AN APPROACH TO RECONSTRUCT TARGET FUNCTION OF THE OPTIMIZATION PROBLEM WITH PRECEDENT INITIAL INFORMATION

© Ayder Anafiyev, Alim Abdulkhairov

Taurida National V. I. Vernadsky University
Mathematics and Computer Science Department
e-mail: *anafiyev@gmail.com*, *alim.abdulkhairov@gmail.com*

Abstract. *The optimization problem with precedent (training sample) initial information is considered. Some approaches for reconstruction of the target function of such optimization problem are proposed. The open problems that must be solved to obtain better quality solutions of this problem are highlighted.*

1. Formulation of the problem

Let X, Y and W are the spaces of object (feature), target function value and admissible function value respectively, $f : X \to Y$ is a target function and Ω is an admissible set of objects. Consider the optimization problem

$$f(x) \to \max_{x \in \Omega \subseteq X} \tag{1}$$

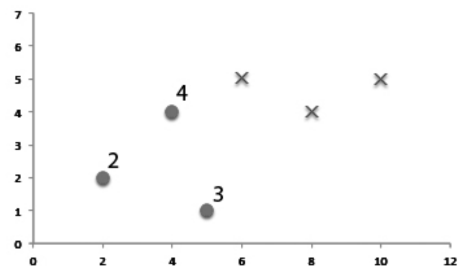
with initial information is represented by the set of triple $X^\ell = (x_i, y_i, w_i)_{i=1}^{\ell}$, where $x_i \in X$, $y_i \in Y$ and $w_i \in W$. The triple (x_i, y_i, w_i) will be called *precedent* (training sample). If $W = \{0, 1\}$ then $w_i = 0$ means that the object $x_i \notin \Omega$, otherwise ($w_i = 1$) means that the object $x_i \in \Omega$. If $W = [0, 1]$ then $w_i \in W$ could be interpreted as the probability that the object x_i belongs to the set Ω.

The problem (1) will be called the optimization problem with precedent initial information [1, 2, 3]. This is a problem with incomplete information. For solving this problem it's necessary to construct an algorithm which finds in Ω the optimal object(s) of the target function f or reduces the problem to a certain optimization problem with a fully defined data and which allows to find an effective decision.

The optimization problem could be divided into two problems: the problem of reconstructing of the target function f (regression problem) and the problem of reconstructing the admissible object set Ω (classification problem). There are many approaches for solving the regression and classification problems. *However it's still open the problem of synthesis of these two problems to get the better quality solution of the given optimization problem.*

Example 1. Lets consider the maximization problem with precedent initial information: $X = \mathbb{R}^2$, $Y = \mathbb{R}$, $W = \{0, 1\}$. The training sample X^ℓ is set as a training table:

x_1	x_2	y	w
2	2	2	1
5	1	3	1
4	4	4	1
6	5	-	0
8	4	-	0
10	5	-	0



Admissible objects are marked by circles, inadmissible — by crosses. In additional the objects with known values are labeled by the target function values.

As you can see from the figure when we come near to the imaginary border of the space Ω the value of the target function grows. Obviously this information would be useful for the decision making. The location of objects of different classes ("admissible" and "inadmissible" objects) may be very important during the target function reconstruction.

2. Reconstruction of the target function

2.1. **Linear regression.** Lets consider an input space $X = \mathbb{R}^n$ and output space $Y = \mathbb{R}$. The linear regression model $\phi(x, \alpha)$ is represented by

$$\phi(x, \alpha) = \sum_{i=1}^{n} \alpha_j x^j, \quad \alpha_j \in \mathbb{R}, \quad j = \overline{1, n}.$$

The optimal value of parameter α is selected from solution of the optimization problem

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^{\ell} L(\alpha, x_i) = \arg\min_{\alpha \in \mathbb{R}^n} Q(a, X^\ell) \qquad (2)$$

where $L(\alpha, x)$ is a loss function which is used to determine loss on the object x and $Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} L(\alpha, x_i)$ is an empirical risk.

When defining the loss function it must be considered the fact that we reconstruct the target function of the maximization problem (1). In this case we need more "detailed" study the objects on which the target function takes large values. Therefore the object importance depends on the value of the target function for this object. Thus the loss function will be considered as

$$L_\gamma(\alpha, x) = \gamma(x) L(\alpha, x)$$

where $\gamma(x)$ is the weight function which defines an importance of the object x for the optimization problem.

