František Štulajter

Projection pursuit quadratic regression - the normal case

Aplikace matematiky, Vol. 33 (1988), No. 3, 204–212

PROJECTION PURSUIT QUADRATIC REGRESSION – – THE NORMAL CASE

FRANTISEK ŠTULAJTER

(Received July 6, 1987)

Summary. The model of quadratic regression is studied by means of the projection pursuit method. This method leads to a decomposition of the matrix of quadratic regression, which can be used for an estimation of this matrix from the data observed.

Keywords: quadratic regression, projection pursuit, gaussian random vector, quadratic estimators, conditional mean value.

AMS Classification: 62H12.

1. INTRODUCTION

The projection pursuit (PP) methods belong to those methods of analysis of data which have been developed for data analysis by computers. These methods are widely used in applications of some nonlinear statistical methods. The aim of this paper is to study the PP-method for the solution of problem of quadratic regression. The theory of quadratic statistics (estimators), in particular for normally distributed random vector, which has been developed recently, can also be used for the solution of the problem of quadratic regression by the PP-method. As we will see, under the assumption of normality, the exact mathematical model for this problem exists and provides a guide for the special methods of data analysis by computers.

2. PP-METHOD IN REGRESSION ANALYSIS

This part of the paper is based on the work of Huber [2]. Let us assume that $\mathbf{X} = (X_1, \ldots, X_d)'$ is a d-dimensional ($d \geqq 2$) random vector and let Y be a random variable. It is well known that $f(\mathbf{X}) = \mathsf{E}[Y \mid \mathbf{X}]$ – the conditional expectation of Y given $\mathbf{X}$ is the best (in the mean square error sense) nonlinear estimator of Y based on the random vector $\mathbf{X}$. The main statistical problem is that the function f defining

$\mathsf{E}[Y \mid \boldsymbol{X}]$ is a nonlinear function of d variables which depends on the distribution of the random vector $(\boldsymbol{X}', Y)'$. However, in statistical problems we do not know this distribution and consequently we do not know the function f.

The projection pursuit method in regression analysis is based on successive approximations of the random variable $f(\boldsymbol{X})$ by random variables which are computed from the projections of the random vector $\boldsymbol{X}$ on (one-dimensional) subspaces of R^d.

Let $\boldsymbol{a} \in R^d$ be a d-dimensional vector. The random variable $\boldsymbol{a}'\boldsymbol{X} = \sum_{i=1}^{d} a_i X_i$ can be regarded as a projection of the random vector $\boldsymbol{X}$ to the (one-dimensional) subspace of R^d generated by the vector $\boldsymbol{a}$. The random variable $\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}] = g_{\boldsymbol{a}}(\boldsymbol{a}'\boldsymbol{X})$, i.e. the conditional expectation of Y given $\boldsymbol{a}'\boldsymbol{X}$, is the best approximation of the random variable $\mathsf{E}[Y \mid \boldsymbol{X}]$ given the random variable $\boldsymbol{a}'\boldsymbol{X}$.

Let $\boldsymbol{a}_1 \in R^d$ be such a vector that

$$\mathsf{E}[f(\boldsymbol{X}) - \mathsf{E}[Y \mid \boldsymbol{a}_1'\boldsymbol{X}]]^2 = \min_{\boldsymbol{a} \in R^d} \mathsf{E}[f(\boldsymbol{X}) - g_{\boldsymbol{a}}(\boldsymbol{a}'\boldsymbol{X})]^2 .$$

Now, let $R_1 = Y - \mathsf{E}[Y \mid \boldsymbol{a}_1'\boldsymbol{X}] = Y - g_1(\boldsymbol{a}_1'\boldsymbol{X})$, where we have used the notation g_1 instead of $g_{\boldsymbol{a}_1}$, and let $f_1(\boldsymbol{X}) = \mathsf{E}[R_1 \mid \boldsymbol{X}]$. Then we have

$$f_1(\boldsymbol{X}) = \mathsf{E}[Y \mid \boldsymbol{X}] - \mathsf{E}[\mathsf{E}[Y \mid \boldsymbol{a}_1'\boldsymbol{X}] \mid \boldsymbol{X}] = f(\boldsymbol{X}) - g_1(\boldsymbol{a}_1'\boldsymbol{X}) .$$

