World Applied Sciences Journal 21 (4): 572-577, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.21.4.1112

The Use of Partial Least Squares Regression and Feed Forward Artificial Neural Networks Methods in Prediction Vertical and Broad Jumping of Young Football Players

[1]E. Bulut, [2]U. Yolcu, [3]M.Y. Tasmektepligil and [4]E. Egrioğlu

[1]Department of Business, Ondokuz Mayýs University, Samsun, Turkey
[2]Department of Statistics, Giresun University, Giresun, Turkey
[3]Department of Physical Education and Sports, Ondokuz Mayýs University, Samsun, Turkey
[4]Department of Statistics, Ondokuz Mayýs University, Samsun, Turkey

Abstract: Many different parts of body participate in jumping. Especially, arms and lengths are the most important parts of the body that effect vertical and broad jumping. Out of these parts muscle structure is also an effective factor. Reproduction of variables is possible. In these situations the number of variables can exceed the number of observation unit. In such a case it is possible to get a model that fits the data but this model will fail in predicting new data sets. For such a data set, there are so many reasons that fail to work with MLR (multiple linear regression) analysis. There are many prediction methods that can be used as an alternative to MLR. PLSR (partial least squares regression) and FFANN (feed forward artificial neural networks) are two of them. In this study the performance of these two methods were compared for predicting vertical and broad jumping depending on young football players body measurements. The data used in this study was about a total of 30 young football players enrolled in the league of "Football Players who are Candidates of Professional Leagues. This study suggests an idea about the impression of every part of the body on performance of jumping. These statistical analyses can be easily used in all sport sciences in making prediction and obtaining the importance of the variables.

Key words: Partial least square regression · Feed forward artificial neural networks · Prediction · Vertical and broad jumping

INTRODUCTION

PLS's origin lies in the sixties, seventies and eighties of the previous century, when Herman O. A. Wold vigorously pursued the creation and construction of models and methods for the social sciences, where "soft models and soft data" were the rule rather than the exception and where approaches strongly oriented at prediction would be of great value. The author was fortunate to witness the development firsthand for a few years. Herman Wold suggested (in 1977) [1] to write a PhD-thesis on LISREL versus PLS in the context of latent variable models, more specifically of "the basic design" [2].

The widespread uses of PLSR method have begun with son Svante Wold in chemometrics. The PLS calibration methods as used in chemometrics, have recently obtained some attention in the statistical literature [3-4], in theoretical contributions elsewhere [5-9] are the other major names studied in this field.

The activation functions of artificial neural networks are used in PLS method. Because the activation functions provide highly nonlinear transformations, they solve multicollinearity problem. Moreover, PLS method has non-linear modeling ability. [10] propose non-linear PLS method based on feed forward artificial neural networks whereas [11] propose non-linear PLS algorithm based on radial bases activation functions, [12] propose non-linear PLS method based on logistic activation function and particle swarm optimization methods. [13] suggests different non-linear PLS algorithm that differently used feed forward neural networks. [14] compared Counter propagation neural network and PLS-DA algorithm.

Corresponding Author: E. Bulut, Department of Business, Ondokuz Mayis University, Samsun, Turkey

There are many studies in the literature made in predicting the performances of athletes. Some of them that used advanced knowledge of statistics are [15-19].

In this study, FFANN and PLS methods were compared on a sport data and the results were discussed to obtain the best prediction model. Contents of the paper are as follows: A brief description of FFANN and PLS were given in Material and Methods part. Results of the analysis were given in the following section and finally these results were discussed in conclusion part.

MATERIALS AND METHODS

The data used in this study was about a total of 30 young football players enrolled in the league of "Football Players who are Candidates of Professional Leagues". In this data set, the number of observation units (young football players) is 30. Explanatory variables are taken from the right side (R) and left side (L) of the body such as width (W) of the circumference (C) for right and left arm (A), width of circumference for right and left forearm (Fa), width of circumference for right and left hand (Ha). These calculations were done also for thigh (T), knee (K), hip (H) and foot (F). At the same time the length (Le) of the arm, forearm, hand, thigh, foot and leg for the right and left side of the body was calculated. The thickness (Th) of skinfold (S) of abdomen (A), the skinfold of triceps (Tr), subscapular (Ss), biceps (B) patella (P) and extremities (upper: UE, lower: LE) values were taken, too. So the number of explanatory variables is about 73.

