УДК 519.6 R.A.Eleouva, N.B.Zakariyanova, M.K.Bekbaeva

APPLICATION OF THE EIGENVALUES AND EIGENVECTORS OF MATRICES WITH STATISTICAL DATA PROCESSING

The article discusses a useful observation, which has a clear meaning and is useful for statistical data processing. Them arterials presented without superfluous mathematical wisdoms and is available to economists, sociologists and specialists in other fields who use statistical methods.

In the statistical analysis of the data table, which consists of several attributes, it is necessary to bear in mind the effect of a substantial multidimensionality, because of which we can come to a faithful conclusion while taking into account the entire set of interrelated features. For example, attempt to distinguish between two types of consumer behaviour first on the one feature (the cost of food), then on the other (the cost of manufactured goods and services) did not give the result, while the count of both signs at the same time allowed to detect significant difference between the analyzed sets of families.

If the number of signs - a sufficiently large number, then the partition of the set of the objects in compact groups (the so-called clusters) can be challenging. This is the problem of classification or cluster - analysis. Once objects are grouped into homogeneous groups (classes), the problem arises of studying the relationship attributes within a single class. If a homogeneous group forms a "cloud" of elliptic type, then the methods of correlation analysis is used. When objects are placed in the vicinity of a curve, it is necessary to apply techniques of regression analysis.

The theory of the eigenvectors of matrices and their use in correlation analysis.

Assume that each of n objects described by к attributes (height, weight, length of skull, length and width of the upper jaw, and so on), and imagine the data's for a particular class of objects in a table form A = $X = \|x_{il}\|_{n\times k}$. Compute for each sign a mean value $\bar{x}_l = \frac{1}{n}\sum_{i=1}^{n} x_{il}$ and centres the data: $y_{il} = x_{il} - \bar{x}_l$. Then $\bar{y}_l = 0$, $l = 1, \dots, k$. Denote with $S = \|c_{il}\|_{k\times k}$ selective covariance matrix of features: $S = \frac{1}{n} Y^T Y$, that is c_{il} - selective covariance of the i-th and l-th column of the matrix Y. The fact that the covariance matrix $S = \|c_{il}\|_{k\times k}$, is nonnegative definite matrix, in other words, a self-ad joint matrix should be reducible to diagonal form. Consequently, there is an orthogonal matrix U, which leading S to the principal axes: $U^T S U = \Lambda$. Here is Λ - diagonal matrix with nonnegative elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ on the main diagonal, which are the roots of the equation $\det(S - \lambda \ E) = 0$. They are called the Eigen values of the matrix S Assumethatall $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ positive and different. For Experimental data this condition is almost always performed. Note also that the columnsu_1 , u_2 ,...,u_k of the matrix$U$ represent the principal axes and determined uniquely up to choice of direction of the axis. They form an orthonormal basis in R^n, which has important properties.

1. The projections of objects on the first principal axis u_1, have the largest selective

dispersion among the all projections for all sorts of directions in space R^n, and this maximum is equal to λ_1 .

2. The projections of objects on the second main axis u_2, have the largest selective variance of the projections for all sorts of directions in space R^n, which are orthogonal to the vecton u_1 . Moreover, this maximum is equal toλ_2 .
3. The sum of the selective variances of the initial features $trS = \frac{1}{n}\sum_{l=1}^{k} c_{ll}$ to the similarity of matrixes$S$ d $\wedge$ ise qualto $\wedge = \lambda_1 + \cdots + \lambda_k$, i.e. the sum of sample variances of the projections of objects on the main axis.

4. This value can be regarded as a measure of the total scatter of the objects about their mass center. Interesting is there lative proportion of variation which is attributable to the first l main axes,

$$\gamma_1 = \frac{\lambda_1 + \cdots + \lambda_l}{\lambda_1 + \cdots + \lambda_k}$$

If this value for some /is sufficiently close to 1, then the dimension reduce of space of attribute sis possible due to the transition from κ the initial features to the l new features. Practice it is often possible to limit byt woor three components with out significant loss of information.

