

Jiří Polách

The possibilities of the linear discriminant method for decision

Kybernetika, Vol. 10 (1974), No. 1, (44)--48

Persistent URL: <http://dml.cz/dmlcz/125745>

Terms of use:

© Institute of Information Theory and Automation AS CR, 1974

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these

Terms of use.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*
<http://project.dml.cz>

The Possibilities of the Linear Discriminant Method for Decision

JIRÍ POLÁCH

In the paper the question of the reduction of the initial set of variables in medical decision problem is solved. The discriminant measure, as the criterion for ordering variables, is introduced and its properties are derived.

INTRODUCTION

Medical diagnosis is a special recognition problem characterized on the one hand by its statistical character and on the other hand by unequal costs of measured variables.

Especially the second feature is significant for diagnostic decision making. Variables, by them we mean any data about a patient gained by history-taking, laboratory examinations, therapeutic results or even by surgery, have not the same information value. Also costs of the measurements of these variables are different and they can be hardly determined, because they are connected with payload, the patient's discomfort, the risk of this health aggravation, etc. These circumstances mean for a diagnostician to choose such a decision strategy that will guarantee sufficient information for making a diagnosis and, at the same time, the number of variables will be the minimum.

The presented method of ordering and selection of the most important variables is based on properties of well-known linear discriminant method for the case of multivariate distribution of variables, that need not be normal, in which the criterion for construction of the discriminant function is the maximum of the between-class variance in the transformed space. Ordering of variables is made possible by the existence of monotonicity of a certain characteristic of the mutual distinction of diagnostic classes.

Let each patient, that can have one from s diagnoses, be described by a column vector \mathbf{x} with k components. Let us call the set of all n patients described by vector \mathbf{x} the training sample from the mixed class, the set of n_γ patients with diagnosis γ the training sample from the γ -th class. Further, let us denote by U^w the sample within-class covariance matrix, by U^b the sample between-class covariance matrix and by U the sample covariance matrix of the mixed class. These matrices of order k have the elements:

$$U^w: u_{ij}^w = \frac{1}{n} \sum_{\gamma=1}^s \sum_{\xi=1}^{n_\gamma} (x_{i\xi}^{(\gamma)} - \bar{x}_i^{(\gamma)}) (x_{j\xi}^{(\gamma)} - \bar{x}_j^{(\gamma)}),$$

$$U^b: u_{ij}^b = \frac{1}{n} \sum_{\gamma=1}^s n_\gamma (\bar{x}_i^{(\gamma)} - \bar{x}_i) (\bar{x}_j^{(\gamma)} - \bar{x}_j),$$

$$U: u_{ij} = \frac{1}{n} \sum_{\xi=1}^n (x_{i\xi} - \bar{x}_i) (x_{j\xi} - \bar{x}_j),$$

where by $x_{i\xi}^{(\gamma)}$ we denote the values of the i -th variable measured on the ξ -th patient from the γ -th class, by $\bar{x}_i^{(\gamma)}$ the mean of the i -th variable of the γ -th class, by \bar{x}_i the mean of the i -th variable of the mixed class, i.e. the class consists of all $n = n_1 + \dots + n_s$ patients of all s classes.

The method of linear discrimination is based on looking for such a linear transformation (with a square matrix C of full rank)

$$(1) \quad \mathbf{z} = C^T \mathbf{x}$$

that maximizes the multivariate between-class variance, at the same time the within-class variance is constant, in the transformed space. The solution is looked for by the method of Lagrange's multipliers and leads to solving the eigenvalue problem in the non-standard form [2]:

$$(2) \quad (U^b - \lambda U^w) \mathbf{c} = 0,$$

where \mathbf{c} 's are column vectors of the matrix C and λ 's Lagrange's multipliers.

For continuous variables x_i the matrix U^w is positive definite and U^b is positive semidefinite of rank $(s - 1)$ with probability 1 under the condition: $(n - s) \geq k$, $s \leq k$, where n is the number of all patients of training sample from the mixed class, s is the number of diagnostic classes and k is the number of all measured variables (see [2]). If some of the variables x_i are discrete, then non-singularity of U^w must be attested or eventually secured.

Under the condition of non-singularity of U^w we can decompose it into the product of the upper and the lower triangular matrices

$$U^w = LL^T$$

46 and the equation (2) will have the form

$$(3) \quad \mathbf{P}\mathbf{y} = \lambda\mathbf{y},$$

where

$$\mathbf{P} = \mathbf{L}^{-1}\mathbf{U}^b(\mathbf{L}^T)^{-1}, \quad \mathbf{y} = \mathbf{L}^T\mathbf{c}.$$

Now let us introduce a characteristic of the mutual distinction of diagnostic classes in the transformed space:

$$(4) \quad \eta_* = 1 - \frac{|\mathbf{U}_*^w|}{|\mathbf{U}_*|},$$

where $|\mathbf{U}_*^w| = \det \mathbf{U}_*^w$ is the within-class variance and $|\mathbf{U}_*| = \det \mathbf{U}_*$ is the variance of the mixed class in the transformed space. It holds:

$$\mathbf{U}_*^w = \mathbf{C}^T\mathbf{U}^w\mathbf{C}, \quad \mathbf{U}_* = \mathbf{C}^T\mathbf{U}\mathbf{C}, \quad \mathbf{U}_* = \mathbf{U}_*^w + \mathbf{U}_*^b.$$

It can be shown (see [2]) that

$$(5) \quad \eta_* = 1 - \frac{1}{(1 + \lambda_1) \dots (1 + \lambda_{s-1})},$$

where $\lambda_i \neq 0$ are eigenvalues of the matrix \mathbf{P} . From (5) we can see that values of η_* are from the interval $(0, 1)$. The higher values of η_* correspond with better discrimination, i.e. diagnostic classes are mutually more separated in the transformed space. Therefore we shall call η_* the *discriminant measure*.