The best approximation of $\mathsf{E}[R_1 \mid \boldsymbol{X}]$ based on the random variable $\boldsymbol{a}'\boldsymbol{X}$, $\boldsymbol{a} \in R^d$ is the random variable $\mathsf{E}[R_1 \mid \boldsymbol{a}_2'\boldsymbol{X}] = g_2(\boldsymbol{a}_2'\boldsymbol{X})$, which satisfies

$$\mathsf{E}[f_1(\boldsymbol{X}) - g_2(\boldsymbol{a}_2'\boldsymbol{X})]^2 = \min_{\boldsymbol{a} \in R^d} \mathsf{E}[f_1(\boldsymbol{X}) - \mathsf{E}[R_1 \mid \boldsymbol{a}'\boldsymbol{X}]]^2 .$$

We proceed further by analogy, and in the m-th step we get the random variable $R_m = R_{m-1} - \mathsf{E}[R_{m-1} \mid \boldsymbol{a}_m'\boldsymbol{X}] = R_{m-1} - g_m(\boldsymbol{a}_m'\boldsymbol{X}) = Y - \sum_{j=1}^{m} g_j(\boldsymbol{a}_j'\boldsymbol{X})$. Let $f_m(\boldsymbol{X}) = \mathsf{E}[R_m \mid \boldsymbol{X}]$. Then we have

$$f_m(\boldsymbol{X}) = f(\boldsymbol{X}) - \sum_{j=1}^{m} g_j(\boldsymbol{a}_j'\boldsymbol{X}) \quad \text{or} \quad f(\boldsymbol{X}) = f_m(\boldsymbol{X}) + \sum_{j=1}^{m} g_j(\boldsymbol{a}_j'\boldsymbol{X}) .$$

Under the assumption that for some m either

$$\mathsf{E}[f_m(\boldsymbol{X}) - g_{m+1}(\boldsymbol{a}_{m+1}'\boldsymbol{X})]^2 = 0$$

or

$$\mathsf{E}[R_m \mid \boldsymbol{X}] = \mathsf{E}[R_m \mid \boldsymbol{a}_{m+1}'\boldsymbol{X}] \quad \text{we can write} \quad f(\boldsymbol{X}) = \sum_{j=1}^{m+1} g_j(\boldsymbol{a}_j'\boldsymbol{X}) .$$

In this case $\mathsf{E}[Y \mid \boldsymbol{X}]$ can be expressed as the sum of the best approximations $g_j(\boldsymbol{a}_j'\boldsymbol{X})$, $j = 1, 2, \ldots, m + 1$ which are computed from the orthogonal projections $\boldsymbol{a}_j'\boldsymbol{X}$ of the vector $\boldsymbol{X}$.

We shall now apply the PP-method to the problem of quadratic regression. Let $\boldsymbol{X}$ be a d-dimensional $N_d(\boldsymbol{O}, \boldsymbol{\Sigma})$ (normally) distributed random vector with $\mathsf{E}[\boldsymbol{X}] = \boldsymbol{O}$ and with a regular covariance matrix $\boldsymbol{\Sigma}$, Let

$$Y = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} + \varepsilon = \sum_{i,j=1}^{d} X_i X_j A_{ij} + \varepsilon\,,$$

where $\boldsymbol{A}$ is a $d \times d$ symmetric matrix and ε is a random variable with $\mathsf{E}[\varepsilon] = 0$ and $D[\varepsilon] = \sigma^2$; let $\boldsymbol{X}$ and ε be independent. It is clear that under these assumptions $f(\boldsymbol{X}) = \mathsf{E}[Y \mid \boldsymbol{X}] = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$. We will show using the PP-method that the equality $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \sum_{j=1}^{d} g_j(\boldsymbol{a}_j'\boldsymbol{X})$ holds.

