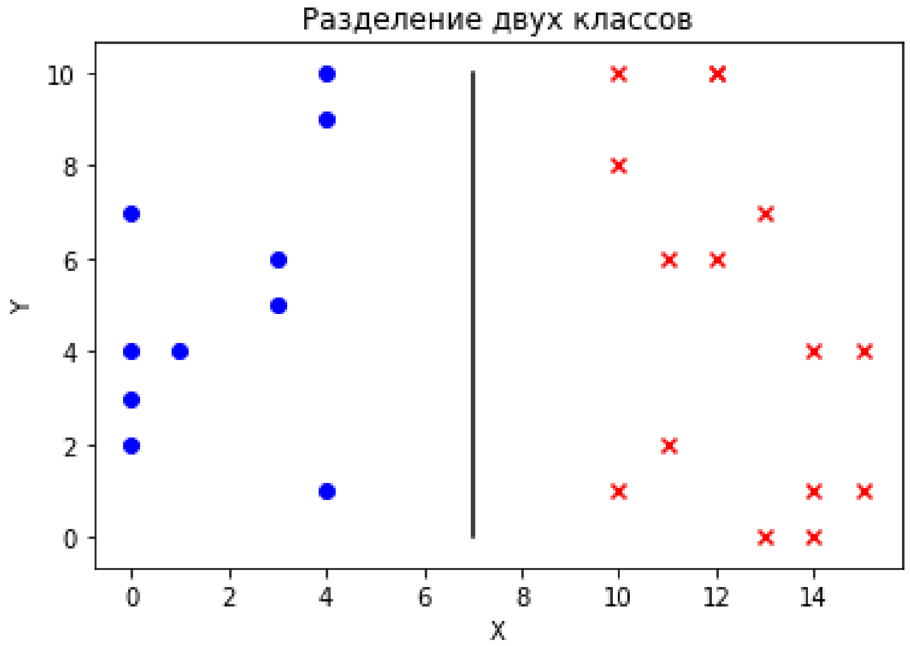
Introduction

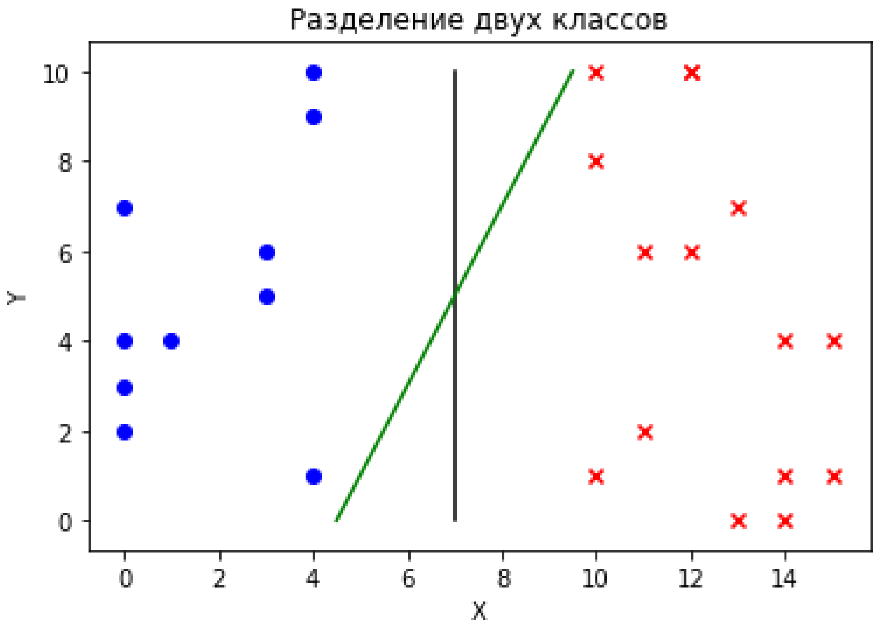


dataset

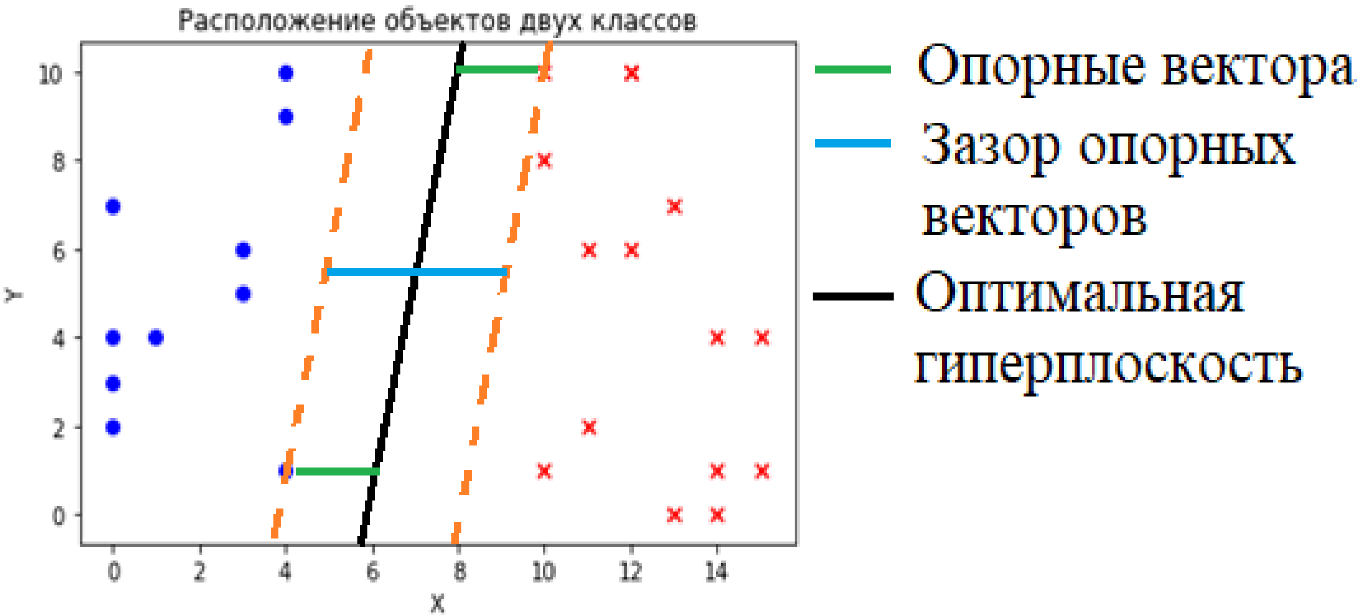


desirable result

Introduction



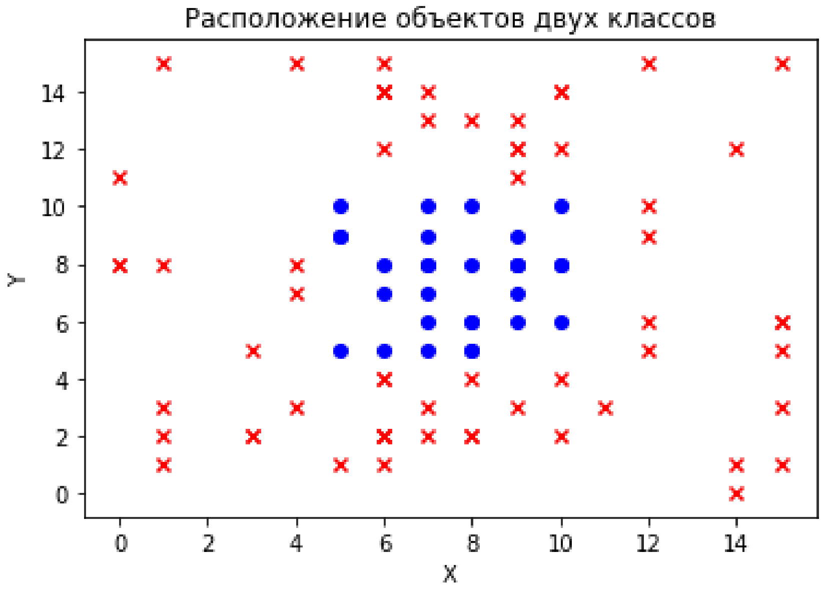
dataset



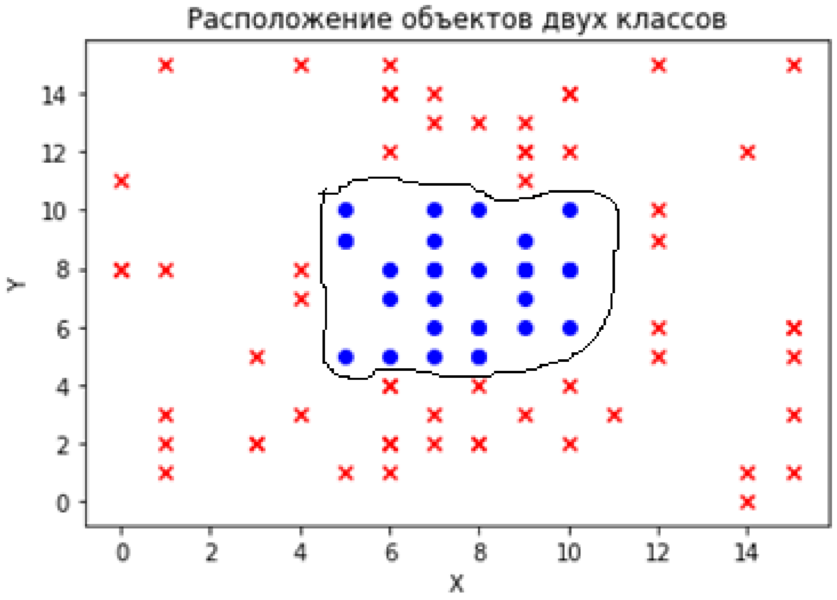
desirable result

Introduction

Non-linear solution



dataset



desirable result

Introduction

Applications:

- intrusion detection
- medical diagnosis
- fraud detection
- surveillance

ONE-CLASS SVMs

Motivation

$$g(x) = w^T \phi(x) - \rho$$

$\phi(\cdot)$ – an implicit transformation function

w – the vector perpendicular to the decision boundary

ρ – the bias term

Objective

The primary objective of one-class SVMs:

$$\min_{w,\xi,\rho} \frac{\|w\|^2}{2} - \rho + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i$$

$$\text{subject to: } w^T\phi(x_i) \geq \rho - \xi_i,\ \xi_i \geq 0$$

ξ_i – the slack variable for point i that allows it to lie on the other side of the decision boundary

n – the size of the training dataset

ν – the regularization parameter

Objective

The decision boundary is defined as:

$$g(x) = 0.$$

The distance of any arbitrary data point to the decision boundary can be computed as:

$$d(x) = \frac{|g(x)|}{||w||}$$

Objective

The primary objective is transformed into a dual objective:

$$\min_{\alpha} \frac{\alpha^T Q \alpha}{2}$$

$$\text{subject to: } 0 \leq \alpha_i \leq \frac{1}{\nu n}, \sum_{i=1}^{n} \alpha_i = 1$$

