Decision and classification problems using Mahalanobis statistical distance

Carlos Criado
Departamento de Fisica Aplicada I
Universidad de Malaga
29071 Malaga, Spain
c_criado@uma.es

Nieves Alamo
Departamento de Algebra, Geometria y Topologia
Universidad de Malaga
29071 Malaga, Spain

Hector Rabal and Nelly Cap
Centro de Investigaciones Opticas
CONICET La Plata-CIC
Facultad de Ingenieria
Universidad Nacional de la Plata
La Plata, Argentina

Abstract—We present a method that uses Mahalanobis distance to study some decision and classification problems. Four examples are used to show the diversity of possible applications of our method. Based on two simulated sets of data, we have calculated the corresponding Mahalanobis distances and their sum and differences Holodiagrams. These diagrams are valuable visualization tools for studying decision and classification problems in the proposed applications.

Keywords: Mahalanobis distance; Holodiagram; Classification; Decision analysis.

I. INTRODUCTION

In this work, we present a method that uses Mahalanobis distance [1] to study some decision and classification problems. The Mahalanobis distance is a generalization of the Euclidean distance and is useful in our method due to its statistical character. We represent the sum and difference Holodiagrams (HDs) ([2]; [3]) because they are valuable visualization tools for studying decision and classification problems in the proposed applications.

We briefly review the main aspects of the already known HDs and the concept of Mahalanobis distance. We then propose the concept of a Mahalanobis-based Holodiagram (MHD) and give some possible general applications.

The sum or difference can be considered in the MHDs, that is, the locus of points where the sum \( S \) of the Mahalanobis distances to two fixed points called foci is a constant, or the locus of the points where the difference \( D \) of those distances is a constant. We show these loci through plots of \( \cos(kS) \) and \( \cos(kD) \) with \( k \) a scale constant, such that they appear as fringes, and their geometry and spacing indicate, in each particular case, the corresponding properties. We also show sensitivity to changes in the variables in both cases. Two examples of data are simulated to illustrate our applications of the Mahalanobis distance and the associated Holodiagrams. To show the utility of the MHD concept in decision and classification problems, we provide four examples from different fields to which, among others, it can be applied: (a) the manufacture of commercial products; (b) investment funds; (c) the composition of pharmaceutical products; and (d) ecology.

II. HOLODIAGRAMS AND MAHALANOBIS DISTANCE

A. Holodiagrams

Holodiagrams HD is a plot of the loci of equal sum of the (Euclidean) distances \( d_1 \) and \( d_2 \) to two fixed points \( F_1 \) and \( F_2 \). It can be shown as the \( \cos k(d_1 + d_2) \) in the shape of fringes. The total distance \( S = d_1 + d_2 \) is a constant along each of such fringes, such that changing the variables along it does not modify the value of \( S \), whereas in the orthogonal direction to the fringes, the change in variables produces the maximal variation in \( S \). These properties make the HD an appropriate tool for graphically solving variational problems in optics ([4]; [5]), and as we show, it can also be useful, together with the Mahalanobis distance, in studying problems in decision optimization and in setting classifications.

B. Mahalanobis distance

To motivate the definition of the Mahalanobis distance, assume that among a set of points characterized by two parameters \( x = (x_1, x_2) \), we know a random sample of \( N \) points and want to estimate the probability that a given point belongs to that set. It seems natural to assume that the probability should be greater the closer the vector of means \( \mu = (\mu_1, \mu_2) \), that is, the shorter the distance \( d(x, \mu) = \sqrt{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2} \) is.