To find the functions g_j, $j = 1, 2, \ldots, d$ we have to compute $\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}]$ and $\mathsf{E}[R_j \mid \boldsymbol{a}'\boldsymbol{X}]$ for any $\boldsymbol{a} \in R^d$ and $j = 1, 2, \ldots, d - 1$. This can be easily done by virtue of the assumption that the vector $\boldsymbol{X}$ has the normal distribution.

It is well known (see [5], [6]) that if U is a $N(0, 1)$-distributed random variable, then any random variable V satisfies

$$\mathsf{E}[V \mid U] = \sum_{i=0}^{\infty} \mathsf{E}[V h_i(U)] \,.\, h_i(U)\,,$$

where h_i are Hermite's polynomials, for which we have the identities

$$\mathsf{E}[h_i(U)\, h_j(U)] = \int_{-\infty}^{\infty} h_i(u)\, h_j(u) \frac{1}{\sqrt{2\pi}} e^{(-1/2)u^2}\, \mathrm{d}u = \delta_{ij}\,, \quad i, j = 1, 2, \ldots$$

It is well known that $h_0(u) = 1$, $h_1(u) = u$, $h_2(u) = (1/\sqrt{2})(u^2 - 1), \ldots$. Using this result for the random variables $V = Y$ and

$$U = \frac{\boldsymbol{a}'\boldsymbol{X}}{\sqrt{D[\boldsymbol{a}'\boldsymbol{X}]}} = \frac{\boldsymbol{a}'\boldsymbol{X}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}}\,,$$

which has the N(0, 1) distribution, and taking into account the independence of $\boldsymbol{X}$ and ε, we get the equalities

$$\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}] = \sum_{i=0}^{\infty} \mathsf{E}\left[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} + \varepsilon) \,.\, h_i\left(\frac{(\boldsymbol{a}'\boldsymbol{X})}{\sqrt{D[\boldsymbol{a}'\boldsymbol{X}]}}\right)\right] \,.\, h_i\left(\frac{(\boldsymbol{a}'\boldsymbol{X})}{\sqrt{D[\boldsymbol{a}'\boldsymbol{X}]}}\right) =$$

$$= \sum_{i=0}^{\infty} \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \,.\, h_i(U)] \,.\, h_i(U)\,.$$

However, $\mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \,.\, h_i(U)] = \mathsf{E}[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} - \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}]) \,.\, h_i(U)] + \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] \,.\, \mathsf{E}[h_i(U)]$ is different from zero only for $i = 0$ and $i = 2$ if $U = \boldsymbol{a}'\boldsymbol{X}/\sqrt{D[\boldsymbol{a}'\boldsymbol{X}]}$ (see [6]) and thus we conclude

$$\mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \,.\, h_0(U)] = \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}]\,,$$

$$\mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \, . \, h_2(U)] = \mathsf{E}\left[(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} - \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}]) \, . \, \frac{1}{\sqrt{2}}\left(\frac{(\boldsymbol{a}'\boldsymbol{X})^2}{D[\boldsymbol{a}'\boldsymbol{X}]} - 1\right)\right] =$$

$$= \frac{1}{\sqrt{(2)}\, D[\boldsymbol{a}'\boldsymbol{X}]} \operatorname{Cov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \, ; \, (\boldsymbol{a}'\boldsymbol{X})^2) \, .$$

Using these results we obtain

$$\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}] = \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] + \tfrac{1}{2} \frac{\operatorname{Cov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}; (\boldsymbol{a}'\boldsymbol{X})^2)}{(D[\boldsymbol{a}'\boldsymbol{X}])^2} ((\boldsymbol{a}'\boldsymbol{X})^2 - D[\boldsymbol{a}'\boldsymbol{X}]) \, .$$

From the theory of quadratic estimation (see [3], [4], under the assumption that $\boldsymbol{X}$ is an $N_d(\boldsymbol{O}, \boldsymbol{\Sigma})$ distributed random vector, it is well known that $\mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] = \operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma})$ and $\operatorname{Cov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}; \boldsymbol{X}'B\boldsymbol{X}) = 2 \operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma})$ for any $d \times d$ symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, where tr denotes the trace of a matrix.