Q – the kernel matrix

α – the Lagrange multipliers

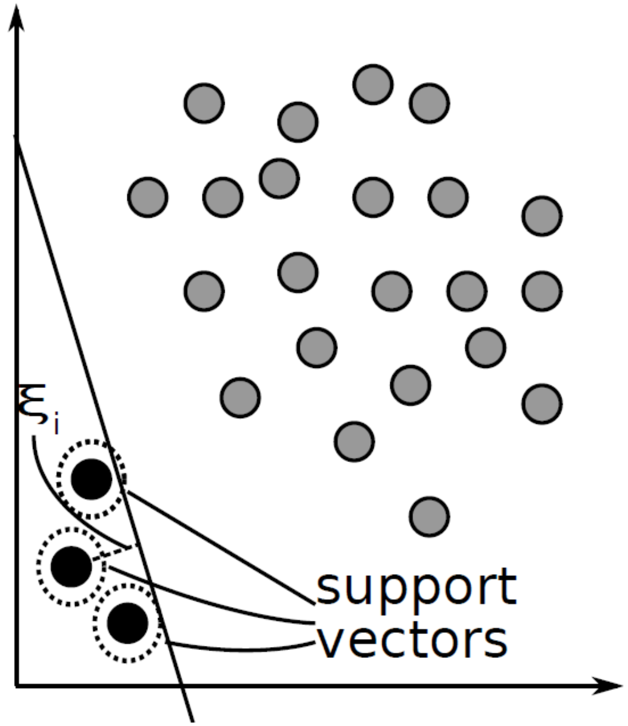
Outlier Score

$$f(x) = \frac{g_{max} - g(x)}{g_{max}}$$

g_{max} – the maximum directed distance between the dataset points and the decision boundary

$f(x) > 1.0$ – indicates that the point is a potential outlier

Influence of Outliers



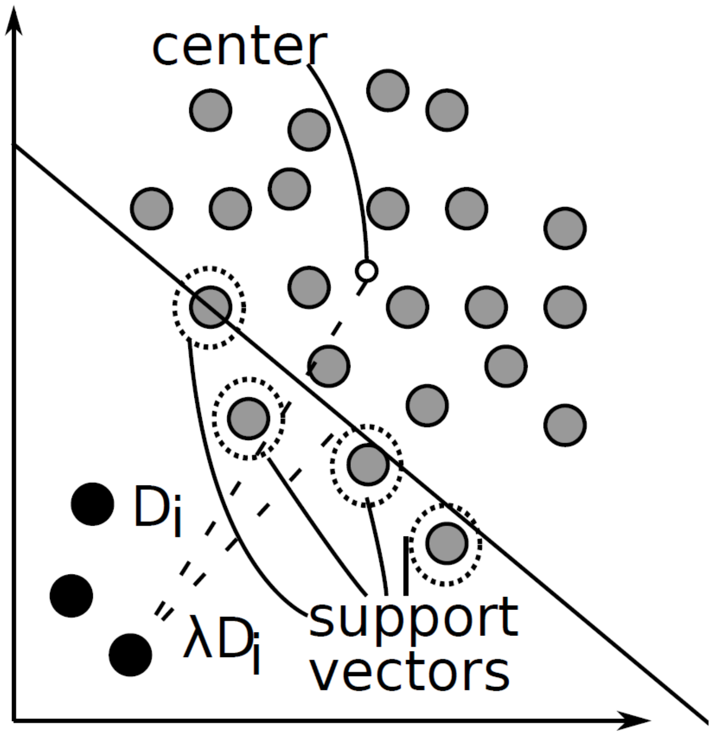
A 2-dimensional example of the decision boundary in the kernel space learned by a one-class SVM.

Whilst the shifting of the decision boundary might not have a great influence on the overall rank of the points when using $f(x) = \frac{g_{max} - g(x)}{g_{max}}$, the shape of the decision boundary will.

Robust One-class SVMs

Motivation

The key idea is the minimization of MSE for tackling outliers using the center of class as an averaged information.



Modifying the slack variables for robust one-class SVMs. Each slack variable is proportional to the distance to the centroid.

Objective

The objective of the proposed robust one-class SVMs:

$$\min_{w,\rho} \frac{||w||^2}{2} - \rho$$
$$\text{subject to } w^T\phi(x_i) \geq \rho - \lambda * \hat{D}_i$$

ξ_i – the slack variable for point i that allows it to lie on the other side of the decision boundary

D_i – the slack variable

λ – the regularization parameter

Objective

The slack variable D_i represents the distance to the centroid in the kernel space.

$$D_i = \|\phi(x_i) - \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\|^2$$

$$\hat{D}_i = \frac{D_i}{D_{max}}$$

Since the transformation function is implicitly defined by the kernel, D_i can not directly be used.