On the other hand, it also seems natural to take into account the standard deviations \( \sigma_1 \) and \( \sigma_2 \) of the random variables, and then to weigh the contribution of a random variable through the inverse of the corresponding standard deviation. In this way, we should obtain for the distance
\[
d(x, \mu) = \sqrt{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2}.
\]
Finally, some consideration should be given to whether these variables are correlated, such that a suitable definition for the distance might be \( d(x, \mu) = \sqrt{(x - \mu)\Sigma^{-1}(x - \mu)^T} \), where \( \Sigma \) is the covariance matrix, \( \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \). This is precisely the definition given by Mahalanobis in [1], which, for any two points \( x \) and \( y \), takes the form
\[
d(x, y) = \sqrt{(x - y)\Sigma^{-1}(x - y)^T}. \tag{1}
\]
It is easy to see that this definition satisfies the characteristic properties of a distance. On the other hand, it is clear that
the previous expression can be used in a space of \( N \) random variables, for any natural number \( N \geq 2 \). In two dimensions this distance can also be written in terms of the correlation coefficient \( r = \frac{\sigma_{12}}{\sigma_1 \sigma_2} \), taking the form:

\[
(1 - r^2)d(x, y)^2 = \left( \frac{x_1 - y_1}{\sigma_1} \right)^2 + \left( \frac{x_2 - y_2}{\sigma_2} \right)^2 - 2r \left( \frac{x_1 - y_1}{\sigma_1} \right) \left( \frac{x_2 - y_2}{\sigma_2} \right).
\]

(2)

In Euclidean geometry, the locus of the points bearing the same distance to a center is a circumference. This is a particular case of Mahalanobis distance when both variables have the same standard deviations and are not correlated (see Figure 1(a)). According to Eq. (2), the locus of points in which the Mahalanobis distance to a fixed point has the same value is, in general, an ellipse (see Figures 1(b) and 1(c)). In Figure 1(b), the variables are not correlated, whereas in Figure 1(c) they are.

\[\text{C. Simulation of two examples for a two-dimensional space of parameters.}\]

In order to have data to graphic representations of the examples presented in the following sections, we are going to provide a method to simulate a set of \( N \) points in a two-dimensional space of parameters with a mean vector \( \mu = (\mu_1, \mu_2) \), standard deviation \( \sigma_1 \) and \( \sigma_2 \) of the two coordinates, and correlation coefficient \( r \) between them. For that, we first obtain two sets of \( N \) numbers with normal distributions \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\) and with the correlation coefficient \( r \). These sets can be obtained from two sets \( \{x_1\} \) and \( \{x_2\} \) of \( N \) numbers with uniform distributions in the interval \((0, 1)\), to which we apply any method (see, for example, [6] or [7]) to find (pseudo)-random numbers with the standard normal distribution. For example, Box-Muller transformations give

\[z_1 = \sqrt{-2 \ln x_1 \cos(2\pi x_2)}, \quad z_2 = \sqrt{-2 \ln x_1 \sin(2\pi x_2)},\]

(3)

which both have the standard normal distribution \((\mu, \sigma) = (0, 1)\). Then the numbers \( \{z_3 = r z_1 + \sqrt{1 - r^2} z_2\} \) also have the standard normal distribution and are correlated with the \( N \) numbers \( \{z_1\} \) by the correlation coefficient \( r \). Finally, the numbers

\[z'_1 = z_1 \sigma_1 + \mu_1, \quad \text{and} \quad z'_3 = z_3 \sigma_2 + \mu_2,\]

(4)

have normal distributions \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\) and their correlation coefficient is \( r \).

Following the above method, we have simulated two examples of two sets of points with normal distributions which are represented in Figure 2. For Figure 2(a), in the first set, marked with black disks, \( N = 100, (\mu_1, \mu_2) = (1, 1.5), (\sigma_1, \sigma_2) = (1, 3), \) and \( r = 0.7 \). In the second set, marked with grey squares, \( N = 156, (\mu_1, \mu_2) = (4, 2), (\sigma_1, \sigma_2) = (2, 2), \) and \( r = -0.6 \). For Figure 2(b), in the first set, marked with black disks, \( N = 100, (\mu_1, \mu_2) = (1, 1.5), (\sigma_1, \sigma_2) = (1, 1.5), \) and \( r = 0.7 \). In the second set, marked with grey squares, \( N = 80, (\mu_1, \mu_2) = (3, -1), (\sigma_1, \sigma_2) = (2, 1), \) and \( r = 0.6 \). We have also represented the sets (ellipses) of points \( x \) such that \( d(x, \mu) = 2 \) for each vector of means \( \mu \) and its corresponding Mahalanobis distance \( d \).