Hence we can write

$$2(D[\boldsymbol{a}'\boldsymbol{X}])^2 = 2(\boldsymbol{a}'\Sigma\boldsymbol{a})^2 = 2 \operatorname{tr}(\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\Sigma}) = D[\boldsymbol{X}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{X}] = D[(\boldsymbol{a}'\boldsymbol{X})^2] \, .$$

Using this result and the independence of $\boldsymbol{X}$ and ε we get *the final form for* $\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}]$:

$$\mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}] = \mathsf{E}[Y] + \frac{\operatorname{Cov}(Y; (\boldsymbol{a}'\boldsymbol{X})^2)}{D[(\boldsymbol{a}'\boldsymbol{X})^2]} ((\boldsymbol{a}'\boldsymbol{X})^2 - \mathsf{E}[(\boldsymbol{a}'\boldsymbol{X})^2]) \, .$$

Further, we have

$$\mathsf{E}[\mathsf{E}[Y \mid \boldsymbol{X}] - \mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}]]^2 = D[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] - \frac{(\operatorname{Cov}(Y; (\boldsymbol{a}'\boldsymbol{X})^2))^2}{D[(\boldsymbol{a}'\boldsymbol{X})^2]} =$$

$$= D\,[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}] - 2\left(\frac{\boldsymbol{a}'\Sigma\boldsymbol{A}\Sigma\boldsymbol{a}}{\boldsymbol{a}'\Sigma\boldsymbol{a}}\right)^2 ,$$

where we have used the equalities $\operatorname{Cov}(Y; (\boldsymbol{a}'\boldsymbol{X})^2) = \operatorname{Cov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}; \boldsymbol{X}'\boldsymbol{a}\boldsymbol{a}'\boldsymbol{X}) = 2 \operatorname{tr}(\boldsymbol{A}\Sigma\boldsymbol{a}\boldsymbol{a}'\boldsymbol{\Sigma}) = 2(\boldsymbol{a}'\Sigma\boldsymbol{A}\Sigma\boldsymbol{a})$.

From this expression we can see that the problem of finding

$$\boldsymbol{a}_1 = \arg\min_{\boldsymbol{a}} \mathsf{E}[\mathsf{E}[Y \mid \boldsymbol{X}] - \mathsf{E}[Y \mid \boldsymbol{a}'\boldsymbol{X}]^2 =$$

$$= \arg\min_{\boldsymbol{a}} \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} - \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \mid \boldsymbol{a}'\boldsymbol{X}]]^2 ,$$

which is the first step of the PP-method in the regression analysis, coincides with the problem of finding

$$\arg\max_{\boldsymbol{a} \in R^d} \left(\frac{\boldsymbol{a}'\Sigma\boldsymbol{A}\Sigma\boldsymbol{a}}{\boldsymbol{a}'\Sigma\boldsymbol{a}}\right)^2 .$$

Let us denote $\Sigma^{1/2}\boldsymbol{a} = \boldsymbol{b}$. Then we have

$$\max_{\boldsymbol{a} \in R^d} \left(\frac{\boldsymbol{a}'\Sigma\boldsymbol{A}\Sigma\boldsymbol{a}}{\boldsymbol{a}'\Sigma\boldsymbol{a}}\right)^2 = \max_{\boldsymbol{b} \in R^d} \left(\frac{\boldsymbol{b}'\Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2}\boldsymbol{b}}{\|\boldsymbol{b}\|^2}\right) = \max_{\|\boldsymbol{b}\|^2 = 1} (\boldsymbol{b}'\Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2}\boldsymbol{b})^2 \, .$$