Thus, an approximation is computed instead.

Objective

An approximation of D_i:

$$D_i = \|\phi(x_i) - \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\|^2$$

$$= Q(x_i, x_i) - \frac{2}{n}\sum_{j=1}^{n}Q(x_i, x_j) - \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$

$$\approx Q(x_i, x_i) - \frac{2}{n}\sum_{j=1}^{n}Q(x_i, x_j)$$

Objective

The dual objective of the robust one-class SVM can be summarized as follows:

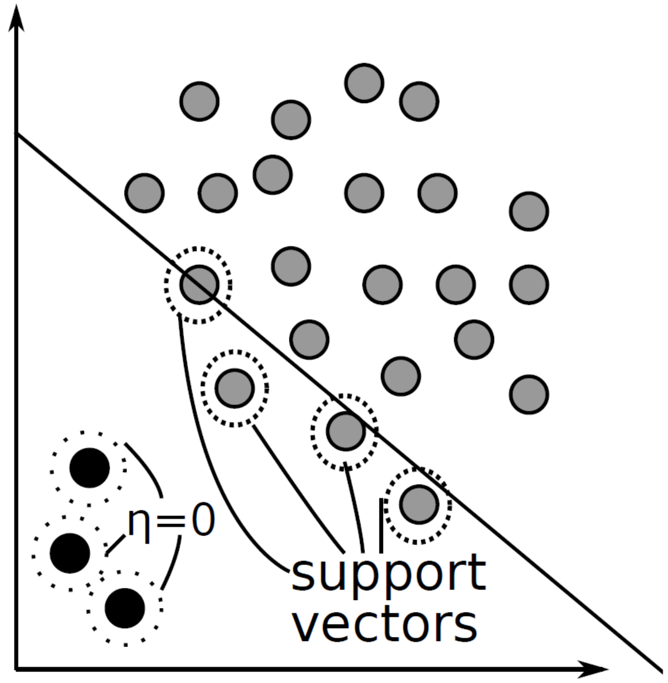
$$\min_{\alpha} \frac{\alpha^T Q \alpha}{2} + \lambda D^T \alpha$$

$$\text{subject to } 0 \leq \alpha \leq 1, e^T \alpha = 1$$

Eta One-class SVMs

Motivation

The key idea: outliers have small values for and do thus not contribute to the decision boundary



Objective

The objective of the eta one-class SVM:

$$\min_{w,\rho} \min_{\eta_i \in \{0,1\}} \frac{\|w\|^2}{2} - \rho + \sum_{i=1}^{n} \eta_i max(0, \rho - w^T * \phi(x_i))$$

$$\text{subject to } e^T \eta \geq \beta n$$

η – variable, represents an estimate that a point is normal

β – variable, controls the maximum number of points that are allowed to be outlying

Semi-Definite Programming Problem

The non-convex optimization objective of previous equation can be relaxed by relaxing the constraints on η

$$\min_{0 \leq \eta \leq 1, M = \eta * \eta^T} \max_{0 \leq \alpha \leq 1} \frac{\alpha^T Q \cdot M \alpha}{2},$$

$$\text{subject to } e^T * \eta \geq \beta n,\ \alpha^T \eta = 1,\ 0 \leq \alpha \leq 1$$

Iterative Relaxation

The SDP solution is expensive to compute and hence an alternative approach was proposed

$$\min_{w,\rho,\eta} E_{vex} + E_{cave}$$

$$E_{vex} = \frac{\|w\|^2}{2} - \rho + \eta^T h(w), \; E_{cave} = g^*(\eta)$$

Iterative Relaxation

The previous equation can be solved by iteratively minimizing E_{vex} and E_{cave}

$$\min_\alpha \frac{\alpha^T Q \cdot N\alpha}{2},$$
$$\text{where } N = \eta * \eta^T,$$
$$\text{subject to } \alpha^T \eta = 1,\ 0 \le \alpha \le 1$$

$$u_i = max(0, \rho - w^T \phi(x_i))$$
$$\eta_i = I(\beta n - s(i))$$