The number of dependent variables is 2. They are vertical and broad jumping with two legs refer to y1 and y2, respectively. So, X : (30×73), Y : (30×2). Vertical jumping was measured in centimeter, broad jumping was measured in meter. Length and circumference measurements were measured in centimeters and skinfold was measured in millimeter. Variables are defined as follows. *(Note: Some organs and terms were displayed with italic type.)*

Feed Forward Neural Networks: Artificial neural network is a data processing mechanism generated by the simulation of human nerve cells and nervous system in a computer environment. The most important feature of artificial neural network is its ability to learn from the examples. Despite having a simpler structure in comparison with the human nervous system, artificial neural networks provide successful results in solving problems such as forecasting, pattern recognition and classification.

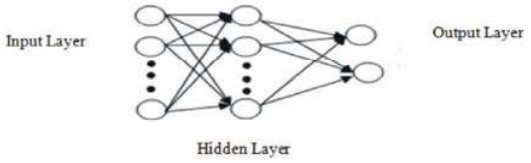

Fig. 1: Multilayer feed forward artificial neural network with two output neuron

Although there are many types of artificial neural networks in literature, feed forward artificial neural networks are frequently used for many problems. Feed forward artificial neural networks consist of input layer, hidden layer(s) and output layer. An example of feed forward artificial neural network architecture is shown in Figure 1. Each layer consists of units called neuron and there is no connection between neurons which belong to same layer. Neurons from different layers are connected to each other with their weights. Each weight is shown with directional arrows in Figure 1. Bindings shown with directional arrows in feed forward artificial neural networks are forward and unidirectional. Single activation function is used for each neuron in hidden layer and output layer of feed forward artificial neuron network. Inputs incoming to neurons in hidden and output layer are made up multiplication and addition of neuron outputs in the previous layers with the related weights. Data from these neurons pass through the activation function and neuron output are formed. Activation function enables curvilinear match-up. Therefore, non-linear activation functions are used for hidden layer units. In addition to a non-linear activation function, linear (pure linear) activation function can be used in output layer neuron.

In feed forward artificial neural networks, learning is the determination of weights generating the closest outputs to the target values that correspond with the inputs of artificial neural network. Learning is achieved by optimizing the total errors with respect to weights. There are several types of training algorithms in literature used for learning feed forward artificial neural networks. One of the widely used training algorithms is Levenberg-Marquardt (LM) algorithm which was also used in this study.

Partial Least Squares Regression Method: Many explanatory variables in a data set cause an increase in the probability of being multicollinearity problem among explanatory variables. Moreover, more explanatory variables than observation unit make ordinary least square regression unavailable. PLSR is one of the

alternative methods that can be used in these situations. It is a statistical multivariate method consisting of PLS and Multiple linear regression (MLR).

The aim of PLS is to form components that capture most of the information in the X variables with dimension that is useful for predicting $Y_1,...,Y_k$ while reducing the dimensionality of the regression problem by using fewer components than the number of X variables [8]. These components obtained by PLS algorithms as orthogonal variables did not have collinearity. Also, they explain most of the variability in the covariance matrix X′Y having dimension M×K. Basic algorithm of PLS is NIPALS algorithm. SIMPLS, UNIPALS, SAMPLS and Kernel algorithms are based on NIPALS.

Data matrices X with dimension N×K and Y with dimension N×K can be modeled separately by these components as given below.

$$X = TP' + E$$
$$Y = UC' + F$$

Here, E and F are error terms, T and U are the N×A matrices of the A derived components for X and Y. P and C represents loading and weight matrices with dimensions N×A and K×A, respectively. PLSR model can be written as a multivariate regression $Y=XB_{PLSR} + F$. Here regression coefficients for PLSR are obtained from $B_{PLSR} = W(P'W)^{-1} C'$.

SIMPLS Algorithm: This algorithm was developed by Sijmen De Jong in 1993. This name was given since it's being a straightforward implementation of a statistically inspired modification of the PLS method [21]. As mentioned in the paper [22], it is much faster than NIPALS and needs less computer memory according to NIPALS algorithm.

Both of the algorithms depend on maximizing S =X′Y covariance matrix but in NIPALS algorithm data matrices are deflated in each step and the components are the linear combinations of the deflated matrix rather than being the original matrix. For that reason the interpretation of the component matrix T is not straightforward. SIMPLS calculates the PLS components directly as linear combinations of the original variables because of deflating S =X′Y.

RESULTS

For this data set, it was observed that there were high correlations among variables because of measurements were taken on both side of the body. Except body measurements, high correlation coefficients were observed, too. For experience and age r=0.462 (p=0.010), for height and weight r=-0.551 (p=0.02) for weight and age r=0.508 (p=0.04). All these correlations are significant in p<0.05. So, multicollinearity is an inevitable problem.