The solution of the problem of finding $\arg\max_{\|b\|^2=1} (b'\Sigma^{1/2}A\Sigma^{1/2}b)^2$ is now very simple. Let us assume that $\{\lambda_j\}_{j=1}^d$ are the eigenvalues and $\{b_j\}_{j=1}^d$ the orthonormal eigenvectors of the symmetric matrix $\Sigma^{1/2}A\Sigma^{1/2}$, and let $|\lambda_1| \geqq |\lambda_2| \geqq \ldots \geqq |\lambda_d|$. Then we have

$$\Sigma^{1/2}A\Sigma^{1/2} = \sum_{j=1}^d \lambda_j b_j b_j' , \quad \max_{\|b\|^2=1} (b'\Sigma^{1/2}A\Sigma^{1/2}b)^2 = \lambda_1^2 ,$$

$$\arg\max_{\|b\|^2=1} (b'\Sigma^{1/2}A\Sigma^{1/2}b)^2 = b_1 ,$$

which implies

$$a_1 = \arg\max_{a\in R^d} \left(\frac{a'\Sigma A\Sigma a}{a'\Sigma a}\right)^2 = \Sigma^{-1/2}b_1$$

and

$$\mathsf{E}[\mathsf{E}[Y | X] - \mathsf{E}[Y | a_1'X]]^2 = D[X'AX] - 2\lambda_1^2 =$$
$$= 2\,\mathrm{tr}\,((\Sigma^{1/2}\, A\, \Sigma^{1/2})^2) - \lambda_1^2 = 2\sum_{j=2}^d \lambda_j^2 .$$

The next step of the PP-method is to consider the random variable $R_1 = Y - \mathsf{E}[Y | a_1'X]$.

We can write

$$\mathsf{E}[Y | a_1'X] = \mathsf{E}[Y] + \lambda_1((a_1'X)^2 - \mathsf{E}[(a_1'X)^2]) ,$$

since

$$\frac{\mathrm{Cov}\,(Y; (a_1'X)^2)}{D[(a_1'X)^2]} = \frac{a_1'\Sigma A\Sigma a_1}{(a_1'\Sigma a_1)^2} = b_1'\Sigma^{1/2}A\Sigma^{1/2}b_1 = \lambda_1 .$$

Next, we have

$$R_1 = X'AX - \lambda_1 \,.\, X'a_1a_1'X - (\mathsf{E}[Y] - \lambda_1\, \mathsf{E}[(a_1'X)^2]) + \varepsilon =$$
$$= X'(A - \lambda_1 a_1 a_1')\, X - (\mathsf{E}[Y] - \lambda_1\, \mathsf{E}[(a_1'X)^2]) + \varepsilon ,$$

where $\mathsf{E}[Y] = \mathrm{tr}\,(A\Sigma) = \mathrm{tr}\,(\Sigma^{1/2}A\Sigma^{1/2}) = \sum_{j=1}^d \lambda_j$, $\lambda_1\, \mathsf{E}[(a_1'X)^2] = \lambda_1(a_1'\Sigma a_1) = \lambda_1$ and thus

$$\mathsf{E}[Y] - \lambda_1\, \mathsf{E}[(a_1'X)^2] = \sum_{j=2}^d \lambda_j = \mathrm{tr}\,(\Sigma^{1/2}A_1\Sigma^{1/2}) = \mathrm{tr}\,(A_1\Sigma) = \mathsf{E}[X'A_1X] ,$$

where $A_1 = A - \lambda_1 a_1 a_1'$.