Let us consider $L(\alpha, x) = (\phi(\alpha, x) - f(x))^2$ and $\gamma(x) = f(x)$.

Introduce the matrix notations $F = (x_i^j)_{\ell\times n}$, $y = (y_i)_{\ell\times 1}$, $\alpha = (\alpha_j)_{n\times 1}$ and $\gamma = \operatorname{diag}(\gamma(x_1), \ldots, \gamma(x_\ell)) = \operatorname{diag}(y_1, \ldots, y_\ell)$.

Let us write the optimization problem (2) in matrix form

$$Q(\alpha, X^\ell) = \gamma\|F\alpha - y\|^2 \to \min_\alpha .$$

The standard way to solve this optimization problem is to use a necessary condition of minimum

$$\frac{\partial Q}{\partial \alpha} = 2F^T\gamma(F\alpha - y) = 0.$$

Therefore

$$F^T\gamma F\alpha = F^T\gamma y.$$

If $F^T\gamma F$ is a nonsingular matrix then the solution of the system will

$$\alpha^* = (F^T\gamma F)^{-1}F^T\gamma y.$$

The result of the using the linear regression method for reconstruction of the target function f of the optimization problem (1) is illustrated on the figure 1. As you can see a more detailed learning of the optimal objects could improve the solution[1].

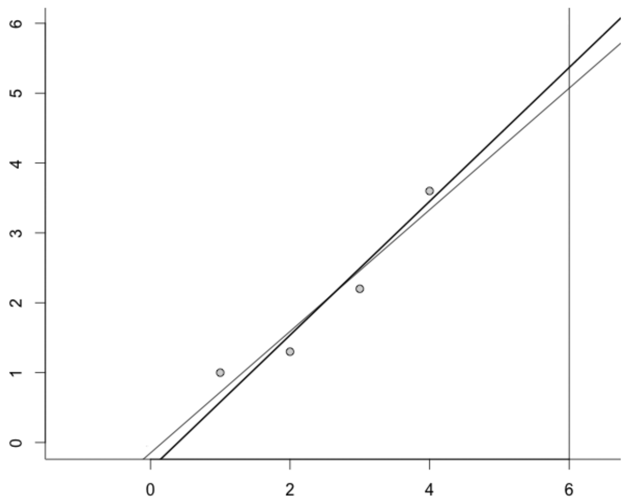


Fig. 1. Application of linear regression method for reconstruction of a target function of an optimization problem (bold line — with proposed loss function)

[1]Of course this approach requires more detailed research.

2.2. **Support Vector Regression.** The SVR [4] (Support Vector Regression) could be used to solve the problem of reconstructing the target function.The SVM (Support Vector Machines) method is used for the reconstruction of the admissible object set Ω. The support vector machines is one of the best classification algorithm nowadays. Learning of SVM leads to solving a quadratic (linear) programming problem. The position of the discriminant hyperplane depends only from a few support objects. In addition the use of kernel functions allows the efficient using this method for both: linearly separable and inseparable samples.

The figure 2 shows an example of the optimization problem with the precedent initial information with six objects: two are inadmissible and marked with a cross symbol and the other four belong to the space Ω and marked as a circle (the big radius circles have the larger target function value).

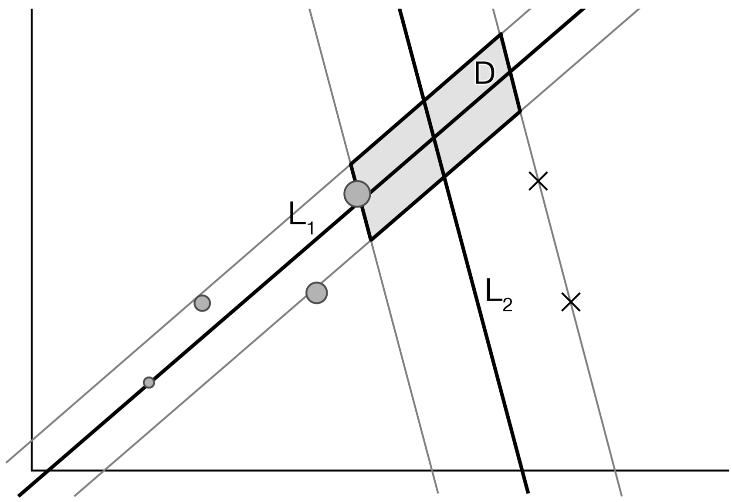


Fig. 2. SVR and SVM for solving the optimization problem with precedent initial information.