It can be proved that η_* is invariant under non-singular linear transformations (i.e. the matrix of transformation is a square matrix of full rank). So, it holds namely

$$(6) \quad \eta_* = 1 - \frac{|\mathbf{U}_*^w|}{|\mathbf{U}_*|} = 1 - \frac{|\mathbf{U}^w|}{|\mathbf{U}|} = \eta,$$

i.e. the discriminant measure in the transformed and the original spaces are the same.

Thus η can be computed from (6) without solving the eigenvalue problem. Further, as we want to compute η for the set of $(k - 1)$ variables, when i -th variable is eliminated, then it is enough to omit the i -th row and i -th column in the matrix \mathbf{U}^w and \mathbf{U} and to replace $|\mathbf{U}^w|$ and $|\mathbf{U}|$ by determinants of obtained minors of order $(k - 1)$. So, if we succeeded to prove that η is a monotonic function of the number of variables, then we would obtain the suitable criterion for ordering and selecting variables according to their discriminant ability.

Indeed, there exists a theorem on the base of which we can prove the monotonicity of η .

Theorem. Let matrices \mathbf{U}^w and \mathbf{U}^b be symmetric and \mathbf{U}^w positive definite. Then the roots λ'_j of the equation

$$\det(\mathbf{U}_r^b - \lambda'\mathbf{U}_r^w) = 0$$

separate the roots λ_j of the equation

$$\det(\mathbf{U}_{r+1}^b - \lambda \mathbf{U}_{r+1}^w) = 0,$$

where \mathbf{U}_i^w , \mathbf{U}_i^b are the leading principle minors of \mathbf{U}^w and \mathbf{U}^b of order i .

The proof of the theorem is given by Wilkinson [1].

In our case the rank of \mathbf{U}^b (and therefore of \mathbf{P} too) is, under above mentioned assumptions, equal to $(s - 1)$. Thus in view of the Theorem we shall have

$$(7) \quad \lambda_1 \geq \lambda'_1 \geq \dots \geq \lambda_{s-1} \geq \lambda'_{s-1}.$$

This gives

$$\frac{1}{(1 + \lambda_1) \dots (1 + \lambda_{s-1})} \leq \frac{1}{(1 + \lambda'_1) \dots (1 + \lambda'_{s-1})}$$

and hence

$$(8) \quad \eta_{r+1} \geq \eta_r,$$

where η_i denotes the discriminant measure calculated for the set of i variables.

Now we can summarize the properties of the discriminant measure η :

- 1) values of η are from the interval $(0, 1)$,
- 2) it is invariant under linear non-singular transformations,
- 3) it is a monotonic function of the number of variables.

We see that η is similar to an information measure of dependence [3].

ALGORITHM OF ORDERING AND SELECTING OF VARIABLES

Let us consider the case of two classes ($s = 2$).

- 1) Construct the sample matrices \mathbf{U}_k^w , \mathbf{U}_k^b , \mathbf{U}_k and compute η_k for the complete set of variables x_i :

$$\eta_k = 1 - \frac{|\mathbf{U}_k^w|}{|\mathbf{U}_k|}.$$

- 2) Choose a value of coefficient α that will determine assigned reduction of η_k and denote

$$\eta_0 = (1 - \alpha) \eta_k.$$

- 3) Set $q = 1$.
- 4) Compute η_{qj} for combinations of q variables ($q - 1$) of them have been selected in the preceding steps; j denotes indexes of variables that have not been selected in the preceding steps. Let be

$$\max_j \eta_{qj} = \eta_{q1}.$$

48 If

$$\eta_{qt} \geq \eta_0$$

then choose the variable x_t as the best in the given step and interrupt the process of selecting of variables.

5) If

$$\eta_{qt} < \eta_0$$

choose the variable x_t as the best in the given step, set $q = q + 1$ and go to 4).

If we index variables according to the number of the step in which they have been selected we get the following sequence of combinations of variables:

$$x_1, x_1x_2, \dots, x_1 \dots x_t, \quad t \leq k - 1.$$

The remaining $(k - t)$ variables have not any substantial discriminant power (the discriminant measure can increase at most by $\alpha\eta_k$), so we can eliminate them as nonsubstantial. For the resulting combination of t variables we use the linear discriminant method.

CONCLUSION

The advantages of the presented method for the reduction of the initial set of variables in a diagnostic decision problem are:

- 1) a normal distribution of variables needs not be assumed,
- 2) an independence of variables needs not be assumed,
- 3) variables can be continuous and discrete,
- 4) estimations of the discriminant measure are better than those of measures based on probabilities namely for sets of a large number of variables.

Disadvantages of the method are given by restrictions of linear models.

After the information-theoretic approach of Perez [4] to the problem of the reduction of a system of parameters the proposed method shows other possibility of solving the problem.

(Received December 7, 1972.)

REFERENCES

- [1] J. H. Wilkinson: The Algebraic Eigenvalue Problem, Clarendon Press, Oxford 1965.
- [2] S. S. Wilks: Mathematical Statistics, John Wiley & Sons, New York 1962.
- [3] J. Nikl, A. Perez: Aplikace method teorie informace při studiu závislosti v biologických systémech. *Čs. fyziologie* 10 (1961), 5, 454–460.
- [4] A. Perez: Information-Theoretic Approach to Measurement Reduction Problems. *Kybernetika* 6 (1970), 2, 90–98.

RNDr. Jiří Polách, Katedra sociálního lékařství LF University J. E. Purkyně (Department of Social Medicine, Medical Faculty), Obránců míru 10, 662 43 Brno, Czechoslovakia.