

A New Student Ranking System by Identifying Important Parameters Using Principal Component Analysis : A Case Study For Class 10 Level Students

¹Abhoy Chand Mondal, ²Sumanta Ray

^{1,2}Dept. of Computer Science, The University of Burdwan, Golapbag, Burdwan, West Bengal, India

Abstract

In order to improve the academic performance of class 10 level students, here we have assessed 50 number of students of class 10 level. Assessment of class 10 level students is the main theme of this research paper. Our focus in this paper is that how class 10 level students will improve their academic performance. In general, a class 10 level student has been assessed due to the marks obtained in different subjects. Apart from that, we have selected many different parameters that are closely related to a class 10 level student to improve their academic performance. We also retrieved the more important parameters and less important parameters. If the students will give the extra preference of the more important parameters than the less important parameters, then obviously they will improve their academic performance. We also established a different rank among 50 students, those students have been passed class 10 examination.

Keyword

Principal Component Analysis, Student Assessment, Academic Performance, Important Parameter, Students New Ranking.

I. Introduction

In today's competitive environment outcome, assessment is now a primary focus of higher education. No work has been carried out for the assessment of class 10 level students in West Bengal. Generally, in class 10 Examination a student has been assessed by his/her academic score in different subjects. Apart from this convention, we have discussed about different parameters that are closely related to a class 10 level student. We have chosen nine parameters and collected the details of 25 students of 2011 class 10 batch and 25 students 2012 class 10 batch from a class 10 standard school in West Bengal. We have also taken their total marks in class 10 examination as a different indicator. In the first stage, we find out the more important and less important parameters. In the second stage, we have established a new ranking among 50 students according to individual important parameters. In the final stage, we have also established a different rank among students taking all important parameters. The assessment and new ranking has been carried out by using Principal Component Analysis.

II. PCA

Principal component analysis (PCA) is the oldest best known techniques of multivariate analysis. PCA was first introduced by Pearson in 1901, and developed by Hotelling in 1933. PCA was not widely used until the advent of electronic computers, but it is now widely used in every Statistical Computing Package. It is a statistical method that reduces the data dimensionality from a large number of interrelated data [1]. We can apply PCA technique everywhere where the amount of data are very large. The main goal of Principal Component Analysis are i) Variable Selection ii) Simplification iii) Modeling iv) Classification v) Data Reduction vi) Mixing etc.

III. Different parameters used in this paper

We have selected nine different indicators by which a class 10 level student is not assessed. The indicators are (x1) Attendance in School, (x2) Discipline in school, (x3) Result in different class test, (x4) Result in different surprise test, (x5) Financial background, (x6) Communication skill, (x7) Comprehensive skill, (x8) Professional skill, (x9) Dress code. We have allotted 10 marks for each indicator

such as if a student has 100% attendance in one academic year then he/she will score 10 out of 10. Financial support is also required for a class 10 level student, if a student got the financial support for his/her education then he/she will get 10 out of 10, otherwise less than that. If a student is well disciplined then he/she will get 10 out of 10. If a student scored outstanding in different class test and different surprise test in one academic year then he/she will score 10 out of 10 for each indicator. If the communication skill, comprehensive skill and professional skill of a student is good then he/she will get 10 out of 10. Dress code means he/she comes in School wearing school uniform or not.

IV. Data Collection

We have collected the data (i.e. different parameter values along with the final marks obtained in class 10 level) from Belar-Bhurkunda High School which is a class 10 standard school in Burdwan district of West Bengal state. We have randomly selected the details of 25 students in 2011 class 10 batch and 25 students of 2012 class 10 batch. The data has been stored in Table-1.

V. Method, Result and Discussion

Step 1: Corresponding Correlation matrix of Table-1 has been calculated by using Matlab and stored in Table-2.

Step 2: Then the eigenvalues, Variance Contribution rate, % of variance, cumulative % of variance of Table-2 has been calculated by using Matlab and furnished in Table-3.

Step-3: We have plotted all the eigenvalues in a graph (Fig.-1) and first two components have been selected as a principal component.

Step-4: The Eigenvectors of Table-2 has been calculated by using Matlab and stored in Table-4.

It has been shown that the first component is most correlated with the parameter 'Result in surprise test', So we can use the first component to measure this parameter. We also conclude that 'Result in surprise test' this parameter is one of the most important parameters to assess a class 10 level student. It has been also shown from Table-4 that the second component is most correlated with the parameter 'Financial Background', So we can

use the second component to measure this parameter. We also conclude that 'Financial Background' this parameter is one of the most important parameter to assess a class 10 level student[2]. So we can conclude that the parameters 'Result in Surprise Test' and 'Financial Background', are more important parameters than the other seven parameters which we have taken earlier[3].