Line L_1 is the result of the solving regression problem and corresponds to the target function. Line L_2 is the result of solving classification problem on two classes: "admissible" and "inadmissible" objects. The set D is very interesting from scientific point and needs to be researched.

Consider the applying of SVR method to the reconstruction of the target function of the optimization problem 1. The target function $f(x)$ is represented as

$$f(x) = \langle \alpha, x \rangle + \alpha_0,$$

where $\langle \cdot\, , \cdot \rangle$ is the scalar product.

The optimal set of parameters $(\alpha_1, \dots, \alpha_n, \alpha_0)$ is the solution of the optimization problem

$$\begin{cases} \frac{1}{2}\|\alpha\|^2 + C\sum_{i=1}^{\ell} L_\gamma(\alpha, x_i) \to \min\limits_{\alpha,\alpha_0}, \\ L_\gamma(\alpha, x_i) \leqslant \varepsilon + \xi_i, \\ \xi_i \geqslant 0, i = \overline{1, \ell}. \end{cases}$$

where ξ_i is an error on the object x_i.

It's proposed to use $L_\gamma(\alpha, x)$ as a loss function

$$L_{\varepsilon,\gamma}(\alpha, x) = \begin{cases} 0, & \gamma(x)L(\alpha, x) \leqslant \varepsilon \\ \gamma(x)L(\alpha, x) - \varepsilon, & \gamma(x)L(\alpha, x) > \varepsilon, \end{cases}$$

where $\gamma(x)$ is the weight function which defines the importance of the object x for optimization problem (1).

Let us $\xi_x = \gamma(x)L(\alpha, x) - \varepsilon$. Then

$$L_{\varepsilon,\gamma}(\alpha, x) = \begin{cases} 0, & \xi_x \leqslant 0, \\ \xi_x, & \xi_x > 0, \end{cases}$$

and

$$L_{\varepsilon,\gamma}(\alpha, x) = \frac{1}{2}\Big(|\xi_x| + \xi_x\Big).$$

Introduce additional variables ξ_x^+ and ξ_x^-:

$$\xi_x^+ = \frac{|\xi_x| + \xi_x}{2}, \quad \xi_x^- = \frac{|\xi_x| - \xi_x}{2}, \quad \xi_x^+ \geqslant 0, \quad \xi_x^- \geqslant 0.$$

Note that

$$\xi_x = \xi_x^+ - \xi_x^- \qquad \textit{and} \qquad |\xi_x| = \xi_x^+ + \xi_x^-.$$

As a result the optimization problem is got in the form below

$$\begin{cases} \frac{1}{2}\|\alpha\|^2 + C\sum_{i=1}^{\ell} \xi_{x_i}^+ \to \min\limits_{\alpha,\alpha_0}, \\ \gamma(x_i)L(\alpha, x_i) \leqslant \varepsilon + \xi_{x_i}^+ - \xi_{x_i}^-, \\ \xi_{x_i}^+ \geqslant 0, \quad \xi_{x_i}^- \geqslant 0, \quad i = \overline{1, \ell}, \end{cases}$$

which reduces to the problem of quadratic (linear) programming.

Conclusion

The optimization problem with precedent (training sample) initial information is considered. The open problems that must be solved to obtain better quality solutions of this problem are highlighted. Proposed the weighted loss function which uses the importance of the object for the optimization problem. It's shown how to use such function for reconstruction of the target function using linear regression and SVR methods. It should be noted that the using of different weight loss functions (with weights depending on an optimization problem) provide more accuracy formalize the optimization problem and obtain better solutions.

References

1. Anafiyev, A. and Blyschik, V. 2011. Optimization problem with precedent initial information. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 51–57.
2. Anafiyev, A. 2012. The approach to solving optimization problems with precedent initial information. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 7–12.
3. Taratynova, N. 2006. Building an optimization model for precedent initial information as a problem of nonlinear regression. *Iskusstvennyi intellekt*, 2, pp. 238–241.
4. Smola, A. 2004. Tutorial on Support Vector Regression. *Statistics and Computing*, 14, pp. 199–222.