In this study 73 explanatory variables, having multicollinearity among them, were reduced to 10 components by maximizing S =X′Y. by SIMPLS algorithm in MATLAB statistical software. By proceeding algorithm, regression coefficients were obtained from these independent components. Regression coefficients for two dependent variables are given (Table 2).

Randomly selected 27 observations were used to obtain the models. 3 observations were used as test set (n_{test}=3). That is, 27 observations were used in modeling, 3 were used in prediction. As a comparison criterion RMSE (root mean square error) was used. RMSE values for test set by predicting with PLSR method appear in Table 4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{test}} (y_i - \hat{y}_i)^2}{n_{test}}}$$

Secondly, prediction was made with FFANN method in MATLAB. Input number of FFANN is the number of explanatory variables. On the other hand, the numbers of hidden layer neurons vary between 1 and 73, the 73 different FFANN architectures are used for obtaining predictions. The FFANN was trained by using

Table 1: Explanatory Variables

X1:DoninantlegR	X12:CWFaL	X23:ClegL	X34:LeLER	X45:SBL	X56:SPL	X67:flexibility2
X2:DominantlegL	X13:CHaR	X24:CFR	X35:LeLEL	X46:STrR	X57:legTrR	X68:30meter1
X3:Experience	X14:CHaL	X25:CFL	X36:LeTR	X47:STrL	X58: legTrL	X69:30meter2
X4:Age(years)	X15:WHaR	X26:LeUER	X37:LeTL	X48:noseS	X59:KflexionR	X70:elbowextensionR
X5:Height(cm)	X16:WHaL	X27:LeUEL	X38:LelegR	X49:CSsR	X60:KflexionL	X71:elbowextensionL
X6:Weight(kg)	X17:CH	X28:LeAR	X39:LelegL	X50:CSsL	X61:KextensionR	X72: elbowflexionR
X7:CWAR	X18:CTR	X29:LeAL	X40:LeFR	X51:ThAR	X62: KextensionL	X73: elbowflexionL
X8:CWAL	X19:CTL	X30:LeFaR	X41:LeFL	X52:ThAL	X63:HflexionR	-
X9:ThCAR	X20:CKR	X31:LeFaL	X42:faceSR	X53:ScristailiacR	X64:HflexionL	-
X10:ThCAL	X21:CKL	X32:LeHR	X43:faceSL	X54:ScristailiacL	X65:legstrenght	-
X11:CWFaR	X22:ClegR	X33:LeHL	X44:SBR	X55:SPR	X66:flexibility1	-

Table 2: Coefficients of PLSR model

	Beta			Beta	
Number	y1	y2	Number	y1	y2
0	99,2291	0,2183	37	0,1460	0,0027
1	0,0209	0,0002	38	0,0315	0,0007
2	-0,0209	-0,0002	39	-0,0180	0,0007
3	-0,2452	0,0000	40	0,0575	0,0000
4	0,0197	0,0008	41	-0,1643	-0,0004
5	-0,1413	0,0008	42	-0,1900	-0,0001
6	-0,2046	0,0035	43	-0,2244	-0,0004
7	0,0772	0,0003	44	0,1681	-0,0006
8	-0,0040	0,0002	45	0,0109	-0,0007
9	0,2755	0,0018	46	-0,2998	-0,0010
10	-0,0387	0,0014	47	-0,5445	-0,0009
11	0,0572	0,0007	48	0,2374	0,0004
12	-0,2198	0,0005	49	-0,1808	-0,0014
13	0,0327	0,0006	50	-0,0989	-0,0020
14	-0,1438	0,0005	51	0,0930	-0,0011
15	-0,0084	-0,0001	52	-0,1367	-0,0011
16	-0,1414	0,0000	53	0,0151	-0,0006
17	0,1844	-0,0001	54	0,0435	-0,0010
18	0,0940	0,0023	55	-0,2301	-0,0013
19	-0,1054	0,0034	56	-0,2461	-0,0010
20	-0,4181	-0,0004	57	-0,3028	-0,0005
21	-0,1596	0,0002	58	-0,0286	-0,0006
22	-0,2624	0,0020	59	-0,0096	-0,0007
23	0,2246	0,0012	60	0,0914	0,0016
24	0,0146	0,0009	61	-0,0972	0,0012
25	-0,5558	0,0001	62	-0,1414	0,0019
26	0,7973	-0,0004	63	-0,0414	0,0011
27	-0,0569	-0,0013	64	-0,1277	-0,0020
28	-0,3026	-0,0012	65	0,1744	0,0023
29	-0,4725	-0,0013	66	0,1249	0,0045
30	0,2415	0,0019	67	-0,5887	0,0045
31	0,0917	0,0017	68	-0,0662	0,0000
32	0,1906	0,0013	69	-0,0542	0,0000
33	0,1284	0,0014	70	0,4718	0,0026
34	-0,3313	0,0009	71	-0,5585	-0,0033
35	-0,1001	0,0013	72	-0,1208	-0,0035
36	0,3611	0,0036	73	0,6217	0,0019

