EXACT DISCRIMINANT FUNCTION DESIGN USING SOME OPTIMIZATION TECHNIQUES

Yury Laptin, Alexander Vinogradov

Abstract: Some aspects of design of the discriminant functions that in the best way separate points of predefined final sets are considered. The concept is introduced of the nested discriminant functions which allow to separate correctly points of any of the final sets. It is proposed to apply some methods of non-smooth optimization to solve arising extremal problems efficiently.

Keywords: cluster, solving rule, discriminant function, linear and non-linear programming, non-smooth optimization

ACM Classification Keywords: G.1.6 Optimization - Gradient methods, I.5 Pattern Recognition; I.5.2 Design Methodology - Classifier design and evaluation

Conference: The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009

Introduction

Linear decision rule (LDR) keep relative simplicity at high computational efficiency. At use of the algorithms realizing LDR, the raised speeds of recognition can be reached that is important for the decision of various problems concerned to mass data processing. At the same time, construction of the best LDR quite often leads to posing complex optimization problems. Situation with strongly overlapped classes under condition of weakness of stochastic components in data can serve here as an example, when search exact LDR with a zero mistake on training sample is justified, but encounters difficulties of strictly combinatory character [1]. Similar difficulties arise also when each pair of classes is easily separable by means of LDR, but the number of classes is great. In such situations crucial importance gets a choice of an adequate method of solving the optimization problem. Researches on the given direction are carried out all over the world and continue to remain actual, since are based and supported from two parties, as by progress in the field of creation of new methods of optimization, as by successes of the theory of recognition [2-6]. In this work some applications of methods of non-smooth optimization are considered in problems of search of linear discriminant functions (linear classifiers) correctly separating clusters as final sets in R^n .

1. Simple discriminant functions

Let's consider as predefined some collection of final sets $\Omega_i = \left\{ p^t \in R^n ,\, t \in T_i \right\},\, i = 1,\ldots,\, m$, where T_i is the set of point indices in Ω_i . We use the term *discriminant function* for any function $\pi : R^n \to \{1,\ldots,\, m\}$.

Let functions $f_i : R^n \to R,\; i = 1,\ldots, m$, be set. In the further we consider discriminant functions of the following kind

$$\pi(x) = \arg\max_i \left\{ f_i(x) :\; i = 1,\ldots,\, m \right\}. \qquad (1)$$

We say that discriminant function $\pi(x)$ correctly divides points from Ω_i , $i = 1,\ldots,\, m$, if $\pi(x) = i$, for all $x \in \Omega_i$, $i = 1,\ldots,\, m$. Set $K_i = \left\{ x \in R^n : \pi(x) = i \right\}$ is referred to as *class* K_i *generated by function* $\pi(x)$.

Remark 1. Function $\pi(x)$ is invariant concerning to multiplication of all functions f_i by positive value, and to addition of any value to all of f_i.

Function $\pi(x)$ of a kind (1) is named *simple discriminant function* if all functions f_i are linear. Let $m = 2$. It is easy to see, that if simple discriminant function correctly divides points of two final sets, a hyperplane defined by condition

$$(a^1, x) + b_1 = (a^2, x) + b_2, \tag{2}$$

separates sets Ω_1, Ω_2.

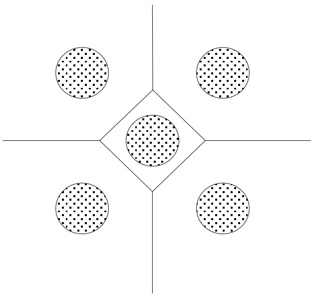

Fig. 1.

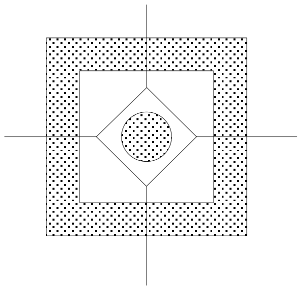

Fig. 2.

