CREATIVE MATH.
14 (2005), 71 - 73

A better grading of the student's performance using the first principal component of the covariance matrix of the scores

Cristian Marinoiu and Iuliana Dobre

Abstract. In this paper we describe an application of the first principal component of the empirical covariance matrix of the scores in educational testing theory. An algorithm for better grading the students' performance is implemented in Matlab and the results are provided.

1. Introduction

The average of the scores of the administrated tests is usually used as a means of grading the students' performance. This is equivalent to assigning equal positive weights $1/n$ for each score of the n tests.

Another remarkable number, that can better summarize each student's performance, is obtained as a linear combination of the scores, where the weights are the values of the first principal component of the empirical covariance matrix of the scores.

2. The first principal component of the matrix of the scores

Suppose that m students each took n tests in different subjects. Let us denote by $X = (X_1, X_2, ..., X_n)$ a random vector, where the component X_j represents the score for the subject j, $j = 1, 2, ..., n$. Also, let us denote by $(x_{i1}, x_{i2}, ..., x_{in})$ the obtained scores by the student i, $i = 1, 2, ..., m$ for the considered subjects. Then, the empirical covariance matrix of the random vector X is $V = (V_{ij})\ \ 1 \leqslant i, j \leqslant n$, where

$$V_{ij} = \frac{1}{m}\sum_{k=1}^{m}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

$$\bar{x}_i = \frac{1}{m}\sum_{k=1}^{m} x_{ki}$$

Since the matrix V is symmetric and positive definite, it has n positive eigenvalues [2]. The eigenvector corresponding to the largest eigenvalue is called the first principal component of V. The remarkable property of the first principal component is that the projection of X on this eigenvector has the maximal variance equal to the largest eigenvalue. Thus, if $V^1 = (v_{11}, v_{12}, ..., v_{1n})$ is the first principal component of V, then the linear combination

Received: 10.05.2004. In revised form: 20.05.2005.
2000 *Mathematics Subject Classification.* 97B40.
Key words and phrases. *Educational testing theory, eigenvalues.*

$$Z_i = \sum_{j=1}^{n} v_{1j} x_{ij} \quad i = 1,\, 2,\, ...,\, m$$

is the best in the sense that it has the maximum variance among all possible linear combinations of the obtained scores by the student i, $i = 1, 2, ..., m$ [3]. Since Z_i $i = 1,\, 2,\, ...,\, m$ captures the largest amount of variation of the scores, using Z_i for grading, the very grading of the students' performance is more realistic.

3. The algorithm

The steps to obtain the desired results are as follows:

1. Read the scores $(x_{i1},\, x_{i2},\, \dots,\, x_{in})\,,\ i = 1,\, 2,\, ...,\, m$
2. Form the empirical covariance matrix V
3. Calculate the eigenvalues and the eigenvectors of V
4. If $V^1 = (v_{11},\, v_{12},\, ...,\, v_{1n})$ is the first principal component of V, calculate

$$Z_i = \sum_{j=1}^{n} v_{1j} x_{ij},\ i = 1,\, 2,\, ...,\, m$$

5. Sort descending the vector $Z_1,\, Z_2,\, ...,\, Z_m$; the obtained list provides a better classification of students performance
6. Stop.

We have implemented the algorithm in Matlab, because this software offers many facilities in matrix computation.

4. Results

For the entrance data presented in Table 1 one obtains a sorted list (Z_i criterion) given in Table 2.

Table 1. The entrance data

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Informatics Bases	8	7	9	7	10	8	4	7	6	8	10	7	4	4	10	8	8	7	6
Algebra	4	4	4	5	6	5	4	4	5	7	5	4	4	4	4	5	4	4	4
Logic and Theory of sets	9	9	8	8	9	9	2	8	9	8	10	5	7	8	9	7	9	8	7

Table 2. The sorted list

No.	11	5	15	6	3	17	1	10	2	16
Z_i	14.69	14.16	13.88	12.56	12.48	12.42	12.42	12.17	11.69	11.22
Average	8.3	8.3	7.6	7.3	7.00	7.00	7.00	7.6	6.6	6.6

Table 2. The sorted list (continuation)

4	9	18	8	19	12	14	13	7
11.16	11.11	11.02	11.02	9.62	9.01	8.84	8.17	4.81
6.3	6.6	6.3	6.3	5.6	5.3	5.3	5.00	3.3

Remark. Generally, for the same average (for example 7.00, 6.6, 5.3) there are different Z_i . This fact illustrates the power of the Z_i criterion.

References

[1] E. Bradley, R. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, 1993, USA

[2] B.P. Demidovich, L.A. Maron, *Computational mathematics*, Mir Publishers, Moscow, 1981

[3] V. Pugachev, *Theory de probabilités et statistique mathematique*, Editions Mir, Moscou, 1982

University of Ploieşti
Petroleum-Gas Faculty
Department of Informatics
Bd. Bucureşti 39, 100680 Ploieşti, Romania
E-mail address: marinoiu_c@yahoo.com
E-mail address: iulianadobre@yahoo.com