Step-5: In this step, we have established a rank among fifty students according to first principal component and also established a different rank among fifty students according to second principal component(that has been shown in Table-5.1 and Table-5.2).[2][4][5]

Here, we have applied the first two principal components expression as follows:

$$C1=0.3488x1+0.3663x2+0.3638x3+0.3797x4+0.0926x5+0.3352x6+0.3419x7+0.3274x8+0.3501x9$$

$$C2=0.1439x1+0.2215x2+0.1761x3-0.1250x4+0.8512x5-0.3303x6-0.1799x7-0.1280x8-0.0361x9$$

Step-6:Finally,we established a overall rank according to all principal components and it has been shown in Table-6.

Here, we have made the comprehensive evaluation of 50 students by using the following formula[5]:

$$C=\sum_{i=1}^n Vi Ci$$

where Vi=Variance contribution rate and Ci=Principal component

In this problem C1 and C2 are the major components, So other components are neglected ,Therefore the formula in this problem is

$$C=V1C1+V2C2$$

This implies that $C=0.5267C1+0.11817772778C2$

Table-1

Index	X1	X2	X3	X4	X5	X6	X7	X8	X9	Total Marks in class 10 Examination (800)
1	9	8	9	6	5	6	7	8	7	555
2	9	9	9	7	4	9	9	9	8	480
3	8	8	8	6	8	6	8	7	7	434
4	9	9	8	7	9	7	7	7	7	480
5	7	9	8	7	7	7	7	7	9	480
6	8	8	8	6	9	7	7	8	7	320
7	7	8	8	7	4	7	8	7	8	375
8	8	7	7	6	7	6	7	8	7	359
9	8	8	7	7	5	8	7	8	7	301
10	7	7	8	6	5	7	8	7	8	327
11	8	8	7	7	4	7	8	7	7	311
12	8	7	8	6	4	8	7	8	7	274
13	8	8	7	7	4	7	7	8	8	324
14	7	7	8	6	7	6	8	7	8	270
15	8	8	7	7	4	7	7	8	8	273
16	9	9	9	9	8	9	9	9	9	715
17	9	9	9	9	7	8	8	9	9	633
18	9	8	8	8	8	8	8	8	9	630
19	8	8	8	8	7	7	8	8	7	491

20	8	8	7	7	6	8	8	7	8	482
21	8	7	7	8	6	7	8	8	7	500
22	7	7	7	6	7	7	7	7	7	338
23	7	7	7	7	4	8	8	7	7	392
24	8	8	8	7	8	8	8	7	8	387
25	7	7	7	8	7	7	7	8	7	305
26	8	7	8	8	4	8	8	9	9	282
27	8	9	9	9	8	8	9	9	9	493
28	8	8	8	8	7	7	8	8	8	348
29	9	9	9	9	4	9	9	9	9	420
30	9	9	8	9	4	8	8	8	8	420
31	8	9	9	9	9	9	8	7	8	522
32	8	8	8	7	4	7	7	7	7	261
33	8	8	8	8	7	8	8	8	8	377
34	8	8	8	8	5	8	8	8	8	376
35	8	7	7	8	7	8	8	7	8	236
36	9	9	9	9	4	8	8	8	8	233
37	8	8	7	7	4	7	7	8	7	324
38	8	8	8	8	4	8	8	7	7	211
39	9	9	9	9	7	8	8	8	8	379
40	7	7	8	7	4	8	8	7	7	233
41	9	9	9	9	9	9	9	9	9	422
42	9	9	8	8	8	8	8	8	8	229
43	8	8	7	7	4	8	8	8	8	276
44	8	8	8	8	8	8	8	8	8	377
45	8	7	7	7	7	7	8	8	8	270
46	9	9	9	9	6	8	9	9	9	383
47	8	8	8	8	4	8	8	7	7	235
48	9	9	9	9	4	8	8	9	9	360
49	8	8	8	8	7	8	8	7	7	238
50	7	7	7	7	4	7	7	7	7	208

Table 2 : (Correlation matrix of Table-1)