On Fig.1 an example of sets in R^2 and the division of a plane into classes by simple discriminant function is presented. Sets 1, 2, …, 5 are circles of radius 1 placed, accordingly, in points (-2,2), (2,2), (2,-2), (-2,-2), (0,0). Linear functions $l_i(x) = (a^i, x) + b_i$: $a^1 = (-1,1)$, $a^2 = (1,1)$, $a^3 = (1,-1)$, $a^4 = (-1,-1)$, $a^5 = (0,0)$; $b_i = 0,\ i = 1,\dots,4$, $b_5 = 2$.

Generally (for any m) there is a question on existence of the discriminant function $\pi(x)$ correctly separating points from Ω_i, $i = 1,\dots, m$.

Theorem 1. Let around of each set Ω_i the sphere S_i, $i = 1,\dots, m$, can be constructed, so that $S_i \cap S_j = \varnothing,\ i \neq j$. Then there is a simple discriminant function $\pi(x)$ separating points from Ω_i, $i = 1,\dots, m$ correctly.

Proof. We shall consider all over again a case when each set Ω_i consists of one point. Let $F(x)$ be strictly convex smooth function such that all points from Ω_i, $i = 1,\dots, m$ belong to domain of $F(x)$. To each set $\Omega_i = \{p^i\}$ we shall put in correspondence the function $f_i(x) = F(p^i) + (\nabla F(p^i), x - p^i),\ i = 1,\dots, m$. By the strict convexity it is forced that $f_i(p^i) = F(p^i) > F(p^j) + (\nabla F(p^j), p^i - p^j) = f_j(p^i),\ \ j \neq i$. Whence it follows, that discriminant function $\pi(x)$ correctly separates points from Ω_i, $i = 1,\dots, m$.

Let's pass to the general case. As function $F(x)$ we shall choose a hemisphere of enough the big radius r in space R^{n+1} which center is located in a point (x^0, r), where x^0 is fixed, and r we shall vary (if necessary). For each set Ω_i we shall select linear function $f_i(x) = (a^i, x) + b_i$. We shall designate $E_i = \left\{x \in R^n : (a^i, x) + b_i \geq F(x)\right\}$. The set E_i is a projection of crossing of a plane and a semicircle in R^{n+1} on space R^n. We shall consider such linear functions $f_i(x) = (a^i, x) + b_i$, for which E_i is an ellipsoid. It is easy to see, that if radius r is big enough then always it is possible to choose function $f_i(x) = (a^i, x) + b_i$ so that $S_i \subseteq E_i$ will be valid. We shall choose functions $f_i(x)$ so that corresponded to them ellipsoids E_i had the minimal size (with the minimal small axis) and $S_i \subseteq E_i$ was still valid. It is easy to see, that increasing radius *r* of a hemisphere it is possible always to achieve that ellipsoids E_i, $i = 1,\dots, m$, were not crossed.

Let such functions $f_i(x) = (a^i, x) + b_i$ are constructed, ellipsoids E_i corresponding to them are not crossed and $S_i \subseteq E_i$ holds for all $i = 1,\dots, m$. It is easy to see, that at construction we have $F(x) > f_j(x),\ x \notin E_j$ and $f_i(x) \geq F(x) > f_j(x),\ x \in E_i,\ i \neq j$. Thus, $f_i(x) > f_j(x),\ x \in E_i,\ i \neq j$, and the discriminant function $\pi(x)$ does separate correctly points from Ω_i, $i = 1,\dots, m$. Theorem is proved ■.

It should be noticed, that conditions of the Theorem 1 are rather rigid. It is possible to find many examples where these conditions don't hold, but the correct discriminant function for Ω_i, $i = 1,\dots, m$ does exist.