III. USE OF MAHALANOBIS DISTANCE IN THE

HOLODIAGRAM

To be able to display a graphic description of how are the HDs using Mahalanobis distances, we require for the number of variables to be two. If more than two variables are involved, we will only be able to plot cuts of the surfaces that are generated by setting the sum or the difference of the distances to the same value.

We define the Mahalanobis-distance-based Holodiagram (MHD) as the family of curves that is obtained when the parameter \( C \) in the following expression is continuously varied:

\[d(P) = p_1 d_1(P, F_1) + p_2 d_2(P, F_2) = C,\]

(5)

where \( d_1(P, F_1) \) is a Mahalanobis distance from a generic point \( P \) in the plane of the variables to a focus \( F_1 \), \( d_2(P, F_2) \) is another Mahalanobis distance from the same point to the other focus \( F_2 \), and \( p_1 \) and \( p_2 \) are positive coefficients.

For the particular case that \( d_1 \) and \( d_2 \) are both the Euclidean distance, we obtain a family of Descartes ovals (see, for example, [5]). If, in addition, \( p_1 = p_2 \), we obtain families of
of ellipses or hyperbolas, according to taking the sign + or −, respectively, in Eq. (5). These curves are shown in Figure 3 as plots of \( \cos kd(P) \), for \( k \) a scale constant.

IV. APPLICATIONS OF THE MHD

A. First example: the manufacture of commercial products

Let us consider a product (for instance, a car or computer) that is characterized by \( m_p \) parameters (price, size, etc) and that admits \( m_u \) possible uses (work, family, etc). To simplify the explanation, assume the case of \( m_p = 2 \) and \( m_u = 2 \). Let \( x_1 \) and \( x_2 \) denote the parameters and \( A \) and \( B \) denote the possible uses or types of uses. The manufacturer has data on a number \( n_A \) of items bought with intended use \( A \): \( x^A = (x^A_1, x^A_2); i = 1, \ldots, n_A \) with mean value \( \mu^A = (\mu^A_1, \mu^A_2) \) and covariance matrix \( \Sigma^A \), and of a number \( n_B \) of items bought with intended use \( B \): \( x^B = (x^B_1, x^B_2); i = 1, \ldots, n_B \) with mean value \( \mu^B = (\mu^B_1, \mu^B_2) \) and covariance matrix \( \Sigma^B \). Let

\[
\begin{align*}
d_A(x, \mu^A) &= \sqrt{(x - \mu^A)^T \Sigma^{-1}_A (x - \mu^A)}, \\
d_B(x, \mu^B) &= \sqrt{(x - \mu^B)^T \Sigma^{-1}_B (x - \mu^B)}
\end{align*}
\]

be the corresponding Mahalanobis distances from a point \( x \) to \( \mu^A \) and \( \mu^B \) respectively.

If the manufacturer wants to manufacture a product that serves both types of users “in the best possible way”, it seems natural to consider points that minimize the sum of the above-mentioned distances, each them multiplied by adequate weights \( p_A \) and \( p_B \). These weights can depend on different factors associated with the interest of the manufacturer for each type of user. In some cases they could depend on \( n_A \) and \( n_B \) and a simple choice could be

\[
\begin{align*}
p_A &= \frac{n_A}{n_A + n_B}, \\
p_B &= \frac{n_B}{n_A + n_B}.
\end{align*}
\]

The function to be optimized is

\[
S(x) = p_A d_A(x, \mu^A) + p_B d_B(x, \mu^B).
\]

The points defined by \( S(x) = \text{constant} \) would all have the same “commercial value” or “multiple-use value”. In the case that \( \Sigma_A = \Sigma_B = I \) (identity matrix), that is, \( A \) and \( B \) are uncorrelated and with the same standard deviation, the distances would be Euclidean and the set would be a Descartes oval. If, in addition, \( n_A = n_B \), then the set would be an ellipse.