INDEX	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	1.0000	0.7102	0.5821	0.5409	0.1397	0.4234	0.3801	0.6202	0.3984
2	0.7102	1.0000	0.6987	0.6104	0.1983	0.4768	0.4077	0.4231	0.5150
3	0.5821	0.6987	1.0000	0.5357	0.1977	0.4513	0.5588	0.4696	0.5255
4	0.5409	0.6104	0.5357	1.0000	0.1202	0.6743	0.6021	0.5142	0.5754
5	0.1397	0.1983	0.1977	0.1202	1.0000	-0.0029	0.1203	0.0219	0.1720
6	0.4234	0.4768	0.4513	0.6743	-0.0029	1.0000	0.6245	0.3939	0.4830
7	0.3801	0.4077	0.5588	0.6021	0.1203	0.6245	1.0000	0.4261	0.5853
8	0.6202	0.4231	0.4696	0.5142	0.0219	0.3939	0.4261	1.0000	0.5930
9	0.3984	0.5150	0.5255	0.5754	0.1720	0.4830	0.5853	0.5930	1.0000

Table 3 : (Eigenvalues of Table-2)

Eigenvalues	Variation contribution rate	% of variance	Cumulative %
4.7403	0.5267	52.67	52.67
1.0636	0.118177778	11.81777778	64.48777778
0.8777	0.097522222	9.752222222	74.24
0.6856	0.76177778	7.61777778	81.85777778
0.1561	0.017344444	1.734444444	83.59222222
0.5098	0.056644444	5.664444444	89.25666667
0.4123	0.045811111	4.581111111	93.83777778
0.2682	0.0298	2.98	96.81777778
0.2864	0.031822222	3.182222222	100

Table 4 : (Eigenvector of Table-2)

INDICATORS	C1	C2	C3	C4	C5	C6	C7	C8	C9
X1	0.3488	0.1439	0.5216	0.0894	-0.4937	-0.2685	0.2860	0.4128	0.0978
X2	0.3663	0.2215	0.2847	0.3731	0.6051	0.1317	-0.3902	0.2347	0.0754
X3	0.3638	0.1761	0.1015	0.1923	-0.2357	0.6224	0.2062	-0.5501	-0.0360
X4	0.3797	-0.1250	-0.1466	0.1528	-0.1432	-0.3744	-0.2722	-0.1792	-0.7270
X5	0.0926	0.8512	-0.3845	-0.1256	0.0398	-0.2815	0.0999	-0.0841	0.0736
X6	0.3352	-0.3303	-0.3109	0.3268	-0.0157	-0.3646	0.0248	-0.2635	0.6098
X7	0.3419	-0.1799	-0.4688	-0.0196	0.2148	0.2461	0.5214	0.4687	-0.1799
X8	0.3274	-0.1280	0.3258	-0.6486	0.4024	-0.2034	0.2226	-0.3108	0.0097
X9	0.3501	-0.0361	-0.2115	-0.4988	-0.3227	0.2725	-0.5633	0.2167	0.2127

Table 5.1: Rank of fifty students according to first principal component

New Order	C1
31	21.559401
9	24.579700
37	21.139099
23	22.319799
25	22.137199
32	21.552401
34	21.484900
46	20.301901
33	21.533100
43	20.831499
38	21.119801
41	21.058300
35	21.455400
45	20.681499
35	21.455400
2	26.059601
6	25.289900
14	23.945299
22	22.488500
28	21.990299
30	21.645800
48	19.960899
44	20.739901
20	22.539301
42	21.047701

16	23.187201
5	25.375601
19	22.818600
3	25.689199
12	23.970800
10	24.456600
36	21.141701
17	23.153799
18	22.968599
27	22.096300
11	24.334600
39	21.105301
24	22.198500
8	24.612400
40	21.103701
1	26.152201
13	23.961500
26	22.132500
15	23.246401
29	21.708799
4	25.539200
24	22.198500
7	25.012100
21	22.476299
47	20.062799

Table 5.2: Rank of fifty students according to second principal component

New Order	C2
24	3.640200
43	1.370700
3	5.821900
1	6.763100
10	4.700700
2	6.394700
36	1.781800
12	4.625000
30	2.358500
29	2.536500
35	1.785700
40	1.586900
34	1.801500
13	4.599200
34	1.801500
14	4.489400
19	4.148400
8	4.855000
17	4.262400
26	3.121700
28	3.013600
16	4.278800
46	1.090000
7	5.000200
22	3.900800
48	0.956800
11	4.675800
18	4.226300
47	1.084600

41	1.582800
4	5.668700
32	2.141700
23	3.896000
31	2.193600
25	3.626400
37	1.758900
33	1.837600
42	1.506500
15	4.312500
45	1.266100
5	5.340600
6	5.112600
44	1.291300
9	4.747200
21	3.953700
27	3.117300
42	1.506500
39	1.594800
20	4.060100
38	1.600200