Let's introduce a criterion of quality of function concerning to collection $\Omega_i \subset R^n$, $i = 1,\dots, m$

$$\delta(\pi) = \min\left\{ f_i(x) - f_j(x):\ j \in \{1,\dots, m\} \setminus i,\ x \in \Omega_i,\ i = 1,\dots, m \right\}, \tag{3}$$

The criterion $\delta(\pi)$ characterizes how much values of functions $f_j(x),\ j \in \{1,\dots, m\} \setminus i$ differ from values $f_i(x)$ in points $x \in \Omega_i$. It is obvious, that if $\delta(\pi) > 0$ holds then the function $\pi(x)$ correctly separates points from $\Omega_i \subset R^n$, $i = 1,\dots, m$. Design of simple discriminantal function $\pi(x)$ is equivalent to a choice of values of vectors a^i and parameters b_i, $i = 1,\dots, m$. In view of the Remark 1 the problem of choosing the best simple discriminant function for criterion $\delta(\pi)$ we shall present in the form of a problem of linear programming: to find

$$\delta^* = \max_{a,b,\delta} \ \delta, \tag{4}$$

at restrictions

$$(a^i - a^k, p^t) + b_i - b_k \geq \delta,\ \ t \in T_i,\ k \in \{1,\dots, m\} \setminus i,\ \ i = 1,\dots, m, \tag{5}$$

$$-1 \leq a_j^i \leq 1,\ i = 1,\dots, m,\ \ j = 1,\dots, n. \tag{6}$$

$$b_1 = 0. \tag{7}$$

Restriction (7) is added in view of invariance of functions $\pi(x)$ concerning addition of any number to all f_i. Restrictions (6) are the normalizing conditions. These conditions can be written as restrictions put on the norms:

$$\left\| a^i \right\|^2 \leq 1,\ i = 1,\dots, m. \tag{8}$$

In this case the problem (4), (5), (7), (8) will be a problem of quadratic programming.

It is easy to see, that if there exists the simple discriminant function $\pi(x)$ correctly separating points from Ω_i, $i = 1,\dots, m$, then $\delta^* > 0$ and the decision of the problem (4) - (7) defines optimum discriminant function. Otherwise, any set for which $a^i = a^k$, $b_i = b_k$, $i, k \in \{1,\dots, m\}$, is optimum, $\delta^* = 0$, and the decision of problem (4) - (7) does not contain useful information.

Variables number of problem (4)-(7) is equal to $m(n+1)+1$, number of restrictions (5) – $\mathrm{N}(m-1)+1$, where N – total number of points in sets Ω_i, $i = 1,\dots, m$.

For large N it is advisable to consider the problem (4), (5), (7), (8) and to represent it in the form: find

$$\delta^* = \max_{a,b} \ \left\{ \min\left\{ (a^i - a^k, p^t) + b_i - b_k :\ \ t \in T_i,\ k \in \{1,\dots, m\} \setminus i,\ \ i = 1,\dots, m \right\} \right\}, \tag{9}$$

subject to (7), (8). Objective function of this problem is peace-wise linear, so, non-smooth optimization methods [Error! Reference source not found.] could be used to solve this problem.

In the case, when $\delta^* = 0$ for the problem (4)-(7), finding good simple discriminant function will be realized in two stages. Analogous approaches were considered in [7, 8]. At the first stage it is proposed to exclude some points from the sets Ω_i, $i = 1,..., m$ in such a way that for other points inequality $\delta^* \geq \overline{\delta}$ be satisfied for the problem (4)-(7), where $\overline{\delta}$ is a parameter. On the second stage the values of b_i, $i = 1,..., m$ have to be chosen to improve the discriminant function.

Denote $T = \bigcup_{i=1}^{m} T_i$. Let associate with every point p^t, $t \in T$ a variable $y_t = 0 \vee 1$ such that $y_t = 1$, if a point p^t should be considered while formulating the problem (4)-(7), and $y_t = 0$ otherwise. Let parameter $\overline{\delta} > 0$ and large positive number M be given.