Thus we have shown that

$$R_1 = X'A_1X - \mathsf{E}[X'A_1X] + \varepsilon ; \quad A_1 = A - \lambda_1 a_1 a_1' .$$

Making use of this expression for R_1 we can write

$$\mathsf{E}[R_1 | X] - \mathsf{E}[R_1 | a'X] = X'A_1X - \mathsf{E}[X'A_1X | a'X] ,$$

from which it can be seen that the second step of the PP-method, which consists in finding the vector

$$\boldsymbol{a}_2 = \arg\min_{\boldsymbol{a}\in R^d} \mathsf{E}[\mathsf{E}[R_1 \mid \boldsymbol{X}] - \mathsf{E}[R_1 \mid \boldsymbol{a}'\boldsymbol{X}]]^2 =$$
$$= \arg\min_{\boldsymbol{a}\in R^d} \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}_1\boldsymbol{X} - \mathsf{E}[\boldsymbol{X}'\boldsymbol{A}_1\boldsymbol{X} \mid \boldsymbol{a}'\boldsymbol{X}]]^2,$$

coincides with the first step with the matrix $\boldsymbol{A}_1$ instead of the matrix $\boldsymbol{A}$. We must find the eigenvalues and the eigenvectors of the matrix $\Sigma^{1/2}\boldsymbol{A}_1\Sigma^{1/2}$. However,

$$\Sigma^{1/2}\boldsymbol{A}_1\Sigma^{1/2} = \Sigma^{1/2}(\boldsymbol{A} - \lambda_1\boldsymbol{a}_1\boldsymbol{a}_1')\,\Sigma^{1/2} = \Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2} - \lambda_1\Sigma^{1/2}\boldsymbol{a}_1\boldsymbol{a}_1'\Sigma^{1/2} =$$
$$= \sum_{j=1}^{d}\lambda_j\boldsymbol{b}_j\boldsymbol{b}_1' - \lambda_1\boldsymbol{b}_1\boldsymbol{b}_1' = \sum_{j=2}^{d}\lambda_j\boldsymbol{b}_j\boldsymbol{b}_j',$$

from which it is clear that $\lambda_2, \ldots, \lambda_d$ ($|\lambda_2| \geqq |\lambda_3| \geqq \ldots \geqq |\lambda_d|$) are the eigenvalues and $b_2, \ldots, b_d$ are the eigenvectors of the matrix $\boldsymbol{A}_1$, and we conclude

$$\boldsymbol{a}_2 = \arg\min_{\boldsymbol{a}} \mathsf{E}[\mathsf{E}[R_1 \mid \boldsymbol{X}] - \mathsf{E}[R_1 \mid \boldsymbol{a}'\boldsymbol{X}]]^2 = \Sigma^{-1/2}\boldsymbol{b}_2.$$

We proceed by analogy in the subsequent steps. It is clear that the PP-method of quadratic regression has only d steps. The equality

$$\Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2} = \sum_{j=1}^{d}\lambda_j\boldsymbol{b}_j\boldsymbol{b}_j'$$

yields the following expression for $\boldsymbol{A}$:

$$\boldsymbol{A} = \sum_{j=1}^{d}\lambda_j\Sigma^{-1/2}\boldsymbol{b}_j\boldsymbol{b}_j'\Sigma^{-1/2} = \sum_{j=1}^{d}\lambda_j\boldsymbol{a}_j\boldsymbol{a}_j',$$

wheter the vectors $\boldsymbol{a}_j = \Sigma^{-1/2}\boldsymbol{b}_j$ satisfy the equalities $\boldsymbol{a}_j'\Sigma\boldsymbol{a}_j = 1$.

Example. Let $d = 2$,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}.$$

The matrix $\boldsymbol{A}$ has the eigenvalues $\alpha_1 = 4$ and $\alpha_2 = 2$, and the orthonormal eigenvectors

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We can write $\boldsymbol{A} = \sum_{j=1}^{2}\alpha_j\boldsymbol{v}_j\boldsymbol{v}_j'$, $\boldsymbol{v}_i'\boldsymbol{v}_j = \delta_{ij}$, $i, j = 1, 2$.