To obtain the parameters for which the product has optimal values we have to find the minima of \( S(x) \), that is, the points \( x \) such that the gradient of \( S(x) \) is zero. Thus, they are given by the equations

\[
\frac{\partial S(x)}{\partial x_1} = \frac{\partial S(x)}{\partial x_2} = 0.
\]

B. Second example: investment funds

The method referred to above can be applied to financial companies. For example, let us consider the case of a set of investment funds, each of them characterized by \( m_p \) parameters (for example, amount invested, term investment, profitability, etc.), and \( m_u \) profiles of investors (for example, those whose main interest is profit, those whose main interest is premium safety, etc).

The financial company has data on \( n_j \) investors of profile \( j \), for \( j = 1, \ldots, m_u \), and it wants to characterize investment funds according to their utility for all of these types of investor profiles.

It seems natural to calculate the means and covariance matrices for the different investor profiles to define the following function that is the sum of the different Mahalanobis distances \( d_j \) to the respective mean values \( \mu^j = (\mu^j_1, \ldots, \mu^j_{m_p}); j = 1, \ldots, m_u \) weighted by certain factors \( p_j \) which must be set by the financial company:

\[
S(x) = \sum_{j=1}^{m_u} p_j d_j(x, \mu^j); \quad x = (x_1, \ldots, x_{m_p}).
\]

For each constant \( C \), the points for which \( S(x) = C \) would form a class with equivalent values regarding the optimization grade of the product. The products with optimal values would be those for which

\[
\frac{\partial S(x)}{\partial x_k} = 0; \quad k = 1, \ldots, m_p.
\]

C. Third example: composition of pharmaceutical products

The following is an application for the medical sciences. Suppose we have a medicament consisting of a number \( m_p \) of active compounds, and which is used in the treatment of \( m_u \) diseases.

The pharmaceutical company that produces the medicament has data on the most appropriate concentration of the different active compounds for \( n_j \) patients for each of the \( m_u \) diseases.

With these data, the company can find the mean values and the covariance matrices needed to calculate the different Mahalanobis distances corresponding to each different type of disease. The choice of the weights \( p_j, j = 1, \ldots, m_u \) will depend on medical considerations regarding the different diseases. Thus, the function \( S(x) \) can be obtained according to Eq. (9), and therefore used to characterize and optimize the products as in the previous example.

D. Fourth example: ecology

Suppose we have two species of microorganisms \( A \) and \( B \) that do not interact with each other and whose habitats are determined by humidity, \( x_1 \), and temperature, \( x_2 \). We have the data on these parameters for \( n_A \) individuals of
Fig. 4. (a) and (b) Holodiagrams of the sum $S$ associated with the point distributions of Figure 2(a) and (b), respectively, with their respective means (disks and squares), and optimal values of $S$ (triangles). Some constraints sets are also shown and the extremal values on them.

Fig. 5. (a) and (b) Holodiagrams of the difference $D$ associated with the distribution of points of Figure 2(a) and 2(b), respectively.

species $A$: $(x_1^A, x_2^A); i = 1, \ldots, n_A$, and for $n_B$ of species $B$: $(x_1^B, x_2^B); i = 1, \ldots, n_B$.

With these data we calculate the means $\mu_A$ and $\mu_B$, standard deviations, and coefficient of correlation, which allows us to calculate the respective Mahalanobis distances $d_A$ and $d_B$. The weights $p_A$ and $p_B$ should be determined by ecological considerations. Thus, we can calculate the optimal values of humidity and temperature for the maximal survival of individuals of both species through the minima of $S(x) = p_A d_A(x, \mu^A) + p_B d_B(x, \mu^B)$.

We can also look for the extremes on a constraint set of the space of parameters. This example can be generalized to any number of species and parameters that characterize the habitat.

V. STUDY OF MHD DIAGRAMS FOR THE SUM

We can plot the corresponding Holodiagram for any of the above applications. Figures 4(a) and 4(b) show these diagrams for the distribution points of Figures 2(a) and 2(b). In these figures the means of the two distributions $\mu^A, \mu^B$, are represented by a black disk and a black square, respectively. The optimal points correspond to the points $x$ for which the sum $S(x)$ of the respective weighted distances from $x$ to $\mu^A$ and $\mu^B$, as given in Eq. (7), is minimal. In Figure 4 these points are triangles. Note that these points are not in the segment joining $\mu^A$ and $\mu^B$, contradicting naive intuition.