Table 3: RMSE values of ANN for test set

Number of Hidden			Number of Hidden			Number of Hidden		
Layer Neurons	RMSE(y1)	RMSE(y2)	Layer Neurons	RMSE(y1)	RMSE(y2)	Layer Neurons	RMSE(y1)	RMSE(y2)
1	21,09887	0,247362	26	19,28291	0,52644	51	33,92205	0,472842
2	36,06257	0,352278	27	34,07456	0,468002	52	34,44804	0,733553
3	36,04038	0,411007	28	34,81132	0,43745	53	27,89245	0,256019
4	41,70834	0,534798	29	34,57826	0,519486	54	29,37088	0,526499
5	31,4048	0,7302	30	32,86597	0,537342	55	28,0606	0,733553
6	14,04772	0,624171	31	23,96558	0,446673	56	34,39794	0,560524
7	28,35364	0,614316	32	31,91676	0,513076	57	31,18838	0,296774
8	35,15700	0,721477	33	35,11431	0,520347	58	18,4081	0,565385
9	21,29077	0,525349	34	34,14693	0,464362	59	26,79236	0,221734
10	34,92026	0,513752	35	21,94723	0,60893	60	31,66982	0,523972
11	18,66147	0,559778	36	27,68912	0,520499	61	32,40722	0,531895
12	35,90748	0,593502	37	33,20489	0,422956	62	33,58802	0,522592

Table 3: Continue

Number of Hidden Layer Neurons	RMSE(y1)	RMSE(y2)	Number of Hidden Layer Neurons	RMSE(y1)	RMSE(y2)	Number of Hidden Layer Neurons	RMSE(y1)	RMSE(y2)
13	18,2517	0,474109	38	25,07418	0,510399	63	28,12255	0,516704
14	35,87588	0,38443	39	22,97573	0,529364	64	34,56713	0,534191
15	25,75353	0,309386	40	34,90739	0,521	65	31,92042	0,534926
16	29,09698	0,523513	41	32,7953	0,316307	66	34,28279	0,533283
17	20,21182	0,52177	42	25,7285	0,543902	67	28,66531	0,605271
18	34,6535	0,560915	43	17	0,519632	68	20,03895	0,451688
19	31,85123	0,532544	44	35,46836	0,300042	69	23,06382	0,577582
20	17,91545	0,574188	45	35,24735	0,527726	70	34,59053	0,386239
21	22,92825	0,583705	46	34,00764	0,276911	71	19,93855	0,402519
22	36,00759	0,510811	47	28,14869	0,534591	72	30,803	0,250066
23	29,52492	0,285631	48	20,16465	0,396209	73	30,35177	0,404274
24	23,63508	0,532672	49	32,44909	0,733553			
25	31,8933	0,212521	50	30,54776	0,509444			

Table 4: Best results of ANN and PLSR

	PLSR	ANN- (Best Architecure)
RMSE(y1)	8,5319	14,0477 - (73-8-2)
RMSE(y2)	0,1247	0,6241 - (73-8-2)

Levenberg-Marquardt algorithm with 500 maximum number of iterations. The RMSE values for two dependent variables are given in (Table 4). When the Table 3 is examined, the best result of FFANN is the architecture (73-8-2) which has 73 inputs, 8 hidden layer neurons and two outputs.

CONCLUSION

In this study, FFANN which provides non-linear modeling without any assumption and PLSR which is used for constructing predictive models when the variables are many and highly collinear were compared according to the performances of modeling vertical and broad jumping. Among two test sets with different extents PLSR gives better prediction results according to FFANN results obtained from RMSE. For this study it is better to make prediction with PLSR method since it has smaller RMSE values. With this method, for prediction with samples do not appear in the data set, it makes prediction with 8.5319 mean of error for vertical jumping and 0.1247 mean of error for jumping forward. For this study, PLSR method makes it easier to predict jumping with many explanatory variables with less prediction error. The regression coefficients for PLSR in Table 2 show the impression of the explanatory variable on response variable when the other explanatory variables get fixed. For example regression coefficient for the length of the upper extremities for right (X26:LeUER) and left side (X27:LeUEL) are β_{26}=0.7973 and β_{27}= 0.0569, respectively. That is, the impression of the length of the upper extremities to vertical jumping is 0.7973 and -0.0569 centimeter per one centimeter. Also, coefficients for length of thigh for right and left side (X36:LeTR, X37:LeTL) are for vertical jumping 0.3611 and 0.1460, for broad jumping 0.0036 and 0.0027, respectively. So, thigh is one of the high contributions of those to jumpings. Although, some measurements such as thickness of skinfold are not directly related to vertical and broad jumping, considering the whole body it is meaningful in sporty aspect to form an opinion on the performance of football players. Thereby this study suggests an idea about the impression of every part of the body on performance of jumping.