An example of use of the Eigen vectors of the matrix in correlation analysis.

The table shows the size of the j aws and teeth of thirty dogs (numbers 1 - 30), twelve wolves (numbers 31 - 42) and the fossil skull of unknown animal (No. 43), found in the quarter nary layer (according to De Bonis [1]). The figure shows measurable characteristics: 1 - the length of the skull, 2 - the length of the upper jaw, 3 - the width of the upper jaw, the following measure ments belong to the teeth: 4 - Length of the upper Carnivore, 5 - length offirst upper molars, 6 - the width of the upper first molar. Need to find, to which of the classes (dogs or wolves) should be included an unknown animal.

Here we will deal with more modest goal: find and interpret the main components for this example.

The algorithm for determining the principal axes.

1. Ine achcolumn of the table we find the mean value.
2. From the column we sub tract the found average. There suit will be denoted by Table 2.
3. Then drawup a new Table 3 from the square soft he elements in Table 2. There suit will be donated by Table 3.
4. Ine achcolumn of Table 3 we find the new average.
5. The columns of Table 2 we divide by the square roots of the corresponding mean step 4. There suit is presented in the form of Table 4.
6. able 4 is an oblong matrix (rows 43, columns 6). Multiply it by its trans poses as obtain a matrix of dimension 6 by 6.
7. There suit of step 6 divide to 43. See Table 7.

Table 1						
	1	2	3	4	5	6
1	129	64	95	17,5	11,2	13,8
2	154	74	76	20	14,2	16,5
3	170	87	71	17,9	12,3	15,9
4	188	94	73	19,5	13,3	14,8
5	161	81	55	17,1	12,1	13
6	164	90	58	17,5	12,7	14,7
7	203	109	65	20,7	14	16,8
8	178	97	57	17,3	12,8	14,3
9	212	114	65	20,5	14,3	15,5
10	221	123	62	21,2	15,2	17
11	183	97	52	19,3	12,9	13,5
12	212	112	65	19,7	14,2	16
13	220	117	70	19,8	14,3	15,6
14	216	113	72	20,5	14,4	17,7
15	216	112	75	19,6	14	16,4
16	205	110	68	20,8	14,1	16,4
17	228	122	78	22,5	14,2	17,8
18	218	112	65	20,3	13,9	17
19	190	93	78	19,7	132	14
20	212	111	73	20,5	13,7	16,6
21	201	405	70	19,8	14,3	15,9
22	196	106	67	18,5	12,6	14,2
23	158	71	71	16,7	12,5	13,3
24	255	126	86	21,4	15	18
25	234	113	83	21,3	14,8	17
26	205	105	70	19	12,4	14,9
27	186	97	62	19	13,2	14,2
28	241	119	87	21	14,7	18,3
29	220	111	88	22,5	15,4	18
30	242	120	85	19,9	15,3	17,6
31	199	105	73	23,4	15	19,1
32	227	117	77	25	15,3	18,6
33	228	122	82	24,7	15	18,5