12.066682	25	37
13.024898	13	39
11.333326	41	25
11.869985	29	41
13.47299	8	16
11.264944	44	39
14.405505	1	12
13.224719	10	40
11.809791	31	31
12.804893	15	16
11.901265	28	34
13.819893	4	15
11.869985	29	38
13.362344	9	20
12.318081	22	36
10.756185	48	42

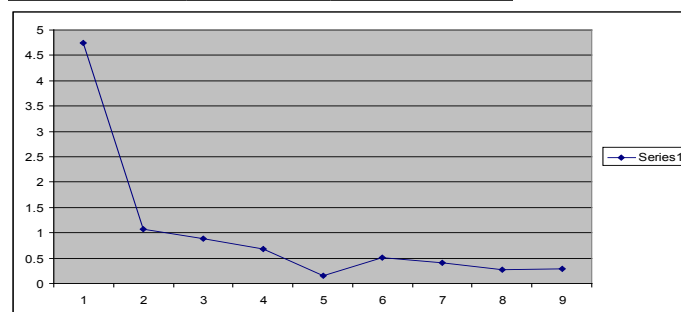


Fig.1

Table 6: Final ranking using $C=V1*C1+V2*C2$

Score	Final Order	Original Order
11.785528	32	4
13.108115	12	10
11.821984	30	11
12.555088	17	10
12.215182	23	10
12.107362	24	26
11.526666	36	19
11.23958	45	21
11.620207	34	29
11.271709	43	24
11.334829	40	27
11.278943	42	32
11.513457	37	25
11.43647	38	34
11.513457	37	33
14.25614	2	1
13.81044	5	2
13.185743	11	3
12.348415	20	8
11.951207	27	9
11.756984	33	6
11.019066	47	23
11.05252	46	13
12.462363	18	14
11.546813	35	28
12.325771	21	30
13.917906	3	7
11.991312	26	22
13.658677	6	13
12.812472	14	13
13.551207	7	5
11.388435	39	35
12.655527	16	17
12.356796	19	18

VI. Conclusion

In this technique, it has been proved that all the indicators related to the class 10 level students are not more important indicators to assess them. By applying this technique we can capable to generate a different rank apart from the conventional ranking. Our moto was to identify the students those students has the capability but they do not utilize their capability due to the lack of the more important indicator(s). This accounts should gives education system researchers new tools and ideas on how to approach proposed problems using PCA based neural network methods.

References

[1] I.T.Jolliffe, "Principal Component Analysis", Springer,
 [2] T.Zhang, J.Guo, N.Zhang, "Application of PCA in Quality Assessment of College Students" International Conference on Information Technology, Computer Engineering and Management Science, IEEE Computer Society, 2011, pp. 134-137.
 [3] S.N.Mandal, J.Pal Choudhury, S.R.Bhadra Choudhury, "Neuro-PCA-Factor Analysis in Prediction of Time Series Data," American Journal of Intelligent Systems, 2012, pp. 45-52
 [4] G.Hong-bo, L.Xuan, "Research on Optimizing Quality of Higher Vocational Education Based on Principal Component Analysis of Data Mining", International Conference on Intelligence Science and Information Engineering, IEEE Computer Society, 2011, pp. 534-537
 [5] Z.Sumei, "The Comprehensive Evaluation of Teaching Quality Based on Principal Component Analysis", International Conference on Advanced Computer Theory and Engineering (ICACTE), IEEE, 2010, pp. v3-314-v3-317.

- [6] S.Ray,A.C.Mondal,A.Neogi, "A Study of the Development of Principal Component Analysis using Neural Network in Assessment of Higher Education System", Proceedings of NACCS-2012, Burdwan University,West Bengal, India, March, 2012, pp.

Author's Profile



He is an Associate Professor and Head of Dept. of Computer Science, The University of Burdwan, from 2008. He was born in 27/02/1964.He received his B.Sc.(Math-Hons.) degree from The University of Burdwan in 1987, M.Sc.(Math) and MCA from Jadavpur University in 1989 and 1992 respectively. He received his Ph.D. degree from Burdwan University in 2004.

His research interest is in Soft Computing, Document Processing, Web Mining, NLP etc. He has 1 year industry experience and 18 years of teaching and research experience. No. of papers published is 60(No. of Journal papers 20).So far two students awarded Ph.D. degree under his guidance. Currently 8 students are undergoing their Ph.D. work under his supervision.



Sumanta Ray is a Ph.D. Research Scholar of The Department of Computer Science, The University of Burdwan. He obtained his Bachelor's degree in Computer Applications from The University of Burdwan and Master's degree in Computer Applications from West Bengal University of Technology in 2005 and 2008 respectively. He has about 5 years of teaching experience. He worked in

different Institutions. His research interest is in the related areas of Soft Computing.