Let

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

Then

$$\Sigma^{1/2} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2} = \begin{pmatrix} 1 & 6 \\ 6 & 4 \end{pmatrix}.$$

This matrix has the eigenvalues $\lambda_1 = 8{\cdot}6847$ and $\lambda_2 = -3{\cdot}6847$ and the orthonormal eigenvectors

$$\boldsymbol{b}_1 = \begin{pmatrix} 0{\cdot}6150 \\ 0{\cdot}7885 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}_2 = \begin{pmatrix} 0{\cdot}7885 \\ -0{\cdot}6150 \end{pmatrix}.$$

We can easily compute

$$\boldsymbol{a}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{b}_1 = \begin{pmatrix} 0{\cdot}6150 \\ 0{\cdot}3942 \end{pmatrix} \quad \text{and} \quad \boldsymbol{a}_2 = \begin{pmatrix} 0{\cdot}7885 \\ -0{\cdot}3075 \end{pmatrix}.$$

We can see that $\lambda_i \neq \alpha_i$, $\boldsymbol{a}_i \neq \boldsymbol{v}_i$, $i = 1, 2$, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are not orthogonal, but

$$\boldsymbol{A} = \sum_{i=1}^{2} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i'.$$

4. APPLICATION OF THE PP-METHOD OF QUADRATIC REGRESSION TO REAL DATA

Let us assume that $(\mathbf{x}_i', y_i)$, $i = 1, 2, \ldots, n$ are independent observations of a $(d + 1)$-dimensional random vector $(\boldsymbol{X}', Y)'$. Let

$$Y = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} + \varepsilon,$$

where $\boldsymbol{A}$ is an unknown symmetric $d \times d$ matrix, $\boldsymbol{X}$ is an $N_d\ (\boldsymbol{O}, \boldsymbol{\Sigma})$ distributed random vector with $\boldsymbol{\Sigma} > 0$, ε is a random variable with $\mathsf{E}[\varepsilon] = 0$, $D[\varepsilon] = \sigma^2$, $\boldsymbol{X}$ and ε are independent. According to the results of the previous section we have $\boldsymbol{A} = \sum_{j=1}^{d} \lambda_j \boldsymbol{a}_j \boldsymbol{a}_j'$, where $\boldsymbol{a}_j'\boldsymbol{\Sigma}\boldsymbol{a}_j = 1$, $j = 1, 2, \ldots, d$. By the PP-method we can estimate the numbers λ_j and the vectors $\boldsymbol{a}_j$, $j = 1, 2, \ldots, d$ from the data $(\mathbf{x}_i', y_i)'$, $i = 1, 2, \ldots, n$ in the following way:

Let

$$\boldsymbol{S} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$$

and let $\boldsymbol{a} \in R^d$ be a vector for which the equality $\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a} = 1$ holds. Let us compute the data $(\boldsymbol{a}'\mathbf{x}_i)^2$, $i = 1, 2, \ldots, n$. We approximate the data y_i, $i = 1, 2, \ldots, n$ by the usual regression line with respect to the data $(\boldsymbol{a}'\mathbf{x}_i)^2$, $i = 1, 2, \ldots, n$. Using a computer we find such a vector ${}^\wedge\boldsymbol{a}_1$ with ${}^\wedge\boldsymbol{a}_1'\boldsymbol{S}\,{}^\wedge\boldsymbol{a}_1 = 1$ and a real number ${}^\wedge\lambda_1$ that the equality

$$\sum_{i=1}^{n} [y_i - \bar{y} - {}^\wedge\lambda_1(({}^\wedge\boldsymbol{a}_1'\mathbf{x}_i)^2 - ({}^-\boldsymbol{a}_1'\mathbf{x})^2)]^2 =$$

$$= \min_{\{\boldsymbol{a}:\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}=1\}} \sum_{i=1}^{n} [y_i - \bar{y} - {}^\wedge\lambda_{\boldsymbol{a}}((\boldsymbol{a}'\mathbf{x}_i)^2 - ({}^-\boldsymbol{a}'\mathbf{x})^2)]^2 \quad \text{holds.}$$