34	232	123	83	25,3	16,8	15,5
35	231	121	78	23,5	16,5	19,6
36	215	118	74	25,7	15,7	19
37	184	100	69	23,3	15,8	19,7
38	175	94	73	22,2	14,8	17
39	239	124	77	25	16,8	27
40	203	109	70	23,3	15	18,7
41	226	118	72	26	16	19,4
42	226	119	77	26,5	16,8	19,3
43	210	103	72	20,5	14	16,7
average	204,9535	106,4651	72,53488	21,05581	17,05814	16,8093
Table 4						
	1	2	3	4	5	6
1	-2,81171	-2,86441	2,491943	-1,3857	-0,32938	-1,23658
2	-1,88624	-2,18987	0,384368	-0,41145	-0,1607	-0,1271
3	-1,29394	-1,31298	-0,17026	-1,22982	-0,26753	-0,37365
4	-0,6276	-0,84081	0,051593	-0,6063	-0,21131	-0,82566
5	-1,62711	-1,7177	-1,94506	-1,54158	-0,27878	-1,56532
6	-1,51605	-1,11062	-1,61228	-1,3857	-0,24504	-0,86675
7	-0,07232	0,170986	-0,83581	-0,13866	-0,17195	-0,00382
8	-0,99779	-0,63845	-1,72321	-1,46364	-0,23942	-1,03112
9	0,260853	0,508252	-0,83581	-0,2166	-0,15508	-0,53802
10	0,594022	1,11533	-1,16858	0,056189	-0,10448	0,078361
11	-0,81269	-0,63845	-2,27783	-0,68424	-0,2338	-1,35986
12	0,260853	0,373345	-0,83581	-0,52836	-0,1607	-0,33256
13	0,557004	0,710611	-0,28118	-0,48939	-0,15508	-0,49693
14	0,408929	0,440799	-0,05933	-0,2166	-0,14946	0,366005
15	0,408929	0,373345	0,273443	-0,56733	-0,17195	-0,16819
16	0,001722	0,238439	-0,50303	-0,09969	-0,16633	-0,16819
17	0,853154	1,047877	0,606218	0,562797	-0,1607	0,407097
18	0,482966	0,373345	-0,83581	-0,29454	-0,17757	0,078361
19	-0,55356	-0,90826	0,606218	-0,52836	6,462765	-1,1544
20	0,260853	0,305892	0,051593	-0,2166	-0,18882	-0,08601
21	-0,14635	-0,09883	-0,28118	-0,48939 j-0,15508	-0,37365	

22	-0,33145	-0,03137	-0,61396	-0,996	-0,25067	-1,07221
23	-1,73816	-2,39223 | -0,17026	-1,69746	-0,25629	-1,44204	
24	1,852661	1,31769	1,493618	0,134129	-0,11572	0,489281
25	1,075267	0,440799	1,160843	0,095159	-0,12697	0,078361
26	0,001722	-0,09883	-0,28118	-0,80115	-0,26191	-0,78457
27	-0,70164	-0,63845	-1,16858	-0,80115	-0,21693	-1,07221
28	1,334398	0,845518	1,604543	-0,02175	-0,13259	0,612557
29	0,557004	0,305892	1,715468	0,562797	-0,09323	0,489281
30	1,371417	0,912971	1,382693	-0,45042	-0,09885	0,324913
31	-0,22039	-0,09883	0,051593	0,913526	-0,11572	0,941293
32	0,816135	0,710611	0,495293	1,537044	-0,09885	0,735833
33	0,853154	1,047877	1,049918	1,420135	-0,11572	0,694741
34	1,001229	1,11533	1,160843	1,653954	-0,01451	-0,53802
35	0,96421	0,980424	0,606218	0,952496	-0,03138	1,146753
36	0,37191	0,778064	0,162518	1,809833	-0,07636	0,900201
37	-0,77567	-0,43609	-0,39211	0,874556	-0,07074	1,187845
38	-1,10884	-0,84081	0,051593	0,445888	-0,12697	0,078361
39	1,260361	1,182783	0,495293	1,537044	-0,01451	4,187559
40	-0,07232	0,170986	-0,28118	0,874556	-0,11572	0,776925
41	0,779116	0,778064	-0,05933	1,926743	-0,0595	1,064569
42	0,779116	0,845518	0,495293	2,121592	-0,01451	1,023477
43	0,186816	-0,23373	-0,05933	-0,2166	-0,17195	-0,04491
Table 7						
1	0,958741	0,348183	0,612949	-0,032121	0,587251	
0,958741	1	0,200333	0,661002	-0,085869	0,594653	
0,348183	0,200333		0,369962	0,120454	0,354777	
0,612949	0,661002	0,369962	1	-0,015032	0,762643	
-0,03212	-0,085869	0,120454	-0,015032	1	-0,120108	
0,587251	0,594653	0,354777	0,762643	-0,120108	1	