Each of the fringes indicates the locus of points where the weighted sum of the Mahalanobis distance to the means is constant. Therefore, according to the different points of view presented in the examples in Section IV, all the points of the same fringe have the same commercial value (example IV-A), the same financial value (example IV-B), the same grade of optimization for the pharmaceutical product (example IV-C), or the same ecological value (example IV-D).

If we have some constraints on the space of parameters, given, for example, by a level set $f(x) = \text{constant}$ of a function on this space, or by any other subset of the space of parameters, we can look for the maxima and minima of $S(x)$ over this set. Figure 4(a) shows an example of constraint given by a function (grey curve) relating the parameters of the two-dimensional space of parameters. Note that the points where the curve is tangent to the MHD identify the extreme values of the function $S(x)$ (maxima and minima) that satisfy the constraints. In the constraint set of Figure 4(a), there are two maxima and two minima, one of which is also the absolute minimum, which is marked by a white circle. The constraint set of Figure 4(b) is a subset of the space of parameters bounded by an ellipse. It contains one maximum and one minimum, both absolute extremes.

VI. STUDY OF MHD DIAGRAMS FOR THE DIFFERENCE

The difference of the weighted Mahalanobis distances from a point $x$ of the parameters space to the means $\mu^A, \mu^B$ of two distributions of points,

$$D(x) = p_A d_A(x, \mu^A) - p_B d_B(x, \mu^B),$$

is also a useful concept, because the locus of the points for which this difference is null is the decision boundary for ascribing a point to the class $A$ or $B$.

Figure 5 shows the corresponding holodiagrams for the two simulated cases of Figure 2. Here the difference of the weighted Mahalanobis distances from each point of a fringe to the means $\mu^A$ and $\mu^B$ is constant, i.e. $D(x) = \text{constant}$.

Figure 6 shows the locus of points for which the difference is null. In Figure 6(a) this locus limits the two regions in black and white. The points in the black (or white) region should be classified in the class with mean $\mu^A (\mu^B$, respectively). Notice that one of these regions is a disconnected set, what is also a surprising fact derived from the properties of Mahalanobis distance. Figure 6(b) shows how this locus classifies the points of the initial points distributions.

VII. SENSITIVITY DEGREE VECTOR FIELD

In an MHD, for any point $P$, the fringe that contains it defines the direction where variations in the parameters do not improve or worsen the classification decision. If the point $P$ moves orthogonally to the fringes it changes the $S$ value in the most rapid manner. Then, the maximal sensitivity of $S$ or $D$ to changes in the parameters occurs in the direction orthogonal to the fringes, and the magnitude of that change is determined by how tightly packed the fringes are at that point.
Fig. 6. (a) Locus of points for which the difference $D$ is null. In black and white the two regions of decision. (b) The same locus with the corresponding points distribution.

Fig. 7. Diagrams of the sensitivity of the sum $S$ in (a), and difference $D$ in (b), through the gradient vector fields of $S$ and $D$, respectively.

The directions of the gradient vector fields of $S(x)$ and $D(x)$ give the directions of maximal variation of $S(x)$ and $D(x)$, respectively, and their modulus gives the speed of this variation. Therefore, the gradient vector fields of $S(x)$ and $D(x)$ are a measure of the sensitivity for classifications and decisions.

We have used the data of the distributions presented in Figure 2(a) to show, in Figures 7(a) and 7(b), a representation of the gradient vector fields for the sum and the difference, respectively.

VIII. CONCLUSIONS

We have shown how Mahalanobis distance and sum and difference Holodiagrams can be used to study some decision and classification problems. To this end, we have provided four examples illustrating how our method can be applied. The range of its potential use is highlighted by the diversity of these applications: the manufacture of commercial products; investment funds; the composition of pharmaceutical products; and ecology.

We have also simulated two sets of data from which we have calculated the corresponding Mahalanobis distances and their sum and differences Holodiagrams. By using these diagrams we have seen how it is possible to clearly visualize the information needed to study decision and classification problems in the proposed applications.