REFERENCES

1. Wold, H., 1977. Open path models with latent variables. In Albach, Helmstedter and Henn (eds), Kuantitative Wirtschaftsforschung; Wilhelm Krelle zum 60 Geburtstag, Tübingen: Mohr.
2. Dijkstra, T.K., 2010. H and book of Partial Least Squares. Concepts, Methods and Applications, Springer, pp: 24.
3. Martens, H. and T. Naes, 1989. Multivariate Calibration. John Wiley and Sons,
4. Helland, I.S., 1988. On the structure of partial least squares regression. Commun. Statist. B-Simulation Comput., 17: 581-607.
5. Höskuldson, A., 1988. PLS Regression Methods. Journal of Chemometrics, 2: 211-28.
6. Helland, I.S., 1990. Partial Least Squares Regression and Statistical Models. Scandinavian Journal of Statistics, 17: 97-114.

7. Geladi, P. and R. Kowalski, 1986. Partial Least Squares Regression: A Tutorial. Analytica Chimica Acta, 185: 1-17.
8. Garthwaite, P.H., 1994. An Interpretation of Partial Least Squares. Journal of the American Statistical Association, 89: 122-27.
9. Wold, S., M. Sjöström and L. Eriksson, 2001. PLS-regression:a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58: 109-30.
10. Qin, S.J. and T.J. McAvoy, 1992. Nonlinear PLS modeling using neural networks, Comp. Chem. Eng., 16: 379-91.
11. Yan, X.F., D.Z. Chen and S.X. Hu, 2003. Chaos-genetic algorithms for optimizing the operating conditions based on RBF-PLS model, Comp. Chem. Eng., 27: 1393-1404.
12. Zhou, Y.P., J.H. Jiang, W.Q. Lin, L. Xu, H.L. Wu, G.L. Shen and R.Q. Yu, 2007. Artificial neural network based transformation for nonlinear partial least square regression with application to QSAR studies, Talanta, 71: 848-53.
13. Xufeng, Y., 2010. Hybrid artificial neural networks based on BP-PLSR and its application in development of soft sensors, Chemometrics and Intelligent Laboratory Systems, 103: 152-59.
14. Alvarez-Guerra, M., D. Ballabio, J.M. Amigo, R. Bro and J.R. Viguri, 2010. Development of models for predicting toxicity from sediment chemistry by partial least squares-discriminant analysis and counter propagation artificial neural networks, Environmental Pollution, 158: 607-14.
15. Kimbrough, S.K., L. De Bolt and R. Balkin, 2007. Use of the Athletic Coping Skills Inventory for Prediction of Performence in Collegiate Baseball. The Sport Journal, 10: 1.
16. José Silva, A., A.M. Costa, P.M. Oliveira, V.M. Reis, J. Saavedra, J. Perl, A. Rouboa and D.A. Marinho, 2007. The use of neural network technology to model swimming performance. Journal of Sports Science and Medicine, 6: 117-25.
17. Reeves, R.A., O.D. Hicks and J.W. Navalta, 2008. The Relationship Between Upper Arm Anthropometrical Measures and Vertical Jump Displacement. International Journal of Exercise Science, 1(1), Iss. 1, Article 4.
18. Young, J.F., L.B. Larsen, A. Malmendal, N.C. Nielsen, I.K. Straadt, N. Oksbjerg and H.C. Bertram, 2010. Creatine-induced activation of antioxidative defence in myotube cultures revealed by explorative NMR-based metabonomics and proteomics. Journal of the International Society of Sports Nutrition, 7: 9.
19. Almagro, B., P. Sáenz-López and J.A. Moreno, 2010. Prediction of sport adherence through the influence of autonomy-supportive coaching among Spanish adolescent athletes. Journal of Sports Science and Medicine, 9: 8-14.
20. De Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 18: 251-63.
21. Lindgren, F. and S. Rannar, 1988. Alternative Partial Least-Squares (PLS) Algorithm. Perspective in Drug Discovery and Design, 1988;12/13/14:105-13.