The problem of exclusion some points from the sets Ω_i, $i = 1,..., m$ has the form: find

$$\max_{a,b,y} \left\{ \sum_{t \in T} y_t \right\}, \quad (10)$$

subject to

$$(a^i - a^k, p^t) + b_i - b_k + M(1 - y_t) \geq \overline{\delta}, \quad t \in T_i,\ k \in \{1,..., m\} \setminus i,\ i = 1,..., m, \quad (11)$$

$$-1 \leq a_j^i \leq 1,\ i = 1,..., m,\ j = 1,..., n, \quad (12)$$

$$\sum_{t \in T_i} y_t \geq 1, \quad i = 1,..., m, \quad (13)$$

$$0 \leq y_t \leq 1,\ t \in T, \quad (14)$$

$$b_1 = 0. \quad (15)$$

$$y_t = 0 \vee 1,\ t \in T, \quad (16)$$

It is evident that if $y_t = 0$, then for sufficiently large M corresponding inequality of form (11) will be satisfied for any a^i, b_i, i.e. the point p^t is excluded from the problem.

Constraints (13) specify the condition that at least one point from every set Ω_i must be included in the problem.

Let an approximate solution $\overline{a}^i$, $\overline{b}_i$, $i \in \{1,..., m\}$, $\overline{y}_t$, $t \in T$ of the problem (10)-(16) is found. At the second stage to improve the discriminant function we solve the problem (4)-(7) under fixed variables $a^i = \overline{a}^i$, $i \in \{1,..., m\}$.

It should be noted that the resulting discriminant function does not guarantee proper separating of points from sets Ω_i, $i = 1,..., m$.

2. Nested discriminant functions

Partitioning the sets Ω_i into non-overlapping sets $\Omega_i = \bigcup_{j \in J_i} \Omega_i^j$ will be referred to be effective, if it is possible to build a simple discriminant function for the whole Ω_i^j, $j \in J_i$, $i = 1,..., m$, properly separating the points of these sets. Such discriminant function may not exist for initial sets Ω_i, $i = 1,..., m$.

Nevertheless, effective partitioning always exists, for example, when every set Ω_i^j consists from one point.

Let an effective partitioning $\Omega_i = \bigcup_{j \in J_i} \Omega_i^j$, $i = 1,..., m$ be given. Denote $\pi^*(x)$ an optimal simple discriminant function for the sets Ω_i^j, $j \in J_i$, $i = 1,..., m$,

$$\pi^*(x) = \arg\max_{ij} \left\{ (a^{ij}, x) + b_{ij} : i = 1,..., m, \ j \in J_i \right\}. \tag{17}$$

The function $\pi^*(x)$ returns a pair $\left(i^*(x), j^*(x)\right)$, giving a maximum in (17). It is evident, that $i^*(x)$ is a discriminant function properly separating points from $\Omega_i \subset R^n$, $i = 1,..., m$.

Denote

$$\psi_i^*(x) = \max \left\{ (a^{ij}, x) + b_{ij} : j \in J_i \right\}, \ i = 1,..., m. \tag{18}$$

It is easy to see that

$$i^*(x) = \arg\max_{i} \left\{ \psi_i^*(x) : \ i = 1,..., m \right\}. \tag{19}$$

Functions (19) will be named *nested discriminant function*. The use of nested discriminant function allows us to improve the quality of the best approximation of sets Ω_i, $i = 1,..., m$.

Let we consider two sets in Fig. 2. The nested discriminant function has a form $i^*(x) = \arg\max \left\{ \psi_i^*(x) : \ i = 1, 2 \right\}$, where $\psi_1^*(x) = l_5(x)$, $\psi_2^*(x) = \max \left\{ l_i(x) : i = 1,..., 4 \right\}$, functions $l_i(x), \ i = 1,...,5$ are determined for Fig. 1.

Heuristic scheme for finding a nested discriminant function consists from finite number of steps of handling the current partitioning $\Omega_i = \bigcup_{j \in J_i} \Omega_i^j$, $i = 1,..., m$, and looks as follows:

1) On the first step $k = 1$, take Ω_i, $i = 1,..., m$ as a current partition of $\Omega_i = \bigcup_{j \in J_i} \Omega_i^j$, $i = 1,..., m$.