Here

$${}^\wedge\lambda_{\boldsymbol{a}} = \frac{\sum_{i=1}^{n} (y_i - \bar{y}) \,.\, ((\boldsymbol{a}'\mathbf{x}_i)^2 - ({}^-\boldsymbol{a}'\mathbf{x})^2)}{\sum_{i=1}^{n} ((\boldsymbol{a}'\mathbf{x}_i)^2 - ({}^-\boldsymbol{a}'\mathbf{x})^2)^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

and

$$(\bar{\mathbf{a}'\mathbf{x}})^2 = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{a}'\mathbf{x}_i)^2 .$$

Then we repeat the same procedure with the data $(\mathbf{x}'_i, r_i^{(1)})'$, $i = 1, 2, \ldots, n$, where $r_i^{(1)} = y_i - \bar{y} - \hat{\lambda}_1((\mathbf{a}'_1\mathbf{x}_i)^2 - (\bar{\mathbf{a}'_1\mathbf{x}})^2)$, $i = 1, 2, \ldots, n$. We get the vector $\hat{\mathbf{a}}_2$ and the number $\hat{\lambda}_2$ and the data

$$r_i^{(2)} = r_i^{(1)} - \bar{r}^{(1)} - \hat{\lambda}_2((\mathbf{a}'_2\mathbf{x}_i)^2 - (\bar{\mathbf{a}'_2\mathbf{x}})^2), \quad i = 1, 2, \ldots, n .$$

These data together with $\mathbf{x}_i$, $i = 1, 2, \ldots, n$ are used in the next step of the PP-method.

After performing d steps we get the estimators $\hat{\lambda}_1, \ldots, \hat{\lambda}_d$ of $\lambda_1, \ldots, \lambda_d$ and $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_d$ of $\mathbf{a}_1, \ldots, \mathbf{a}_d$. Using these estimators we can construct the estimate

$$\hat{\mathbf{A}} = \sum_{i=1}^{d} \hat{\lambda}_i \, \hat{\mathbf{a}}_i \, \hat{\mathbf{a}}'_i$$

of the unknown matrix $\mathbf{A}$. In real situations the random variable $\mathbf{X}' \hat{\mathbf{A}}\mathbf{X}$ can be used as the first (quadratic) approximation of the unknown nonlinear dependence of Y on $\mathbf{X}$.

References

[1] *J. H. Friedman, W. Stuetzle:* Projection pursuit regression. J. Amer. Stat. Assoc. 76 (1981), 817—823.

[2] *P. J. Huber:* Projection pursuit. Ann. Statist. 13 (1985), 435—475.

[3] *L. R. La Motte:* Quadratic estimation of variance components. Biometrika 29 (1973), 311 to 330.

[4] *C. R. Rao:* Estimation of variance and covariance components — MINQUE theory. J. Mult. Analysis 1, 1971, 445—456.

[5] J. Rozanov: Gaussian random processes (Russian). Nauka, Moskva 1970.

[6] *F. Štulajter:* RKHS approach to nonlinear estimation of random variables. Trans. VIII-th Prague Conf. Volume B (1978), 239—246.

Súhrn

METÓDY ANALÝZY PROJEKCIÍ V KVADRATICKEJ REGRESII ZA PREDPOKLADOV NORMALITY

FRANTIŠEK ŠTULAJTER

V článku je študovaný model kvadratickej regresie metódou analýzy projekcií. Pomocou tejto metódy je odvodený rozklad pre maticu kvadratickej regresie. Rozklad je použitý pre odhad tejto matice z pozorovaných dát.

Резюме

МЕТОД АНАЛИЗА ПРОЕКЦИЙ В КВАДРАТИЧЕСКОЙ РЕГРЕСИИ ДЛЯ ГАУССОВСКОГО СЛУЧАЯ

FRANTISEK ŠTULAJTER

В статье изучается модель квадратической регрессии на основе метода анализа проекций. С помощью этого метода найдено разложение для матрицы квадратической регрессии, которое позволяет дать оценку этой матрицы, исходя из наблюдаемых данных.

Author's address: Doc. RNDr. *František Štulajter*, CSc., Katedra teórie pravdepod. a mat. štat. MFF UK, Mlynská dolina, 842 16 Bratislva.