Tables 4 and 7 were calculated on the popular program on the use of the spread sheet Microsoft Excel. The Eigen vectors and Eigen values of the matrix, shown in Table7, were calculated using variation methods. In the dissertation work [2], we proposed different algorithms for computing Eigen values and Eigen vectors of matrices on the base of variation method. In work[3], these methods were applied to some problems in the economy. In this paper, we propose the use of these algorithms to some problems of statistical data.

$$\vec{c}_1 = \begin{bmatrix} 0.43 \\ 0.43 \\ 0.23 \\ 0.44 \\ 0.46 \\ 0.42 \end{bmatrix}, \vec{c}_2 = \begin{bmatrix} 0.23 \\ 0.38 \\ -0.89 \\ -0.07 \\ 0.02 \\ -0.10 \end{bmatrix}, \vec{c}_3 = \begin{bmatrix} 0.53 \\ 0.39 \\ 0.38 \\ -0.40 \\ -0.27 \\ -0.44 \end{bmatrix}, \vec{c}_4 = \begin{bmatrix} 0.11 \\ 0.01 \\ -0.02 \\ -0.52 \\ -0.31 \\ 0.72 \end{bmatrix}, \vec{c}_5 = \begin{bmatrix} 0.05 \\ -0.20 \\ 0.00 \\ -0.58 \\ 0.78 \\ -0.09 \end{bmatrix}, \vec{c}_6 = \begin{bmatrix} 0.68 \\ -0.69 \\ -0.13 \\ 0.18 \\ -0.09 \\ -0.01 \end{bmatrix}$$

$$\lambda_1 = 4.100, \lambda_2 = 0.883, \lambda_3 = 0.639, \lambda_4 = 0.259, \lambda_5 = 0.097, \lambda_6 = 0.022$$

Trace of a matrixes qualto 6, here with the first eigenvalue is 68.3% of the trace, the amount of the first two eigenvalues is 83.0%, the amount of the first three eigen values is 93.7%.

Discussion and inter pretention of obtained results. To the first three components account 93.7% of the total variance of the "cloud." In this case the first component of the overall size makes sense. It follows from the fact that all components of c, have one sign and approximately identical in size, that is, during the designing on this axis coordinates of the normalized features are added.

These cond component is main ly responsible for the width of the upper jaw(characteristic 3) because the third coordinate off, with absolute values qualto 0.89(almost 1), and these cond - 0.38. Since the sign soft hese coordinates are different, these symptoms reflect a difference in the proportions of the jaws, and are distinguished from the elongated shape to truncated (hound sand collies from bull dog sand boxers). These condand third features of wolves and German Shepherds are almost identical. The third axis contrasts the size ofjawsizeof the teeth: the first three coordinates off, are approximately equal to the sum of the last three unsigned, but opposite in sign. This axis allows to distinguish between animals with developed teeth (wolf, German Shepherds, Dobermans) from dogs of other breeds (St. Bernard’s, Setters).

Conclusion

The above method of main components can be used in a variety of problems where appear symmetric matrices. For example, when the initial in formation about the objects serve as expert information about the differences between them, which expressed in numbers.

REFERENCES

1. Zhambyu M. Hierarchical cluster analysis and compliance. -M.: finance and statistics. 1988

2. Eleuov A. A., Otelbaev A. J., Akzhalova A.Zh., Rysbajuly B. Calculating eigenvalues and eigenvectors of matrices. //Eurasian mathematical journal ENU. L.n. Gumilev

and Moscow State University. Lomonosov. Astana, 2005- №l-57-78page.

Eleuov A.A. The algorithms of the count of eigenvalues and eigenvectors of matrices. // Vestnik KazNPU named after Abay. The seria of physics, mathematic, computer science. - 2007. - №1(17). -23-28 page