2) On k th step solve the problem (4)-(7) for the current partitioning $\Omega_i = \bigcup_{j \in J_i} \Omega_i^j$, $i = 1,..., m$. If optimal value $\delta^* > 0$, the process is finished. Otherwise find an approximate solution of (10)-(16). On the basis of this solution every set Ω_j^j is divided into two subsets: points with $y_t = 0$ and points with $y_t = 1$. Then define the current partition more precisely, put $k = k + 1$ and go to 2).

It is easy to see that the process is finite, and as a result we get the nested discriminant function, properly separating points from Ω_i, $i = 1,..., m$.

Conclusions

Approaches for finding discriminant function separating points from given sets $\Omega_i \subset R^n$, $i = 1,..., m$ are considered. The problem of finding an optimal discriminant function is formulated as a linear (4)-(7) or quadratic (4), (5), (7), (8) programming problems. However this problem has a sense only in the case when there exists simple discriminant function, properly separating points from Ω_i, $i = 1,..., m$.

In the case, when proper separating points from Ω_i, $i = 1,..., m$ is impossible, a two-stage procedure for finding a simple discriminant function is proposed. At the first stage it is proposed to exclude some points from the sets Ω_i, $i = 1,..., m$, and at the second stage the resulting discriminant function can be improved.

The notion of nested discriminators allowing to make properly separating of points from any disjoint sets $\Omega_i \subset R^n$, $i = 1,..., m$ is introduced. An heuristic scheme for finding nested discriminator is proposed.

Optimization problems arising in the considered approaches are large-scale problems and have a great number of constraints. These problems can be reduced to the problem of maximization a concave piece-wise linear function with a great number of pieces under simple constraints. To solve them it is advisory to use non-smooth optimization methods [6] – generalized subgradient descent methods for large number of variables or methods with space transformation, if the number of variables does no exceed 300.

Acknowledgements

This work was done in the framework of Joint project of the National Academy of Sciences of Ukraine and the Russian Foundation for Basic Research No 08-01-90427 "Methods of automatic intellectual data analysis in tasks of recognition objects with complex relations".

Bibliography

1. Гупал А.М., Сергиенко И.В. Оптимальные процедуры распознавания. - Киев: Наук.думка, 2008. - 232 с.
2. Koel Das, Zoran Nenadic. An efficient discriminant-based solution for small sample size problem // Pattern Recognition – Volume 42, Issue 5, 2009, Pages 857-866.
3. Juliang Zhang, Yong Shi, Peng Zhang. Several multi-criteria programming methods for classification // Computers & Operations Research – Volume 36, Issue 3, 2009, Pages 823-836.
4. E. Dogantekin, A. Dogantekin, D. Avci Automatic Hepatitis Diagnosis System based on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System // Expert Systems with Applications, In Press, 2009.
5. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. – К.: Наукова думка, 2004. – 545 с.
6. Shor N.Z. Nondifferentiable Optimization and Polynomial Problems. – Dordrecht, Kluwer, 1998. – 394 p.
7. Bennett K.P., Mangasarian O.L. Robust Linear Programming Discrimination of Two Linearly Inseparable Sets // Optimization Methods and Software. – 1996. –№5. – P. 23-34.
8. Журбенко Н.Г., Саимбетов Д.Х. К численному решению одного класса задач робастного разделения двух множеств // Методы исследования экстремальных задач. – К.: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 1994. – С. 52–55.

Authors' Information

Yury Laptin –senior researcher, V.M.Glushkov Institute of Cybernetics of the NASU, Prospekt Academika Glushkova, 40, 03650 Kyiv, Ukraine; e-mail: laptin_yu_p@mail.ru

Alexander Vinogradov – senior researcher, Dorodnicyn Computing Centre of the RAS, Vavilova 40, 119333 Moscow, Russian Federation; e-mail: vngrccas@mail.ru